

Metodología de la Investigación

Estadística aplicada en la Investigación



Lic. Nel Quezada Lucio

Temario:

Paso previo

Paso 1: Planteamiento del problema

Paso 2: Diseño Muestral

Paso 3: Recolectar Datos

Paso 4: Procesamiento de Datos

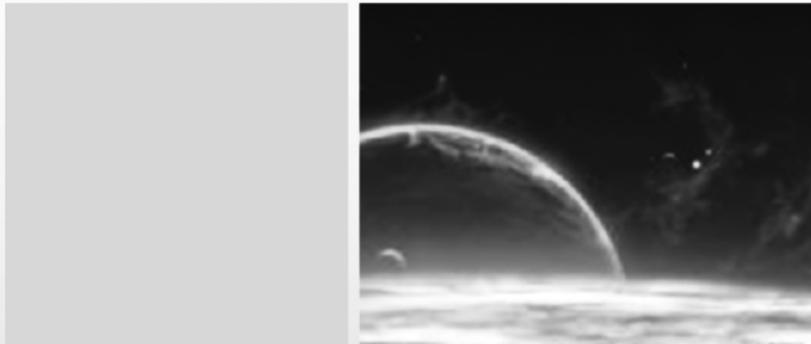
Paso 5: Análisis de Resultados y conclusiones



EMPRESA EDITORA
MACRO

Metodología de la Investigación

Estadística aplicada en la Investigación



 EMPRESA EDITORA
MACRO®

Metodología de la investigación

Autor: Nel Quezada Lucio

© Derecho de autor reservado

Empresa Editora Macro E.I.R.L.

© Derecho de edición, arte gráfico y diagramación reservados

Empresa Editora Macro E.I.R.L.

Edición a cargo de:

Empresa Editora Macro E.I.R.L.

Av. Paseo de la República 5613 – Miraflores

Lima - Perú

☎ (511) 719-9700

✉ ventas@editorialmacro.com

<http://www.editorialmacro.com>

Primera edición: Abril 2010 - 1000 ejemplares

Impreso en los Talleres Gráficos de

Empresa Editora Macro E.I.R.L.

Lima - Perú

ISBN Nº 978-612-4034-50-3

Hecho el Depósito Legal en la Biblioteca Nacional del Perú Nº 2010-04740

Prohibida la reproducción parcial o total, por cualquier medio o método de este libro sin previa autorización de la Empresa Editora Macro E.I.R.L.

Nel Quezada Lucio

Licenciado en Estadística de la Universidad Nacional de Ingeniería, con Maestría en Ciencias y mención en Ingeniería de Sistemas-UNI. Catedrático de la Escuela Profesional de Ingeniería Económica e Ingeniería Estadística de la Universidad Nacional de Ingeniería. Asesor y consultor de empresas.

Trabajos realizados en: Diseño y desarrollo de productos con estudios de panel de consumidores, tamaño de mercado y segmentación de mercado; en Sistema de Información Estadística, así como en la Elaboración, Crítica, Codificación, Procesamiento, Control de Calidad y Análisis; en el desarrollo y gestión de estudios e investigaciones en diferentes campos; en Teoría de Muestreo aplicado en encuestas Psicológicas y Socio Económicas, para efectos de evaluar el comportamiento de la población

Dedicatoria

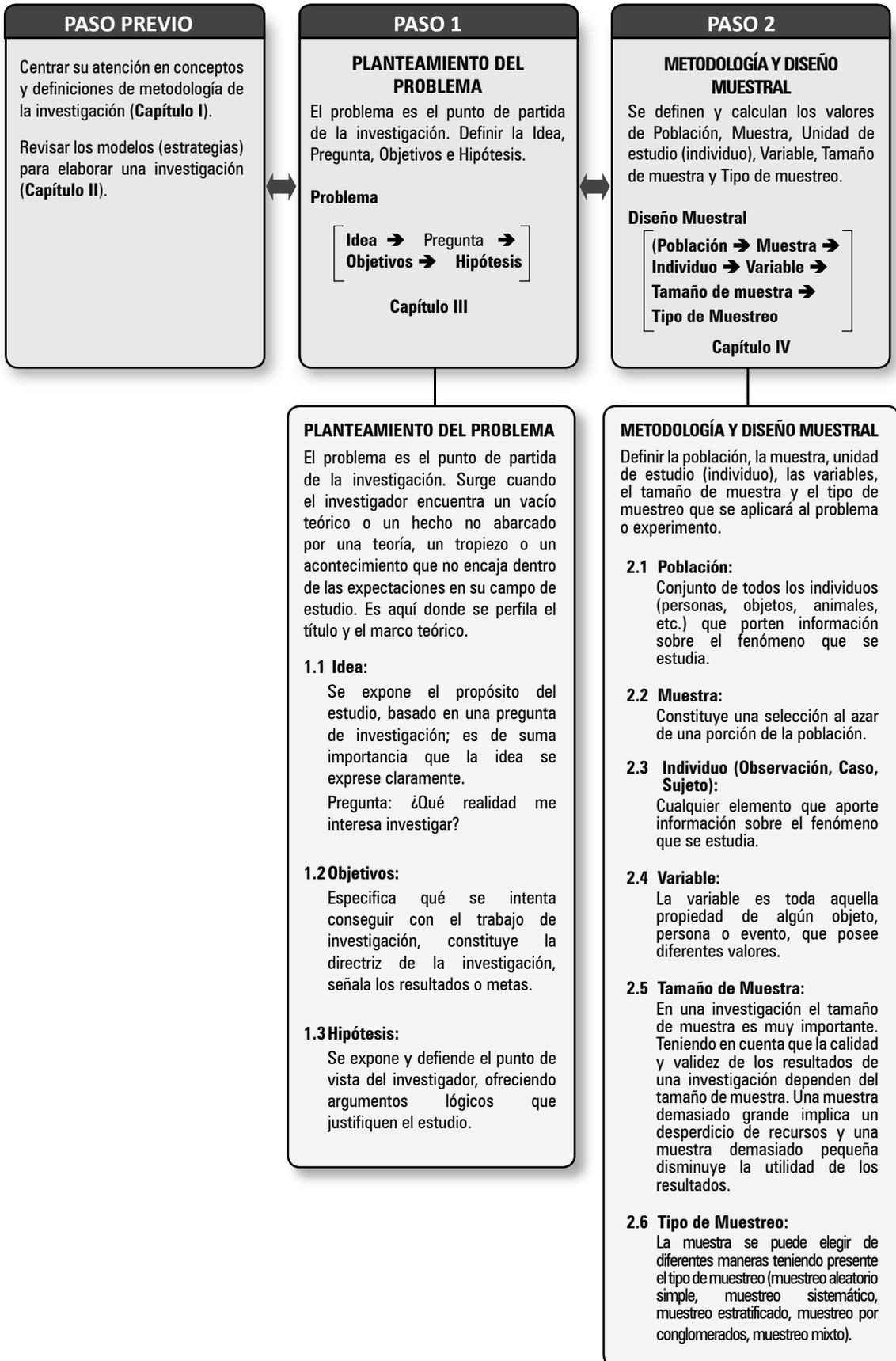
A mi madre Francisca Lucio Azaña.

Introducción

La investigación, en términos operativos, orienta al investigador en su razonamiento y aproximación a la realidad, ordena sus acciones y aporta criterios de rigor científico de supervisión de todo el proceso. En tanto que, investigar supone la responsabilidad de producir una lectura real de las cuestiones de la investigación y demostrar la contribución efectiva de ella.

El presente libro permite una orientación metodológica para la elaboración de proyectos y trabajos de investigación. El objetivo es brindar una orientación didáctica respecto a los criterios científicos y técnicos para la elaboración y desarrollo de una investigación, que debe cumplir con los requisitos y condiciones conceptuales y metodológicas que permitan la validez científica y social de los trabajos de investigación. Se desarrollará puntualmente los contenidos mínimos necesarios en un investigación, describiendo con la mayor claridad posible lo que debe incluirse en cada sección del la investigación.

La investigación está siempre vinculada a la realidad, al campo de conocimiento disciplinar de aplicación, al contexto cultural, social y político, cuya difusión aproxima a científicos de diferentes campos disciplinares, en que se enriquece la formación universitaria y orienta a actores sociales relevantes.



PASO 3

RECOLECTAR DATOS O INFORMACIÓN

Implica seleccionar un instrumento de medición disponible o desarrollar uno propio, es decir, elaborar el cuestionario y recolectar la información.

Recolectar datos

Cuestionario → Equipos de medición → Codificación

Capítulo V

PASO 4

PROCESAMIENTO DE DATOS O INFORMACIÓN

Consiste en ingresar los datos recolectados en el cuestionario a un software de su preferencia.

Capítulo VI

PASO 5

ANÁLISIS DE RESULTADOS Y CONCLUSIONES

Ingresado los datos, es hora de sacar los resultados. Los resultados muestran los principales hallazgos de la investigación aplicando técnicas estadísticas.

Capítulos: VII – VIII – IX – X – XI – XII - XIII - XIV - XV

RECOLECTAR DATOS O INFORMACIÓN

Recolectar los datos implica seleccionar un instrumento de medición disponible o desarrollar uno propio. Es decir, elaborar el cuestionario y recolectar la información.

3.1 Instrumento de medición (Cuestionario):

Cualquier recurso de que se vale el investigador para acercarse a los fenómenos y extraer de ellos información.

Cuestionario: Representa un conjunto de preguntas respecto a una o más variables a medir, utilizada para sacar información de cualquier fenómeno.

3.2 Equipos de Medición:

Cualquier equipo que permita apoyar la medición de variables (grabadora de audio ó vídeo, computador, satélite, gráficos, informes, expedientes, estereoscopio, etc.).

3.3 Codificación de datos:

Codificar consiste en asignarle un número o letra diferente a las posibles respuestas (alternativas) de una pregunta.

PROCESAMIENTO DE DATOS O INFORMACIÓN

Consiste en ingresar los datos recolectados en el cuestionario a un software de su preferencia. Si los datos son de un cuestionario desarrollado, llamaremos a este proceso introducción directa de datos. Una investigación también se realiza con información ya existente en un soporte electrónico de datos (archivos de Excel, SQL, Lotus, texto, etc.), a este proceso lo llamaremos extrayendo un archivo de disco.

ANÁLISIS DE RESULTADOS Y CONCLUSIONES

Ingresado los datos, es hora de sacar los resultados. Los resultados muestran los principales hallazgos de la investigación aplicando técnicas estadísticas (tablas, cuadros, gráficas, etc.) y presenta una potente interpretación teórica que demuestra el dominio técnico del investigador. Las conclusiones extraen lo esencial de todo el proceso enfatizando especialmente la riqueza de la evidencia empírica aportada. Las técnicas estadísticas utilizadas son: Estadísticas Descriptivas, Correlación, Regresión, Prueba de hipótesis, ANOVA (Modelos Lineales), Prueba no paramétricas, etc.

Para la obtención de resultados (Paso 5) usted puede utilizar una técnica estadística expuesta en este libro.

Contenido

Capítulo 1

| | |
|---|-----------|
| METODOLOGÍA DE LA INVESTIGACIÓN CIENTÍFICA | 15 |
| Conocimiento y Ciencia | 17 |
| 1. Origen del Conocimiento | 17 |
| 1.1 Sujeto cognoscente | 17 |
| 1.2 Objeto del conocimiento | 17 |
| 1.3 El conocimiento | 18 |
| 2. Principales corrientes en la obtención del conocimiento científico | 18 |
| 2.1 Positivismo y Neopositivismo | 18 |
| 2.2 Hermenéutica | 18 |
| 2.3 Dialéctica | 19 |
| 2.4 Racionalismo crítico | 19 |
| 3. Objetividad y subjetividad del conocimiento | 19 |
| 4. Teoría y práctica | 20 |
| 5. Ciencia | 20 |
| Investigación | 21 |
| 1. Importancia | 22 |
| 2. Elementos | 22 |
| 3. Clasificación | 22 |
| 3.1 Por el propósito o finalidades perseguidas: básica o aplicada. | 22 |
| 3.2 Por la clase de medios utilizados para obtener los datos: documental, de campo o experimental | 23 |
| 3.3 Por el nivel de conocimientos que se adquieren: exploratoria, descriptiva o explicativa | 23 |
| 4. Características | 24 |
| 5. El objeto | 25 |
| 6. Formas | 25 |
| 6.1 Investigación pura | 25 |
| 6.2 Investigación aplicada | 25 |
| 7. Tipos | 26 |
| 7.1 Primer tratado | 26 |
| 7.2 Segundo Tratado | 27 |
| 8. Proceso de la investigación | 28 |
| Métodos y técnicas de investigación | 32 |
| 1. Métodos del conocimiento | 32 |
| 2. Técnicas de investigación | 35 |
| 2.1 Técnica documental | 35 |
| 2.2 Técnica de campo | 39 |
| El diseño de investigación | 42 |
| 1. Objetivos del diseño de investigación | 42 |
| 2. Características del plan de investigación | 43 |
| 3. Defectos del plan de investigación | 43 |
| 4. Modelos en diseño de investigación | 43 |
| El método científico | 45 |
| 1. Etapas del método científico | 45 |

| | |
|--|----|
| 1.1 Elección y enunciado del problema..... | 45 |
| 1.2 Estructuración del marco teórico | 47 |
| 1.3 Establecimiento de hipótesis..... | 47 |
| 1.4 Prueba de hipótesis | 49 |
| 1.5 Resultados | 51 |

| | |
|------------------------------|-----------|
| Recomendaciones | 51 |
|------------------------------|-----------|

Capítulo 2

| | |
|--|-----------|
| MODELOS DE INVESTIGACIÓN | 53 |
| Modelos de investigación | 55 |
| Contenidos del Proyecto de Investigación | 55 |
| Primera estrategia de un proyecto de investigación | 55 |
| Segunda estrategia de un proyecto de investigación | 58 |
| Tercera estrategia de un proyecto de investigación | 60 |
| Contenidos del informe final de investigación | 75 |
| Primera estrategia de investigación | 75 |
| Segunda estrategia de investigación | 77 |
| Tercera estrategia de investigación | 78 |
| Estrategia del autor..... | 78 |
| Observación | 80 |

Capítulo 3

| | |
|---|-----------|
| PLANTEAMIENTO DEL PROBLEMA | 81 |
| Problema..... | 83 |
| 1. Idea..... | 83 |
| 2. Objetivos | 83 |
| 3. Hipótesis | 84 |
| 4. Seleccionar un tema de estudio limitado | 84 |
| Diagrama del Problema..... | 84 |
| 1. Ejercicios resueltos sobre problemas de investigación..... | 84 |
| Otro diagrama de formulación de hipótesis y objetivos | 88 |
| Lluvia de ideas..... | 89 |
| 1. Ejercicios sobre Lluvias de Ideas | 89 |
| Verbos y definiciones | 90 |

Capítulo 4

| | |
|---|-----------|
| METODOLOGÍA Y DISEÑO MUESTRAL | 93 |
| Pasos del diseño muestral..... | 95 |
| Población | 95 |
| Muestra..... | 95 |
| Individuo (Observación, Caso, Sujeto) | 95 |
| Variables..... | 95 |
| Ejercicios de Población y muestra..... | 96 |
| Tamaño de muestra | 98 |
| 1. El estudio de proporciones..... | 99 |
| 2. Error muestral (E) o Error de Estimación | 99 |

| | |
|---|-----|
| 3. El nivel de confianza | 99 |
| 4. Ejemplos de nivel de confianza | 99 |
| Ejemplos de tamaño de muestra | 101 |
| Tipos de Muestreo | 103 |
| Muestreo aleatorio..... | 103 |
| 1. Muestreo aleatorio simple | 103 |
| 2. Muestreo estratificado | 104 |
| 2.1 Asignación proporcional | 104 |
| 2.2 Asignación óptima | 104 |
| 3. Muestreo sistemático | 104 |
| 4. Muestreo por conglomerados..... | 104 |
| 5. Muestreo mixto..... | 105 |
| 6. Muestreo por estadios múltiples..... | 105 |
| 7. Muestreo por cuotas | 105 |
| 8. Muestreo de "bola de nieve" | 105 |
| 9. Muestreo subjetivo por decisión razonada | 105 |
| Ejercicios de tipos de muestreo (elegir la muestra) | 106 |
| Piloto..... | 112 |

Capítulo 5

| | |
|---|------------|
| RECOLECTAR DATOS O INFORMACIÓN | 113 |
| Recolectar datos..... | 115 |
| Instrumento de medición (Cuestionario)..... | 115 |
| Soporte Electrónico de Datos | 115 |
| Cuestionario | 115 |
| 1. Diseñar un cuestionario | 115 |
| 1.1 Tamaño de un cuestionario..... | 116 |
| 1.2 Tipos de pregunta del cuestionario..... | 116 |
| 1.3 Inquietudes en torno a las preguntas que fórmula..... | 116 |
| 1.4 Recomendaciones para formular las preguntas..... | 117 |
| 1.5 Sugerencias para aplicar un cuestionario..... | 117 |
| 1.6 Desarrollar un Cuestionario | 117 |
| Observaciones..... | 118 |
| Equipos de Medición..... | 118 |
| Codificación de datos | 118 |
| 1. Codificación pregunta cerrada..... | 118 |
| 2. Codificación pregunta abierta | 118 |
| Ejercicios de recolectar datos | 119 |
| Definiciones importantes..... | 123 |
| 1. Instrumentos de Medición | 123 |
| 2. Confiabilidad de los instrumentos de medición | 123 |
| 3. Construir un instrumento de medición | 123 |
| 4. Tipos de instrumentos de Medición | 124 |
| 4.1 Encuesta..... | 124 |
| 4.2 Entrevista | 124 |
| 4.3 Pruebas | 127 |

| | |
|-----------------------------------|-----|
| 4.4 Cuestionarios | 130 |
| 4.5 Observación | 130 |
| 5. Escala de medición | 131 |
| 5.1 Construcción de Escala | 132 |
| 5.2 Construcción de Índice..... | 135 |
| 5.3 Ejemplos de Indicadores | 136 |
| 6. Equipos de medición | 139 |
| 7. Codificación | 139 |
| Ejemplos de cuestionarios | 140 |

Capítulo 6

| | |
|--|------------|
| PROCESAMIENTO DE DATOS O INFORMACIÓN..... | 147 |
| Proceso de Datos | 149 |
| Introducción directa..... | 149 |
| 1. Vista de datos | 149 |
| 2. Vista de Variable (definir variables | 150 |
| 3. Introducir datos | 155 |
| 4. Uso del Editor de datos y variables | 156 |
| Extrayendo un archivo de disco | 156 |
| Ejercicio 1: Manipulación de datos | 157 |
| Ejercicio 2: Manipulación de datos | 161 |

Capítulo 7

| | |
|--|------------|
| ESTADÍSTICAS DESCRIPTIVAS | 165 |
| ESTADÍSTICAS DESCRIPTIVAS | 167 |
| Distribución de Frecuencias | 167 |
| Ejemplo 1..... | 168 |
| Ejercicios de distribución frecuencias | 169 |
| Representaciones gráficas..... | 171 |
| Diagrama de tallos y hojas | 172 |
| Medidas de posición y de tendencia central..... | 174 |
| 1. Medidas de posición central (media, mediana, moda) | 174 |
| Ejercicio de medidas de tendencia | 176 |
| 2. Medidas de tendencia no central (percentiles, cuartiles)..... | 178 |
| Ejercicio de medidas de tendencia no central | 179 |
| 3. Medidas de variación | 180 |
| Ejercicio medidas de variación..... | 181 |
| 4. Gráficos de cajas y bigotes..... | 183 |
| 5. Asimetría y Curtosis | 183 |
| 6. Tablas de Contingencia | 184 |
| Ejercicio 1..... | 187 |
| Estadísticas Descriptivas | 188 |
| Frecuencias..... | 189 |
| P-P Plots..... | 190 |
| Q-Q Plots | 191 |

Capítulo 8

| | |
|--|------------|
| CORRELACIÓN | 193 |
| Correlación..... | 195 |
| Coefficiente de correlación lineal..... | 195 |
| Bivariadas..... | 198 |
| La matriz de varianzas-covarianzas | 199 |
| La matriz de correlaciones | 199 |
| Combinación lineal de variables | 201 |
| Ejercicio 1..... | 202 |
| Correlaciones..... | 203 |
| Correlación en gráficos..... | 205 |

Capítulo 9

| | |
|---|------------|
| REGRESIÓN | 207 |
| Regresión Lineal Simple | 209 |
| Análisis de regresión lineal simple | 212 |
| Diagrama de dispersión | 212 |
| Ecuación de regresión | 214 |
| Bondad de ajuste del modelo | 217 |
| Análisis de Regresión Lineal Múltiple | 221 |
| Ejercicio 1..... | 223 |

Capítulo 10

| | |
|---|------------|
| PRUEBA DE HIPÓTESIS | 227 |
| Prueba de hipótesis (Comparar medias)..... | 229 |
| Medias..... | 229 |
| Prueba T | 232 |
| Prueba T para una muestra | 232 |
| Prueba T para muestras independientes | 235 |
| Prueba T para muestras relacionadas | 237 |
| ANOVA de un factor | 239 |
| ANOVA A – EF – CA..... | 239 |
| Comparaciones múltiples entre medias..... | 241 |
| Comparaciones a posteriori | 243 |

Capítulo 11

| | |
|--|------------|
| MODELOS LINEAL GENERAL - ANOVA | 245 |
| Modelo lineal general | 247 |
| Análisis de varianza de dos factores..... | 247 |
| ANOVA AB – EF – CA | 247 |
| Medidas Repetidas..... | 252 |
| Análisis de varianza de un factor con medidas repetidas..... | 252 |
| ANOVA A- EF- MR..... | 252 |

Capítulo 12

| | |
|--|------------|
| PRUEBAS NO PARAMÉTRICAS | 255 |
| Pruebas no paramétricas | 257 |
| Contrastes no paramétricos | 257 |
| Pruebas de los signos (binomial) | 258 |
| Contrastes para dos muestras | 260 |
| Dos muestras independientes, prueba de Mann-Whitney | 260 |
| Dos muestras relacionadas, prueba de Wilcoxon..... | 261 |
| Contrastes para más de dos muestras..... | 263 |
| Más de dos muestras independientes, prueba de Kruskal Wallis | 263 |
| Más de dos muestras relacionadas, prueba de Friedman..... | 264 |
| Contrastes sobre proporciones | 265 |
| Contraste sobre una proporción..... | 265 |
| Dos proporciones relacionadas | 266 |
| Más de dos proporciones relacionadas..... | 268 |
| Dos proporciones independientes | 269 |
| Prueba X^2 de Pearson | 269 |
| X^2 sobre Bondad de ajuste..... | 269 |
| Tablas de contingencia | 272 |

Capítulo 13

| | |
|--|------------|
| CONGLOMERADOS DE K MEDIAS | 275 |
| Conglomerados de K medias..... | 277 |
| Conglomerados jerárquicos | 281 |
| Análisis Discriminante | 285 |
| Análisis factorial | 289 |

Capítulo 14

| | |
|----------------------------------|------------|
| SERIES TEMPORALES | 293 |
| Suavizado exponencial | 295 |
| Autorregresión | 295 |
| Sequence Charts | 295 |
| ARIMA | 295 |
| Descomposición estacional..... | 296 |
| Series temporales con PASW | 297 |

Capítulo 15

| | |
|---|------------|
| SUPERVIVENCIA Y RESPUESTAS MÚLTIPLES | 311 |
| Supervivencia..... | 313 |
| Respuestas múltiples | 315 |
| Tablas de Contingencia..... | 319 |

| | |
|---------------------------|------------|
| BIBLIOGRAFÍA | 323 |
|---------------------------|------------|

| | |
|-----------------------|------------|
| APÉNDICE | 325 |
|-----------------------|------------|

Capítulo

1

Metodología de la investigación científica

Centramos su atención en los conceptos y definiciones ampliamente conocidos con la finalidad que el lector interesado pueda tener a la mano los conceptos que orienta al investigador en su razonamiento y aproximación a la realidad, permitiendo ordenar sus acciones y criterios de rigor científico. Porque, investigar supone la responsabilidad de producir una lectura real del problema de investigación y demostrar la contribución efectiva.

Los aspectos metodológicos de la ciencia están en constante evolución. Por ello, el presente capítulo sólo introduce al lector en el tema.

CONOCIMIENTO Y CIENCIA

Conocimiento: Es la comprensión y entendimiento de hechos, verdades o principios.

Ciencia: Conjunto de conocimientos obtenidos mediante la observación y el razonamiento, sistemáticamente estructurados y de los que se deducen principios y leyes generales.

Ciencia pura: Estudio de los fenómenos naturales y otros aspectos del saber por sí mismo, sin tener en cuenta sus aplicaciones. *Ciencias humanas:* Las que, como la psicología, antropología, sociología, historia, filosofía, etc., se ocupan de aspectos del hombre no estudiados en las ciencias naturales.

Ciencias naturales: Las que tienen por objeto el estudio de la naturaleza, como la geología, la zoología, la botánica, etc. A veces se incluyen la física, la química, etc.

1. Origen del conocimiento

En la obtención del conocimiento intervienen tres elementos: un sujeto que quiere conocer (sujeto cognoscente), un objeto del estudio (objeto de conocimiento) y el conocimiento.

1.1. Sujeto cognoscente

Sujeto que busca saber o conocer, entra en contacto con el objeto del conocimiento a través de sus sentidos. Luego sus percepciones son procesadas mentalmente. Al razonarlas se forman imágenes (conceptos) que son abstracciones de la realidad. El proceso concluye al conformar las propias ideas con la realidad objetiva. El proceso mental es subjetivo, porque no puede verse, tocarse, olerse, oírse, saborearse.

Las corrientes que intentan explicar el origen del conocimiento son variadas. Algunos autores afirman que su origen está en los sentidos; otros, que se inicia al razonarse, y otros, que dependen del azar.

1.2. Objeto del conocimiento

En relación con la teoría del conocimiento, existen dos posturas diametralmente opuestas: el *materialismo* y el *idealismo*. Conforme a una postura materialista, el objeto de conocimiento existe independientemente de la razón o de las ideas del hombre. Según una postura idealista, el objeto de conocimiento es estructurado por las ideas del hombre, no existe sin ellas. El hecho de que tanto el sujeto como el objeto son influidos en el proceso de conocimiento, es la tesis de diversas corrientes en la teoría del conocimiento.

- *Corriente materialista:* Sostiene que la materia es infinita y eterna; el hombre es materia. Para esta corriente, el conocimiento es objetivo y da énfasis al objeto de conocimiento del cual, según se afirma, existe independientemente de que sea o no conocido por el sujeto cognoscente.
- *Corriente idealista:* Sostiene que lo primario es la conciencia cognoscente, que es el sujeto quien recrea el objeto a través de sus abstracciones. El objeto de conocimiento existe a partir de que el sujeto le da vida.

1.3. El conocimiento

Los tipos de conocimiento se reducen básicamente a dos; conocimiento empírico y conocimiento científico.

- *Conocimiento empírico*: El conocimiento empírico se desprende de la experiencia y a través de los sentidos. Permite al hombre interactuar con su ambiente; es generacional, sin un razonamiento elaborado, ni una crítica al procedimiento de obtención ni a las fuentes de información. Los conceptos empíricos son imprecisos e inciertos, se producen por ideas preconcebidas, tienden a aceptar explicaciones metafísicas y son dogmáticos. Sin embargo, el conocimiento empírico sirve de base al conocimiento científico, al extraerse con método de la realidad.
- *Conocimiento científico*: El conocimiento empírico se convierte en científico al extraerlo de la realidad con métodos y herramientas precisas. Se integra en un sistema de conceptos, teorías y leyes. El conocimiento científico rebasa los hechos empíricos. Puede generalizarse. Puede pronosticarse. El conocimiento científico resiste la confrontación con la realidad, descarta explicaciones metafísicas y utiliza fuentes de primera mano.

2. Principales corrientes en la obtención del conocimiento científico

La forma de interpretar la información de la realidad en un contexto científico deriva de cuatro posturas al respecto:

- Positivismo y neopositivismo.
- Hermenéutica.
- Dialéctica.
- Racionalismo crítico.

2.1. Positivismo y neopositivismo

Sus raíces se remontan al siglo XIX, con autores relevantes como: Bacon, Locke, Newton, Lavoisier, Saint-Simon, Comte, Russell y Nagel, etc. El principio fundamental del positivismo y el neopositivismo es: no existe conocimiento que no provenga de la percepción. La importancia se da al rigor en las técnicas para obtener el conocimiento; se insiste en la cuantificación y se afirma la unidad de la ciencia, es decir, la semejanza de todas las ciencias del hombre.

El positivismo concibe al conocimiento como separado de la acción, a la ciencia como “neutra”. El científico debe desprenderse de sus conceptos e ideas, de la influencia del ambiente.

2.2. Hermenéutica

Surge en forma difusa con autores historicistas en Alemania, particularmente Dilthey y Weber, con el desarrollo de la fenomenología. El planteamiento fundamental de la hermenéutica es opuesto al positivismo. Afirma que no existe un lenguaje observacional puro; todo lenguaje es interpretación, todo conocimiento es interpretación. Lo más importante, según esta corriente, es la teorización, el interés en datos cualitativos, el rechazo a lo cuantitativo, el manejo crítico del lenguaje y la diferencia importante entre las ciencias naturales y las sociales.

La hermenéutica y la dialéctica coinciden en su oposición al positivismo clásico. En el positivismo los científicos hacen ciencia “neutra”; en la hermenéutica, la ciencia se supedita a la práctica social de acuerdo con las determinantes ambientales, En una, es el conocimiento por el conocimiento; en otra, es el conocimiento para la acción.

2.3. Dialéctica

Los representantes más relevantes de esta corriente son Marx, Engels, Lenin y Stalin. El énfasis está en la realidad como una “totalidad”, de la que no puede aislarse ningún elemento sin que deje de tener sentido. Se rechaza el análisis cuantitativo. En la práctica, las actividades científicas no se disocian de las posiciones de los científicos, las cuales influyen de manera decisiva en el conocimiento.

Supuestos de la dialéctica:

- La realidad existe independientemente y con anterioridad al ser humano. No ha sido creada por ningún espíritu universal.
- La realidad es cognoscible.
- Los procesos y los objetos están relacionados y son interdependientes.
- Todos los conocimientos son verdades relativas.
- La realidad está en continuo cambio, movimiento y transformación.
- La realidad se presenta a diversos grados y niveles.
- La práctica es el criterio de verdad del conocimiento.

2.4. Racionalismo crítico

El racionalismo crítico tiene su origen en el decenio de 1930. Su principal exponente fue Karl L. Popper. Constituye una postura intermedia entre el positivismo y la hermenéutica clásicos. En esta postura, el conocimiento es la interpretación de las experiencias. Pone énfasis en lo cuantitativo y subraya la importancia de las técnicas para obtener datos. Describe una relación entre las ciencias naturales y las sociales.

La corriente del racionalismo crítico pretende establecer un equilibrio entre el conocimiento y la acción, y critica a la dialéctica por su “alto grado de confusión en cuanto a la teoría para la acción”. La ciencia tiene criterios propios diferentes e independientes de las condicionantes ambientales. Se acepta el valor del conocimiento empírico en la construcción del conocimiento científico.

3. Objetividad y subjetividad del conocimiento

Todo conocimiento es objetivo y subjetivo a la vez. Objetivo, porque corresponde a la realidad; subjetivo, porque está impregnado de elementos pensantes implícitos en el acto cognoscitivo. El conocimiento vincula procedimientos mentales (*subjetivos*) con actividades prácticas (*objetivas*).

Sin embargo, la ciencia busca el predominio de lo objetivo a través de explicaciones congruentes, predicciones y control de los fenómenos naturales.

Si bien el proceso mental es subjetivo, el contenido de las teorías no lo es en modo alguno. La verdad se presenta con teorías, la realidad se *aprehende* con herramientas teórico metodológicas.

A medida que se profundiza en el conocimiento, aparecen nuevos elementos originados por las contradicciones internas. Todo conocimiento es aproximado y relativo, de tal modo que nunca se considera acabado (*verdad absoluta*), ya que éste es condicionado por la realidad histórica.

4. Teoría y Práctica

El término *teoría* deriva del griego *theoria*, que significa observación, investigación. En las reflexiones filosóficas se le da otra acepción; se le explica como un saber generalizado.

Una de las definiciones más difundidas es la de Popper: las teorías científicas son enunciados universales, consisten en apresar aquello que llamamos el mundo, para racionalizarlo, explicarlo y dominarlo.

Kerlinger afirma: la teoría es un conjunto de proposiciones hipotéticas, conceptos relacionados entre sí, que ofrecen un punto de vista sistemático de los fenómenos, al explicar relaciones existentes entre variables, con el objeto de explicar y predecir los fenómenos.

La teoría es finalmente la meta de la ciencia; todo lo demás se deduce de ella. Los tres momentos en el proceso del conocimiento son: descripción, explicación y predicción.

Los principales aspectos que se deberán considerar sobre la teoría son:

- La teoría es un saber generalizado.
- Se forma con conceptos, categorías y leyes.
- La teoría es el reflejo de la realidad objetiva.
- La teoría y la práctica forman un todo y son insolubles.

Marx afirma que la teoría, por sí misma, no puede transformar la realidad. Por tanto, es necesario aplicar el conocimiento en la práctica. Esa concepción fue superada al afirmar que la teoría adquiere su dimensión propia al margen de la práctica. La teoría crítica es un intento de rescatar los elementos fundamentales de la propuesta de Marx sobre la unión de la teoría con la práctica.

5. Ciencia

La ciencia es el quehacer humano que consiste en la actitud de observar y experimentar dentro de un orden particular de conocimientos, los cuales se organizan de manera sistemática mediante determinados métodos, partiendo de un núcleo de conceptos o principios básicos, a fin de alcanzar un saber de validez universal.

La ciencia es un conjunto de conocimientos en desarrollo, que se obtienen mediante los métodos cognoscitivos y que se reflejan en conceptos exactos cuya verdad se comprueba y demuestra a través de la práctica social.

Un tipo de clasificación distingue entre: *ciencias empíricas* y *ciencias formales*.

- **Ciencias empíricas:** Estas concentran su atención en la realidad material, objetiva, tangible. Son ejemplos de ellas las ciencias naturales y las ciencias sociales.
- **Ciencias formales:** Las ciencias no empíricas o formales concentran su interés en aspectos simbólicos, derivados del pensamiento del hombre. Son ejemplos las matemáticas y la filosofía, entre muchas otras.

Características de las ciencias puras y aplicadas

El sistema de conocimientos teóricos representado únicamente por ideas, es conocido como ciencia formal o ciencia pura. El sistema de conocimientos teóricos relativo a hechos de la práctica social es conocido como ciencia factual (fáctica) o ciencia aplicada. Algunas de las características que diferencian a dichos sistemas se resumen en el cuadro:

| Ciencia formal | Ciencia factual |
|--|---|
| Objeto de estudio: Ideas | Objeto de estudio: Hechos |
| Representación: Símbolos, signos | Representación: Palabras |
| Método de análisis: Inducción, deducción, lógica | Método del análisis: Método científico |
| Comprobación: Razonamiento | Comprobación: En la práctica |
| Tipos: Lógica, matemática | Tipos: Física, psicología, administración, historia, etc. |

La teoría, el método y las técnicas, forman la ciencia. Es posible resumir que la ciencia permite conocer el mundo y colocar al hombre frente a la realidad.

INVESTIGACIÓN

Es la actividad de búsqueda que se caracteriza por ser reflexiva, sistemática y metódica; tiene por finalidad obtener conocimientos y solucionar problemas científicos, filosóficos o empírico-técnicos, y se desarrolla mediante un proceso. La investigación científica es la búsqueda intencionada de conocimientos o de soluciones a problemas de carácter científico; el método científico indica el camino que se ha de transitar en esa indagación y las técnicas precisan la manera de recorrerlo.

Investigar supone aplicar la inteligencia a la exacta comprensión de la realidad objetiva, a fin de dominarla. Sólo al captar la esencia de las cosas, al confrontarla con la realidad, se cumple la labor del investigador. La consecuencia de tal proceso incrementará los conocimientos científicos.

Investigación en el proceso del conocimiento:

Para obtener el conocimiento se parte de una etapa sensorial; es decir, captar por medio de los sentidos los fenómenos, procesos y objetos de la naturaleza. Una segunda etapa supone las conceptualizaciones, el conocimiento racional del mundo exterior. Por último, para calificar como verdaderos los juicios teóricos, éstos deben comprobarse en la práctica.

El enfrentamiento con la realidad se realiza con base en métodos e instrumentos debidamente seleccionados. La investigación es esencial en el proceso del conocimiento, porque no basta con percibir. Es necesario comprender y explicar, para poder predecir. Al investigar se parte del supuesto de que no puede haber conocimiento al margen de la práctica.

Investigación Científica

La investigación es un proceso que, mediante la aplicación del método científico, procura obtener información relevante y fidedigna, con el fin de entender, verificar, corregir o aplicar el conocimiento.

Para obtener un resultado de manera clara y precisa es necesario aplicar algún tipo de investigación, que está muy ligada a los seres humanos; esta posee una serie de pasos para lograr el objetivo planteado o para llegar a la información solicitada.

La investigación tiene como base el método científico y este es el método de estudio sistemático de la naturaleza que incluye las técnicas de observación, reglas para el razonamiento y la predicción, ideas sobre la experimentación planificada y los modos de comunicar los resultados experimentales y teóricos. Ayuda a mejorar el estudio porque nos permite establecer contacto con la realidad a fin de que la conozcamos mejor; la finalidad de esta radica en formular nuevas teorías o modificar las existentes, en incrementar los conocimientos.

La actividad investigadora se conduce eficazmente mediante una serie de elementos que hacen accesible el objeto al conocimiento y de cuya sabia elección y aplicación va a depender en gran medida el éxito del trabajo del investigador.

1. Importancia

La investigación nos ayuda a mejorar el estudio porque nos permite establecer contacto con la realidad a fin de que la conozcamos mejor. Constituye un estímulo para la actividad intelectual creadora. Ayuda a desarrollar una curiosidad creciente acerca de la solución de problemas. Además, contribuye al progreso de la lectura crítica y solución de problemas, mejora el estudio porque nos permite establecer contacto con la realidad a fin de que la conozcamos mejor, constituye un estímulo para la actividad intelectual creadora, ayuda a desarrollar una curiosidad creciente acerca de la solución de problemas.

2. Elementos

Desde un punto de vista estructural reconocemos cuatro elementos presentes en toda investigación: sujeto, objeto, medio y fin. Se entiende por:

Sujeto.- El que desarrolla la actividad, el investigador.

Objeto.- Lo que se indaga, esto es, la materia o el tema.

Medio.- Lo que se requiere para llevar a cabo la actividad, es decir, el conjunto de métodos y técnicas adecuados.

Fin.- Lo que se persigue, los propósitos de la actividad de búsqueda que radican en la solución de una problemática detectada.

3. Clasificación

Es conveniente señalar que en la realidad la investigación no se puede clasificar exclusivamente en alguno de los tipos que se mencionan, sino que generalmente en toda investigación se persigue un propósito señalado, se busca un determinado nivel de conocimiento y se basa en una estrategia particular o combinada.

3.1. Por el propósito o finalidades perseguidas: básica o aplicada

Investigación básica.- También recibe el nombre de investigación pura, teórica o dogmática. Se caracteriza porque parte de un marco teórico y permanece en él; la finalidad radica en formular nuevas teorías o modificar las existentes, en incrementar los conocimientos científicos o filosóficos, pero sin contrastarlos con ningún aspecto práctico.

Investigación aplicada.- También recibe el nombre de práctica o empírica. Se caracteriza porque busca la aplicación o utilización de los conocimientos que se adquieren. La investigación aplicada se encuentra estrechamente vinculada con la investigación básica, pues depende de los resultados y avances de esta última; esto queda aclarado si nos percatamos de que toda investigación aplicada requiere de un marco teórico. Sin embargo, en una investigación empírica, lo que le interesa al investigador, primordialmente, son las consecuencias prácticas. Si una investigación involucra problemas tanto teóricos como prácticos, recibe el nombre de mixta.

3.2. Por la clase de medios utilizados para obtener los datos: documental, de campo o experimental

Investigación documental.- Es la que se realiza, como su nombre lo indica, apoyándose en fuentes de carácter documental, esto es, en documentos de cualquier especie. Como subtipos de esta investigación encontramos la investigación bibliográfica, la hemerográfica y la archivística; la primera se basa en la consulta de libros, la segunda en artículos o ensayos de revistas y periódicos, y la tercera en documentos que se encuentran en los archivos como cartas, oficios, circulares, expedientes, etc.

Investigación de campo.- Este tipo de investigación se apoya en informaciones que provienen entre otras, de entrevistas, cuestionarios, encuestas y observaciones. Como es compatible desarrollar este tipo de investigación junto a la investigación de carácter documental, se recomienda que primero se consulten las fuentes de la de carácter documental, a fin de evitar una duplicidad de trabajos.

Investigación experimental: Recibe este nombre la investigación que obtiene su información de la actividad intencional realizada por el investigador y que se encuentra dirigida a modificar la realidad con el propósito de crear el fenómeno mismo que se indaga y así poder observarlo.

3.3. Por el nivel de conocimientos que se adquieren: exploratoria, descriptiva o explicativa

Investigación exploratoria.- Es aquella que se realiza con el propósito de destacar los aspectos fundamentales de una problemática determinada y encontrar los procedimientos adecuados para elaborar una investigación posterior. Es útil desarrollar este tipo de investigación porque, al contar con sus resultados, se simplifica abrir líneas de investigación y proceder a su consecuente comprobación.

Investigación descriptiva.- Mediante este tipo de investigación, que utiliza el método de análisis, se logra caracterizar un objeto de estudio o una situación concreta, señalar sus características y propiedades. Combinada con ciertos criterios de clasificación sirve para ordenar, agrupar o sistematizar los objetos involucrados en el trabajo indagatorio. Al igual que la investigación que hemos descrito anteriormente, puede servir de base para investigaciones que requieran un mayor nivel de profundidad.

Investigación explicativa.- Mediante este tipo de investigación, que requiere la combinación de los métodos analítico y sintético, en conjugación con el deductivo y el inductivo, se trata de responder o dar cuenta del porqué del objeto que se investiga.

4. Características

La investigación recoge conocimientos o datos de fuentes primarias y los sistematiza para el logro de nuevos conocimientos. No es investigación confirmar o recopilar lo que ya es conocido o ha sido escrito o investigado por otros. La característica fundamental de la investigación es el descubrimiento de principios generales. El investigador parte de resultados anteriores, planteamientos, proposiciones o respuestas en torno al problema que le ocupa.

Para ello debe:

- Planear cuidadosamente una metodología.
- Recoger, registrar y analizar los datos obtenidos.
- De no existir estos instrumentos, debe crearlos.

La investigación debe ser objetiva, es decir, eliminar en el investigador preferencias y sentimientos personales, así como resistir a buscar únicamente aquellos datos que le confirmen su hipótesis; de ahí que emplea todas las pruebas posibles para el control crítico de los datos recogidos y los procedimientos empleados.

Finalmente, una vez sistematizados los datos son registrados y expresados mediante un informe o documento de investigación, en el cual se indican la metodología utilizada y los procedimientos empleados para llegar a las conclusiones presentadas, las cuales se sustentan por la misma investigación realizada.

En la investigación deben darse una serie de características para que sea en realidad científica:

- a. Estar planificada, es decir, tener una previa organización, establecimiento de objetivos, formas de recolección y elaboración de datos y de realización de informe.
- b. Contar con los instrumentos de recolección de datos que respondan a los criterios de validez, confiabilidad y discriminación, como mínimos requisitos para lograr un informe científicamente válido.
- c. Ser original, esto es, apuntar a un conocimiento que no se posee o que este en duda y sea necesario verificar y no a una repetición reorganización de conocimientos que ya posean.
- d. Ser objetiva, vale decir que el investigador debe tratar de eliminar las preferencias personales y los sentimientos que podrían desempeñar o enmascarar el resultado del trabajo de investigación.
- e. Disponer de tiempo necesario a los efectos de no apresurar una información que no responda, objetivamente, al análisis de los datos que se dispone.
- f. Apuntar a medidas numéricas en el informe, tratando de transformar los resultados en datos cuantitativos más fácilmente representables y comprensibles y más objetivos en la valoración final.
- g. Ofrecer resultados comprobables y verificarles en las mismas circunstancias en las se realizó la investigación.
- h. Apuntar a principios generales trascendiendo los grupos o situaciones particulares investigados, para los que se requiere una técnica de muestreo con el necesario rigor científico, tanto en el método de selección como en la cantidad de la muestra, en relación con la población de que se trate.

5. El objeto

El objeto de la investigación científica es aquello a lo que se aplica el pensamiento. Cuando se trata de obtener nuevo conocimiento científico el objeto se erige en fortaleza que hay que conquistar con métodos que aseguren la garantía de obtención de una verdad contrastable por toda la comunidad científica. Este objeto de conocimiento científico, de investigación, está constituido por los vestigios que plantean un conjunto de problemas epistemológicos en el tema de la investigación cuya resolución se persigue.

El investigador debe tener conciencia asumida de que el objeto de conocimiento se le opone por naturaleza, no se deja conocer fácilmente, plantea numerosas dificultades. La investigación es, en consecuencia, ejercicio intelectual dificultoso, lleno de obstáculos y, en consecuencia, factor formativo para la persona que lo ejerce.

Sin embargo, la actividad investigadora se conduce eficazmente mediante una serie de elementos que hacen accesible el objeto al conocimiento y de cuya sabia elección y aplicación va a depender en gran medida el éxito del trabajo a investigador. Sin perjuicio de que estos elementos deban especializarse en la propia naturaleza de los problemas concretos objeto de resolución por parte del investigador.

Un ambiente favorable puede estimular al investigador en los momentos de desánimo: es precisamente el clima científico, el ambiente de trabajo en facultades, departamentos y centros oficiales de investigación lo que, con más frecuencia, suple con creces otras carencias.

6. Formas

La Investigación Científica posee dos formas: pura y aplicada.

6.1. Investigación pura

Se le da también el nombre de básica o fundamental. Se apoya dentro de un contexto teórico y su propósito fundamental es el de desarrollar teoría mediante el descubrimiento de amplias generalizaciones o principios. Esta forma de investigación emplea cuidadosamente el procedimiento de muestreo, a fin de extender sus hallazgos más allá del grupo o situaciones estudiadas.

Poco se preocupa de la aplicación de los hallazgos, por considerar que ello corresponde a otra persona y no al investigador. No obstante la carencia de aplicación inmediata, esta forma de investigación busca el progreso científico y su importancia reside en que presente amplias generalizaciones y niveles de abstracciones con miras a formulaciones hipotéticas de posible aplicación posterior.

Persigue igualmente el desarrollo de teorías basada en principios y leyes. La investigación fundamental es un proceso formal y sistemático de coordinar el método científico de análisis y generalización con las fases deductivas e inductivas del razonamiento. Pardinas nos dice que la investigación pura “tiene como objeto el estudio de un problema destinado exclusivamente al progreso o a la simple búsqueda del conocimiento”.

6.2. Investigación aplicada

Se le denomina también activa o dinámica y se encuentra íntimamente ligada a la anterior ya que depende de sus descubrimientos y aportes teóricos. Busca confrontar la teoría con la realidad. Es el estudio y aplicación de la investigación a problemas concretos, en circunstancias

y características concretas. Esta forma de investigación se dirige a su aplicación inmediata y no al desarrollo de teorías.

La investigación aplicada, movida por el espíritu de la investigación fundamental, ha enfocado la atención sobre la solución de teorías. Conciene a un grupo particular más bien que a todos en general. Se refiere a resultados inmediatos y se halla interesada en el perfeccionamiento de los individuos implicados en el proceso de la investigación.

7. Tipos

Existen muy diversos tratados sobre las tipologías de la investigación. Las controversias para aceptar las diferentes tipologías sugieren situaciones confusas en estilos, formas, enfoques y modalidades.

En rigor, y desde un punto de vista semántico, los tipos son sistemas definidos para obtener el conocimiento.

Se presentan dos tratados de los tipos, mostrados por diferentes autores.

7.1. Primer tratado

Conviene anotar que los tipos de investigación difícilmente se presentan puros; generalmente se combinan entre sí y obedecen sistemáticamente a la aplicación de la investigación.

Tradicionalmente se presentan tres tipos de investigación.

| | |
|--------------------|-----------------------|
| Histórica | Describe lo que era. |
| Descriptiva | Interpreta lo que es. |
| Experimental | Describe lo que será. |

Histórica.- Trata de la experiencia pasada; se aplica no sólo a la historia sino también a las ciencias de la naturaleza, al derecho, la medicina o a cualquier otra disciplina científica. En la actualidad, la investigación histórica se presenta como una búsqueda crítica de la verdad que sustenta los acontecimientos de pasado, este tipo de investigación tiene las siguientes etapas:

- Formas y Tipos de Investigación.
- Enunciación del Problema.
- Recolección de información.
- Crítica de Datos y Fuentes.
- Formulación de Hipótesis.
- Interpretación e Informe.

Descriptiva.- Comprende la descripción, registro, análisis e interpretación de la naturaleza actual, y la composición o procesos de los fenómenos. El enfoque se hace sobre conclusiones dominantes o sobre cómo una persona, grupo o cosa se conduce o funciona en el presente. La investigación descriptiva trabaja sobre realidades de hechos y su característica fundamental es la de presentarnos una interpretación correcta. Este tipo de investigación tiene las siguientes etapas:

- Descripción del Problema.
- Definición y Formulación de la Hipótesis.
- Supuestos en que se basa la Hipótesis.

- Marco Teórico.
- Selección de Técnicas de Recolección de Datos.
- Categorías de Datos, a fin de facilitar relaciones.
- Verificación de validez del instrumento.
- Descripción, Análisis e Interpretación de Datos.

Experimental.- Se presenta mediante la manipulación de una variable experimental no comprobada, en condiciones rigurosamente controladas, con el fin de describir el modo o por qué causa se produce una situación o acontecimiento particular. El experimento es una situación provocada por el investigador para introducir determinadas variables de estudio manipulada por él, para controlar el aumento o disminución de esas variables y su efecto en las conductas observadas. El investigador maneja de manera deliberada la variable experimental y luego observa lo que ocurre en condiciones controladas. Presenta las siguientes etapas:

- Presencia del Problema, del cual se ha realizado una revisión bibliográfica.
- Identificación y Definición del Problema.
- Definición de Hipótesis y variables. Y la operatividad de las mismas.
- Diseño del plan experimental.
- Prueba de confiabilidad de datos.
- Realización de experimento.
- Tratamiento de datos. Aquí, en este punto, hay que tener en cuenta que una cosa es el dato bruto, otro el dato procesado y otro el dato que hay que dar como definitivo.

7.2. Segundo tratado

Cuando se va a resolver un problema en forma científica, es muy conveniente tener un conocimiento detallado de los posibles tipos de investigación que se pueden seguir.

Según la fuente de información:

- Investigación documental.
- Investigación de campo.

Según la extensión del estudio:

- Investigación censal.
- Investigación de caso.
- Encuesta.

Según las variables:

- Experimental.
- Casi experimental.
- Simple y compleja.

Según el nivel de medición y análisis de la información:

- Investigación cuantitativa.
- Investigación cualitativa.
- Investigación cuali-cuantitativa.
- Investigación descriptiva.
- Investigación explicativa.

- Investigación inferencial.
- Investigación predictiva.

Según las técnicas de obtención de datos:

- Investigación de alta y baja estructuración.
- Investigación participante.
- Investigación participativa.
- Investigación proyectiva.
- Investigación de alta o baja interferencia.

Según su ubicación temporal:

- Investigación histórica.
- Investigación longitudinal o transversal.
- Investigación dinámica o estática.

Según el objeto de estudio:

- Investigación pura.
- Investigación aplicada.

8. Proceso de la investigación

La investigación es un proceso sistemático y controlado de constante exploración y descubrimiento, fundamentado en el método científico y, comúnmente, dirigido hacia la solución de un problema o contestar una incógnita. La investigación tiene un proceso muy riguroso que contiene los siguientes pasos:

- Elección del tema.
- Objetivos.
- Delimitación del tema.
- Planteamiento del problema.
- Marco teórico.
- Metodología.
- Informe.

8.1 Elección del Tema

A nivel de procesos investigativos no debe suponerse conocido el tema y arrancar con el problema, lo importante es elegir el tema ya que el problema se deriva de éste. Cuando se selecciona el tema de mueve un marco de generalidades, cuando se selecciona el problema se reduce la misma. A partir de la realidad surge una problemática, la cual está integrada por una serie de factores. La realidad de la investigación es problemática, de la cual debe elegirse un factor que se determine como tema de investigación y dentro del cual debe seleccionarse un problema investigable.

Si se comienza por la selección del problema se pierde de vista la ubicación contextual del tema. La elección del tema es el primer caso en la realización de una investigación. Consiste en determinar con claridad y precisión el contenido del trabajo a presentar. La elección del tema corresponde necesariamente al alumno investigador, que presentará por escrito a la persona indicada por la universidad, o centro docente, para su aceptación.

8.2 Objetivos

Cuando se ha seleccionado el tema de la investigación debe procederse a formular los objetivos de investigación, que deben estar armonizados con los del investigador y los de la investigación.

El objetivo de la investigación es el enunciado claro y preciso de las metas que se persiguen. El objetivo del investigador es llegar a tomar decisiones y a una teoría que le permita generalizar y resolver en la misma forma problemas semejantes en el futuro.

Los métodos que se elijan deben ser los más apropiados para el logro de los objetivos. Todo trabajo de investigación es evaluado por el logro de los objetivos mediante un proceso sistemático, los cuales deben haber sido previamente señalados y seleccionados al comienzo de la investigación.

La sistematización hace posible el planeamiento de estrategias válidas para el logro de objetivos. Por esta razón los objetivos tienen que ser revisados en cada una de las etapas del proceso; el no hacerlo puede ocasionar fallas en la investigación con la misma intensidad en que se presentan fallas en los objetivos.

La evaluación de la investigación se realiza en base a los objetivos propuestos y pueden ser sumativa, es decir progresiva; esto lleva a clasificar los distintos niveles de resultados que se quieren lograr en la investigación. Si la investigación es planeada científicamente, debe tener validez en cada una de sus etapas en razón de objetivos y el logro de este en cada etapa es lo que permite pasar a la siguiente.

Al finalizar la investigación, los objetivos han de ser identificables con los resultados; es decir, toda la investigación deberá estar respondiendo a los objetivos propuestos. Los objetivos generales dan origen a objetivos específicos que indica lo que se pretende realizar en cada una de las etapas de la investigación.

Estos objetivos deben ser evaluados en cada paso para conocer los distintos niveles de resultados. La suma de los objetivos específicos es igual al objetivo general y por tanto a los resultados esperados de la investigación. Conviene anotar que son los objetivos específicos los que se investigan y no el objetivo general, ya que este se logra de los resultados.

8.3 Delimitación del tema

Delimitar el tema es ver la viabilidad para su desarrollo. Unida a esta delimitación es necesaria la justificación del mismo; es decir, indicar las características que llevan el investigador a escoger el tema para desarrollarlo, las cuales deben ser de orden externo u objetivo, y de orden interno o subjetivo.

Una de las fallas más comunes en la investigación consiste en la ausencia de delimitación del tema; el 80% de las investigaciones fracasan por carecer de delimitación del tema, es decir, por ambición del tema. Delimitar el tema quiere decir poner límite a la investigación y especificar el alcance de esos límites. Es preferible señalar, de acuerdo a las propias inclinaciones y preferencias, un tema reducido en extensión. No importa que el tema constituya un grano de arena dentro de la ciencia.

Además, por reducido ilimitado que puede parecer un tema, si se explora convenientemente pronto surge toda una serie de ramificaciones que le dan importancia y valor. Al delimitar, se aclara si el tema de investigación será de tipo exploratoria, descriptivo o experimental. La aclaración sobre el tipo de estudio permite tener una visión general sobre la validez y el grado de confianza que puede tener como resultado. Esto supone determinar el alcance y los límites del tema.

8.4 El Problema

El problema es el punto de partida de la investigación. Surge cuando el investigador encuentra una laguna teórica dentro de un conjunto de datos conocidos, o un hecho no abarcado por una teoría, un tropiezo o un acontecimiento que no encaja dentro de las expectativas en su campo de estudio. Todo problema aparece a raíz de una dificultad, la cual se origina a partir de una necesidad en la cual aparece dificultades sin resolver.

Diariamente se presentan situaciones de muy diverso orden. Una situación determinada puede presentarse como una dificultad la cual requiere una solución mayor o menor plazo. El título del problema es la presentación racional de lo que se va a investigar, precede al plan de la investigación y debe presentar una idea clara y precisa del problema, es decir, en forma rápida y sintética nos presenta el problema a tratar y debe realizarse con el siguiente criterio, a mayor extensión menor comprensión y viceversa.

Por tal razón, si el título es muy largo conviene reducirlo a pocas palabras y clarificarlo con un subtítulo. Decíamos que todo problema aparece a raíz de una dificultad; ésta se origina a partir de una necesidad en la cual aparecen dificultades sin resolver. De ahí, la necesidad de hacer un planteamiento adecuado del problema a fin de no confundir efectos secundarios del problema a investigar con la realidad del problema que se investiga.

8.5 Marco Teórico

El marco teórico nos amplía la descripción del problema. Integra la teoría con la investigación y sus relaciones mutuas. Es la teoría del problema, por lo tanto, conviene relacionar el marco teórico con el problema y no con la problemática de donde éste surge. No puede haber un marco teórico que no tenga relación con el problema.

Toda ciencia está estipulada por dos elementos básicos: la teoría y el método del trabajo. Toda investigación requiere un conocimiento presente de la teoría que explica el área de fenómenos de estudio. Todo hecho anterior a la formulación del problema que sirve para aclarar, juzgar e interpretar el problema planteado, constituye los antecedentes del problema.

Establecer los antecedentes del problema de ninguna manera es hacer un recuento histórico del problema, o presentar fuentes bibliográficas que se utilizarán, o los datos recolectados los cuales no sabemos en donde ubicar, o la descripción de las causas del problema a no ser que la investigación sea causal.

En los antecedentes trata de hacer una síntesis conceptual de las investigaciones o trabajos realizados sobre el problema formulado con el fin de determinar en enfoque metodológico de la misma investigación. El antecedente puede indicar conclusiones existentes en torno al problema planteado. La hipótesis es el eslabón necesario entre la teoría y la investigación que nos lleva al descubrimiento de nuevos hechos. Por tal, sugiere explicación a ciertos hechos y orienta la investigación a otros.

La hipótesis puede ser desarrollada desde distinto puntos de vista. Puede estar basada en una conjetura, en el resultado de otros estudios, en la posibilidad de una relación semejante entre dos o más variables representadas en un estudio. O puede estar basada en una teoría mediante la cual una suposición de proceso deductivo nos lleva a la pretensión de que si se dan ciertas condiciones se pueden obtener ciertos resultados, es decir, la relación causa-efecto.

Una hipótesis sirve de guía para la obtención de datos en función del interrogante presentado en el problema, o también para indicar la forma cómo debe ser organizado según el tipo de estudio. Todo investigador debe hacer uso de conceptos para poder organizar sus datos y percibir las relaciones que hay entre ellos. Un concepto es una abstracción obtenida de la realidad y, por tanto, su finalidad es simplificar resumiendo una serie de observaciones que se pueden clasificar bajo un mismo nombre.

Algunos conceptos están estrechamente ligados a objetos de lo que representan, por eso cuando se define se busca asegurar que las personas que lleguen a una investigación determinada conozcan perfectamente el significado con el cual se va a utilizar el término o concepto a través de toda la investigación.

8.6 Metodología

Para toda investigación es de importancia fundamental que los hechos y relaciones que establece, los resultados obtenidos o nuevos conocimientos, tengan el grado máximo de exactitud y confiabilidad. Para ello planea una metodología o procedimiento ordenado que se sigue para establecer lo significativo de los hechos y fenómenos hacia los cuales está encaminado el significado de la investigación.

Científicamente la metodología es un procedimiento general para lograr de una manera precisa el objetivo de la investigación. De ahí, que la metodología en la investigación nos presenta los métodos y técnicas para la investigación. Es necesario tener en cuenta el tipo de investigación o de estudio que se va a realizar, ya que cada uno de estos tiene una estrategia diferente para su tratamiento metodológico. Por ello, se debe indicar el tipo de investigación, si es histórica, descriptiva o experimental. Si es un estudio causal, exploratorio o productivo.

Una población está determinada por sus características definitorias, por tanto, el conjunto de elementos que posea ésta característica se denomina población o universo. Población es la y totalidad del fenómeno a estudiar en donde las unidades de población poseen una característica común, la cual se estudia y da origen a los datos de la investigación.

Una vez recopilado los datos por los instrumentos diseñados para este fin es necesario procesarlos, es decir, elaborarlos matemáticamente, ya que la cuantificación y su tratamiento estadístico nos permitirán llegar a construcciones en relación con la hipótesis planteada. El procesamiento de datos, antes dispendioso mediante métodos manuales, es hoy realizado por computadoras electrónicas las cuales han eliminado, por así decirlo, gran parte del trabajo matemático y estadístico que antes se realizaba.

8.7 El Informe

La estructura del informe de investigación es sencilla y sigue fielmente los pasos fundamentales del diseño de la investigación; en ningún momento debe ser contraria al diseño, ya que el informe debe ser la respuesta de lo planteado al diseño de la investigación. Para la presentación

del informe debe seguirse las normas de la metodología formal de presentación de trabajos cinéticos, los cuales se han considerado en diversas obras por los tratadistas de la metodología formal.

MÉTODO Y TÉCNICAS DE INVESTIGACIÓN

El vocablo *método*, proviene de las raíces: *meth*, que significa meta, y *odos*, que significa vía. Por tanto, el método es la vía para llegar a la meta. Método y metodología son dos conceptos diferentes. El método es el *procedimiento* para lograr los objetivos. Metodología es el *estudio del método*.

Conceptos de método científico:

El método científico es una abstracción de las actividades que los investigadores realizan, concentrando su atención en el proceso de adquisición del conocimiento (Efi de Gortari -1980).

El materialismo histórico es el único método certero para estudiar los fenómenos sociales. Konstantinov (1980).

El método es un camino, un orden, conectado directamente a la objetividad de lo que se desea estudiar. Las demostraciones metodológicas llevan siempre de por medio una afirmación relativa a las leyes del conocimiento humano en general (S. Iglesias - 1976).

El método científico es la estrategia de la investigación para buscar leyes (Mario Bunge - 1969).

El método científico como: La manera sistemática en que se aplica el pensamiento al investigar, y es de índole "reflexiva" (Kerlinger - 1981).

1. Métodos del conocimiento

La amplitud de criterios en las formas de investigar ha producido diferentes métodos. Algunos de los más usuales son:

- Inducción-deducción.
- Análisis-síntesis.
- Experimento.
- Explicación.
- Axiomas.
- Estructura.
- Dialéctica.
- Matemática.
- Mecanicismo.
- Funcionalismo.
- Materialismo histórico.
- Sistemas.

1.1 Inducción-deducción

La inducción consiste en ir de los casos particulares a la generalización. La deducción, en ir de lo general a lo particular. El proceso deductivo no es suficiente por sí mismo para explicar el conocimiento. Es útil principalmente para la lógica y las matemáticas, donde los conocimientos de las ciencias pueden aceptarse como verdaderos por definición.

Algo similar ocurre con la inducción, que solamente puede utilizarse cuando a partir de la validez del enunciado particular se puede demostrar el valor de verdad del enunciado general.

La combinación de ambos métodos significa la aplicación de la deducción en la elaboración de hipótesis y la aplicación de la inducción en los hallazgos. Inducción y deducción tienen mayor objetividad cuando son consideradas como probabilísticas.

1.2 Análisis y síntesis

El análisis maneja juicios. La síntesis considera los objetos como un todo. El método que emplea el análisis y la síntesis consiste en separar el objeto de estudio en dos partes y, una vez comprendida su esencia, construir un todo.

1.3 Experimentación

El método experimental ha sido uno de los que más resultados han dado. Aplica la observación de fenómenos, que en un primer momento es sensorial. Con el pensamiento abstracto se elaboran las hipótesis y se diseña el experimento con el fin de reproducir el objeto de estudio, controlando el fenómeno para probar la validez de las hipótesis.

1.4 Explicación

Consiste en elaborar modelos para explicar el porqué y el cómo del objeto de estudio. Se aplica la explicación sistemática.

1.5 Axiomas

Utiliza símbolos a los cuales asigna valor (el método axiomático). La representación simbolizada de una multiplicidad de objetos permite el análisis de los fenómenos.

1.6 Estructura

Considera como elemento de estudio la estructura de los objetos, la cual es inherente a elementos y sistemas. La estructura tiene un significado propio, independientemente de sus elementos.

1.7 Dialéctica

El método considera los objetos y los fenómenos en proceso de desarrollo. La dialéctica es la ciencia de las leyes generales del movimiento y del desarrollo de la naturaleza, de la sociedad humana y del pensamiento, la ciencia de la concatenación universal de todos los fenómenos que existen en el mundo. Y precisamente por eso es opuesta a toda metafísica".

La verdadera teoría científica fue creada por Marx y Engels... Es por su esencia una teoría revolucionaria. Por tanto, el materialismo y la dialéctica, en su unidad y vinculación.... son la teoría y el método del marxismo.

La investigación con este método lleva implícita una transformación de la realidad. El análisis y la síntesis utilizados en el proceso de abstracción deben ser guiados por las categorías de la dialéctica materialista (contradicciones, contenido y forma, causa y efecto, entre otras), de tal forma que la abstracción mental permita reproducir los procesos y los objetos en su desarrollo y transformación. Según Gortari, en el estudio del método, “éste se particulariza en tantas ramas como disciplinas científicas existen y, dentro de ellas, se especializa hasta singularizarse”.

1.8 Matematicista

El número es la sustancia de la cual se componen todas las cosas, todos los objetos. Dice Gutiérrez: “La representación geométrica lleva a elaborar construcciones mentales que identifican a objetos materiales, para de ahí hacer cálculos sobre la síntesis o reproducción de los objetos. Bajo este supuesto se logra la creación de postulados que, por medio de fórmulas matemáticas, se aplican a casos similares”.

1.9 Mecanicista

Trata de explicar los fenómenos vitales por medio de las leyes de la mecánica. Todo ello entendido como un sistema que se fundamenta en las leyes de Isaac Newton. El sujeto actúa como un espejo que refleja la realidad.

1.10 Funcionalismo

El funcionalismo tiene influencia de la metafísica en cuanto a que acepta el cambio en algunas partes del sistema, para que siga funcionando, pero rechaza el cambio o transformación de todo el sistema. El funcionalismo parte de una interpretación metodológica orientada a explicar y fundamentar el equilibrio social, lo que significa conservar las condiciones sociales establecidas.

1.11 Estructuralismo

Para obtener el conocimiento es necesario observar lo real, construir modelos y analizar la estructura. Una estructura es un sistema de transformación que implica leyes como sistema. Comprende los caracteres de totalidad, transformación y autorregulación.

1.12 Materialismo histórico

El materialismo histórico es la ciencia de las leyes generales que rigen el desarrollo de la sociedad. Da la única solución científica certera a los problemas teóricos y metodológicos de las ciencias sociales. Es el método de mayor importancia; se rige por el movimiento constante y por la transformación. Sostiene que el modo de producción determina la estructura de la sociedad.

1.13 Teoría general de los sistemas

Su objeto de estudio es el sistema, entendido como un conjunto de elementos interrelacionados en un todo. Su autor es Ludwig von Bertalanffy, quien reconoce la influencia filosófica del neopositivismo, el cual utiliza técnicas cuantitativas y procedimientos estadísticos, en especial el cálculo de probabilidades, que fueron determinantes en la formación de la teoría general de los sistemas. La determinación de conceptos en la teoría general de los sistemas no ha seguido.

Una construcción propia. Generalmente, los conceptos fundamentales de ésta son adoptados de otras ciencias, siguiendo el objetivo de propuesta sistémica de unificación de la ciencia y el análisis científico. Se han postulado múltiples enfoques que amplían la perspectiva de la teoría de sistemas, entre ellos: teoría de la comunicación, teoría de los juegos y teoría de las decisiones. En todos ellos se subraya la importancia de reconocer en un grupo social a una organización.

2. Técnicas de la investigación

La técnica es indispensable en el proceso de la investigación científica, ya que integra la estructura por medio de la cual se organiza la investigación.

Pretende los siguientes objetivos:

- Ordenar las etapas de la investigación.
- Aportar instrumentos para manejar la información.
- Llevar un control de los datos.
- Orientar la obtención de conocimientos.

En cuanto a las técnicas de investigación, se estudiarán dos formas generales: técnica documental y técnica de campo.

- *La técnica documental*: Permite la recopilación de información para enunciar las teorías que sustentan el estudio de los fenómenos y procesos. Incluye el uso de instrumentos definidos según la fuente documental a que hacen referencia.
- *La técnica de campo*: Permite la observación en contacto directo con el objeto de estudio y el acopio de testimonios que permitan confrontar la teoría con la práctica en la búsqueda de la verdad objetiva.

2.1. Técnica documental

El objetivo de la investigación documental es elaborar un marco teórico conceptual para formar un cuerpo de ideas sobre el objeto de estudio. Con el propósito de elegir los instrumentos para la recopilación de información es conveniente referirse a las fuentes de información.

a) Fuentes primarias de información:

Estas fuentes son los documentos que registran o corroboran el conocimiento inmediato de la investigación. Incluyen, libros, revistas, informes técnicos y tesis.

Libros: De acuerdo con la UNESCO (1964), se llama libro a aquella publicación que tiene más de 49 páginas, y folleto a la que tiene entre cinco y 48 páginas. Según el tipo de usuarios los libros se clasifican en:

- De tipo general. Expuestos en forma elemental.
- De texto. Para el estudio de alguna disciplina.
- Especializados. Para profesionales o investigadores.

Monografías: Son documentos en los cuales un asunto se trata exhaustivamente. El verdadero punto de partida de una monografía no es la necesidad de cumplir con una exigencia impuesta por las normas universitarias, sino el verdadero entusiasmo por un tema preferido. La elaboración de una monografía, nombre con el que se conoce al primer intento

de escribir un artículo científico, requiere de conocer las técnicas de lectura. No puede ser ella misma base de una futura tesis.

Revistas: Son documentos de información reciente, por lo general publicaciones especializadas. Pueden ser profesionales, técnicas y científicas.

Informes técnicos: En este tipo de informes se incluyen las memorias de conferencias, los informes de congresos, reuniones y otros. Son importantes debido a que la información generalmente versa sobre asuntos de actualidad que afectan a una comunidad particular o son de interés grupal.

Diarios y periódicos: Son fuente de información en cuanto a hechos ocurridos en espacio y tiempo, pasado y presente.

Tesis: La tesis es un documento de tipo académico, que exige una afirmación original acerca de un tema de estudio particular. Se espera una aportación y conocimientos nuevos, presentados como un sistema sólido de pruebas y conclusiones. En la tesis se describe el procedimiento empleado en la investigación, hallazgos y conclusiones.

b) Fuentes secundarias de información

Incluye las enciclopedias, los anuarios, manuales, almanaques, las bibliografías y los índices, entre otros; los datos que integran las fuentes secundarias se basan en documentos primarios.

Depósitos de información: El primer paso para recabar datos es acudir a los centros de información, como biblioteca, hemeroteca, archivos y museos.

Organización bibliotecaria: En la búsqueda de información en la biblioteca es importante tener algunas nociones sobre su organización y funcionamiento, tales como:

- Los libros están ordenados de manera alfabética.
- Se clasifican por fichas.
- Se marca en los lomos del libro el código por fechas.
- Se forman catálogos.
- Se elaboran registros para localizar materiales afines a la obra.
- Se da una clasificación. En general se utiliza la de Dewey, quien divide en:

| 10 clases distintas el acervo cultural, de la siguiente forma: | |
|--|--------------------|
| 000 Obras generales | 500 Ciencias puras |
| 100 Obras filosóficas | 600 Tecnología |
| 200 Obras religiosas | 700 Las artes |
| 300 Obras sociales | 800 Literatura |
| 400 Lingüística | 900 Historia |

c) Instrumentos para investigación documental

Ficha bibliográfica: Por lo común es una tarjeta de 14 x 8 cm. En ella se anotan los datos correspondientes a la obra y el autor, preferentemente con base en un código internacional. Con el *objetivo* de ubicar, registrar y localizar la fuente de información. Con el siguiente *orden de los datos*:

- Nombre del autor, por apellidos paterno, materno y nombre; el apellido paterno se escribe en mayúsculas: por ejemplo: QUEZADA Lucio, Nel. Cuando la obra está escrita por varios autores, se cita al primero y se agrega *et al.* (o bien y col.), que significa y otros (o y *colaboradores*).
- Título del libro, subrayado.
- Serie o colección, entre paréntesis, así como volumen.
- Si se trata de una obra traducida, el nombre del traductor.
- Editorial.
- País en el que fue impreso.
- Año de publicación.
- Número de edición, medio renglón arriba del año de publicación (lo que en tipografía se llama un número “voladito”). Si es primera edición, se omite el dato.
- Número total de páginas del libro.
- En el ángulo superior izquierdo se pueden incluir datos sobre el tema o área a que corresponde la obra.

Ficha hemerográfica: Esta ficha contiene cuatro tipos de datos.

- Título de la publicación.
- Nombre del director.
- Lugar de edición.
- Periodicidad.

Ficha de trabajo: Tiene relevancia especial en la tarea de investigación. Su construcción debe ser creativa. Es el fruto de la reflexión, el análisis, la síntesis y la crítica. Por lo general, se presentan en tarjetas de 21 x 13 cm. Para una mejor conservación se pueden enmarcar y también hacer en hojas de papel con las medidas anteriores. Con el objetivo de ordenar, clasificar y registrar información teórica sobre el objeto de estudio. Con el siguiente orden de los datos:

- Ficha bibliográfica en el ángulo superior izquierdo. Sólo en la primera ficha se anotarán todos los datos. A partir de la segunda del mismo autor, se escribe el apellido paterno, el inicio del título de la obra (seguido de tres puntos suspensivos) y los números de las páginas consultadas.
- En el ángulo superior derecho se incluyen datos para organizar los ficheros. Entre los más comunes están: números y letras; el número corresponde al tema investigado; la letra, al autor.
- Tema, en mayúsculas y al centro de la tarjeta.
- Texto o datos: las anotaciones sobre el tema se escriben en el cuerpo de la tarjeta, dejando los márgenes respectivos. Si se transcribe en forma textual, se debe entrecomillar. Si se anotan observaciones de quien escribe, se utilizan paréntesis o corchetes. Cuando en la cita textual se saltan frases que corresponden al texto, se utilizan puntos suspensivos y se continúan las anotaciones.

Citas: La investigación documental implica plasmar el contenido, utilizando el mismo código de las fichas de trabajo. Las referencias hacen más veraz y serio el trabajo realizado; las citas pueden hacerse dentro del texto y anotarse al pie de la página, o bien, al final de la obra o del capítulo, como “notas de texto”. Con el objetivo de reforzar, clasificar, complementar y señalar fuentes de información.

Para citar las referencias en el texto, puede utilizarse la numeración progresiva, el método de nombre y fecha, o figuras simbólicas. La numeración de las referencias siempre ha de ser progresiva, pudiendo hacerse por página, en cuyo caso se reinicia la numeración en cada página, o por capítulo, en cuyo caso se inicia al citar la primera referencia y termina al citar la última del capítulo.

Notas de pie de página: Los objetivos de las notas de pie de página son:

- Citar referencias bibliográficas o fuentes de información.
- Ampliar explicaciones.
- Aclarar o corregir.
- Sugerir al lector nuevas posibilidades de búsqueda.

Para redactar las notas de pie de página se siguen estos lineamientos:

- *Numeración.* Se coloca un numeral arábigo sin signos de puntuación, medio renglón arriba del que se está escribiendo y después del punto donde termina la oración.
- *Colocación.* La primera vez que aparece una nota de pie de página, se escribe completa, pudiendo omitirse el pie de imprenta, ya que la ficha bibliográfica aparece en la lista de obras consultadas.
- *Diferencias entre la nota de pie de página y la ficha bibliográfica:* el nombre del autor no se invierte (se inicia por el nombre de pila); sólo se anotan las páginas consultadas; dentro de la nota se pueden hacer observaciones, ampliaciones, etcétera.
- *Referencias posteriores.* Cuando en un documento ya se ha hecho la nota de pie de página completa, las referencias posteriores se hacen utilizando abreviaturas (locuciones latinas). Entre ellas están:
 - *Ibid.* Significa ‘en el mismo lugar’. Se utiliza para hacer notar que se trata del mismo documento consultado; se usa cuando no se intercalan notas de otros documentos.
 - *Loc. cit.*, del latín locus *citado*, que significa lugar citado. Se usa para señalar que se trata de referencias del mismo documento y página.
 - *Op. cit.* Significa “obra citada”. Se usa para hacer referencia a un trabajo citado con anterioridad.
- *Notas de contenido* y referencias. Se utilizan para ampliar la información, hacer referencia a otras partes de las obras, y enviar al lector a otras partes del trabajo. Se emplean los siguientes términos y abreviaturas: *Infra*, que significa ‘abajo’. *Supra*: ‘arriba’. *Apud*: ‘citado por’, ‘apoyado en’. *Vid*: ‘véase’. *Sic*: ‘así’ (exactamente como se cita). *Et al*: ‘y otros’.
- Todas las locuciones latinas se subrayan.

2.2. Técnica de campo

El instrumento de observación se diseña según el objeto de estudio.

a) Objetivos de la observación:

- Explorar. Precisar aspectos previos a la observación estructurada y sistemática.
- Reunir información para interpretar hallazgos.
- Describir hechos.

b) Requisitos al observar:

- Delimitar los objetivos de la observación.
- Especificar el procedimiento o instrumentos de observación.
- Comprobación continua.

c) Tipos de observación:

En la observación de campo es menester distinguir dos clases principales: la observación participante y la no participante. En la primera participan varios investigadores dispersos con el fin de recoger reacciones colectivas. Otro tipo de investigación participante es aquella en la que los investigadores participan de la vida del grupo, con fines de acción social.

En la observación no participante el investigador es ajeno al grupo. Solicita autorización para permanecer en él y observar los hechos que requiere. La observación simple no controlada se realiza con el propósito de “explorar” los hechos o fenómenos de estudio que permitan precisar la investigación. La observación sistemática se realiza de acuerdo con un plan de observación preciso, en el que se han establecido variables y su relación, objetivos y procedimientos de observación.

d) Instrumentos para investigación de campo:

Para la *observación simple*, los instrumentos más comunes son:

- Ficha de campo.
- Diario.
- Registros.
- Tarjetas.
- Notas.
- Mapas.
- Diagramas.
- Cámaras.
- Grabadoras.

Para la *observación sistemática*, los instrumentos más comunes son:

- Plan de observación.
- Entrevistas.
- Cuestionarios.
- Inventarios.
- Mapas.
- Registros.

- Formas estadísticas.
 - Medición.
- e) **Ficha de campo:** Esta ficha se utiliza para anotar los datos recogidos mediante la observación. Por lo general se usan tarjetas de 21 x 13 cm. El orden de los datos es:
- Nombre del lugar donde se realizó la observación.
 - Nombre del informante o informantes.
 - Fecha.
 - Aspecto de la guía de observación en el ángulo superior derecho.
 - Tema, al centro.
 - Texto: descripción de la observación.
 - Iniciales del investigador, en el ángulo inferior izquierdo.
- f) **Entrevista:** La encuesta es una pesquisa o averiguación en la que se emplean cuestionarios para conocer la opinión pública. Consiste en el acopio de testimonios orales y escritos de personas vivas. En la investigación de campo, para la recopilación de información pueden utilizarse las entrevistas, los cuestionarios y el muestreo, entre otros. La entrevista es una de las técnicas más usuales en ciencias sociales. Puede definirse como la relación que se establece entre el investigador y los sujetos de estudio. Puede ser individual o grupal, libre o dirigida. Los objetivos de la entrevista son:
- Obtener información sobre el objeto de estudio.
 - Describir con objetividad situaciones o fenómenos.
 - Interpretar hallazgos.
 - Plantear soluciones.
- g) **Pasos de la entrevista: planeación, ejecución, control y cierre:**
- Planeación de la entrevista:*
- Elaborar la guía.
 - Definir de manera clara los propósitos de la misma.
 - Determinar los recursos humanos, tiempo y presupuesto necesarios.
 - Planear las citas con los entrevistados, mostrando respeto por el tiempo del mismo.
- Ejecución de la entrevista:*
- Propiciar durante esta un ambiente positivo.
 - Presentarse y explicar los propósitos de la misma.
 - Mostrar interés y saber escuchar.
 - Actuar con naturalidad.
 - No mostrar prisa.
 - Hacer las preguntas sin una respuesta implícita.
 - Emplear un tono de voz modulado.
 - Ser franco.
 - No extraer información del entrevistado sin su voluntad.

- No apabullar con preguntas.
- Evitar aparecer autocrático.
- Permitirle salidas airoosas.

Control de la entrevista

- Verificar si todas las preguntas han tenido respuesta.
- Detectar contradicciones.
- Detectar mentiras, dándole a conocer lo que se conoce del hecho.
- Evitar desviaciones del tema y oportunidades para distraer la atención.

Cierre de la entrevista

- Concluirla antes que el entrevistado se canse.
- Despedirse, dejando el camino abierto por si es necesario volver.
- Si se requiere de una entrevista larga, programarla en sesiones.

h) Cuestionario: Es un instrumento para recolección de información, que es llenado por el encuestado. Los objetivos son:

- Uniformar la observación.
- Fijar la atención en los aspectos esenciales del objeto de estudio.
- Aislar problemas y precisar los datos requeridos.

Pasos para el diseño de formularios

- Delimitar objetivos.
- Operativizar variables.
- Determinar la unidad de observación.
- Elección del método de aplicación.
- Adiestrar al personal recolector.
- Prueba del cuestionario.
- Diseño propiamente dicho.

Reglas para el diseño de cuestionarios

- Hacerlos cortos.
- Utilizar términos claros y precisos, así como una redacción sencilla.
- El tamaño debe facilitar su manejo.
- Los espacios de llenado deberán ser suficientes para las respuestas.
- Señalar siempre en su cuerpo los objetivos que persigue.
- De preferencia, hacer preguntas cerradas, para facilitar el procesamiento de la información.
- Adjuntar instrucciones para su manejo.

Tipos de cuestionarios

- Por cuadros, con datos objetivos.
- De opinión.

- De organización y funcionamiento.

Se llama *preguntas cerradas* a las que sólo permiten una opción para contestar, y *abiertas* a las que dejan plenas libertad para responder.

Desventajas del cuestionario

1. Sólo se puede aplicar a personas que sepan leer.
2. Las respuestas pueden falsearse.
3. Puede haber preguntas sin respuesta.
4. Debe ser perfectamente estructurado.

Ventajas del cuestionario

1. Económico.
2. Puede enviarse a lugares distantes.
3. Aplicable a grandes grupos de población.

i) Prueba piloto del cuestionario

Una vez redactado el cuestionario, se somete a una prueba de validez, confiabilidad y operatividad. Se aplica en forma experimental a un pequeño grupo de personas. Tendrá validez si en verdad se recogen los datos esperados. Si, independientemente de quien lo aplique, produce el mismo resultado, es confiable. Será operativo cuando los términos empleados generen la misma interpretación.

DISEÑO DE INVESTIGACIÓN

Diseño de investigación es el plan de acción. Indica la secuencia de los pasos a seguir. Permite al investigador precisar los detalles de la tarea de investigación y establecer las estrategias a seguir para obtener resultados positivos, además de definir la forma de encontrar las respuestas a las interrogantes que inducen al estudio. El diseño de investigación se plasma en un documento con características especiales, lenguaje científico, ubicación temporal, lineamientos globales y provisión de recursos.

1. Objetivos del diseño de investigación

El diseño de investigación tiene también otras denominaciones: plan, protocolo, diseño o proyecto de investigación, entre otras. Si bien los términos son diferentes, en esencia son semejantes en cuanto a que se refieren al plan de trabajo. Los objetivos del plan de investigación son:

- Definir el contexto ambiental del objeto de estudio.
- Precisar el objeto de estudio.
- Definir y delimitar el problema de investigación y los aspectos que intervienen.
- Seleccionar el método y las técnicas adecuadas al objeto de estudio.
- Organizar y sistematizar las acciones por desarrollar.
- Describir los recursos necesarios.
- Verificar la factibilidad del estudio.

2. Características del plan de investigación

El plan de investigación, como documento de carácter científico, tiene características especiales que deben tomarse en cuenta al elaborarlo.

Congruencia:

Se refiere al hecho de que exista una concatenación lógica entre los elementos que lo conforman. El título del estudio maneja variables que se encuentran en: el problema, los objetivos, las hipótesis y las técnicas y métodos del análisis seleccionado, los cuales serán en función de dichas variables.

Fundamentación teórica:

El plan incluye las teorías relativas al objeto de estudio en forma exhaustiva.

Flexibilidad:

Es importante que el plan considere la posibilidad de realizar cambios o adaptaciones, sin desorganizar el diseño original.

Lenguaje científico:

El plan debe expresarse considerando la terminología científica. Por ejemplo: El diseño es de tipo experimental y la técnica es de balanceo.

3. Defectos del plan de investigación

Al elaborar el plan de investigación es posible cometer errores u omitir algunos detalles, lo cual se traduce en defectos del plan. Los errores de diseño muchas veces son costosos; en otras, no se logran los propósitos iniciales del estudio. Algunos *defectos del diseño* son los siguientes: *Indefiniciones del tema y los propósitos del estudio*. Se producen cuando los conceptos empleados son nebulosos.

Manejo teórico inadecuado. Es resultado del desconocimiento teórico relativo al objeto de estudio. *Complejidad del marco teórico*. Cuando el nivel de complejidad del marco teórico se debe al enorme volumen de datos sobre el tema y, a su vez, no se produce un análisis de dichas teorías que resuma y relacione con el problema a investigar, el marco teórico se torna complejo.

Imprecisión de métodos y técnicas. Cuando el investigador desconoce o tiene un nivel insuficiente en torno a la investigación, lo más seguro es que las técnicas descritas en el plan sean inadecuadas.

4. Modelos en diseños de investigación

Existe un sinnúmero de modelos para elaborar diseños de investigación. Por ello, se seleccionaron dos formatos:

Modelo Uno

- A. Parte introductoria.
 - a. Título y subtítulos.
 - b. Problemática en que se ubica el tema.
 - c. Cuestiones relevantes que surgen de la problemática.
 - d. Definición específica del tema.

- B. Justificación del propósito investigativo.
 - a. Oportunidad para tratar el problema Amplitud de la población.
 - b. Impacto del estudio.
 - c. Factibilidad.

- C. Contextualización de la labor investigativa.
 - a. Examen de estudios previos.
 - b. Delimitación de la investigación.

- D. Tipo específico de investigación a que pertenece el estudio.
 - a. Objetivos.
 - b. Marco teórico de la investigación.
 - c. Conceptualización específica.
 - d. Hipótesis.
 - e. Criterios de la muestra.
 - f. Tipo específico de la muestra.
 - g. Técnicas en la obtención de datos.
 - h. Elaboración de material.
 - i. Técnicas del análisis.
 - j. Guía de trabajo.

Modelo Dos

- A. Sección preliminar.
 - a. Nombre y currículum del investigador.
 - b. Instituciones que patrocinan el estudio.

- B. Definición del proyecto de investigación.
 - a. Título.
 - b. Antecedentes.
 - c. Hipótesis.
 - d. Objetivos.

- C. Definición de las unidades de análisis.

- D. Diseño del experimento.

- E. Metodología.

- F. Especificación de variables y procedimientos de medición.

- G. Capacitación de personal para la observación.

- H. Ingreso de participantes en el experimento.

- I. Captación de datos.

- J. Universo de estudio.

- K. Recursos.

MÉTODO CIENTÍFICO

Como ya se explicó, el método científico es el procedimiento que se sigue para obtener el conocimiento. Los puntos convergentes de los diversos autores son los relativos a las etapas del método. En general, puede concluirse que son las siguientes:

1. Etapas del método científico

Las principales etapas del método científico son:

- Elección y enunciado del problema que motiva la investigación.
- Estructuración de un marco teórico.
- Establecimiento de hipótesis.
- Prueba de hipótesis.
- Resultados.
- Propuestas derivadas del estudio.

1.1. Elección y enunciado del problema

La interacción constante con el ambiente nos permite observar situaciones que a menudo despiertan una serie de interrogantes. Cada una de las preguntas planteadas delimita un problema amplio. Las preguntas planteadas no son suficientemente precisas para ser investigadas. Si las interrogantes nos inquietan al grado de considerar su estudio, debemos definir y delimitar el problema de investigación. Para ello, consideraremos los siguientes aspectos:

a) Fuentes de los problemas:

- Cuál es el origen del problema.
- Qué intereses profesionales o científicos tiene el investigador para hacer el estudio.
- Qué conocimientos se tienen sobre el tema.
- Qué aplicación daría a los resultados de la investigación.

Al dar respuesta a las interrogantes anteriores, se infiere que los problemas derivan de: el ambiente, la capacidad de razonar, los intereses profesionales y los productos de la investigación.

b) Tipos de problemas:

- Teóricos. Cuyo propósito es generar nuevos conocimientos.
- Prácticos. Con objetivos destinados al progreso.
- Teórico-prácticos. Para obtener información desconocida en la solución de problemas de la práctica.

Sin duda existe un gran número de problemas que nos inquietan, pero quizá la mayor parte de ellos no están al alcance de todos. Los requisitos para elegir un problema de investigación son:

- Experiencia en el tema.
- Importancia del problema.

- Conocimientos para su manejo.
- Relevancia científica.
- Relevancia humana.
- Relevancia contemporánea.

Requisitos para elaborar un problema de investigación son:

- Señalar manifestaciones del problema.
- Manejar dos variables como mínimo.
- Definir con claridad el problema.
- Delimitar los aspectos que abarca el problema.
- Enunciado del problema.

c) Señalar manifestaciones del problema:

Consiste en describir las experiencias empíricas, contexto, determinantes, interrogantes generales, efectos, posibles soluciones y sugerir los propósitos del estudio.

d) Manejar dos variables como mínimo:

Al perfilar el problema, y a la luz de los referentes empíricos, es posible relacionar al menos dos elementos, que pueden ser: posibles causas del problema y efectos del mismo.

e) Definir con claridad el problema:

Los referentes empíricos y el manejo de dos variables como mínimo, nos permiten definir el área problema con precisión de detalles. Los términos utilizados para definir el problema deben ser lo bastante claros para permitir que cualquier persona, con sólo leer el problema, se ubique en lo que se pretende estudiar.

f) Delimitar los aspectos que abarca el problema:

La definición del problema obliga a precisar los aspectos que incluye. La delimitación de los aspectos por estudiar evita las frecuentes divagaciones y centra la atención en los elementos medulares del problema de investigación.

g) Enunciado del problema:

Se manejan dos formas de enunciar el problema de investigación: interrogativo y declarativo. *Interrogativo.* Se expresa a través de una pregunta. *Declarativo.* Se expresa a manera de propósito.

h) Objetivos de investigación

Los objetivos de investigación se construyen tomando como base la operatividad y el alcance de la investigación. Requisitos para plantear los objetivos (Enfocarse en la solución del problema).

- Ser realistas.
- Ser medibles.
- Ser congruentes.
- Ser importantes.
- Redactarse evitando palabras subjetivas.

Para construir los objetivos deben considerarse las siguientes interrogantes: Quién, qué, cómo, cuándo y dónde. La investigación no es un fin por sí misma. La transformación de la realidad operante tiene su base en la investigación, pero sólo la aplicación práctica de los resultados y la confrontación permanente permitirán acrecentar el conocimiento.

Una investigación cuyos resultados se archiven, no cumple con la función de interacción sujeto-objeto, pues no produce una práctica transformadora de la realidad.

1.2. Estructuración del marco teórico

La teoría- da significado a la investigación. Es a partir de las teorías existentes sobre el objeto de estudio que pueden generarse nuevos conocimientos. La validez interna y externa de una investigación se demuestra en las teorías que la apoyan y, en esa medida, los resultados pueden generalizarse. El marco teórico de la investigación considera:

- Conceptos explícitos e implícitos del problema.
- Conceptualización específica operacional.
- Relaciones de teorías v conceptos adoptados.
- Análisis teórico del cual se desprenden las hipótesis.
- Concluir las implicaciones de la teoría con el problema.

1.3. Establecimiento de hipótesis

Las hipótesis son suposiciones conjeturales, en transición hacia su confirmación. Se desprenden del análisis teórico para plantear supuestos con alto grado de certeza. Las hipótesis son el vínculo entre la teoría y la práctica; se construyen con tres elementos:

- El *objeto de estudio*, al cual se denomina unidad de análisis.
- *Las variables*, que se conocen como propiedades de las unidades del análisis.
- La relación, que se describe como los términos lógicos que unen los objetos con sus propiedades.

Hipótesis es una forma de desarrollo de las ciencias naturales, por cuanto son pensamientos (Engels).

La hipótesis es una afirmación tentativa, más que definitiva. Debe ser formulada de tal manera que pueda ser potencialmente aceptada o rechazada por medio de los hallazgos. La teoría sirve de base a la hipótesis y a su vez es modificada por ésta. La hipótesis requiere de la investigación para la comprobación de los postulados que contiene.

a) Requisitos para elaborar una hipótesis

- Construirla con base en la realidad que se pretende explicar.
- Fundamentarla en la teoría referente al hecho que se pretende explicar.
- Establecer relaciones entre variables.
- Ser susceptible de ponerse a prueba, para verificar su validez.
- Dar la mejor respuesta al problema de investigación, con un alto grado de probabilidad.
- No incurrir en nada superfluo en su construcción.

b) Clasificación de las hipótesis

- *Sustantivas*. Se refieren a la realidad social.
- *De generalización*. Se refieren a los datos.

- *Generales*. Relación entre variables básicas.
- *Particulares*. Derivan de una hipótesis básica.
- *Alternativas*. Misma variable independiente, con otras dependientes.
- *Descriptivas*. Señalan la existencia de regularidades empíricas.
- *Tipos ideales complejos*. Ponen a prueba la existencia de relaciones entre un tipo ideal y la realidad.
- *Analíticas*. Formulan relaciones entre variables y explican la relación entre diversos factores.
- *Postfacto*. Se deducen de la observación de un fenómeno.
- *Antefacto*. Inducen a una explicación antes de la observación.
- *Nulas*. Se diseñan para reafirmar que no se ha rechazado una hipótesis verdadera por una falsa.
- *De trabajo*. Provisional y previa a la investigación definitiva, a efecto de hallar otras más sugestivas.

c) Función de las hipótesis

- Indicar el camino para la búsqueda de la verdad objetiva.
- Impulsar el trabajo científico.
- Sistematizar el conocimiento.
- Permiten explicar el objeto de estudio.
- Sirven de enlace entre el conocimiento ya obtenido y el que se busca.

Las hipótesis son intentos de explicación mediante una suposición verosímil que requiere comprobarse.

d) Variables

Son discusiones que pueden darse entre individuos y conjuntos. El término variable significa característica, aspecto, propiedad o dimensión de un fenómeno puede asumir distintos valores. Para operativizar variables, se requiere precisar su valor, traduciéndolas a conceptos susceptibles de medir. Por tanto, conviene considerar su definición nominal, real, operativa: lo que significa el término, la realidad y la práctica.

e) Clasificación de variables

En términos generales, las variables se clasifican según el nivel de medición que representan:

- *Variables cualitativas*. Son aquéllas que se refieren a cualidades o atributos no medibles en números. Por ejemplo, organización, personal y funciones.
- *Variables cuantitativas*. Son las susceptibles de medirse en términos numéricos. Se subdividen a su vez en:
 - o *Cuantitativas continuas*. Pueden asumir cualquier valor. Por ejemplo, peso, edad y talla.
 - o *Cuantitativas discretas*. Asumen sólo valores enteros. Por ejemplo, número de hijos.
- *Variables independientes*. Expresan las causas del fenómeno. Por ejemplo, organización deficiente.

- *Variables dependientes*. Expresan las consecuencias del fenómeno. Por ejemplo, calidad de la enseñanza.

1.4. Prueba de hipótesis

El propósito central de la investigación lo constituye la prueba de hipótesis. Se pretende comprobar si los hechos observados concuerdan con las hipótesis planteadas. En general, comprende dos pasos que son, selección de la técnica y recolección de la información.

a) Selección de la técnica

Para comprobar o refutar las hipótesis es necesario elegir por lo menos dos o tres técnicas de investigación y diferentes tipos de observación de fenómenos. En ciencias sociales, deben aplicarse la técnica documental y la de campo. Es importante hacer las siguientes consideraciones:

- La técnica será acorde al tipo de hipótesis que se desea comprobar.
- Diseñar los instrumentos según la técnica elegida.
- Probar los instrumentos.
- Determinar la muestra.

b) Recolección de la Información o datos

La manera más formal de proceder a la búsqueda de información es seguir los lineamientos del método científico. La estadística resulta de gran utilidad en el manejo de información. El proceso consiste en:

- Recoger la información.
- Tabularla.
- Presentarla.
- Analizarla.

El aspecto medular del manejo de información es la recolección, ya que el procesamiento de datos depende de la confiabilidad que aquélla pueda tener.

Métodos de recolección de datos

- *Encuestas*: La información se recoge por muestras, por lo que no se aplica a la población total.
- *Censos*: La información se recoge en forma general a toda la población.
- *Registros*: La información es continua. Se recoge a medida que se va produciendo.

Técnicas de recolección

- Entrevistas.
- Aplicación de cuestionarios.
- Observación.

Métodos para el recuento

- *Listas*. Cuando son pocas las unidades y no se manejen más de dos escalas.
- *Palotes*. Consiste en poner, en una hoja de trabajo, un “palote” por cada unidad que se cuenta. No se utiliza para gran número de observaciones.

- *Tarjetas simples.* La información por individuos se registra en una tarjeta. El número de unidades es corto. La clasificación por variables se hace rápidamente, ordenando las tarjetas en tantos grupos como categorías resulten. Se recomienda utilizar para menos de 500 casos y menos de 12 variables.
- *Mecanizada.* El cómputo e impresión de resultados pueden obtenerse por computadora.

Presentación de datos.

Consiste en dar a conocer los datos en forma resumida, objetiva y entendible. Las formas más usuales son la tabular (tablas o cuadros) y las gráficas.

Tablas. He aquí las partes principales de una tabla estadística:

1. Título. Contiene el qué, cómo, cuándo y dónde. Debe ser breve y conciso.
2. Cuadro. Contiene filas y columnas en escalas cualitativas y cuantitativas.
3. Fuente y notas explicativas. Deben aparecer al pie del cuadro.

Gráficas. Entre los requisitos que debe reunir una gráfica, están:

1. Ser auto explicativo.
2. Presentar de manera fiel los hechos.
3. Estética.
4. Limpia y de trazos netos.

Los principales *tipos* de gráficas son:

1. *Gráfica de barras.* Se utiliza para presentar distribuciones de frecuencias de variables de escalas nominal y ordinal. El largo de la barra indica la frecuencia: el ancho, la constante.
2. *Barras dobles.* Presentan datos de asociación de dos escalas cualitativas.
3. *Diagrama de sectores.* Se utiliza con fines comparativos. Presenta cifras absolutas o porcentajes, en los que cada 1 % corresponde a 3.61 del círculo.
4. *Histogramas.* Semejante al de barras, pero sin espacios entre una y otra barra. Presenta la distribución de frecuencias en una escala cuantitativa continua.
5. *Polígono de frecuencias.* Se presenta una comparación de distribución de frecuencias en escalas continuas, en las que los puntos se unen con una línea sin interrupción.
6. *Diagrama de correlación.* Se emplea para el estudio de la relación de variables en escala continua. Los ejes son del mismo tamaño y sólo se colocan puntos.
7. *Pictogramas.* Se usan con fines publicitarios y se representan con figuras que indican cantidades.

Descripción y análisis.

Con el propósito de evaluar la información, se utilizan técnicas estadísticas para determinar la validez de los resultados obtenidos.

Medidas de resumen.

Expresan numéricamente los datos en un solo valor. Estas dependen del tipo de variables.

Para las variables cualitativas, las medidas de resumen se extraen de relaciones estadísticas entre conjuntos y subconjuntos que presentan frecuencias relativas.

Para las variables cuantitativas, las medidas se obtienen mediante procesos estadísticos con medidas de tendencia central de dispersión, etc.

Una razón es la relación entre un número de observaciones en una categoría y el número de observaciones de otra.

Una proporción es la relación entre un número de observaciones de una categoría y el total.

Un porcentaje es la relación entre un número de observaciones de una categoría y el total general de un grupo, multiplicado por una constante, que es 100.

Métodos de análisis.

El análisis se realiza mediante computadora y utilizando técnicas estadísticas que se verán con profundidad en los siguientes capítulos.

1.5. Resultados

Los resultados de la investigación se incorporan en un documento que usualmente es el informe de investigación. La difusión extensa de los resultados está implícita en la investigación. En las propuestas de solución es conveniente incluir lo siguiente.

- Argumentación teórica.
- Argumentación metodológica.
- Abordaje de solución.

RECOMENDACIONES

La importancia que tiene la investigación como proceso de aprendizaje, ya que la misma posee una gama de características fundamentales y que se estrechan de manera muy compacta para poder captar la información o para lograr los objetivos propuestos. Es preciso recordar que la investigación científica es un método riguroso en el cual se obtiene una serie de objetivos antes propuestos y de manera muy técnica.

Es muy conveniente tener un conocimiento detallado de los posibles tipos de investigación que se pueden seguir. Este conocimiento hace posible evitar equivocaciones en la elección del método adecuado para un procedimiento específico.

Una de las fallas más comunes en la investigación es la ausencia de delimitación del tema, es decir, por ausencia de ambición del tema, por eso es básico tener muy claro los objetivos y el camino que se va a recorrer con la investigación para que ésta pueda terminar su camino donde debe.

La hipótesis puede ser desarrollada desde distinto puntos de vista. Puede basarse en una conjetura, en el resultado de otros estudios, en la posibilidad de una relación semejante entre dos o más variables representadas en un estudio. O puede estar basada en una teoría mediante la cual una suposición de proceso deductivo nos lleva a la pretensión de que si se dan ciertas condiciones se pueden obtener ciertos resultados.

El objetivo de la investigación es el enunciado claro y preciso de las metas que se persiguen. El objetivo de la persona que investiga es llegar a tomar decisiones y a una teoría que le permita generalizar y resolver en la misma forma problemas semejantes en el futuro.

Capítulo

2

Modelos de investigación

Existe un sin número de modelos para elaborar una investigación, a los que también se le llaman estrategias. En el presente capítulo se muestra un sin número de estrategias tanto de un proyecto de investigación como de una investigación.

Pero lo más importante es que cada investigador puede diseñar su propia estrategia de investigación basándose en los modelos ampliamente conocidos. El autor presenta una estrategia simple que te ayudará a realizar tu investigación de forma fácil y sencilla.



MODELOS DE INVESTIGACIÓN

No hacemos distinción entre trabajo de grado o investigación científica, ya que se parte de que una y otra se enmarcan y participan dentro de los cánones referidos por la ciencia, el método científico y la investigación. En el contexto académico denominamos como tesis al trabajo de grado y como proyecto de trabajo de grado, a la propuesta previa que debe ser aprobada para poder dar inicio a la realización del primero.

Al margen de la propuesta y el resultado de la investigación, el trabajo de grado ahonda en diversas facetas de la personalidad, la habilidad científica y la ética profesional del estudiante. Es, por tanto, una invitación a conocerse a sí mismo, ¿quién soy? ¿qué deseo ser profesionalmente? ¿cuáles son mis intereses a corto, mediano y largo plazo? ¿qué tan responsable y capaz soy? ¿qué tanto compromiso y autonomía tengo?

La investigación es, además, un termómetro para que el estudiante mida sus propias capacidades: ¿en qué soy bueno? ¿puedo realizar el trabajo a cabalidad? ¿sé buscar y consultar bibliografía? ¿sé escribir una idea en forma coherente? ¿qué he aprendido a lo largo de mi carrera? Es también una invitación a mejorar conocimientos básicos necesarios para la vida: lectura de textos en un segundo idioma, uso apropiado del computador abarcados un procesador de texto, una hoja de cálculo, una base de datos o una presentación; o incluso un avance significativo en temas más especializados como técnicas de muestreo, análisis estadístico, paquetes estadísticos, sistemas de información geográfica, legislación, gestión etc.

Los científicos emplean un proceso sistemático, de modo que se asegure un experimento de calidad superior. A estas metodologías científicas se les llamada también Estrategias a seguir a la hora de realizar una investigación. A continuación presentamos las principales estrategias, tanto de un proyecto de investigación, como de una investigación.

CONTENIDOS DEL PROYECTO DE INVESTIGACIÓN

Se han seleccionaron las que, a nuestro juicio, consideran todos los elementos mencionados por diversos autores. Presentando los aspectos más relevantes de la planificación, el diseño y la formulación de la investigación, temática correspondiente al proyecto de investigación.

PRIMERA ESTRATEGIA DE UN PROYECTO DE INVESTIGACIÓN

El proyecto de investigación debe situar las bases de la investigación a realizar. Su valor se establece en la medida en que tiene **plena claridad y concreción** en las razones para analizar el objeto de estudio elegido, la perspectiva teórica desde donde se sitúa el investigador, el paradigma investigativo que sustenta todo el estudio y, por tanto, la metodología de aproximación a la realidad: población, muestra, estrategias de recogida de información, técnicas de análisis de la información y temporalidad de todo el proceso. En suma, el documento demuestra que el investigador conoce suficientemente el tema de investigación y tiene las ideas claras sobre la estructura del proceso y el camino por el que pretende aportar al conocimiento científico.

1. Introducción.
2. Justificación.
3. Planteamiento del Problema.

4. Objeto de Estudio.
5. Preguntas de Investigación.
6. Objetivos.
7. Fundamentación teórica.
8. Metodología de investigación.
9. Población y muestra.
10. Diseño de la Investigación.
11. Cronograma y presupuesto.
12. Bibliografía.

1. Introducción

Contiene una **descripción** clara de la estructura general del proyecto.

2. Justificación

Contiene los argumentos fundamentales que sustentan la investigación a realizar, enfatizando aquellos de carácter **técnico** y **social** principalmente.

La pregunta es: ¿Por qué y para qué realizar la investigación?

3. Planteamiento del Problema

Es resultado de una **profunda** y **serena reflexión** realizada por el investigador después de haber revisado detalladamente la literatura correspondiente (antecedentes teóricos y empíricos) e interiorizado los principales conceptos y proposiciones teóricas que le permiten formular con **toda claridad** y **dominio** el problema que se pretende resolver con la investigación.

La pregunta es: ¿Qué realidad me interesa investigar?

4. Objeto de Estudio

Es consecuencia del planteamiento del problema, **delimita** aquella parte de la realidad que interesa estudiar. La precisión del investigador, en este sentido, se demuestra en la **redacción minuciosa** y **cuidada** con la cual formula el objeto de estudio.

La pregunta es: ¿Qué parte de esa realidad deseo investigar?

5. Preguntas de Investigación

Son las **interrogantes básicas** que se derivan de la justificación y el problema planteado y, consecuentemente, pretenden ser resueltas por la investigación.

La pregunta es: ¿Qué cuestiones me interesan de esa parte de la realidad?

6. Objetivos

Representan las **acciones concretas** que el investigador llevará a cabo para intentar responder a las preguntas de investigación y así resolver el problema de investigación. Se puede notar que todos los subtítulos hasta ahora señalados tienen una **consistencia entre sí** (coherencia interna),

por ello, los objetivos deben ser **concretos, claros, realistas y modestos**, en la medida en que realmente reflejan la contundencia del investigador en su intención de aportar en el conocimiento del objeto de estudio.

La pregunta es: ¿Qué debo hacer para desarrollar la investigación?

En este apartado de objetivos se incluye la **HIPÓTESIS** de investigación, siempre que ésta, **NO CONDICIONE/INDUZCA** el trabajo de investigación pues no siempre los resultados esperados coinciden con los resultados reales. Debe entenderse que la **realidad es una entidad viva** y tiene voz propia y la riqueza de una investigación científica está precisamente en mostrar la realidad tal como es, no en validar una hipótesis. Por tanto, en las **Ciencias Sociales y Humanas** la hipótesis es simplemente una **estimación** de los efectos que creemos se producirán con nuestra intervención. Estimación que es todavía más prudente, si cabe, que las hipótesis formuladas en las Ciencias Experimentales y Tecnológicas.

La pregunta es: ¿Qué espero que suceda con mi intervención sobre la realidad estudiada?

Suele ser este el sentido de las hipótesis de estudios **cuantitativos**. Son hipótesis que anticipan una relación causa efecto y exigen extremo cuidado en el análisis y control de variables extrañas. En este caso suelen ser estudios de largo alcance y suponen tiempo y recursos considerables.

La pregunta es: ¿Qué pienso encontrar en la realidad estudiada?

Este sería el sentido de las hipótesis de estudios cualitativos, que son hipótesis de trabajo que anticipan las cuestiones que el investigador tiene la impresión de encontrar en la realidad, no obstante, prioriza la realidad como tal y ajusta sus hipótesis de trabajo futuras. En este caso actúan como una referencia en mi relación con la realidad estudiada, no como un direccionamiento en esa relación, por tanto, las hipótesis se desestiman según se avanza en el conocimiento cada vez más preciso de la realidad.

7. Fundamentación teórica

Se presentan en este caso las **directrices teóricas** que guían el estudio, describiendo claramente el **panorama global** (completo) de escuelas o teorías que abordan el tema y las principales evidencias teóricas-empíricas existentes en la literatura nacional e internacional.

La pregunta es: ¿En qué me apoyo para estudiar esta realidad?

8. Metodología de la investigación

Contiene la **descripción y argumentación** de las principales decisiones metodológicas adoptadas según el tema de investigación y las posibilidades del investigador. La claridad en el enfoque y estructura metodológica es **condición obligada** para asegurar la validez de la investigación.

La pregunta es: ¿Cómo desarrollo la investigación?

9. Población y Muestra

Describe y argumenta la decisión sobre la **población objetivo** para investigar el objeto de estudio, así como, los criterios básicos de **determinación de la muestra** considerada para la investigación.

Las preguntas son:

¿Quiénes forman parte de la realidad que investigo?

¿De entre ellos, quiénes serán mis informantes?

10. Diseño de la Investigación

Mediante una **representación gráfica** se presenta el panorama metodológico completo que muestra la forma en que se **organiza** todo el proceso de investigación y los aspectos metodológicos esenciales que guían el trabajo del investigador. Es importante la **claridad y precisión** en esta cuestión porque da cuenta del **posicionamiento** del investigador y de su trabajo en el mapa metodológico de la investigación científica.

La pregunta es: ¿Cómo se organiza toda la investigación?

11. Cronograma y Presupuesto

Es una tabla-esquema sencillo–concreto que presenta la estimación del tiempo que tomarán cada una de las etapas de la investigación.

Las preguntas son:

¿Qué tiempo aproximado me supone desarrollar cada etapa de la investigación?

¿Cómo se distribuyen los recursos en cada una de las etapas?

12. Bibliografía

Describe las fuentes documentales consideradas (castellano-inglés) para la elaboración del proyecto. Conviene que no excedan los 5 años de vigencia (actualidad) y su citación cumpla con las normas internacionales/estandarizadas de formato.

SEGUNDA ESTRATEGIA DE UN PROYECTO DE INVESTIGACIÓN

Todo Plan o Proyecto de Tesis debe tener necesariamente la siguiente información:

I. Introducción:

1. Tematización o Marco Situacional:

Descripción del tema o situación que es de interés para el graduando.

2. Problematización:

Subrayar las ideas centrales de la tematización y convertir cada idea en una o más preguntas o temas científicos. Seleccionar el problema de la investigación.

3. Objetivos:

Especificar qué se intenta conseguir con la Tesis o Trabajo de Investigación. Especificar los objetivos o propósitos de trabajo, incluyendo los problemas que se intentan resolver o las hipótesis que se intenta probar.

4. Importancia:

Justificar la importancia o trascendencia de la Tesis o Trabajo de Investigación, precisando los beneficios que se obtendrán.

5. Metodología:

Justificar los métodos de investigación que deben emplearse, determinar el universo del cual se ha recogido la información necesaria, precisar las muestras o unidades de estudio que deben cubrirse en la investigación, el criterio para seleccionar dichas muestras o unidades de estudio, cómo se va a efectuar el análisis de la información. Es posible presentar junto con el Plan las muestras y otras técnicas a ser usadas.

6. Sumario:

Presentar un listado de los diferentes aspectos que comprenderá la Tesis o Trabajo de Investigación. Utilizar una codificación conveniente.

II. Marco Teórico:**1. Antecedentes:**

Localizar e indicar otras investigaciones que se han efectuado en relación del problema o a problemas parecidos que se abordan en la Tesis trabajo de investigación. Si el caso lo requiere, precisar antecedentes legales o de otra orden.

2. Soporte Teórico de la Investigación:

Precisar las fuentes teóricas en que ha de basarse la investigación. En lo posible fundamentar la selección de cada uno de las fuentes.

3. Óptica de la Investigación:

Definir la perspectiva desde la cual ha de realizarse la investigación (punto de vista del graduando); fundamentar esta perspectiva.

4. Selección de Variables:

Indicar las variables que se utilizarán en el análisis con el fin de probar las hipótesis iniciales referidas a los problemas a resolver en la Tesis o Trabajo de Investigación.

III. Marco Empírico:

Especificar el instrumento teórico científico que se utilizará en el desarrollo de la Tesis o Trabajo de Investigación, Estadística, Métodos Cuantitativos, Econometría, etc.

IV. Aspectos Operativos:**1. Cronograma de Trabajo.****2. Asesoramiento:** indicar el asesoramiento que se requiere.**3. Fuentes de información:** especificar la Bibliografía y otras fuentes de información que serán utilizados.

TERCERA ESTRATEGIA DE UN PROYECTO DE INVESTIGACIÓN

Para la *construcción* del proyecto de investigación, se sugiere la siguiente secuencia, que será desarrollada a continuación:

1. Elección de un tema.
2. Consulta bibliográfica.
3. Delimitación del tema.
4. Hipótesis y objetivos.
 - a. Objetivos Generales y Específicos.
 - b. Lluvias de Ideas.
 - c. Diagrama de Modelación de la Variable Objeto del Estudio.
 - d. Pertinencia y Viabilidad.
 - e. Relación con el Marco de Referencia.
 - f. Dificultades en la Definición de Hipótesis u Objetivos.
 - g. Operacionalización de las Variables.
5. Justificación.
6. Metodología.
7. Título.
8. Marco de Referencia.
 - a. Marco de antecedentes.
 - b. Marco Teórico.
 - c. Marco Conceptual.
 - d. Marco Geográfico.
 - e. Marco Demográfico.
 - f. Marco Institucional.
 - g. Marco Legal.
9. Introducción.
10. Cronograma.
11. Materiales y Presupuesto.
 - a. Materiales.
 - b. Presupuesto.
12. Bibliografía.

1. Elección del Tema

El proceso de investigación se inicia con la selección de un tema de estudio. Es importante señalar, sin embargo, que durante esta etapa el tema de investigación tiene carácter provisional y puede dar lugar a cambios leves o radicales. Vale aclarar que el tema por sí mismo no conduce a un trabajo de investigación bueno o malo, sino será su contenido, su originalidad y su rigurosidad los que dictaminen tal calificativo.

La elección del tema es vista por diferentes autores tales como los siguientes:

Salkind (1997):

- Experiencias personales y conocimientos previos.
- Planteamiento de una pregunta de investigación.
- Por profesores o investigadores, con la ventaja de abordar temas de vanguardia.

Muñoz-Razo (1998):

- Área de conocimiento afín.
- Tema de interés.
- Tema a partir de una materia favorita.
- Tema por referencias, amigos, profesores.
- Tema de moda.
- Tema por asignación del mismo.

Tamayo-Tamayo (2000):

- Temas que nos inquietan o son de nuestra preferencia.
- Experiencias personales.
- Consulta a profesores o notas de clase.
- Revisión bibliográfica de diferentes temas.
- Examen de diversas publicaciones sobre un posible tema.
- Información sobre temas afines.
- Conexión con instituciones.
- Considerar en la elección del tema factores de orden *subjetivo* como son el agrado por el tópico, la capacidad para desarrollarlo, el tiempo, los recursos o la disponibilidad del material; como también, de orden *objetivo*, que se refieren a si llena los requisitos exigidos por la institución y su utilidad.

El investigador debe reflexionar también sobre sus propias capacidades, ya que mientras algunos demuestran gran habilidad en el campo o en el fondo del mar, otros hacen lo propio en el laboratorio, frente a un computador, las matemáticas o simplemente expresan amplia comunicación con niños o adultos.

2. Consulta Bibliográfica

Se procede una ardua pesquisa, recopilación y lectura bibliográfica sobre la temática de interés, la cual permite fortalecer los conocimientos teóricos y conceptuales de la ciencia en cuestión, como también, permite reconocer avances metodológicos y perspectivas de análisis. La consulta bibliográfica posibilita conocer las grandes teorías que se han producido, las hipótesis que se han planteado y han salido victoriosas, como también las que se han refutado.

Las mejores revistas científicas se editan en inglés u otros idiomas por lo que el estudiante deberá tener un acercamiento previo con tal idioma, en muchas situaciones la elección del tema puede conducirnos a idiomas particulares donde se han desarrollado los mayores avances científicos.

La revisión bibliográfica influye en la dirección que adoptará la investigación y, a la vez, permite visualizar diversos enfoques que facilitan el camino para plantear la (s) pregunta(s) de investigación.

Añade que la revisión bibliográfica previene el planteamiento de preguntas ya respondidas o triviales. La revisión también ayuda a esclarecer hechos, conceptos, clasificaciones o explicaciones, además, ayuda a estructurar más formalmente la idea de investigación, previniendo errores que se han cometido en otros estudios y provee, también, un marco de referencia para interpretar los resultados finales.

Lerma (1982) señala que los trabajos a considerar principalmente en la revisión bibliográfica deben referirse a:

- Problemas, temas o poblaciones similares.
- Relaciones y variables.
- Trabajos metodológicos propios de la temática elegida-diseño, método de análisis, validez de los instrumentos de medición, entre otros.

Hernández- Sampieri *et al.* (1998), refieren que también es válida la consulta de estudios efectuados en otros países, que si bien podrían diferir del contexto local, pueden aportar información valiosa en muchos aspectos.

Walker (2002) nos advierte que debemos tener prelación por autores reconocidos y por casas editoriales de renombre, como es el caso de revistas científicas y textos publicados por universidades los cuales, generalmente, son revisados por pares expertos antes de su publicación. Esta apreciación la hace extensible a las consultas en la *WEB* las cuales deben recaer en fuentes gubernamentales, no lucrativas, reconocidas o universitarias, y cuya ventaja principal es la posibilidad de actualización inmediata.

3. Delimitación del Tema

Generalmente la idea inicial del investigador se refiere a un *tema amplio o general*, que se va delimitando en población, contenido, espacio y tiempo, a través de diversas fuentes de aproximación, hasta alcanzar el *tema específico*. Esta delimitación se va logrando a medida que se reflexiona sobre él, se revisa bibliografía y se consulta con expertos. La delimitación del tema conlleva, además, una discusión teórica sobre el conjunto de conocimientos ya encontrados.

Este proceso ocurre al ir desmembrando el tema en subtemas y éstos, a su vez, en temas específicos. El proceso generalmente ocurre en forma natural, consciente o no, a medida que el investigador avanza en su revisión bibliográfica. La misma ocurre por afinidad temática, por viabilidad, por tiempo, infraestructura, por consejo de otros investigadores, entre otros.

No obstante, se le sugiere al investigador que a medida que realiza la revisión bibliográfica, realice un esfuerzo consciente y no espere en forma pasiva a concretar la temática de su investigación. En este momento debe esclarecerse *qué voy a hacer, en qué lugar y tiempo y con qué población*.

4. Hipótesis, Objetivos

a. Objetivos Generales y Específicos

El *objetivo general* de una investigación se enmarca en el título de la misma y se descompone, a su vez, en *objetivos específicos* o propósitos de comprobación y descripción del sistema o de las variables de estudio.

El *objetivo general* señala el nivel de conocimiento que se desea obtener del objeto como resultado de la investigación. Por lo regular, cada investigación tiene un único objetivo

general cuyo enunciado se inicia con un verbo en infinitivo, aunque puede escribirse incluso con más de un verbo: *Identificar y describir*. El objetivo general constituye la directriz de la investigación, mientras que el resto del documento se construye como apoyo o soporte al mismo.

Los objetivos específicos, por su parte, reseñan los resultados o metas parciales. Se enuncian e inician también con verbos en infinitivo. Cada objetivo específico debe incluir un solo logro. Una investigación puede tener uno o múltiples objetivos específicos y cada uno de ellos debe aferrarse a una metodología de campo, de laboratorio o de análisis particular, si bien en muchos casos se pueden emplear procedimientos semejantes.

Cuanto más clara sea la pregunta de investigación, más fácil será el planteamiento de la hipótesis. Al respecto, Salkind (1997) nos advierte de dos puntos de interés:

- No es lo mismo expresar una idea que postular una pregunta de investigación, ya que en la primera hay una actividad mental desbocada.
- La formulación escrita del proyecto, conlleva un proceso de orden mental, de reflexión y de compromiso.

Haciendo la salvedad que algunos autores denotan un orden diferente, el proceso formal que da lugar a los objetivos es el siguiente:

IDEA - PREGUNTA - HIPÓTESIS - OBJETIVO

Algunas de las hipótesis y objetivos surgen de las siguientes situaciones:

- Se detectan problemas en las teorías vigentes.
- Se quieren poner a prueba teorías en desarrollo.
- Se desea fortalecer una teoría colocándola a prueba sobre nuevas poblaciones o comunidades.
- Se pretende caracterizar o evaluar la dimensión de un fenómeno.
- Se busca la causalidad de un fenómeno.
- Se desea corroborar una tendencia.

La calidad de las hipótesis está relacionada directamente con el grado de profundidad con que se haya revisado la literatura. Nos recuerda también que pueden tener una función descriptiva o explicativa y que mientras algunas investigaciones contienen gran variedad de hipótesis otras contienen una o dos, sin que la calidad de la investigación esté asociada al número de ellas.

Ahora bien, la forma precisa en que se formule o escriba una hipótesis y de allí un objetivo, incide de manera contundente en cómo deberá abordarse la metodología. Este punto es esencial porque el objetivo define el modo en que deberán tomarse los datos y, a su vez, condiciona lo que se hará con ellos.

Ideas, preguntas, hipótesis y objetivos, constituyen la vía principal de delimitación de la investigación y, a la vez, le dan el norte definitivo a la misma.

b. Lluvia de Ideas

Las preguntas de investigación pueden surgir espontáneamente antes o durante la revisión bibliográfica, a partir de las dudas, vacíos, contradicciones o tendencias encontradas en la literatura. Señalan que las ideas pueden surgir también de las experiencias individuales, los materiales escritos, descubrimientos de otras investigaciones, conversaciones, observaciones de hechos, creencias e incluso presentimientos. Estos autores añaden, que la mayoría de las ideas iniciales son vagas y deben trabajarse hacia planteamientos más estructurados lo cual se logra, con mayor facilidad, en cuanto mejor se conoce el tema.

En caso de que el estudiante exprese dificultades para plantear preguntas de investigación, puede acudir a formular una lluvia de ideas propias, que podría enriquecer con compañeros, profesores o tutores.

La lluvia de ideas trata de formular numerosas ideas, teniendo en cuenta que por tratarse de ideas sueltas aún no analizadas en profundidad, algunas pueden resultar prácticas, objetivas y viables, pero otras ir realizables, fantasiosas o poco importantes. El paso siguiente es, entonces, filtrar o tamizar las mismas respecto a su pertinencia y viabilidad.

c. Diagrama de Modelación de la Variable Objeto del Estudio

Para la formulación de las hipótesis y los objetivos, también es de gran ayuda la diagramación de un modelo teórico que dé razón de las más importantes variables de incidencia del fenómeno que se estudia. Tal modelo permitirá reconocer las variables de mayor importancia y las relaciones existentes entre ellas.

d. Pertinencia y Viabilidad

Algunas de las variables que se quieren estudiar pueden resultar muy costosas; asimismo, otras pueden requerir de equipos especiales que no poseemos y debemos comprarlos, alquilarlos, importarlos, etc. y, por tanto, estas limitaciones deben también ser sopesadas.

Por las vías previamente planteadas, es posible obtener un listado de objetivos tentativos que requieren de su inmediata confrontación frente a factores como justificación, equipos, materiales, personal, logística, costos y tiempo requerido. A este tamiz sobrevive un número menor de propósitos que constituirán los objetivos específicos de la investigación.

e. Relación con el Marco de Referencia

Hay que recordar que los marcos teórico y conceptual están asociados a los objetivos de la investigación por lo que es más conveniente escribirlos una vez formulados éstos e igual ocurre con los marcos demográfico y geográfico. No obstante, los marcos también pueden escribirse antes de los objetivos, dependiendo del punto de entrada que haya tenido la investigación. Situación similar puede ocurrir con los marcos legal e institucional.

f. Dificultades en la Definición de Hipótesis u Objetivos

Las dificultades que experimenta el estudiante o el investigador en la precisión de las hipótesis y los objetivos del proyecto se resumen en los siguientes puntos:

- Falta de revisión bibliográfica.
- Falta de la utilización del marco teórico o de antecedentes.

- Falta de conocimiento respecto a las relaciones entre las variables.
- Falta de conocimientos de los procesos metodológicos-campo, laboratorio, análisis estadísticos.
- Falta delimitar el tema.
- Desconocimiento de procedimientos que le ayuden a formular hipótesis.
- Poca dedicación al proyecto.

g. Operacionalización de las Variables

Las variables se utilizan para designar cualquier característica o cualidad de la unidad de análisis y son los elementos principales de los objetivos y las hipótesis.

La operacionalización consiste en definir las variables a medir y la forma en que vamos a hacerlo, como también, las unidades y los indicadores de interpretación que le vamos a dar: cantidad o cualidad del atributo, unidad y forma de medida, o expresión matemática. Durante este proceso las variables se transforman de un nivel abstracto a un nivel empírico, observable y medible. Para tal proceso se elaboran definiciones de los conceptos o *marco conceptual*, en términos de posibilitar la medición y observación de los mismos.

5. Justificación

La validación de los objetivos se da a partir del *por qué*, el *para qué*, o *qué importancia* tiene el estudio. Las investigaciones generalmente no obedecen tan sólo a un capricho de una persona y, por tanto, debe haber una razón lo suficientemente fuerte que justifique su realización. Debe describirse entonces, por qué es conveniente la investigación y cuáles son los beneficios que se podrían derivar de ella.

Se resumen algunos criterios a tener en cuenta a la hora de elegir el tema:

- Conveniencia.
- Relevancia social.
- Implicaciones prácticas.
- Valor teórico.
- Utilidad metodológica.
- Interés del investigador.
- Necesidades concretas del medio, y/o interrogantes de la ciencia.
- Nuevo enfoque del tema.
- Los resultados serán utilizados por otros investigadores o ayudarán en la toma de decisiones.

6. Metodología

En este aparte de la propuesta se deben plantear las técnicas y los instrumentos que utilizará en la investigación, tanto en el campo, en el laboratorio y en el análisis de la información. Ésta es quizá, la parte más importante de toda la propuesta, ya que obliga establecer el enlace entre lo que quiere lograr y el cómo lograrlo. Es fundamental, entonces, que en la investigación se integren, amarren y relacionen los siguientes tres constructos:

OBJETIVOS - TOMA DE DATOS - ANÁLISIS DE INFORMACIÓN

Con frecuencia los estudiantes se fían de los programas de computador que permiten procesar información cuantitativa o que realizan análisis estadísticos y, por ello, no reparan con el suficiente juicio en los procedimientos de análisis. Por tal razón, en forma recurrente se encuentran con gran cantidad de datos sin saber qué hacer con ellos, o peor aún, corriendo múltiples opciones en tales programas sin tener remota idea de lo que hacen. Por tanto, ni el computador ni el *software*, garantizan una buena investigación. Es común también encontrar que formulan un objetivo pero no capturan los datos de forma correcta para dar cumplimiento al mismo.

Es fundamental exigir a los investigadores, antes de iniciar su estudio, total claridad y transparencia en su proceso metodológico, incluidos los análisis estadísticos. De hecho, en el análisis de información se puede incurrir en múltiples errores como cálculos mal realizados, uso inadecuado de la estadística, mala interpretación de resultados, elección inapropiada de la prueba o inferencias no válidas de muestra a población.

La metodología debe ser lo suficientemente clara y explícita para que otro investigador pueda replicar tal estudio. Así mismo, debe dar cuenta de cuál es la población y describirla, el tipo de muestreo que se habrá de implementar, el tamaño de la muestra, su nivel de confianza y el error de muestreo. Además, se debe especificar si los instrumentos y procedimientos que se emplearán han sido validados por otros autores o si se están diseñando y construyendo para dar cuenta de esta investigación particular.

Lo dicho nos alerta sobre la naturaleza y la cantidad de las variables que habremos de estudiar, su nivel de profundidad, las técnicas, los instrumentos y los análisis cualitativos o cuantitativos a implementar. El *diseño metodológico* constituye la mejor estrategia a seguir para dar solución a los objetivos planteados y comprende la definición y secuenciación de un conjunto de actividades particulares.

Se sugiere, en la medida de lo posible, realizar una visita previa al lugar de estudio para reconocer los elementos más importantes del sistema, así como para coordinar aspectos logísticos. Además, el investigador deberá tener gran familiaridad con las técnicas y los instrumentos de muestreo.

Se recomienda ampliamente no iniciar la investigación, la toma de datos, hasta que no se hayan resuelto todas y cada una de las inquietudes expuestas a continuación:

- ¿Qué variables se van a medir o a estimar?
- ¿Cómo se va a tomar la información de campo?
- ¿Con qué instrumentos se van a medir las diferentes variables?
- ¿Cuáles son los procedimientos de laboratorio?
- ¿El estudio es de carácter espacial, temporal o espacio-temporal?
- ¿Cuál es la muestra y cuál es la población?
- ¿Se va a realizar un censo o un muestreo?
- ¿Cómo evaluar la representatividad de la muestra?
- ¿Cuál es la unidad de estudio?
- ¿Es un estudio experimental, cuasiexperimental, preexperimental o no experimental?
- ¿Qué análisis estadísticos o numéricos se van a realizar con los datos?
- ¿Responden los datos a los objetivos planteados?
- ¿Qué cambios metodológicos se deben contemplar entre unos y otros objetivos?
- ¿Qué tipo de transformaciones se deben realizar sobre los datos crudos?

Para el diseño metodológico incluidos períodos, estaciones, muestras, instrumentos o análisis de información, el investigador cuenta con los antecedentes del tema que, aunque no lo obligan a emular fielmente los mismos, sí le dan amplias luces como punto de partida.

La población es el conjunto de todos los elementos que se quiere estudiar, mientras que la muestra son los elementos que efectivamente se estudian; a partir de esta última se hacen inferencias sobre la población cuando la muestra es representativa. Cuando se evalúa toda la población corresponde a un *censo* y los parámetros poblacionales quedan *determinados*. Cuando, por el contrario, se hace un levantamiento parcial de la información o se estudia una porción representativa que refleja las características de la población, se trata de una *muestra* y los parámetros poblacionales son *estimados*.

En este punto es de gran importancia señalar, que la muestra es representativa de la población en tanto se verifique que tal afirmación es cierta. Por ende, la validez de la generalización depende de la representatividad de la muestra y los estudios que no evalúen la misma, se constituyen en *casos de estudio* que no deben pretender inferencias por fuera de los individuos evaluados.

Ello no significa que tales estudios sean inválidos, sino que los resultados competen solamente al grupo estudiado. En muchas situaciones más que una generalidad, el propósito de una investigación puede recaer en el estudio de elementos con características específicas y, de hecho, el conocimiento se ha construido a partir de pequeñas investigaciones que al replicarse en múltiples sistemas, han permitido generar bases y cimientos sólidos para la ciencia.

Es importante que el investigador incorpore en su proyecto el tipo de tablas, encuestas o entrevistas que habrá de utilizar en su toma de información, así como los estadísticos que aplicará en cada caso.

7. Título

La elaboración del título solamente es posible cuando ya se han definido las principales características del proyecto, en lo que respecta a hipótesis-objetivos, metodología y marcos geográfico y demográfico. El título debe escribirse en forma declarativa, a su vez, debe ser atractivo, lo más explícito posible y sin adornos literarios ni nombres fastuosos. Debe referirse solamente al tema principal evitando iniciar con artículos y preposiciones. Se debe tratar de iniciar la primera frase o palabra por el tema central.

El título de un proyecto debe ser ante todo, corto y claro; por tanto, debe evitar contener información irrelevante a la vez que refleja *qué se va a hacer, dónde* y, si es pertinente, *cuándo o con quién*. El título semeja en gran medida al objetivo general de la investigación, aunque no sigue la norma de iniciar con un verbo en infinitivo sino con un sustantivo derivado de aquél, como:

- Determinación de ... -determinar-
- Uso de ... -usar-
- Análisis de ... -analizar-, etc.

Otros ejemplos que no necesariamente siguen la regla anterior son los siguientes:

- Efecto de ... - Estructura de ... - Tasas de ... - Plan de ...
- Incidencia de ... - Crecimiento de ... - Tendencias de ... - Propuesta de ...
- Factores que ... - Monitoreo de ... - Indicadores de ... - Diagnóstico ...
- Cambios de ... - Distribución de ... - Alternativas de ...

Con frecuencia el título del proyecto se modifica entre el proyecto de investigación y la investigación, dado que sobre ésta se reevalúan hipótesis, objetivos o metodologías que modifican el propósito original. Un marco de referencia pobre, un método poco estudiado o el desconocimiento de la calidad o la población objeto, pueden llevar a modificaciones importantes en el estudio y, con ello, en el título del mismo.

Para la elaboración del título conviene iniciar identificando y escribiendo algunas de las palabras *claves* que se considera describen más el tema de estudio; luego, éstas se precisan y se jerarquizan, eligiéndose las más relevantes y omitiendo las restantes; finalmente, se redacta adecuadamente con las palabras elegidas. Una vez que se define el título tentativo, se debe tratar de reducir su extensión sin perder información relevante.

8. Marco de Referencia

El marco de referencia inscribe el estudio dentro del conjunto de conocimientos, variables, conceptos, hipótesis y teorías desarrolladas por otros investigadores.

Este párrafo suele ser trabajado con diferentes nombres como marco de referencia, marco teórico o marco conceptual. En este documento se ha elegido *marco de referencia* el cual es más universal e involucra los siguientes ítems:

- Marco de antecedentes.
- Marco teórico.
- Marco conceptual.
- Marco demográfico.
- Marco geográfico.
- Marco institucional.
- Marco legal.

a. Marco de Antecedentes

El investigador recopila información sobre investigaciones previas que ya han abordado dicho tema desde múltiples y diversas perspectivas. Es decir, reconstruye los antecedentes históricos con el nombre de los investigadores, los enfoques que han dado, las hipótesis que han planteado, sus resultados, métodos y técnicas utilizadas, a la vez que permite reconocer problemas que se han suscitado en el desarrollo de las mismas.

La revisión bibliográfica arroja información sobre las poblaciones y las condiciones geográficas, el tipo de variables medidas y las relaciones empíricas supuestas o demostradas. Este marco de referencia, al cual llamamos antecedentes, nos sitúa a la vanguardia del conocimiento en el tema de estudio. Cabe destacar que no toda la información consultada será parte de nuestro estudio, por lo que debemos seleccionar aquella que resulte pertinente.

El marco de antecedentes debe incluir información relativa a:

- Autores que trataron el tema.
- Investigaciones en torno al tema.
- Épocas y enfoques dados al problema.
- Abundancia o escasez de la literatura en relación con el problema de investigación.

Los antecedentes constituyen los hechos anteriores que sirven para aclarar, juzgar e interpretar el problema planteado; no se trata de un recuento histórico, ni de presentar fuentes bibliográficas posibles o volúmenes de datos recolectados. Se trata de hacer una síntesis de las investigaciones o trabajos realizados sobre el tema de estudio, con el fin de determinar el enfoque metodológico.

Lo que pretenden los antecedentes de un estudio es que el investigador exprese qué tanto sabe y entiende del tema, a fin de que demuestre si cuenta con los elementos de juicio necesarios para la realización de la investigación propuesta. Para la escritura de los antecedentes se sugiere iniciar con el tema general e ir, poco a poco, procurando más especificidad sobre el tema de estudio.

b. Marco Teórico

Constituye una descripción detallada de los elementos esenciales de la(s) teoría(s). Cada ciencia formula sus propias teorías de acuerdo con el avance que demuestre en ellas. Las mismas varían en generalidad y si bien algunas son universales, otras son individuales. Las teorías universales que cobijan a la ciencia de estudio no son tan necesarias de plasmar, mientras que aquellas más restringidas, conflictivas y pertinentes al tema de estudio si deben presentarse.

A medida que se avanza en la revisión bibliográfica, el investigador va conociendo dichas teorías y debe ir profundizando en ellas, es decir, debe dirigir parte de la revisión bibliográfica a consultar material que presente o analice las mismas. La literatura revisada puede revelar que hay *teorías completamente desarrolladas* y con abundante evidencia empírica; hay fragmentos con *apoyo empírico* moderado o hay *solamente* ideas vagas sobre el tema.

Si la teoría está muy consolidada, debemos darle un nuevo enfoque a nuestro estudio planteando interrogantes que no se han podido resolver, o podemos contribuir con base empírica en aquellas teorías que parecen sólidas pero que carecen de suficientes investigaciones de campo, o que requieren de su comprobación en otras condiciones. Si se toman partes de diferentes teorías, hay que tener precaución de no generar contradicciones lógicas entre ellas.

En muchas áreas del conocimiento hoy nos encontramos con generalizaciones empíricas más que con teorías, por lo que se debe construir una *perspectiva teórica* que incluya información relativa a los resultados y a las conclusiones a los que han llegado otros estudios anteriores.

El marco teórico establece la frontera o límite hasta dónde llegará la investigación. Este marco integra la teoría existente con nuestra investigación. Sus funciones son:

- Delimitar el área de la investigación.
- Sugerir guías de investigación.
- Compendiar conocimientos del tema a investigar.
- Expresar proposiciones teóricas y postulados que servirán como base para formular hipótesis, operacionalizar variables y esbozar técnicas y procedimientos.

El marco teórico debe incluir información relativa a: Teorías básicas, Solidez de las teorías que sustentan la investigación, Claridad y coherencia de los principios postulados y supuestos sobre los cuales se apoya la investigación.

c. Marco Conceptual

Tiene como objetivo suministrar información sobre las principales definiciones con las cuales se abordará el estudio o problema de investigación, consiste en presentar los rasgos principales de la estructura de un concepto para hacerlo más preciso, delimitándolo de otros conceptos, a fin de hacer posible una exploración sistemática del objetivo que representa.

El marco conceptual debe incluir información relativa a:

- Definición de términos básicos
- Definición de conceptos
- Definición de las variables

En él aparecen las definiciones de las variables y de los objetivos contemplados en el problema de investigación, así como de los términos que van a ser utilizados con mayor frecuencia o que pueden suscitar confusión. Tales definiciones las hace el investigador de acuerdo a su criterio, a definiciones propuestas por otros investigadores o a la teoría en la que se apoya la investigación.

Se recomienda, para la elaboración del mismo, hacer un listado de las palabras claves que enmarcan el estudio y consultar textos especializados para los términos básicos. Por tanto, es esencial en el desarrollo de cualquier investigación, que el investigador tenga una apropiación suficiente sobre el vocabulario científico de su área de conocimiento.

d. Marco Geográfico

La *zona geográfica* se define como la circunscripción específica del lugar, espacio físico o ámbito donde se realizará la investigación. Antes de iniciar el estudio, el investigador debe establecer y, en lo posible, visitar el espacio geográfico y físico en el que se desarrollará el estudio, por cuanto la misma habrá de incidir notablemente en los objetivos y en la metodología.

La demarcación de la zona geográfica suele presentar las coordenadas geográficas, un mapa o ambas. Dependiendo de la disciplina de estudio, se incluirá un mapa con escalas entre 1:1'000.000 y 1:50.000 o menos, presentando, en cada caso, escala, coordenadas, norte y área de observación, más aspectos relevantes propios de cada estudio, vías, ríos, ciudades, barrios, cotas de nivel, hospitales, centros educativos, etc.

En algunas ciencias quizá no tenga mucha importancia describir con mayor detalle este marco, por cuanto la investigación se lleva a cabo, por ejemplo, de modo experimental en un laboratorio.

e. Marco Demográfico

Este marco contiene las características demográficas concernientes a la población que se va a estudiar.

En ciencias humanas, describe las principales características sociales, económicas, de edad y de género. Las mismas, generalmente, son tomadas de información secundaria relativa a censos u otros estudios previos. En caso de que tal información no se encuentre descrita, es labor del investigador construirla e incluir la, naturalmente dentro de unos límites de costos y tiempo apropiados. Es decir, si el estudio se va a centrar en indagar sobre una condición

social particular, el proyecto no debe describir tal condición como resultado de encuestas, estadísticas u otros procesos costosos y arduos, porque precisamente ese será el propósito de la investigación; no obstante, sí puede preanunciar la condición observada en la región, la localidad u otro.

Aspectos como si se trata de una población urbana o rural, con ciertos niveles de escolaridad, con tales o cuales tipos de empleo, con ingresos altos, medios o bajos, etc., aportarán la información requerida por este marco.

En caso de estudios sobre instituciones específicas, se denotarán los aspectos que se puedan derivar de charlas con personas que trabajen en dicha institución, y que incluyan, por ejemplo, el número de personas que allí estudian o trabajan, sus edades, distribución por género u otros.

En estudios biológicos de campo, este marco suele referirse a los ecosistemas propiamente y no a la población humana. Si no hay información puntual de dichos ecosistemas, se puede recurrir a descripciones más globales en latitudes próximas.

f. Marco Institucional

En este párrafo se enuncia la institución o instituciones donde se enmarca el estudio. Puede tratarse tan sólo de una requisición académica para optar a un título y en tal caso de describe así un par de líneas, o puede tratarse de un proyecto en el que participan diferentes instituciones aportando fondos, personal o infraestructura y deberá describir se la participación y los compromisos de cada una de ellas.

g. Marco Legal

La realización de algunos proyectos de investigación como, por ejemplo, aquellos que se realizan en áreas protegidas, resguardos indígenas o parques nacionales, deben incluir las Leyes, Decretos o Normas que aplican en cada caso. Algunos permisos adicionales pueden ser requeridos y debe exponerse el marco legal que sustenta los mismos, así como también, si tales permisos ya han sido solicitados u otorgados. De igual modo, estudios al interior de instituciones o empresas requieren, muy seguramente, de autorizaciones para ello, por lo que se recomienda no dejar para último momento tales permisos, pues la negación haría perder todo el proceso avanzado.

9. Introducción

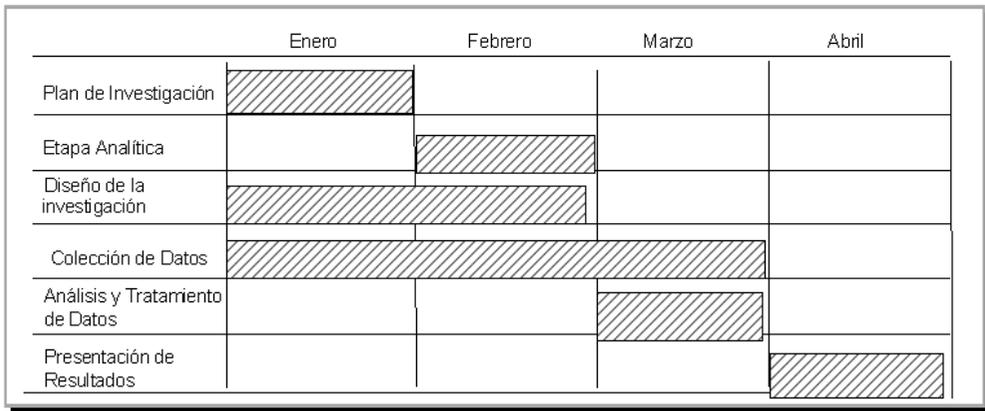
Introduce al lector en el tema de investigación presentándole en forma rápida y sintética la problemática a tratar. Debe procurarse un lenguaje sencillo, coherente y cautivador, ya que este párrafo es la puerta de entrada al documento, por lo que define la primera impresión del lector. Este ítem, no obstante ser el primero del proyecto y de la investigación, se escribe solamente cuando todo el panorama de estudio está claro para el investigador.

Un formato general sugerido para la elaboración de la introducción es el siguiente: iniciar con un párrafo general de la temática, seguido de la importancia de la misma, unos pocos antecedentes relevantes acompañados de los principales teóricos y los problemas teóricos o conceptuales sobre el tema. Se sigue con dificultades temáticas, falencias o necesidades de investigación, para cerrar con un esbozo de la ruta que seguir á el estudio, guardando precaución de que la escritura de este último no semeje el formato de un objetivo.

Ejemplo 2

| ACTIVIDADES | TIEMPO | FECHA | | | | |
|-----------------------------------|------------|-------|---------|-------|-------|------|
| | | ENERO | FEBRERO | MARZO | ABRIL | MAYO |
| PROPOSITO DE INVESTIGACION | PROGRAMADO | | | | | |
| | REAL | | | | | |
| REVISION DE LA LITERATURA | PROGRAMADO | | | | | |
| | REAL | | | | | |
| DISEÑO DE LA METODOLOGIA, ETC. | PROGRAMADO | | | | | |
| | REAL | | | | | |

Ejemplo 3: Gráfica de Gantt.



11. Materiales y Presupuesto

Este párrafo tiene el propósito de que el investigador determine los requerimientos y costos totales del proyecto, así como el proceso de desembolso requerido. Vale recordar que durante la definición y delimitación de los objetivos del estudio, el investigador ha debido sopesar la viabilidad de su ejecución con base en los métodos a seguir y, con ello, los equipos y recursos económicos demandados, incluidos también los costos de transporte, personal, servicios técnicos a contratar, alojamiento y alimentación, principalmente.

Una vez el investigador define el costo total del proyecto, es necesario que establezca si cuenta con tales recursos. Cuando el proyecto resulta muy costoso en el marco de una financiación personal o externa, se deben revisar y replantear los objetivos para amoldar los mismos a un presupuesto más ajustado. Ello puede conllevar a la necesidad de reducir estaciones, muestras, objetivos, variables o períodos de estudio, o puede llevar incluso a cambiar el lugar donde se va a ejecutar el proyecto.

En término general, la reducción de los alcances del proyecto debe hacerse con cautela para que no se debilite, menoscabe o invalide la finalidad del estudio mismo. Se recomienda reducir el número de objetivos manteniendo, sin embargo, un alto nivel de profundidad en los elegidos, en lugar de pretender abarcar muchos propósitos pero con tan poco nivel de confiabilidad en ellos, que no permitan formular conclusiones o inferencias de ningún tipo.

Así por ejemplo, en el estudio del fitoplancton superficial del lago de Titicaca, se podría reducir el número de estaciones a la mitad para disminuir los costos del proyecto; no obstante, si el número de estaciones que subsisten no aporta una representación adecuada del lago, el estudio no podrá inferir sobre la estructura de la comunidad en dicho ecosistema y corre el riesgo de invalidarse por falta de representatividad.

a. Materiales

En este ítem se expone un listado de todos los materiales y equipos que se requieren para la ejecución del estudio, así como la persona o entidad que aportará los mismos. Se sugiere hacerlo en forma de tabla.

b. Presupuesto

Los ítems que deben ser presentados en el proyecto son los correspondientes a los costos directos, tales como:

- *Personal*: auxiliares, servicios técnicos y el costo del investigador.
- *Equipos*: computadores, vehículos, grabadoras, filmadoras, cámaras fotográficas, microscopios, telescopios u otros.
- *Transporte y alojamiento*: incluye gastos de movilización, alimentación y hospedaje.
- *Gastos varios*: fotocopias, planos, fotografías aéreas, software, libros.
- *Material fungible o elementos de consumo*: se refiere a los materiales que se gastan o acaban: papel, bolsas, reactivos, tinta, rollos fotográficos, etc.

El presupuesto del proyecto debe incluir no sólo los ítems que generan gastos, sino también sus cantidades. Se sugiere además, incluir un ítem de *imprevistos* para cubrir aquellos gastos que por descuido no fueron tenidos en cuenta o cuyos costos pueden incrementarse sobre el transcurso de la investigación.

Otros tipos de costos diferentes a los previamente referidos son los *indirectos*, pero éstos son propios de instituciones donde, además de los costos directos del proyecto, se incurre en costos financieros y de administración, servicios, mensajería, impuestos, prestaciones, etc. Es importante diferenciar el costo total del costo en que debe *incurrir el investigador*, ya que este rubro juega un rol importante en la viabilidad del proyecto. Tales costos se deben presentar en forma de tabla.

12. Bibliografía

Presenta solamente los documentos que fueron consultados y referenciados dentro del proyecto, por lo que aquellos que fueron consultados pero no fueron pertinentes al estudio, se excluyen. Algunos de ellos pueden, sin embargo, resultar apropiados para el trabajo final en el marco de la discusión de resultados.

No existe ningún límite mínimo o máximo respecto al número de libros y artículos científicos que deben ser consultados para la formulación de la propuesta, aunque se sugiere que tal cifra no sea inferior a veinte, ello con el fin de que el recorrido del estudiante haya sido suficientemente amplio. El director del proyecto será quien, en primer lugar, determine si la bibliografía consultada es suficiente, así como también, si se deben hacer consultas adicionales de autores importantes en el tema.

La bibliografía se presenta al final del documento y antes de los anexos si los hubiere. Los elementos principales que hacen parte de cada referencia bibliográfica son los siguientes: autor, título de la obra, número de edición, complementarios, traducción, edición, corregida, etc., lugar de edición, ciudad, nombre de la editorial, fecha de edición y número de tomos de la obra. Vale añadir, que en cada referencia se presentan todos los autores y no como ocurre en el cuerpo del documento, donde se expresa *et al.* cuando se trata de más de dos autores.

La forma de presentar una referencia bibliográfica cambia ligeramente según la casa editorial, la revista, la universidad o el país y, por ello, no hay un único formato de cita, si bien el contenido general es muy similar. Las referencias bibliográficas se deben presentar ordenadas alfabéticamente por apellido del primer autor.

CONTENIDOS DEL INFORME FINAL DE INVESTIGACIÓN

Cabe referir, como punto inicial, que el grado de profundidad de la investigación va de acuerdo con el nivel de los estudios o del grado alcanzado, bachiller, pregrado, maestría, doctorado y, por tal razón, no se puede hablar de un nivel de profundidad particular ni de un tiempo específico para ello. En trabajos de grado de nivel de doctorado es esencial que el estudiante demuestre que puede realizar algún aporte a su disciplina. Aun así, unas y otras deben experimentar un proceso semejante enmarcado en un procedimiento juicioso y coherente.

PRIMERA ESTRATEGIA DE INVESTIGACIÓN

El Informe Final es un **resumen** de todo el trabajo de investigación realizado y presenta la estructura global (completa) del proceso, las bases del tema de investigación estudiado, el marco teórico de referencia, la descripción detallada de la metodología aplicada, una síntesis de los principales hallazgos empíricos y su interpretación a la luz del marco teórico y, finalmente, el razonamiento del investigador reflejado en las conclusiones y prospectiva a seguir en procesos de investigación similares o en aplicaciones concretas del conocimiento producido.

Es fundamental que el informe sea elaborado en un **muy cuidado lenguaje**, centrado directamente en las cuestiones clave (concreto y preciso) y demostrando la sistematicidad y rigurosidad propios de todo proceso de investigación científica. Igualmente, teniendo en cuenta los criterios de redacción (formato) estandarizados internacionalmente para que el documento tenga valor y reconocimiento en la comunidad científica y, por tanto, en la sociedad.

1. Introducción.
2. Justificación.
3. Planteamiento del problema.
4. Objeto de estudio.
5. Cuestiones a responder mediante la investigación.
6. Objetivos.
7. Marco teórico.
8. Metodología y desarrollo de la investigación.
9. Presentación y discusión de resultados.
10. Conclusiones.
11. Referencias bibliográficas.

1. Introducción

Contiene una **descripción razonada** de todo el documento, de modo tal, que el lector tiene una visión panorámica completa de la investigación (tema, bases teóricas, hallazgos). Adicionalmente se presenta una descripción de la estructura del informe propiamente como tal.

2. Justificación

Es retomada del proyecto y mejorada desde el punto de vista estético únicamente, pues los argumentos de la investigación se mantienen.

3. Planteamiento del Problema

Considera el texto presentado en el proyecto y es mejorado desde el punto de vista estético en estudios cuantitativos y ajustado si se trata de estudios cualitativos.

4. Objeto de estudio

Considera el texto del proyecto.

5. Cuestiones a responder mediante la investigación

Se derivan de las preguntas de investigación presentadas en el proyecto y sirven de base para la formulación de los objetivos de investigación (coherencia).

6. Objetivos

Considera exactamente el texto del proyecto en caso de investigación **cuantitativa**; en una investigación **cualitativa** en cambio, los objetivos han podido ser ajustados a lo largo del proceso de investigación respecto de los objetivos presentados en el proyecto aunque no modificados en su intención original.

7. Marco teórico

Contiene los capítulos que **resumen** la amplia revisión documental realizada por el investigador mediante la cual ha caracterizado el tema de investigación y establece su posicionamiento teórico.

Igualmente aporta los antecedentes teórico conceptuales y empíricos que en conjunto sostienen la investigación y sirven de referencia para interpretar los datos recogidos. En una investigación **cuantitativa** el marco teórico sirve de base para el establecimiento de variables de estudio, en un estudio **cualitativo** el marco teórico es referencial sobre la realidad estudiada y no funge como base de operacionalización de variables.

8. Metodología y desarrollo de la investigación

Describe **pacientemente** el tipo de investigación elegida (**cuantitativa o cualitativa**), su sólida explicación-fundamentación, la metodología del estudio, población y muestra, estrategias de recogida de información y las técnicas de análisis de datos (SPSS, Atlas ti, etc.).

9. Presentación y discusión de resultados

Sintetiza los principales hallazgos de la investigación aplicando técnicas didácticas de presentación de la información (gráficas, tablas, cuadros, etc.) y presenta una potente interpretación teórica que demuestra el dominio técnico del investigador, la utilidad del marco teórico en la comprensión de la realidad y la ilustración de caminos a seguir en posteriores estudios y aplicaciones prácticas.

10. Conclusiones

Destila **lo esencial de todo el proceso** enfatizando especialmente la riqueza de la evidencia empírica aportada y, a partir de ello, presenta la perspectiva que el investigador contribuye desde su propia reflexión.

11. Referencias bibliográficas

Expone las fuentes documentales teóricas y metodológicas utilizadas en el estudio y cumple con las **normas internacionales/estandarizadas** de manejo y citación de referencias bibliográficas.

SEGUNDA ESTRATEGIA DE INVESTIGACIÓN

1. Indagar sobre literatura relacionada con el posible tema de estudio

Esto ayudará a formular una pregunta o hipótesis.

2. Seleccionar un tema de estudio limitado

Una vez se posee un posible tópico, es vital de delimitarlo hasta donde se pueda. El fin es poder establecer y definir efectivamente la interrogante investigativa.

3. Plantear el problema de investigación

Se expone el propósito del estudio, basado en una pregunta de investigación. Es de suma importancia que el problema se exprese claramente y se establezca una relación entre las variables a ser estudiadas.

4. Se determina la hipótesis

Se expone y defiende el punto de vista del investigador, ofreciendo argumentos lógicos que justifiquen el estudio.

5. Se definen las variables del estudio

El investigador instituye las variables independientes y dependientes.

6. Observación y registro de las variables durante el estudio

Se inicia el experimento científico y se colectan los datos de la muestra mediante los dispositivos de medición, previamente calibrados y validados.

7. Ordenación, clasificación, organización y análisis de la información obtenida

Los datos crudos son tabulados y analizados. Se generan inferencias, conclusiones, recomendaciones y se establece si la hipótesis es aceptada o rechazada.

TERCERA ESTRATEGIA DE INVESTIGACIÓN**1. Título**

Tema de la investigación.

2. Marco teórico

Justificación, antecedentes y pregunta de investigación.

3. Objetivos

General y específicos.

4. Hipótesis

Depende del tipo de estudio.

5. Material y método

Metodología que se lleva a cabo para la ejecución de la investigación, es decir, se describe la forma paso a paso.

- Definición operacional de variables (según tipo: discreta, continua, categoría, también como dependiente, independiente) En esta parte también se incluye algo muy importante de la estadística.
- Cálculo de tamaño de muestra.
- Esquema de muestreo.
- Plan de análisis estadístico.

6. Resultados**7. Conclusiones****ESTRATEGIA DEL AUTOR**

La investigación en la actualidad va más allá de lo científico y académico como es el grado alcanzado, bachiller, pregrado, maestría, doctorado la investigación es importante en el campo empresarial por que constituye la herramienta ideal para resolver los problemas que surgen diariamente en las empresas y organismos oficiales permitiendo tomar un acertada y oportuna decisión.

El modelo o estrategia está desarrollado de forma teórica y práctica, con ejercicios fáciles de entender.

Para poder desarrollar un trabajo de investigación se recomienda seguir los pasos del modelo o estrategia siguiente.

1. Planteamiento del problema:

El problema es el punto de partida de la investigación. Surge cuando el investigador encuentra un vacío teórico, o un hecho no abarcado por una teoría, un tropiezo o un acontecimiento que no encaja dentro de las expectativas en su campo de estudio. Es aquí donde se perfila el título y el marco teórico. El tema se desarrollara ampliamente en el Capítulo III.

1.1 Idea: Se expone el propósito del estudio, basado en una pregunta de investigación. Es de suma importancia que la idea se exprese claramente.

Pregunta: ¿Qué realidad me interesa investigar?

1.2 Objetivos: Especifica qué se intenta conseguir con el trabajo de investigación; constituye la directriz de la investigación, señala los resultados o metas.

1.3 Hipótesis: Se expone y defiende el punto de vista del investigador, ofreciendo argumentos lógicos que justifiquen el estudio.

2. Metodología y Diseño muestral (Población y muestra):

Definir y calcular la población, la muestra, unidad de estudio (individuo), las variables, el tamaño de muestra y el tipo de muestreo que se aplicará al problema o experimento.

2.1 Población: Conjunto de todos los individuos (personas, objetos, animales, etc.) que porten información sobre el fenómeno que se estudia.

2.2 Muestra: Constituye una selección al azar de una porción de la población.

2.3 Individuo: (Observación, Caso, Sujeto): Cualquier elemento que aporte información sobre el fenómeno que se estudia.

2.4 Variable: La variable es toda aquella propiedad de algún objeto, persona o evento que posee diferentes valores.

2.5 Tamaño de Muestra: En una investigación el tamaño de muestra es muy importante. Teniendo en cuenta que la *calidad y validez de los resultados de una investigación depende del tamaño de muestra*. Una muestra demasiado grande implica un desperdicio de recursos y una muestra demasiado pequeña disminuye la utilidad de los resultados.

2.6 Tipo de Muestreo: La muestra se puede elegir de diferentes maneras teniendo presente el tipo de muestreo (muestreo aleatorio simple, muestreo sistemático, muestreo estratificado, muestreo por conglomerados, muestreo mixto).

3. Recolectar Datos o información

Recolectar los datos implica seleccionar un instrumento de medición disponible o desarrollar uno propio. Es decir, elaborar el cuestionario y recolectar la información.

3.1 Instrumento de medición (Cuestionario): Cualquier recurso de que se vale el investigador para acercarse a los fenómenos y extraer de ellos información.

Cuestionario: Representa un conjunto de preguntas respecto a una o más variables a medir, utilizada para sacar información de cualquier fenómeno.

3.2 Equipos de Medición: Cualquier equipo que permita apoyar la medición de variables (grabadora de audio o vídeo, computador, satélite, gráficos, informes, expedientes, estereoscopio, etc.).

3.3 Codificación de datos: Codificar consiste en asignarle un número o letra diferente a las posibles respuestas (alternativas) de una pregunta.

4. Procesamiento de Datos: Consiste en ingresar los datos recolectados en el cuestionario a un software de su preferencia. Si los datos son de un cuestionario desarrollado, llamaremos a este proceso *introducción directa de datos*. Una investigación también se realiza con información ya existente en un soporte electrónico de datos (archivos de excel, sql, lotus, texto, etc.); a este proceso lo llamaremos *extrayendo un archivo de disco*.

5. Análisis de Resultados y conclusiones: Ingresados los datos es hora de sacar los resultados. Los Resultados muestran los principales hallazgos de la investigación aplicando técnicas estadísticas (tablas, cuadros, gráficas, etc.) y presenta una potente interpretación teórica que demuestra el dominio técnico del investigador.

Las **conclusiones** extraen lo esencial de todo el proceso enfatizando especialmente la riqueza de la evidencia empírica aportada.

Las técnicas estadísticas utilizadas son: Estadísticas Descriptivas, Correlación, Regresión, Prueba de hipótesis, ANOVA (Modelos Lineales), Prueba no paramétrica, etc.

OBSERVACIÓN

Como verá la estadística es un elemento muy importante de la investigación científica. Comúnmente, los experimentos científicos producen observaciones o hallazgos que se expresan como datos numéricos.

Las mediciones que se obtienen de los sujetos bajo estudio constituyen los datos. Generalmente, los datos son medidas de la variable. Los datos crudos u originales consisten de aquellos datos obtenidos directamente del experimento, es decir, no han sido sometidos a tratamiento estadístico. El propósito de las investigaciones es poder formular generalizaciones basado en un conjunto de observaciones que se extraen del estudio (el laboratorio). Desarrolla y se convierte en la fuente de generación de pensamiento libre y útil. Esto significa que el científico generaliza hacia la población bajo estudio.

Capítulo

3

Planteamiento del problema

En el presente capítulo aprenderemos a plantear un problema de investigación, resolveremos ejercicios prácticos utilizando el diagrama siguiente.

Problema { **Idea** ⇨ **Pregunta** ⇨ **Objetivos** ⇨ **Hipótesis**

El problema es el punto de partida de la investigación. Surge cuando el investigador encuentra un vacío teórico, o un hecho no abarcado por una teoría, un tropiezo o un acontecimiento que no encaja dentro de las expectativas en su campo de estudio. Es aquí donde se perfila el título y el marco teórico.

- **Idea:** Se expone el propósito del estudio, basado en una pregunta de investigación, es de suma importancia que la idea se exprese claramente.
- **Pregunta:** ¿Qué realidad me interesa investigar?
- **Objetivos:** Especifica que se intenta conseguir con el trabajo de investigación, constituye la directriz de la investigación, señala los resultados o metas.
- **Hipótesis:** Se expone y defiende el punto de vista del investigador, ofreciendo argumentos lógicos que justifiquen el estudio.



PROBLEMA

El problema es el punto de partida de la investigación. Surge cuando el investigador encuentra un vacío teórico, dentro de un conjunto de datos conocidos, o un hecho no abarcado por una teoría, un tropiezo o un acontecimiento que no encaja dentro de las expectativas en su campo de estudio.

Todo problema aparece a raíz de una dificultad, la cual se origina a partir de una necesidad sin resolver.

La investigación en el campo empresarial y laboral surge cuando se presentan situaciones y dificultades de muy diversos orden, las cuales requieren una solución a mayor o menor plazo.

No importa que la investigación constituya un grano de arena dentro de la ciencia. Es preferible señalar, de acuerdo a las propias inclinaciones y preferencias, un tema reducido en extensión, es decir, poner límite a la investigación. Una de las fallas más comunes en la investigación es por ambición del tema, que consiste en la ausencia de restricción del tema. Por reducido ilimitado que puede parecer un tema, si se explora convenientemente surgirá una serie de ramificaciones que le dan importancia y valor.

El planteamiento del problema comienza cuando se define los objetivos y se plantea la hipótesis de la idea de la investigación.

1. Idea

Subrayar las ideas centrales del problema y convertir cada idea en una o más preguntas. Se expone el propósito del estudio, basado en una pregunta de investigación, es de suma importancia que el problema se exprese claramente.

¿Qué realidad me interesa investigar?

Aquí debe especificar los objetivos o propósitos de trabajo, incluyendo los problemas que se intentan resolver o las hipótesis que se intenta probar.

2. Objetivos

Toda investigación tiene objetivos generales y objetivos específicos

Objetivo general.- Especificar que se intenta conseguir con el trabajo de investigación. Por lo regular, cada investigación tiene un único objetivo general cuyo enunciado se inicia con un verbo en infinitivo, aunque puede escribirse incluso con más de un verbo: Identificar y describir. El objetivo general constituye la directriz de la investigación.

Objetivos específicos.- Señalan los resultados o metas parciales. Se enuncian e inician también con verbos en infinitivo. Cada objetivo específico debe incluir un solo logro. Una investigación puede tener uno o múltiples objetivos específicos y cada uno de ellos debe aferrarse a una metodología de campo, de laboratorio o de análisis particular, si bien en muchos casos se pueden emplear procedimientos semejantes.

¿Qué parte de esa realidad deseo investigar?

¿Qué cuestiones me interesan de esa parte de la realidad?

¿Qué debo hacer para desarrollar la investigación?

3. Hipótesis:

Se expone y defiende el punto de vista del investigador, ofreciendo argumentos lógicos que justifiquen el estudio. Los resultados de la **Hipótesis** de investigación, no siempre coinciden con los resultados reales. Debe entenderse que la **realidad es una entidad viva** y tiene voz propia y la riqueza de una investigación científica está precisamente en mostrar la realidad tal como es, no en validar una hipótesis.

¿Qué espero que suceda con mi intervención sobre la realidad estudiada?

¿Qué pienso encontrar en la realidad estudiada?

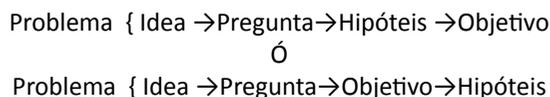
4. Seleccionar un tema de estudio limitado.

Decíamos que todo problema aparece a raíz de una dificultad; ésta se origina a partir de una necesidad en la cual aparecen dificultades sin resolver. De ahí, la necesidad de hacer un planteamiento adecuado del problema a fin de no confundir efectos secundarios del problema a investigar con la realidad del problema que se investiga. Una vez definido el problema se selecciona el posible tópico, es vital de delimitarlo hasta donde se pueda. El fin es poder establecer y definir efectivamente la interrogante investigativa. Además se escoge el posible título del problema siendo este la presentación racional de lo que se va a investigar, si el título es muy largo conviene reducirlo a pocas palabras y clarificarlo con un subtítulo.

DIAGRAMA DEL PROBLEMA

Para la formulación de las hipótesis y los objetivos de la idea de investigación, es de gran ayuda la diagramación de un modelo teórico que dé razón de las más importantes variables de incidencia del fenómeno que se estudia. Tal modelo permitirá reconocer las variables de mayor importancia y las relaciones existentes entre ellas.

Haciendo la salvedad que algunos autores denotan un orden diferente, el proceso formal que da lugar a resolver un *problema de investigación* es la siguiente.



A este proceso lo llamaremos problema de investigación, los siguientes ejercicios representan el proceso descrito:

1. Ejercicios resueltos sobre problemas de investigación

EJERCICIO 1

Problema: Estudiar el costo de la vivienda en la ciudad de Lima.

- **Idea:** Observando la situación de las viviendas en la ciudad de Lima, surge la inquietud de saber si el costo de la vivienda cambia respecto del lugar de ubicación, ya que a mayor o menor incidencia de centros comerciales podría afectar el costo.
- **Pregunta:** ¿el precio de la vivienda cambia de acuerdo al lugar de ubicación?, sirve para plantear hipótesis.

- **Hipótesis:** el precio de la vivienda de la ciudad de Lima cambia de acuerdo al lugar de ubicación. Definida la hipótesis se debe explicar el objetivo.
- **Objetivo:** evaluar el precio de la vivienda en la ciudad de Lima

EJERCICIO 2

Problema: Tácticas en un partido de fútbol.

- **Idea:** mientras contemplo un partido de fútbol me pregunto sobre la táctica que utiliza mi equipo preferido.
- **Pregunta:** ¿ha cambiado la táctica de mi equipo durante la última década?.
- **Hipótesis:** la táctica actual de mi equipo es más defensiva que la expuesta una década atrás.
- En esta hipótesis también se podría compara la táctica de un equipo frente a otro, o a tácticas empleadas en otros países, etc.
- **Objetivo:** determinar si la táctica de mi equipo de fútbol es más defensiva ahora que hace una década.

Hay que recordar que el objetivo debe ser consecuente con la hipótesis y que esta última se escribe a manera de proposición que puede ser demostrada o rechazada. Por tanto, debe tenerse precaución con los verbos utilizados en los objetivos específicos

EJERCICIO 3

Problema: Estudiar las hojas de un árbol de sauce.

- **Idea:** mientras contemplo árbol de sauce, me surge la inquietud de si el área de las hojas situadas en la parte superior cambia respecto a las de la parte inferior, ya que a mayor o menor incidencia solar podría afectar dicha área. Luego formulo la pregunta de investigación.
- **Pregunta:** ¿el área de las hojas de sauce cambia de la parte superior a la inferior? Sirve para plantear la hipótesis.
- **Hipótesis:** el área de las hojas de sauce cambia de la parte superior del árbol a la inferior del árbol. En seguida se explica el objetivo.
- **Objetivo:** determinar si el área de las hojas del sauce cambia de la parte superior a la inferior del árbol.

Observe que, en esencia, se dice lo mismo pero, de forma distinta, por lo que debe tenerse cuidado con la redacción de los mismos.

Mientras que, para otros, autores el problema se refiere solamente a la pregunta, el problema es una pregunta que envuelve intrínsecamente una dificultad teórica o práctica, a la cual se debe hallar una solución. Estos autores añaden, que para formular bien el problema, se debe tener amplios conocimientos sobre el tema y, además, se debe tener una gran imaginación creativa.

EJERCICIO 4

Problema: apareamiento de las ballenas jorobadas.

- **Idea:** observar el ritual de apareamiento de las ballenas jorobadas en las costas Peruanas.
- **Hipótesis:** las ballenas jorobadas se aparean en las costas Peruanas.
- **Objetivo:** observar si existe apareamiento de las ballenas jorobadas en las costas Peruanas

EJERCICIO 5

Problema: Estudiar la situación actual de la justicia en el Perú

- **Idea:** Examinar la situación actual de la justicia Peruana, que se esfuerza por mejorar cada día, gracias a su autonomía.
- **Hipótesis:** La situación de la justicia en el Perú ha mejorado respecto a años anteriores.
- **Objetivo:** Obtener información sobre la opinión de la población de Perú, de 18 a 70 años, acerca de la justicia en el Perú.

EJERCICIO 6

En la formulación de hipótesis y objetivos, también se debe tener presente las variables de mayor importancia y las relaciones existentes entre ellas.

Variables: Se utilizan para designar cualquier característica o cualidad de la unidad de análisis y son los elementos principales de los objetivos y las hipótesis. Es necesario definir las variables a medir y la forma en que vamos a hacerlo, como también, las unidades y los indicadores de interpretación que le vamos a dar: cantidad o cualidad del atributo, unidad y forma de medida, o expresión matemática. Durante este proceso las variables se transforman de un nivel abstracto a un nivel empírico, observable y medible.

Problema: Situación actual de lago Titicaca

- **Idea:** El estado trófico (nutritivo) superficial del lago Titicaca

Las variables más importantes que inciden en nuestro problema son, el estado actual del fitoplancton y la producción primaria, ya que ellas son las que permiten inferir el estado trófico del ecosistema y, además, se trata de variables sencillas y no costosas de medir. Desechando las variables macrófitas, el zooplancton y los registros de vertimientos o afluentes, por su incidencia indirecta y su dificultad para estudiarlas.

- **Objetivos:**

- De terminar la densidad y composición del fitoplancton superficial en diferentes lugares del lago Titicaca.
- Determinar la producción primaria en diferentes lugares del lago Titicaca.

- **Hipótesis:**

H1: El lago Titicaca está eutroficado (Nutrido)

H2: La baja transparencia de las aguas del Titicaca denota alto desarrollo del fitoplancton

Un pequeño cambio en la forma de apreciar el problema, conduce a modificaciones sustanciales en los objetivos y en las variables.

Algunos requisitos para establecer las hipótesis

- Identificar y definir claramente las variables por estudiar
- Identificar las relaciones existentes entre las variables
- Dar consistencia a la comprobación de las hipótesis y las variables a evaluar
- Utilizar pocos supuestos en la corroboración de las hipótesis

Definido la hipótesis y objetivos, el título tentativo es entonces:

Estructura del fitoplancton del lago Titicaca durante la temporada de sequía

EJERCICIO 7

Los siguientes ejemplos ilustran claramente la utilidad del marco conceptual para la manipulación de las variables:

H_o : *Los montones de madera viejas que se acumulan en la margen de los ríos, son utilizadas por mamíferos pequeños para su resguardo*

Una hipótesis como la anterior, requiere que el investigador precise los siguientes conceptos:

- ¿A partir de qué talla se considera un grupo de leños como montón?
- ¿Cómo va a definir o diferenciar los montones viejas de las nuevas?
- ¿Hasta dónde va a considerar margen de los ríos?
- ¿A cuáles mamíferos en particular se refiere y cómo los registrará o evaluará?
- ¿Qué entiende como resguardo?

H_o : *la creatividad en adolescentes de alta capacidad intelectual es mayor que en niños de capacidad intelectual promedio*

Esta hipótesis debe esclarecer en su marco conceptual los siguientes interrogantes:

- ¿Qué se entiende por creatividad y creatividad alta?
- ¿Qué se entiende como capacidad intelectual?
- ¿Qué edades considera el término adolescentes? ¿Los adolescentes que se estudiarán pertenecen a qué país, ciudad o institución?
- ¿Qué se entiende como capacidad intelectual promedio y alto?

Algunos de las hipótesis y objetivos surgen de las siguientes situaciones:

- Se detectan problemas en las teorías vigentes.
- Se quieren poner a prueba teorías en desarrollo.
- Se desea fortalecer una teoría colocándola a prueba sobre nuevas poblaciones o Comunidades.
- Se pretende caracterizar o evaluar la dimensión de un fenómeno
- Se busca la causalidad de un fenómeno
- Se desea corroborar una tendencia

La calidad de las hipótesis está relacionada directamente con el grado de profundidad con que se haya revisado la literatura. Nos recuerda también que pueden tener una función descriptiva o explicativa y que mientras algunas investigaciones contienen gran variedad de hipótesis otras contienen una o dos, sin que la calidad de la investigación esté asociada al número de ellas.

La forma precisa en que se formule o escriba una hipótesis y de allí un objetivo, incide de manera contundente en cómo deberá abordarse la metodología. Este punto es esencial porque el objetivo define el modo en que deberán tomarse los datos y, a su vez, condiciona lo que se hará con ellos.

Ideas, preguntas, hipótesis y objetivos, constituyen la vía principal de delimitación de la investigación y, a la vez, le dan el norte definitivo a la misma.

No debe confundirse una hipótesis de investigación con una hipótesis estadística, si bien puede haber gran correspondencia entre ellas. En la primera se plantea la proposición bajo estudio, mientras que, en las segundas, se plantean dos hipótesis: una nula H_0 que generalmente se refiere a que el resultado de una prueba estadística no es significativa y, una alterna H_1 , que denota que sí lo es.

OTRO DIAGRAMA DE FORMULACIÓN DE HIPÓTESIS Y OBJETIVOS

La secuencia para el desarrollo de una investigación, no es aquella comúnmente dada por los filósofos y que sigue el orden *idea, pregunta, hipótesis y revisión bibliográfica*, porque una investigación no es una secuencia lineal. Porque la pregunta de investigación y la hipótesis son más bien el resultado de una interacción entre la idea original del científico y una revisión continua y exhaustiva de la bibliografía.

Idea - revisión bibliográfica - pregunta de investigación - hipótesis de investigación

Si bien la idea u objeto de estudio puede surgir de la nada, las demás de las veces llega después de una intensa inmersión en el tema que nos permite descubrir las hipótesis resueltas, las dudas teóricas, los planteamientos inconclusos y las diferentes perspectivas utilizadas. Por tanto, las ideas constituyen el punto de partida de los investigadores con amplio dominio temático y una larga experiencia, pero no son, regularmente, el punto de partida de los novatos.

Cuando se inicia un proyecto de investigación con un tema que se desconoce o con el cual ha habido poca aproximación, se corre el riesgo que el mismo se abandone al cabo de un tiempo, por cuanto no llena las expectativas del investigador, o porque resulta muy difícil, costoso o extenso. Por tal razón, conviene que exista algún vínculo emocional, así como de conocimientos previos, para que esto no suceda. Aun así, es común que el tema gire levemente en torno a la temática previamente elegida.

Cuanto más clara sea la pregunta de investigación, más fácil será el planteamiento de la Hipótesis y objetivos. Al respecto nos advierte de dos puntos de interés:

- No es lo mismo expresar una idea que postular una pregunta de investigación, ya que en la primera hay una actividad mental desbocada.
- La formulación escrita del proyecto, conlleva un proceso de orden mental, de reflexión y de compromiso.

LLUVIA DE IDEAS

Surge en el caso de que el estudiante exprese dificultades para plantear preguntas de investigación. Entonces puede acudir a formular una lluvia de ideas propias, que podría enriquecer con compañeros, profesores o tutores.

También podría valerse de las recomendaciones para futuras investigaciones que plantean estudios ya publicados. Se recomienda mesura en el proceso de formulación de hipótesis, ya que con frecuencia el entusiasmo de los estudiantes se traslada a una sobrevaloración de sus ideas que termina por nublar su juicio. Otros estudiantes más, suelen abarcar más de lo que pueden apretar o terminan haciendo algo ya hecho.

Ejercicio de Lluvia de Ideas 1.

EJERCICIO 1

El siguiente ejemplo ilustra el proceso de formulación de una lluvia de ideas, partiendo del supuesto de que el tema de estudio es la *justicia en el Perú*. A partir de esta temática se pueden formular numerosas ideas, algunas de ellas relativas a la institución jueces, y la comunidad judicial:

Tema: Justicia en el Perú

- Cuán informado está sobre el tema de la justicia en el Perú
- Confía en el poder judicial
- Cómo califica la administración de justicia en el Perú
- Calificar el nivel profesional de los jueces en el Perú
- El principal problema de la administración de la justicia en el Perú
- Le preocupa la situación de la justicia en el Perú
- La situación de la justicia en el Perú ha mejorado, se ha mantenido igual o ha empeorado en los últimos años
- Considera que la justicia en el Perú es corrupta
- Considera que el poder judicial es autónomo
- Hay ocasiones, en que los ciudadanos hacen justicia por propia mano
- Ha litigado alguna vez en el poder judicial
- En alguna oportunidad le han pedido coima en el poder judicial
- Es excesivo el tiempo que demandan los procesos judiciales
- Qué es lo que más le incomoda cuando acude a los juzgados
- Los costos del servicio de administración de justicia en el Perú
- La atención que brindan los juzgados es buena

Hay que tener en cuenta que por tratarse de ideas sueltas aún no analizadas en profundidad, algunas pueden resultar prácticas, objetivas y viables, pero otras irrealizables, fantasiosas o poco importantes. El paso siguiente es, entonces, filtrar o tamizar las mismas respecto a su pertinencia y viabilidad.

EJERCICIO 2:

El siguiente ejemplo ilustra el proceso de formulación de una lluvia de ideas, partiendo del supuesto de que el tema de estudio es la *pesca artesanal*. A partir de esta temática se pueden formular numerosas ideas, algunas de ellas relativas a los peces, o tras a la pesca y o tras más a la comunidad de pescadores:

Tema: Pesca Artesanal

- Caracterizar la estructura de las comunidades explotadas
- Comparar las capturas por tipo de arte
- Determinar las especies de mayor importancia económica
- Estimar la biomasa de captura por mes
- Tipificar las especies capturadas
- Caracterizar las artes y embarcaciones de pesca
- Definir la relación entre especies y artes de pesca
- Definir las zonas de pesca
- Establecer la relación entre especies y ecosistemas
- Mapear las zonas de pesca
- Caracterizar la población pesquera involucrada en esta actividad
- Conocerlos canales de comercialización del producto
- Estimar la población que vive de esta actividad en forma directa e indirecta
- Evaluar los ingresos familiares promedios por esta actividad
- Proyectar tendencias en la explotación del recurso
- Verificar si se capturan especies prohibidas o con tallas no permitidas
- Analizar la percepción de los pescadores hacia las legislaciones de pesca
- Determinar si la pesca es una actividad que se trasmite de padres a hijos
- Etcétera

Hay que tener en cuenta que por tratarse de ideas sueltas aún no analizadas en profundidad, algunas pueden resultar prácticas, objetivas y viables, pero otras irrealizables, fantasiosas o poco importantes. El paso siguiente es, entonces, filtrar o tamizar las mismas respecto a su pertinencia y viabilidad.

VERBOS Y DEFINICIONES

Algunos de los verbos que regularmente se emplean en los objetivos de una investigación y sus definiciones se presentan a continuación:

- **Analizar: de Análisis:** distinción y separación de las partes de un todo hasta llegar a conocer sus principios o elementos. Método estadístico usado para cuantificar la importancia de cada uno de los factores actuantes en un fenómeno.
- **Caracterizar:** determinar los atributos peculiares de alguien o de algo, de modo que claramente se distinga de los demás.
- **Comparar:** fijar la atención en dos o más objetos para descubrir sus relaciones o estimar sus diferencias o semejanzas.
- **Comprender:** entender, alcanzar, penetrar.
- **Comprobar:** verificar, confirmar la veracidad o exactitud de algo.
- **Confirmar:** corroborar la verdad, certeza o el grado de probabilidad de algo.
- **Conocer:** averiguar por el ejercicio de las facultades intelectuales la naturaleza, cualidad es y

relaciones de las cosas. Entender, advertir, saber, echar de ver.

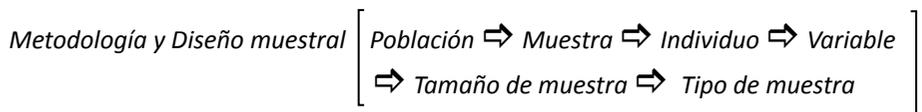
- **Deducir:** sacar consecuencias de un principio, proposición o supuesto.
- **Definir:** fijar con claridad, exactitud y precisión la significación de una palabra o la naturaleza de una persona o cosa. Decidir, determinar, resolver algo dudoso.
- **Delimitar:** determinar o fijar con precisión los límites de algo.
- **Demostrar:** mostrar, hacer ver que una verdad particular está comprendida en otra universal, de la que se tiene entera certeza.
- **Describir:** representar a alguien o algo por medio del lenguaje, refiriendo o explicando sus distintas partes, cualidades o circunstancias.
- **Determinar:** señalar, fijar algo para algún efecto.
- **Diseñar:** hacer un diseño.
- **Especificar:** explicar, declarar con individualidad algo. Fijar o determinar de modo preciso.
- **Establecer:** dejar demostrado y firme un principio, una teoría, una idea, etc.
- **Estimar:** apreciar, poner precio, evaluar algo.
- **Estudiar:** ejercitar el entendimiento para alcanzar o comprender algo.
- **Evaluar:** estimar, apreciar, calcular el valor de algo. Estimar los conocimientos, aptitudes y rendimiento de los alumnos.
- **Explicar:** llegar a comprender la razón de algo, darse cuenta de ello. Dar a conocer la causa o motivo de algo.
- **Fabricar:** producir objetos en serie, generalmente por medios mecánicos. Hacer, disponer o inventar algo no material.
- **Identificar:** dicho de dos o más cosas que pueden parecer o considerarse diferentes: Ser una misma realidad.
- **Inferir:** sacar una consecuencia o deducir algo de otra cosa.
- **Interpretar:** explicar acciones, dichos o sucesos que pueden ser entendidos de diferentes modos. Concebir, ordenar o expresar de un modo personal la realidad.
- **Investigar:** realizar actividades intelectuales y experimentales de modo sistemático con el propósito de aumentar los conocimientos sobre una determinada materia.
- **Mapear:** trasladar a un mapa sistemas o estructuras conceptuales. Hacer mapas.
- **Medir:** comparar una cantidad con su respectiva unidad, con el fin de averiguar cuántas veces la segunda está contenida en la primera.
- **Modelar:** configurar o conformar algo no material. Ajustarse a un modelo.
- **Observar:** examinar atentamente. Mirar con atención y recato, atisbar.
- **Presentar:** hacer manifestación de algo, ponerlo en la presencia de alguien.
- **Producir:** engendrar, procrear, criar. Se usa hablando más propiamente de las obras de la naturaleza, y, por extensión, de las del entendimiento.
- **Proyectar:** idear, trazar o proponer el plan y los medios para la ejecución de algo.
- **Reconocer:** examinar con cuidado algo o a alguien para enterarse de su identidad, naturaleza y circunstancias.
- **Redefinir:** volver a definir algo cuyas características o circunstancia han cambiado.
- **Relacionar:** establecer relación entre personas, cosas, ideas o hechos.
- **Simular:** representar algo, fingiendo o imitando lo que no es.
- **Tipificar:** dicho de una persona o de una cosa. Representar el tipo de la especie o clase a que pertenece.
- **Verificar:** comprobar o examinar la verdad de algo.

Capítulo

4

Metodología y Diseño Muestral

En este capítulo aprenderemos a desarrollar la metodología y el diseño muestral, resolviendo ejercicios prácticos utilizando el diagrama siguiente:



Metodología y Diseño muestral: Se definen y calcularán los valores de:

- **Población:** Conjunto de todos los individuos (personas, objetos, animales, etc.) que porten información sobre el fenómeno que se estudia.
- **Muestra:** Constituye una selección al azar de una porción de la población.
- **Individuo (Observación, Caso):** Cualquier elemento que aporte información sobre el fenómeno que se estudia.
- **Variable:** La variable es toda aquella propiedad de algún objeto, persona o evento que posee diferentes valores.
- **Tamaño de Muestra:** La calidad y validez de los resultados de una investigación depende del tamaño de muestra.
- **Tipo de Muestreo:** La muestra se puede elegir de diferentes maneras teniendo presente el tipo de muestreo, como muestreo aleatorio simple, muestreo sistemático, muestreo estratificado, muestreo por conglomerado.



PASOS DEL DISEÑO MUESTRAL

Consiste en definir y calcular los valores de la población, muestra, unidad de estudio (individuo), las variables, tamaño de muestra, y tipo de muestreo del problema o experimento.

Comúnmente, los experimentos científicos producen observaciones o hallazgos que se expresan como datos numéricos. Las mediciones que se obtienen de los sujetos bajo estudio constituyen los datos.

Generalmente, los datos son medidas de la variable. Los datos crudos u originales son aquellos datos obtenidos directamente del experimento, es decir, no han sido sometidos a tratamiento estadístico.

El propósito de las investigaciones es poder formular generalizaciones basado en un conjunto de observaciones que se extraen del estudio (el laboratorio).

POBLACIÓN

Conjunto de todos los individuos (personas, objetos, animales, etc.) que porten información sobre el fenómeno que se estudia. Representa una colección completa de elementos (sujetos, objetos, fenómenos o datos) que poseen algunas características comunes. Es el conjunto de elementos más grande del cual se puede tomar una muestra representativa para el experimento científico.

La *población* constituye el conjunto de elementos que forma parte del grupo de estudio, por tanto, se refiere a todos los elementos que en forma individual podrían ser cobijados en la investigación. La población la define el objetivo o propósito central del estudio y no estrictamente su ubicación o límites geográficos, u otras características particulares al interior de ella.

MUESTRA

Constituye una selección al azar de una porción de la población, es decir, un subconjunto que seleccionamos de la población.

La *muestra*, por otro lado, consiste en un grupo reducido de elementos de dicha población, al cual se le evalúan características particulares, generalmente, con el propósito de inferir tales características a toda la población.

INDIVIDUO (Observación, Caso, Sujeto)

Cualquier elemento que aporte información sobre el fenómeno que se estudia. El *elemento o individuo muestral* se refiere a la unidad más pequeña en que se puede descomponer una muestra.

VARIABLE

La variable es toda aquella propiedad de algún objeto, persona o evento que posee diferentes valores. El término **variable** implica una característica que cambia, tal como frecuencia cardiaca (FC), presión arterial (PA), masa corporal (MC) etc. Durante el proceso de investigación, se deben definir las variables independiente y dependiente.

Variable independiente es la que habrá de ser manipulada en el experimento, es decir, aquella que cambia y posee el potencial de afectar la variable dependiente. Por lo tanto, este tipo de variable es controlada sistemáticamente por el investigador. Durante el experimento científico se busca hallar el efecto que tiene una o más variables (independientes) sobre otras variables (dependientes).

Variable dependiente se determina antes o después de haber tratado la variable independiente. Representa la medida que se emplea para determinar el efecto de la variable independiente.

Variables cualitativas. Son aquéllas que se refieren a cualidades o atributos no medibles en números. Por ejemplo, organización, personal y funciones.

Variables cuantitativas. Son las susceptibles de medirse en términos numéricos. Se subdividen a su vez en:

- *Cuantitativas continuas.* Pueden asumir cualquier valor (números reales).
- *Cuantitativas discretas.* Asumen sólo valores enteros (números enteros)

EJERCICIOS DE POBLACIÓN Y MUESTRA

Los siguientes ejemplos nos permiten aclarar los conceptos previamente referidos: se debe tener presente que es necesario contar con los objetivos de la investigación para poder precisar la población, muestra, unidad de estudio (individuo) y las variables. Los objetivos ya fueron encontrados en el capítulo anterior.

EJERCICIO 1

Objetivo: Evaluar el precio de la vivienda en una ciudad de Lima.

- **Población:** La población será el total de las viviendas de ciudad de Lima. Así, si se estudia el precio de la vivienda de una ciudad, lo normal será no recoger información sobre todas las viviendas de la ciudad (sería una labor muy compleja).
- **Muestra:** Subgrupo o subconjunto de las viviendas de la ciudad de Lima seleccionadas aleatoriamente, que se entienda que es suficientemente representativo.
- **Individuo** (Observación, Caso, Sujeto): Una vivienda de dicha ciudad de Lima.
- **Variable:** Precio de la vivienda, tamaño, cantidad de habitaciones, ubicación, antigüedad, etc.

Variable independiente: Tamaño de la vivienda, cantidad de habitaciones, ubicación, antigüedad.

Variable dependiente: precio de vivienda (por que el precio depende de la ubicación, antigüedad, tamaño y etc.)

Variables cualitativas: tamaño (grande, pequeño), ubicación (norte, centro, sur)

Variables cuantitativas:

- *Cuantitativas continuas.* Precio (Soles).
- *Cuantitativas discretas.* Antigüedad (años), cantidad de habitaciones (Nº entero).

Una variable independiente se puede convertir en dependiente o viceversa, del mismo modo una variable cualitativa se puede convertir en cuantitativa o viceversa.

EJERCICIO 2

Objetivo: Evaluar la percepción de los estudiantes de un colegio en torno a los niveles de ruido que existen en el plantel.

- **Población:** Está conformada por todos los estudiantes del colegio incluidos aquellos de programas diurnos y nocturnos.
- **Muestra:** La muestra puede tomarse eligiendo al azar un puñado de alumnos con base en un archivo o listado que contenga todos los códigos o nombres de los estudiantes.
- **Individuo (Observación, Caso, Sujeto):** Un alumno del colegio con base en un archivo o listado que contenga todos los códigos o nombres de los estudiantes.
- **Variable:** Edad del alumno, género (masculino, femenino), peso del alumno, niveles de ruido, tipo de ambiente, distancia del ambiente etc.

Variable independiente: Edad del alumno, género (masculino, femenino), peso del alumno, tipo de ambiente, distancia del ambiente.

Variable dependiente: Niveles de ruido.

Variables cualitativas: Género (masculino, femenino), tipo de ambiente (cerrado, despegado), niveles de ruido.

Variables cuantitativas: Edad y peso del alumno, distancia del ambiente (metros).

- *Cuantitativas continuas:* Edad y peso del alumno, distancia del ambiente.
- *Cuantitativas discretas:* Ninguna variable se ajusta a este tipo.

EJERCICIO 3

Objetivo: Obtener información sobre la opinión de la población, de 18 a 75 años, acerca de la justicia en el Perú.

- **Población:** Población electoral del Perú compuesta por todas las personas de 18 a 75 años de edad inscritas en el Registro Nacional de Identidad y Estado Civil (RENIEC), que comprende a hombres y mujeres de todos los niveles socioeconómicos.
- **Muestra:** Un subgrupo seleccionado aleatoriamente de la población electoral del Perú, compuesta por todas las personas de 18 a 75 años de edad inscritas en el Registro Nacional de Identidad y Estado Civil (RENIEC), que comprende a hombres y mujeres de todos los niveles socioeconómicos.
- **Individuo (Observación, Caso, Sujeto):** Una persona de 18 a 75 años de edad inscrita en el Registro Nacional de Identidad y Estado Civil (RENIEC).
- **Variable:** Confianza en la justicia, el tema de justicia, calificación de la administración de justicia, nivel profesional de los jueces, problema de la administración de justicia. La situación de la justicia en el Perú, autonomía del Poder Judicial, los litigantes.

EJERCICIO 4

Objetivo: Determinar si el área de las hojas del eucalipto cambia de la parte superior a la inferior del árbol.

- **Población:** Un bosque con eucaliptos delimitados geográficamente.
- **Muestra:** Un grupo de eucaliptos elegidos al azar del bosque eucaliptos delimitados geográficamente.
- **Individuo (Observación, Caso, Sujeto):** Una hoja de eucalipto del bosque.
- **Variable:** Tamaño de la hoja, altura donde se encuentra la hoja, clima, cambio de la hoja de eucalipto, etc.

EJERCICIO 5

- **Objetivo:** Observar si existe apareamiento de las ballenas jorobadas en las costas peruanas.
- **Población:** Todas las ballenas jorobadas existentes en las costas peruanas.
- **Muestra:** Un grupo de ballenas jorobadas de las costas peruanas. La muestra también puede ser un grupo de parejas (hembra y macho) de ballenas jorobadas.
- **Individuo (Observación, Caso, Sujeto):** Una ballena jorobada. También puede ser una pareja (hembra y macho) de ballenas jorobadas.
- **Variable:** Edad de las ballenas, peso, talla, cantidad de apareamiento de las ballenas, etc.

De los ejemplos anteriores se nota que los trabajos de investigación, los cuales son realizados y financiados por el propio investigador dentro de limitaciones económicas y de tiempo considerables, no deben pretender coberturas amplias en sus objetivos, por cuanto corren el riesgo de desbordar la capacidad e infraestructura del investigar.

El investigador debe estar plenamente familiarizado con el área, la población y los instrumentos de muestreo, cualesquiera que estos sean.

TAMAÑO DE MUESTRA (Muestreo Aleatorio Simple)

En una investigación el tamaño de muestra es muy importante. Teniendo en cuenta que la *calidad y validez de los resultados de una investigación depende del tamaño de muestra*. Una muestra demasiado grande implica un desperdicio de recursos y una muestra demasiado pequeña disminuye la utilidad de los resultados.

En nuestra investigación utilizaremos el **Muestreo Aleatorio Simple** (Raj,D): Teoría del Muestreo pag 32).

El tamaño de muestra en el **Muestreo Aleatorio Simple** se calcula con la fórmula siguiente (Spigel. 1978: 161):

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad \dots\dots (1)$$

$$n_0 = \frac{z_{\alpha}^2 \sigma^2}{E^2} \quad \dots\dots (2)$$

Donde:

- n : Tamaño de muestra.
- n_0 : Tamaño de muestra aproximado.
- N : Tamaño de la población bajo Estudio.
- Z_{α} : Valores correspondiente al nivel de Significancia.
- E : Error de tolerancia de la estimación.
- : Nivel de significancia.
- σ^2 : varianza de la variable.

1. **El estudio de proporciones** la *varianza es igual a* $PQ(\sigma^2)$: P denota la proporción estimada o esperada de la variable; si no se conoce tal valor, se reemplaza por 0.5 ($P=1/2$ y $Q=1-P$). La fórmula quedaría de la siguiente manera:

$$n_0 = \frac{z_{\alpha/2}^2 \sigma^2}{E^2} = \frac{z_{\alpha/2}^2 PQ}{E^2} \dots\dots\dots (3)$$

2. **Error muestral (E) o Error de Estimación:** Es el error a causa de observar una muestra en lugar de la población completa. Este valor depende del investigador y se encuentra entre 0% y 10%; en algunas ocasiones es mayor a 10%.

3. **El nivel de confianza:** Es la probabilidad a priori de que el intervalo de confianza a calcular contenga al verdadero valor del parámetro. Se indica por $1-\alpha$ y habitualmente se da en porcentaje ($1-\alpha$)%. Los valores que se suelen utilizar para el nivel de confianza son el 95%, 99% y 99,9%

4. **Ejemplos de nivel de confianza:**

- a) Para un nivel de confianza del 88%:

Solución:

$1-\alpha=0.88$, $\alpha=0.12$ y $\alpha/2=0.06$. Ver tabla de distribución normal para el valor 0.94 ($1-\alpha/2$). Se observa que para 0.9406 el valor en la fila es 1.5 y el valor en la columna es 0.06 que representa el segundo decimal entonces $Z_{0.06}=1.56$

- b) Para un nivel de confianza del 98%,

Solución:

$1-\alpha=0.98$, $\alpha=0.02$ y $\alpha/2=0.01$. Ver tabla de distribución normal el valor 0.99 ($1-\alpha/2$). Se observa que para 0.9906 el valor en la fila es 2.3 y el valor en la columna es 0.05 que representa el segundo decimal entonces $Z_{0.01}=2.35$.

- c) Para un nivel de confianza del 95%:

Solución:

$1-\alpha=0.95$, $\alpha=0.05$ y $\alpha/2=0.025$. Ver tabla de distribución normal el valor 0.975 ($1-\alpha/2$). Se observa que para 0.975 el valor en la fila es 1.9 y el valor en la columna es 0.06 que representa el segundo decimal entonces $Z_{0.025}=1.96$

En resumen, para hallar el valor de Z_{α} , se tiene que observar el valor de $1-\alpha/2$ (uno menos alfamedias) en la tabla de la distribución normal, el valor de la fila nos da el valor entero y el primer decimal y el valor de la columna nos da el segundo decimal.

Los resultados se pueden ver en la tabla de distribución normal y están resaltados con negrita.

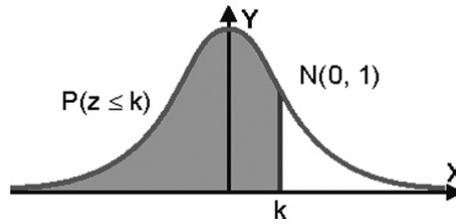


TABLA : DISTRIBUCIÓN DE PROBABILIDAD NORMAL ESTÁNDAR, $N(0, 1)$

| Z | 0 | 0,01 | 0,02 | 0,03 | 0,04 | 0,05 | 0,06 | 0,07 | 0,08 | 0,09 |
|-----|--------|--------|--------|--------|---------------|--------|---------------|--------|--------|--------|
| 0,0 | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1,0 | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,958 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2,0 | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |

| | | | | | | | | | | |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |
| 3,0 | 0,9987 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9990 | 0,9990 |
| 3,1 | 0,9990 | 0,9991 | 0,9991 | 0,9991 | 0,9992 | 0,9992 | 0,9992 | 0,9992 | 0,9993 | 0,9993 |
| 3,2 | 0,9993 | 0,9993 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9994 | 0,9995 | 0,9995 | 0,9995 |
| 3,3 | 0,9995 | 0,9995 | 0,9995 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9996 | 0,9997 |
| 3,4 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9997 | 0,9998 |
| 3,5 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 | 0,9998 |
| 3,6 | 0,9998 | 0,9998 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 |
| 3,7 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 |
| 3,8 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 | 0,9999 |
| 3,9 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 4,0 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |

EJEMPLOS DE TAMAÑO DE MUESTRA: Los ejemplos que se presentan a continuación permiten observar como se calcula el tamaño de muestra.

EJERCICIO 1

Problema: Se intenta estudiar el precio de la vivienda de la ciudad de Lima.

Solución:

Recoger información sobre todas las viviendas de la ciudad sería una labor muy compleja. Por tanto se halla un tamaño muestra:

Datos:

Población (N): Desconocida entonces, $N=\infty$.

P: Desconocido entonces, $P=0.5$ y $Q=0.5$

E: Este valor depende del investigador (recomendable=5%)

α : Desconocido entonces, $\alpha = 5\%$; $1-\alpha = 95\%$; $\alpha/2 = 2.5\%$, por tanto

$Z_{5\%} = 1.96$ (Tabla Normal).

$n = ?$

Reemplazando valores en la fórmula se tiene:

$$n_0 = [(1.96)^2(0.5)^2]/[(0.05)^2] = \mathbf{384.16}$$

Luego:

$$n = (384.16)/(1+384.16/\infty) = (384.16)/(1+0) = 384.16$$

Como nuestra unidad (sujeto, elemento) es una vivienda, se tomarán 384 viviendas de la ciudad de Lima para realizar el estudio.

Observaciones:

- Cuando $N=\infty$ se demuestra que $n=n_0$
- Muchos autores consideran el valor de $Z_{5\%} = 2$, cuando $\alpha = 5\%$;
- Entonces $n_0 = [(2)^2(0.5)^2]/[(0.05)^2] = \mathbf{400}$

Como $N=\infty$ se tiene que $n_0 = n = \mathbf{400}$,

Este motivo origina que muchos estudios sencillos se realicen con tamaño de muestra igual 400.

EJERCICIO 2

Problema: Evaluar la percepción de los estudiantes de un colegio en torno a los niveles de ruido que existen en el plantel conformado por 1,000 estudiantes, incluidos aquellos de programas diurnos y nocturnos.

Solución:

Hallando el tamaño muestra:

Datos:

Población (N): N= 1000.

P: Desconocido entonces, P=0.5 y Q=0.5

E: Este valor depende del investigador (recomendable=5%)

α : Desconocido entonces, $\alpha = 5\%$; $1-\alpha = 95\%$; $\alpha/2 = 2.5\%$, por tanto

$Z_{5\%} = 1.96$ (Tabla Normal).

n=?

Remplazando valores en la fórmula se tiene:

$$n_0 = [(1.96)^2(0.5)^2] / [(0.05)^2] = \mathbf{384.16}$$

Luego:

$$n = (384.16) / (1 + 384.16/1000) = (384.16) / (1.38416) = 277.54$$

Como nuestra unidad (sujeto, elemento) es un alumno, se tomarán 278 alumnos del colegio para realizar el estudio.

EJERCICIO 3

Problema: Se pretende determinar la prevalencia de una enfermedad en una población de 250 ovejas: se supone una prevalencia del 30% y se desea una precisión del 10% para un nivel de confianza del 95%.

Solución:

Datos:

Población: N = 250 ovejas

Unidad de estudio: una oveja

P=30% entonces Q=70%

E = 10%

$\alpha = 95\%$; $\alpha = 5\%$; $\alpha/2 = 2.5\%$ entonces;

$Z_{5\%} = 1.96$ (Tabla Normal)

n= ?

Remplazando valores en la fórmula

$$n_0 = (1.96)^2(0.30)(0.70) / (0.10)^2 = \mathbf{80.67}$$

Luego:

$$n = (80.67) / (1 + 80.67/250) = \mathbf{60.99}$$

El tamaño de muestra será 61 ovejas, entonces se tomarán 61 ovejas a quienes se le aplicará el estudio.

EJERCICIO 4

Problema: Se pretende determinar el tiempo que demora un proceso penal en la Corte Superior de Lima Norte, sabiendo que existen 7470 expedientes con proceso penal en la sede de la Corte Superior.

Solución:

Datos:

Población: $N = 7470$ expedientes

$P=50\%$ entonces $Q=50\%$

$E = 5\%$

$\alpha = 95\%$; $\alpha = 5\%$; $\alpha/2 = 2.5\%$ entonces;

$Z_{5\%} = 1.96$ (Tabla Normal)

$n = ?$

Remplazando valores en la fórmula

$$n_o = [(1.96)^2(0.5)^2]/[(0.5)^2] = \mathbf{384.16}$$

Luego:

$$n = (384.16)/(1+384.16/7470) = \mathbf{365.37}$$

El tamaño de muestra para realizar el estudio será 365 expedientes con proceso penal.

n : Cantidad de Expedientes Penales donde se realizará el estudio.

N : Cantidad total de Expedientes Penales existentes en la sede de la Corte Superior.

TIPOS DE MUESTREO

Existen dos tipos: el muestro aleatorio y el muestreo por conveniencia. El muestreo aleatorio puede realizarse de distintas maneras; las más frecuentes son el muestreo simple, el sistemático, el estratificado, el muestreo por conglomerados, el mixto etc.

La selección intencionada o muestreo por conveniencia es no aleatorio, por lo que suele presentar sesgos y en este libro se tomará en cuenta.

Se discutirá la utilidad del **muestreo aleatorio** y se describen las diferentes clases de muestreo que se pueden aplicar para tomar una muestra de la población.

Muestreo aleatorio

En el muestreo aleatorio todos los elementos tienen la misma probabilidad de ser elegidos. Los individuos que formarán parte de la muestra se elegirán al azar mediante números aleatorios. Existen varios métodos para obtener números aleatorios, los más frecuentes son la utilización de tablas de números aleatorios o generarlos por ordenador.

El muestreo aleatorio puede realizarse de distintas maneras; las más frecuentes son el muestreo simple, el sistemático, el estratificado y el muestreo por conglomerados.

1. Muestreo aleatorio simple

Es el método conceptualmente más simple. Consiste en extraer todos los individuos al azar de una lista (marco de la encuesta). En la práctica, a menos que se trate de poblaciones pequeñas o de estructura muy simple, es difícil de llevar a cabo de forma eficaz.

2. Muestreo estratificado

Consiste en la división previa de la población de estudio en grupos o clases que se suponen homogéneos respecto a la característica a estudiar. A cada uno de estos estratos se le asignaría una cuota que determinaría el número de miembros del mismo que compondrán la muestra. Dentro de cada estrato se suele usar la técnica de muestreo sistemático, ya que con aquella suelen ser las técnicas más usadas en la práctica. Según la cantidad de elementos de la muestra que se han de elegir de cada uno de los estratos, existen dos técnicas de muestreo estratificado:

2.1 Asignación proporcional: El tamaño de cada estrato en la muestra es proporcional a su tamaño en la población.

$$n_i = n \left[\frac{N_i}{N} \right]$$

Donde: N_i : Tamaño de Población del estrato "i".
 n_i : Tamaño de muestra del estrato "i"

2.2 Asignación óptima: La muestra recogerá más individuos de aquellos estratos que tengan más variabilidad. Para ello es necesario un conocimiento previo de la población.

3. Muestreo sistemático:

Se utiliza cuando el universo o población es de gran tamaño, o ha de extenderse en el tiempo. En este caso se elige el primer individuo al azar y el resto viene condicionado por aquél. Luego hay que calcular una constante, que se denomina coeficiente de elevación $K = N/n$; donde N es el tamaño del universo y n el tamaño de la muestra.

Para determinar la primera unidad muestral se debe producir la primera extracción, para ello hay que elegir al azar un número entre 1 y K ; de ahí en adelante tomar uno de cada K a intervalos regulares. Ocasionalmente, es conveniente tener en cuenta la periodicidad del fenómeno.

Esto quiere decir que si tenemos un determinado número de personas que es la población y queremos escoger de ella un número más pequeño, que es la muestra, dividimos el número de la población por el de la muestra que queremos tomar. El resultado de esta operación será el intervalo, entonces escogemos un número al azar desde uno hasta el número del intervalo, y a partir de este número escogemos los demás siguiendo el orden del intervalo.

4. Muestreo por conglomerados

Se divide la población en varios grupos de características parecidas entre ellos y luego se analizan completamente algunos de los grupos, descartando los demás. Dentro de cada conglomerado existe una variación importante, pero los distintos conglomerados son parecidos. Requiere una muestra más grande, pero suele simplificar la recogida de muestras. Frecuentemente los conglomerados se aplican a zonas geográficas.

Técnica similar al muestreo por estadios múltiples. Se utiliza cuando la población se encuentra dividida, de manera natural, en grupos que se supone que contienen toda la variabilidad de la población, es decir, la representan fielmente respecto a la característica a elegir. Pueden seleccionarse sólo algunos de estos grupos o *conglomerados* para la realización del estudio. Cuando, dentro de cada conglomerado, se extraen los individuos que formarán parte de la muestra por muestreo aleatorio simple, el muestreo se llama bietápico.

Las ideas de estratificación y conglomerados son opuestas. El primer método funciona mejor cuanto más homogénea es la población respecto del estrato, aunque más diferentes son éstos entre sí. En el segundo, ocurre lo contrario. Los conglomerados deben presentar toda la variabilidad, aunque deben ser muy parecidos entre sí. Recolección y organización de datos: una vez identificada la población se procede a recoger los datos en muchas ocasiones la población es muy grande y no sería posible realizar la investigación totalmente con el fin de obtener todos los datos asignados a cada uno.

5. Muestreo mixto

Cuando la población es compleja, cualquiera de los métodos descritos puede ser difícil de aplicar, en estos casos se aplica un muestreo mixto que combina dos o más de los anteriores sobre distintas unidades de la encuesta.

6. Muestreo por estadios múltiples

Esta técnica es la única opción cuando no se dispone de lista completa de la población de referencia o bien cuando por medio de la técnica de muestreo simple o estratificado se obtiene una muestra con unidades distribuidas de tal forma que resultan de difícil acceso. En el muestreo a estadios múltiples se subdivide la población en varios niveles ordenados que se extraen sucesivamente por medio de un procedimiento de embudo. El muestreo se desarrolla en varias fases o extracciones sucesivas para cada nivel.

Por ejemplo, si tenemos que construir una muestra de profesores de primaria en un país determinado, éstos pueden subdividirse en unidades primarias representadas por circunscripciones didácticas y unidades secundarias que serían los propios profesores. En primer lugar extraemos una muestra de las unidades primarias (para lo cual debemos tener la lista completa de estas unidades) y en segundo lugar extraemos aleatoriamente una muestra de unidades secundarias de cada una de las primarias seleccionadas en la primera extracción.

7. Muestreo por cuotas

Es la técnica más difundida sobre todo en estudios de mercado y sondeos de opinión. En primer lugar es necesario dividir la población de referencia en varios estratos definidos por algunas variables de distribución conocida (como el género o la edad). Posteriormente se calcula el peso proporcional de cada estrato, es decir, la parte proporcional de población que representan. Finalmente se multiplica cada peso por el tamaño de “n” de la muestra para determinar la cuota precisa en cada estrato. Se diferencia del muestreo estratificado en que una vez determinada la cuota, el investigador es libre de elegir a los sujetos de la muestra dentro de cada estrato.

8. Muestreo de “bola de nieve”

Indicado para estudios de poblaciones clandestinas, minoritarias o muy dispersas pero en contacto entre sí. Consiste en identificar sujetos que se incluirán en la muestra a partir de los propios entrevistados. Partiendo de una pequeña cantidad de individuos que cumplen los requisitos necesarios, estos sirven como localizadores de otros con características análogas.

9. Muestreo subjetivo por decisión razonada

En este caso las unidades de la muestra se eligen en función de algunas de sus características de manera racional y no casual. Una variante de esta técnica es el muestreo compensado o equilibrado, en el que se seleccionan las unidades de tal forma que la media de la muestra para

determinadas variables se acerque a la media de la población.

Observaciones

- Es diferente hallar el tamaño de muestra con elegir las muestras.
- Existen diferentes técnicas de muestreo, por tanto el tamaño de muestra se puede hallar de diferentes maneras teniendo presente el tipo de muestreo.

EJERCICIOS DE TIPOS DE MUESTREO (Elegir la Muestra)

La muestra se puede elegir de diferentes maneras teniendo presente el tipo de muestreo. El muestreo aleatorio puede realizarse de distintas maneras; las más frecuentes son el muestreo simple, el sistemático, el estratificado y el muestreo por conglomerados. De la combinación de las mencionadas anteriormente nacen un sin número de muestreos

EJERCICIO 1

Problema: Se pretende determinar la prevalencia de una enfermedad en una población de 250 ovejas: se supone una prevalencia del 30% y se desea una precisión del 10% para un nivel de confianza del 95%.

Solución

El tamaño de muestra es 61 ovejas: Población(N)=250 ovejas, P=30%, E=10%, $1-\alpha=95\%$, $Z_{5\%}=1.96$, $n_0=(1.96)^2(0.30)^2/(0.10)^2=80.67$. $n=(80.67)/(1+80.67/250) =60.99$. (Ejemplo resuelto en tamaño de muestra)

Calculado el tamaño de muestra. La muestra se puede elegir utilizando los siguientes tipos de muestreo:

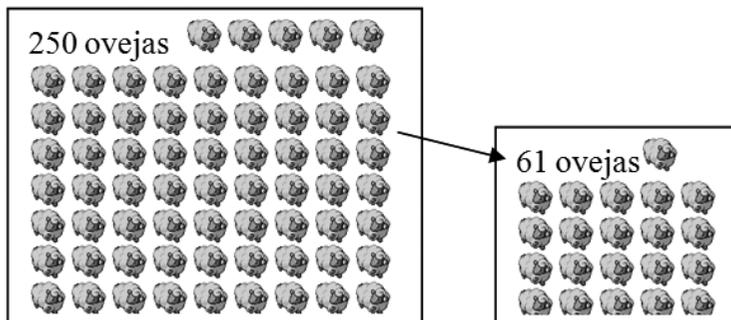
1. Muestreo Aleatorio Simple:

Tamaño de muestra (n) = 61 ovejas

Población (N)=250 ovejas

La elección de la muestra se realiza de la siguiente manera:

A cada oveja se le asigna un número (del 1 al 250), luego se selecciona al azar 61 ovejas entre la oveja número 1 y la oveja 250 utilizando una tabla de números aleatorios o una computadora que contiene, cualquier software para generar números aleatorios.



Representación gráfica del muestreo aleatorio simple: de las 250 ovejas se seleccionan al azar (aleatorio) las 61 ovejas para realizar el estudio.

2. Muestreo Aleatorio Sistemático:

Tamaño de muestra (n) = 61 ovejas

Población (N)=250 ovejas

Primero: Se calcula el valor de la constante “k”, que es el cociente del tamaño de la población (N) entre el tamaño de muestra (n).

Coefficiente de elevación $K= N/n =250/61=4.09$

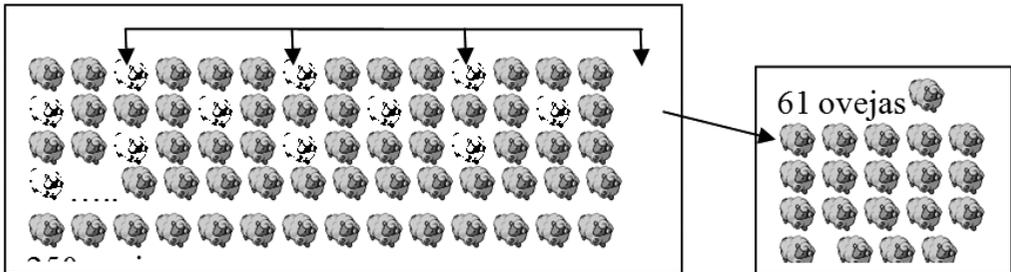
Segundo:

A cada oveja se le asigna un número (del 1 al 250).

Tercero: elegimos la primera muestra:

como el valor de $K=4$.

Entonces elegimos en forma aleatoria un número entre 1 y 4; supongamos que resultó el número 3 (1,2,3,4), por ello tomaremos como primera muestra a la oveja número 3, y a continuación seleccionaremos las demás muestras, cada cuarto oveja de la siguiente manera: 1,2,3,4,5,6,7,8,9, 10,11,12,13,14,15, y así sucesivamente hasta completar la muestra de 61 ovejas.



Representación gráfica del muestreo sistemático: de las 250 ovejas se seleccionan sistemáticamente las 61 ovejas (a las ovejas seleccionadas se les ha cambiado de color).

3. Muestreo Aleatorio Estratificado:

Problema: La probabilidad de que una oveja esté enferma está directamente relacionada con la edad. En el ejemplo anterior, se tiene que el 44% de las ovejas son de menos de 2 años, el 28% de las ovejas son de 3 a 4 años, el 18% de las ovejas son 5 a 6 años y el 10% son ovejas de más de seis años:

Solución:

Población (N)=250 ovejas

Tamaño de muestra (n) = 61 ovejas (Muestra es igual a Tamaño de Muestra)

Primero: Elegimos los estratos con sus respectivas poblaciones y muestras.

Existen 4 estratos y son los siguientes:

E_1 : Ovejas de menos de 2 años, con Población = N_1 y Muestra = n_1

E_2 : Ovejas de 3 a 4 años, con Población = N_2 y Muestra = n_2

E_3 : Ovejas de 5 a 6 años, con Población = N_3 y Muestra = n_3

E_4 : Ovejas de más de seis años, con Población = N_4 y Muestra = n_4

Segundo: Calculamos el tamaño de muestra en cada estrato, a población grande muestra grande, a tamaño de población pequeña tamaño de muestra pequeña.

Aplicando **asignación proporcional**: El tamaño de la muestra en cada estrato es proporcional a su tamaño en la población y la fórmula es la siguiente.

$$n_i = n \left[\frac{N_i}{N} \right]$$

Donde:

N_i : Población del estrato 'i'

n_i : Tamaño de muestra del estrato 'i'

De la información se tiene que:

$$\frac{N_1}{N} = 44\%$$

$$\frac{N_2}{N} = 28\%$$

$$\frac{N_3}{N} = 18\%$$

$$\frac{N_4}{N} = 10\%$$

Reemplazando los valores, el tamaño de muestra en cada estrato es:

$$n_1 = 61(44\%) = 27$$

$$n_2 = 61(28\%) = 17$$

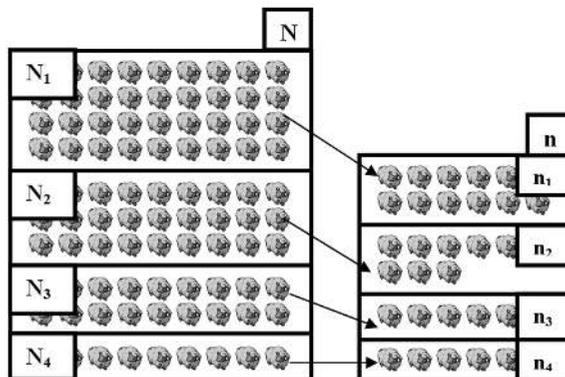
$$n_3 = 61(18\%) = 11$$

$$n_4 = 61(10\%) = 6$$

Esto es, 27 ovejas del estrato 1, 17 ovejas de estrato 2, 11 ovejas de estrato 3 y 6 ovejas del estrato 4.

Con estas muestras seleccionadas en cada estrato se realiza el trabajo de investigación.

Este método evita que se tomen más individuos de un grupo que de los demás y esto pueda condicionar el resultado.



$$n_1 = n \left(\frac{N_1}{N} \right)$$

$$n_2 = n \left(\frac{N_2}{N} \right)$$

$$n_3 = n \left(\frac{N_3}{N} \right)$$

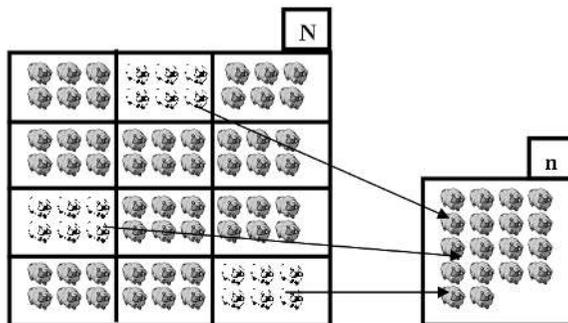
$$n_4 = n \left(\frac{N_4}{N} \right)$$

Representación gráfica del muestreo estratificado con asignación proporcional al tamaño, es decir, a población grande le corresponde tamaño de muestra y a población pequeña le corresponde tamaño de muestra pequeña.

- **Observación:** Para seleccionar la muestra en cada estrato es decir seleccionar n_1, n_2, n_3, n_4 , se puede utilizar el muestreo aleatorio simple o el muestreo sistemático.

Muestreo aleatorio por conglomerados

Se divide la población en varios grupos de características parecidas entre ellos y luego se analizan completamente algunos de los grupos, descartando los demás. Dentro de cada conglomerado existe una variación importante, pero los distintos conglomerados son parecidos. Requiere una muestra más grande, pero suele simplificar la recogida de muestras. Los conglomerados se pueden seleccionar en forma aleatorio simple o en forma sistemática.



Representación gráfica del muestreo por conglomerados.

- **Observación:** Para seleccionar la muestra en cada conglomerado puede utilizar el muestreo aleatorio simple o muestreo sistemático.

Muestreo mixto

Cuando se desea realizar un estudio más preciso, se pueden combinar los muestreos anteriormente descritos. También puede suceder que la población bajo estudio sea compleja entonces se puede aplicar un muestreo mixto que combina dos o más muestreos anteriores.

EJERCICIO 2

Problema: Se pretende determinar el tiempo que demora un proceso penal en la Corte Superior de Lima Norte, sabiendo que existen 7,470 expedientes con proceso penal en la sede de la Corte Superior.

Solución:

Por el tema anterior se sabe que: El tamaño de muestras es: $n = 365$ ($N=7470$; $\alpha=5\%$; $E=5\%$; $p=0.5$; $q=0.5$, $n_0=384$)

Por tanto, el tamaño de muestra (n) = 365 Expedientes

Población (N)= 7470 Expedientes

Además, realizando trabajo de campo se obtuvo la siguiente información de la cantidad de expedientes en cada dependencia (Población de cada dependencia).

| Dependencia | Estratos | Población |
|---------------------------------|----------|-------------|
| 1º JUZGADO PENAL (Reos Libres) | N_1 | 914 |
| 2º JUZGADO PENAL (Reos Libres) | N_2 | 912 |
| 3º JUZGADO PENAL (Reos Libres) | N_3 | 696 |
| 4º JUZGADO PENAL (Reos Libres) | N_4 | 789 |
| 5º JUZGADO PENAL (Reos Libres) | N_5 | 814 |
| 6º JUZGADO PENAL (Reos Libres) | N_6 | 1395 |
| 7º JUZGADO PENAL (Reos Libres) | N_7 | 380 |
| 8º JUZGADO PENAL (Reos Libres) | N_8 | 127 |
| 9º JUZGADO PENAL (Reos Libres) | N_9 | 601 |
| 10º JUZGADO PENAL (Reos Libres) | N_{10} | 275 |
| 11º JUZGADO PENAL (Reos Libres) | N_{11} | 71 |
| 12º JUZGADO PENAL (Reos Libres) | N_{12} | 195 |
| 13º JUZGADO PENAL (Reos Libres) | N_{13} | 301 |
| Total | N | 7470 |

La selección de la muestra se realizará utilizando el muestreo estratificado, porque se requiere una muestra representativa para poder determinar el objetivo de estudio (tiempo de duración de un proceso).

- No se utiliza el muestreo aleatorio simple, porque no es representativa, ya que en la muestra podría seleccionarse más expedientes de determinado juzgado.
- No se utiliza el muestreo por conglomerado, porque no es representativa, ya que al seleccionarse sólo algunos juzgados perderíamos información relevante de los otros no seleccionados.

Muestreo estratificado

Existen 13 estratos desde el primer Juzgado al 13 Juzgado.

La elección de la muestra en cada estrato se realizará aplicando un reparto proporcional, porque se puede observar algunos juzgados con elevado número de expedientes y otros con cantidades pequeñas de expedientes. Para la proporcionalidad se aplica la fórmula siguiente. (Asignación proporcional)

$$n_i = n \left[\frac{N_i}{N} \right]$$

Donde

N_i : Cantidad de expedientes de juzgado "i" (Población "i")

n_i : Tamaño de muestra del juzgado "i"

Aplicando la fórmula para hallar la muestra del primer Juzgado Penal se tiene:
 $n_1 = 365(914/7470) = 44.65 = 45$

Aplicando la fórmula sucesivamente para hallar la muestra de los demás juzgados se tiene el cuadro siguiente:

| 10 | Estrato | Muestra |
|---------------------------------|----------|------------|
| 1º JUZGADO PENAL (Reos Libres) | n_1 | 45 |
| 2º JUZGADO PENAL (Reos Libres) | n_2 | 45 |
| 3º JUZGADO PENAL (Reos Libres) | n_3 | 34 |
| 4º JUZGADO PENAL (Reos Libres) | n_4 | 39 |
| 5º JUZGADO PENAL (Reos Libres) | n_5 | 40 |
| 6º JUZGADO PENAL (Reos Libres) | n_6 | 68 |
| 7º JUZGADO PENAL (Reos Libres) | n_7 | 19 |
| 8º JUZGADO PENAL (Reos Libres) | n_8 | 6 |
| 9º JUZGADO PENAL (Reos Libres) | n_9 | 29 |
| 10º JUZGADO PENAL (Reos Libres) | n_{10} | 13 |
| 11º JUZGADO PENAL (Reos Libres) | n_{11} | 3 |
| 12º JUZGADO PENAL (Reos Libres) | n_{12} | 10 |
| 13º JUZGADO PENAL (Reos Libres) | n_{13} | 15 |
| Total | n | 365 |

Encontrado el tamaño de muestra para cada estrato.

El siguiente paso es elegir las muestras que participarán en el estudio.

Casi siempre se elige las muestras en forma aleatoria, por ejemplo de los 914 expedientes del primer juzgado penal se elige al azar 45 expedientes, a quienes se les aplicará el estudio.

Para hacer el ejercicio más interesante utilizaremos la elección sistemática.

Elección de la muestra en forma Sistemática

La selección de la muestra en forma sistemática se realizará en cada juzgado (estrato), por ejemplo:

El 12º Juzgado Penal tiene 195 expedientes. A cada expediente se le asigna un número: 1,2,3...195. Luego se calcula el coeficiente $K=N/n=195/10=19.5$ ($K=20$).

Entonces elegimos en forma aleatoria un número del 1 al 20, supongamos que resultó 9 (primera muestra elegida; 1,2,3,...8,9,10...20).

Ahora para las demás muestras, a la primera muestra seleccionada (9) se le suma $K=20$, y así hasta completar el tamaño de muestra de 10 expedientes. Del modo siguiente: 1,2,...9,...,27,28,29,30,..., 48,49,50,...188,189,190,...195.

La muestra son los expedientes representados por los siguientes números 9, 29, 49, 69, 89, 109, 129, 149, 169, 189.

De esta forma, se hace la elección de la muestra para los demás juzgados (estratos).

PILOTO

En lo posible debe realizar un piloto antes en campo de estudio. Este consiste en realizar un premuestreo que le permita familiarizarse con el lugar, las condiciones, el equipo, las personas, etc. La función del premuestreo no es tomar datos útiles para la investigación sino validar las condiciones de muestreo frente a los objetivos propuestos: ¿qué variables incidentes hay? ¿qué población se debe evaluar? ¿cuántas muestras deberían tomarse? ¿cuántos sujetos deben entrevistarse? ¿a qué condición social o cultural pertenecen? etc.

Capítulo

5

Recolectar datos o información

Una vez que seleccionamos el diseño de investigación apropiado, aprenderemos a recolectar información, resolviendo ejercicios prácticos utilizando el diagrama siguiente:

Recolectar datos { *Cuestionario* ⇔ *Equipos de medición* ⇔ *Codificación*

Recolectar Datos: Recolectar los datos implica seleccionar un instrumento de medición disponible o desarrollar uno propio.

- **Instrumento de medición (Cuestionario):** Cualquier recurso de que se vale el investigador para acercarse a los fenómenos y extraer de ellos información.
- **Cuestionario:** Representa un conjunto de preguntas respecto a una o más variables a medir, utilizada para sacar información de cualquier fenómeno.
- **Equipos de Medición:** Cualquier equipo que permita apoyar la medición de variables (grabadora de audio o vídeo, computador, satélite, gráficos, informes, expedientes, estereoscopio, etc.).
- **Codificación de datos:** Codificar consiste en asignarle un número o letra diferente a las posibles respuestas (alternativas) de una pregunta.



RECOLECTAR DATOS

Recolectar los datos implica seleccionar un instrumento de medición disponible o desarrollar uno propio.

Para obtener información se realiza un procedimiento sistemático que implica tres actividades estrechamente vinculadas entre sí:

- Instrumento de medición 'cuestionario' (encuesta, entrevista, observación, experimento, prueba, etc.).
- Equipos de medición.
- Codificación de datos.

A continuación presentamos un desarrollo bien definido que nos permitirá conocer cada uno de las actividades para recolectar información.

INSTRUMENTO DE MEDICIÓN (CUESTIONARIO)

Cualquier recurso de que se vale el investigador para acercarse a los fenómenos y extraer de ellos información. Un instrumento de medición debe cubrir dos requisitos: confiabilidad y validez.

El cuestionario es el instrumento de medición más completo y general que se puede aplicar a cualquier tipo de investigación; por tanto, una encuesta, entrevista, observación, prueba (test) u otro instrumento serán tratados como un cuestionario.

Soporte electrónico de datos

Entenderemos por soporte electrónico de datos a todo documento o elemento que conteniendo información es "leído" o interpretado por una unidad periférica de un ordenador (USB, CD, DVD, Disco Magnético Flexible, Disco Duro, PD, etc.).

Así, los datos recogidos en un formato de archivo o en una base de datos de cualquier software existente de recolección de información. También se comporta como un cuestionario, con la única diferencia que alguna información del archivo electrónico no será relevante para el estudio.

También a partir de datos electrónicos ya existentes puede construirse un cuestionario de acuerdo al problema planteado. Para datos electrónicos se utiliza mismo diseño muestral definido anteriormente.

CUESTIONARIO

Representa un conjunto de preguntas respecto a una o más variables a medir. El contenido de las preguntas puede ser tan variado como los aspectos que mida. Constituye un test escrito que el investigador utiliza para sacar información de cualquier fenómeno. Los cuestionarios también pueden ser electrónicos (ver ejemplos de cuestionario).

1. Diseñar un cuestionario

Debemos establecer el propósito del cuestionario acorde con las metas de la investigación. Luego, formular las preguntas en orden lógico sobre la temática a desarrollar y por último, escribirlas de

manera clara y sencilla.

El cuestionario se inicia con el encabezado u hoja de portada donde se recoge información básica que permita identificar la unidad de estudio (individuo). Por ejemplo: nombre, género, edad, ingresos, lugar de nacimiento, vida útil, número de expediente etc., según el objeto de estudio.

1.1 Tamaño de un cuestionario

Muy corto se pierde información, muy largo puede resultar tedioso. En este último caso, las personas pueden negarse a responder, o al menos, lo responderán en forma incompleta. El tamaño depende del número de variables y dimensiones a medir, el interés de los encuestados y la manera como es administrado. No hacer preguntas innecesarias o injustificadas.

1.2 Tipos de preguntas del cuestionario

Preguntas cerradas:

Contienen categorías o alternativas de respuestas que han sido delimitadas. Es decir, se presentan a los sujetos las posibilidades de respuestas y ellos deben circunscribirse a ellas.

Pueden ser dicotómicas (dos alternativas de respuestas) o incluir varias alternativas de respuestas.

Ejemplo. Edad:

niño Adolescente Joven Adulto

Preguntas abiertas:

No delimitan de antemano las alternativas de respuesta. Por lo cual, el número de categorías de respuesta es muy elevado; en teoría, es infinito.

Ejemplo. Edad (Un año, 10 años, 12 años, 15 años, etc)

Ítems de las Preguntas

- () Totalmente de acuerdo.
- () De acuerdo.
- () Indiferente.
- () En desacuerdo.
- () Totalmente en desacuerdo.

1.3 Inquietudes en torno a las preguntas que fórmula

- ¿Se relaciona la pregunta con los objetivos de la investigación?
- ¿Es correcto e idóneo el tipo de pregunta?
- ¿Cada pregunta maneja un solo interrogante?
- ¿Emplea términos subjetivos?
- ¿Requiere la pregunta conocimientos e información de los que carece el sujeto?
- ¿Exige la pregunta información personal o confidencial que el sujeto probablemente se resistirá a proporcionar?

- ¿Está la pregunta impregnada de aceptabilidad social?

1.4 Recomendaciones para formular las preguntas

- Escribirlas en forma corta, clara y sencilla.
- Dar prioridad a preguntas cerradas (falso: verdadero, sí: no: no sé, selección múltiple, marque en una escala de 1 a 5..., marque el rango en que usted..., marque el grado..., totalmente de acuerdo....totalmente en desacuerdo, etc.).
- Evitar el uso de negaciones en las preguntas.
- Deben aparecer en una secuencia lógica.
- La redacción debe ser neutral y no predisponer a un tipo de respuesta.
- No deben dar lugar a varias interpretaciones.
- Evitar términos técnicos.
- Las opciones de respuesta deben ser mutuamente excluyentes.
- Las encuestas o formularios deben ser iguales para toda la población o muestra.

1.5 Sugerencias para aplicar un cuestionario

- Iniciar el desarrollo del cuestionario con una conversación previa con las personas sobre cualquier otro tema para crear un ambiente de confianza.
- Concentrarse en obtener la información deseada.
- Conocer las preguntas apropiadamente como para no estar consultando continuamente el formato.
- Realizar la encuesta en un lugar tranquilo donde no hayan distracciones o interferencias.
- No expresar aprobación o desaprobación ante las respuestas ni emitir juicios sobre ellas.
- Si el entrevistado no proporciona una respuesta satisfactoria, plantearla de otra manera hasta obtenerla.
- Procurar grabar o filmar las sesiones, pero solicitando permiso para ello.
- Practicar la entrevista previamente.
- Permitir la libre expresión del entrevistado.
- Permitir al entrevistado usar una forma narrativa.
- Interrumpirlo lo menos posible.
- Permitir un desarrollo cronológico.
- No apegarse a un guión estricto.
- Formular las preguntas de modo casual.

1.6 Desarrollar un Cuestionario

El cuestionario se puede desarrollar de diferentes maneras:

- Auto administrado*: El cuestionario se proporciona directamente a los encuestados, quienes lo contestan. No hay intermediarios y las respuestas las marcan ellos.
- Entrevista personal*. La entrevista es una confrontación interpersonal, en el cual el entrevistador formula a los encuestados preguntas cuyo fin es conseguir respuestas relacionadas con el problema de investigación. Un entrevistador aplica el cuestionario a los entrevistados.

- c) *Entrevista telefónica*: Esta situación es similar a la anterior, sólo que la entrevista no es “cara a cara” sino a través del teléfono.
- d) *Por correo postal, electrónico o servicio de mensajería*. Los encuestados contestan directamente el cuestionario, ellos marcan o anotan las respuestas.
- e) *Electrónico*: Información en un dispositivo electrónico de datos, es decir base de datos, hojas de cálculo, y cualquier software donde se almacena información (USB, CD, Disco duro, DVD, etc).

Observaciones

- La elaboración de la encuesta requiere de gran conocimiento previo del fenómeno por parte del investigador. Cuando las encuestas son anónimas permiten, además, mayor honestidad en las respuestas.
- Es importante realizar un muestreo o prueba del instrumento (prueba piloto) ante un grupo reducido de individuos (unidad de estudio) lo que permitirá, en forma crítica, ayudar a identificar falencias o dificultades del cuestionario.

EQUIPOS DE MEDICIÓN

En estudios de investigación se emplean múltiples equipos que permiten apoyar la medición de variables como: grabadoras de audio ó vídeo, computador, imágenes satelitales, fotografías aéreas, gráficos, informes, expedientes, estereoscopio, microscopio, espectrofotómetro, cámaras de conteo; en física y química se utiliza medidores de gases, nutrientes, iones o elementos diversos, pH, temperatura, intensidad del sonido, cuadrantes, dragas, botellas muestreadoras, contadores manuales, balanzas, redes, trampas, cámaras de fotografía, binoculares, geoposicionadores, cuerdas metradas, papel milimetrado, etc.

CODIFICACIÓN DE DATOS

1. Codificación pregunta cerrada: Codificar consiste en asignarle un número o letra diferente a las posibles respuestas (alternativas) de una pregunta, por ejemplo:

1. Sexo:

a. Femenino () b. Masculino ()

La pregunta sexo tiene dos alternativas el código de femenino es 1 y de masculino 2, en adelante cuando anotamos 1 nos referimos a femenino y 2 a masculino.

2. Codificación pregunta abierta: A cada respuesta de la pregunta le daremos un código, un número o letra diferente, que servirá para agrupar tras de sí a todas las respuestas u observaciones que sean idénticas o que, al menos, aparezcan como equivalentes. Por ejemplo:

1. La opinión que tiene respecto del SIDA

Las opiniones expuestas son: enfermedad contagiosa, tratamiento caro, utilizar preservativos, medicinas costosas, etc.

Se agrupan las respuestas idénticas y se asigna un código.

Por ejemplo.

Al código 1 corresponde **Costo** por que las respuestas que se refieren a costo son: medicinas costosas y tratamiento caro

Al código 2 le corresponde **Contagio** por que las respuestas que se refieren de alguna manera a contagio son: enfermedad contagiosa, utilizar preservativos.

EJERCICIOS DE RECOLECTAR DATOS

Los cuestionarios llamados formatos de recolección de información constituyen una forma estructurada, sistemática y organizada para registrar los datos.

EJEMPLO 1

Problema: Se quiere realizar un análisis del comportamiento de los consumidores que realizaron sus compras semanales en un Supermercado.

Instrumento de medición (Cuestionario)

Construir un formato de recolección de información o cuestionario obligará al investigador a pensar en términos de sus requerimientos de información, al planeamiento del trabajo de campo o al diseño de su experimento, según sea el caso, y a las técnicas estadísticas para el análisis de los resultados.

a) CUESTIONARIO:

Se diseñó el siguiente cuestionario:

ENCUESTA SOBRE COMPRAS EN EL SUPERMERCADO

Se quiere analizar el comportamiento de los consumidores que hicieron sus compras semanales en un Supermercado, realizando el siguiente cuestionario:

Marque con una “x” la alternativa correcta o escriba en las líneas Punteadas.

1. Edad:

- a) Adolescente b) Joven c) Adulto

2. Sexo:

- a) Varón b) Mujer

3. Grado de Instrucción:

- a) Primaria b) Secundaria c) Superior d) No tiene

4. Estado Civil:

- a) Soltero b) Casado c) Divorciado d) Viudo

5. Horario de Compra:

- a) Mañana b) Tarde c) Noche

6. Compras Semanales en Artículos de Aseo Personal (Nuevo Soles)

7. Compras Semanales en Verduras (Nuevo Soles)

8. Compras Semanales en Abarrotes (Nuevo Soles)
9. Compras Semanales en Bebidas (Nuevo Soles)

EL cuestionario consta de 9 preguntas; las del 1 al 5 son preguntas cerradas y variables de cualidad, y las variables del 6 al 9 son abiertas y de cantidad.

En caso de la variable Edad (L).

También puede trabajar como una pregunta abierta, de la siguiente manera: (1) Edad, porque en las líneas punteadas tiene múltiples opciones como, 12 años, 15 años, 16 años, 37 años, etc.

Finalmente se recopilan los datos correspondientes al cuestionario, es decir, se realiza el llenado de información en la encuesta mostrada anteriormente. Para el ejercicio utilizamos un tamaño de muestra de 30 personas encuestadas en el supermercado.

b) Equipos de Medición

Para el desarrollo de este cuestionario se utilizó la caja registradora, así como equipos de audio y video con la finalidad de no incomodar al cliente.

c) Codificando la encuesta:

Se debe tener presente que cada pregunta se convierte en una variable del cuestionario que será codificada y dicha codificación se muestra a continuación.

Variable Edad:

Adolescente le corresponde el código "0", Joven le corresponde el código "1" y Adulto el código "2".

X1: Edad (0: Adolescente; 1: Joven; 2: Adulto).

Variable Sexo:

Mujer le corresponde el código "0", Varón le corresponde el código "1"

X2: Sexo (0: Mujer; 1: Varón).

Así sucesivamente hasta la variable o pregunta 6:

X3: Grado de Instrucción (1: No tiene; 2: Primaria; 3: Secundaria 4: Superior).

X4: Estado Civil (1: Soltero; 2: Casado; 3: Divorciado 4: Viudo).

X5: Turno (0: Mañana; 1: Tarde; 2: Noche).

Las variables de 6 al 9 no son necesarias codificar ya que son variables de cantidad.

X6: Venta de Artículos de Aseo Personal (Nuevo Soles).

X7: Venta de Verduras (Nuevo Soles).

X8: Venta de Abarrotes (Nuevo Soles).

X9: Venta de Bebidas (Nuevo Soles).

La tabla representa las encuestas codificadas, donde las columnas indican las variables y las filas los casos (30 personas encuestadas).

| .ni | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 |
|-----|----|----|----|----|----|-----|------|------|-----|
| 1 | 0 | 0 | 3 | 2 | 0 | 2,0 | 17,0 | 34,5 | 6,0 |
| 2 | 2 | 0 | 4 | 1 | 2 | 3,0 | 20,0 | 40,0 | 6,0 |
| 3 | 2 | 0 | 2 | 3 | 2 | 0,3 | 10,0 | 31,6 | 6,0 |
| 4 | 1 | 0 | 2 | 4 | 1 | 3,3 | 9,0 | 35,4 | 6,0 |
| 5 | 2 | 0 | 3 | 4 | 2 | 1,3 | 23,0 | 30,0 | 6,0 |
| 6 | 2 | 0 | 4 | 4 | 2 | 0,4 | 13,0 | 32,9 | 6,0 |
| 7 | 1 | 0 | 2 | 2 | 0 | 1,5 | 12,0 | 33,2 | 6,0 |
| 8 | 0 | 0 | 3 | 2 | 0 | 4,5 | 19,0 | 33,1 | 6,0 |
| 9 | 2 | 1 | 4 | 3 | 2 | 2,5 | 18,0 | 35,6 | 6,0 |
| 10 | 1 | 1 | 2 | 2 | 1 | 0,3 | 24,0 | 33,0 | 6,0 |
| 11 | 2 | 1 | 4 | 1 | 2 | 1,0 | 7,0 | 34,5 | 6,0 |
| 12 | 2 | 0 | 3 | 2 | 2 | 6,0 | 10,0 | 33,2 | 6,0 |
| 13 | 2 | 0 | 4 | 4 | 2 | 5,6 | 5,0 | 31,5 | 6,0 |
| 14 | 2 | 1 | 4 | 4 | 2 | 6,0 | 14,0 | 36,2 | 1,0 |
| 15 | 1 | 0 | 1 | 1 | 1 | 1,2 | 15,0 | 36,8 | 1,0 |
| 16 | 2 | 0 | 1 | 2 | 2 | 0,2 | 12,0 | 35,4 | 2,0 |
| 17 | 2 | 0 | 1 | 3 | 2 | 6,0 | 14,0 | 33,2 | 2,0 |
| 18 | 2 | 0 | 2 | 2 | 2 | 5,5 | 6,0 | 35,9 | 2,0 |
| 19 | 2 | 0 | 4 | 4 | 2 | 6,5 | 9,0 | 36,5 | 3,0 |
| 20 | 2 | 0 | 1 | 4 | 2 | 0,2 | 13,0 | 38,2 | 3,0 |
| 21 | 2 | 0 | 3 | 1 | 2 | 2,3 | 6,0 | 34,5 | 3,0 |
| 22 | 2 | 0 | 2 | 2 | 2 | 0,2 | 7,0 | 36,2 | 3,0 |
| 23 | 1 | 0 | 1 | 1 | 1 | 2,3 | 10,0 | 36,8 | 3,0 |
| 24 | 2 | 0 | 2 | 4 | 2 | 1,5 | 13,0 | 30,1 | 3,0 |
| 25 | 2 | 0 | 3 | 4 | 2 | 5,0 | 7,0 | 36,0 | 3,0 |
| 26 | 2 | 0 | 2 | 2 | 2 | 4,5 | 18,0 | 35,4 | 3,0 |
| 27 | 0 | 0 | 3 | 2 | 0 | 0,3 | 24,0 | 31,5 | 3,0 |
| 28 | 2 | 0 | 2 | 3 | 2 | 5,5 | 7,0 | 36,2 | 4,0 |
| 29 | 2 | 0 | 2 | 2 | 2 | 0,5 | 10,0 | 33,9 | 4,0 |
| 30 | 2 | 0 | 2 | 2 | 2 | 0,2 | 4,0 | 34,9 | 4,0 |

EJEMPLO 2

Problema: Estudiar el comportamiento socioeconómico de un grupo de personas de una determinada ciudad.

a) Instrumento de medición: Cuestionario

Se diseñó el siguiente cuestionario:

ENCUESTA: SOCIO ECONÓMICA

Estudiar el comportamiento socioeconómico de un grupo de personas de una determinada ciudad.

Con esta finalidad se realizó el siguiente cuestionario:

Marque con una "x" la alternativa correcta o escriba en las líneas Punteadas.

1. Sexo:

a) Varón b) Mujer

2. Nivel Cultural:

a) Bajo b) Alto

3. Ingresos Mensuales: Nuevo Soles.

4. Peso: Kilogramos.

EL cuestionario consta de 4 preguntas, de las cuales 1 y 2 son preguntas cerradas y variables de cualidad, mientras las variables del 6 al 9 son abiertas y de cantidad.

Finalmente se recopilan los datos correspondientes al cuestionario. Para el ejercicio utilizamos un tamaño de muestra de 5 personas entrevistadas.

b) Equipos de Medición

Para el desarrollo de este cuestionario se utilizó: boletas de pago, una balanza, certificados de estudio.

c) Codificando la encuesta

Se debe tener presente que cada pregunta se convierte en una variable del cuestionario que será codificada.

Variable Sexo:

Mujer le corresponde el código "0", Varón le corresponde el código "1"

Sexo (0: mujer y 1: varón)

Nivel cultural (1: bajo y 2: alto)

Las variables Ingresos en nuevos soles y Peso en kilogramos no necesitan codificarse.

Los datos corresponden a 5 sujetos (casos) medidos en cuatro variables: Sexo (0: mujer y 1: varón), Nivel cultural (1: bajo y 2: alto), Ingresos en nuevos soles y Peso en kilogramos.

| Sujetos | Sexo | Nivel Cultural | Ingresos | Peso |
|---------|------|----------------|----------|------|
| 1 | 0 | 1 | 1000 | 40 |
| 2 | 1 | 2 | 2000 | 60 |
| 3 | 0 | 1 | 1500 | 50 |
| 4 | 1 | 1 | 3000 | 70 |
| 5 | 0 | 2 | 1800 | 40 |

DEFINICIONES IMPORTANTES

1. Instrumentos de Medición

Un instrumento es cualquier recurso de que se vale el investigador para acercarse a los fenómenos y extraer de ellos información. Dentro de cada instrumento concreto pueden distinguirse dos aspectos diferentes: forma y contenido.

Forma: se refiere al tipo de aproximación que establecemos con lo empírico, a las técnicas que utilizamos para esta tarea.

Contenido: éste queda expresado en la especificación de los datos que necesitamos conseguir; se concreta en una serie de ítems que no son otra cosa que los mismos indicadores que permiten medir las variables, pero que asumen ahora la forma de pregunta, puntos a observar, elementos a registrar, etc.

Una adecuada construcción de los instrumentos permite que la investigación alcance una correspondencia entre teoría y hechos. Si los instrumentos son defectuosos los datos recogidos no servirán para satisfacer las interrogantes, no se podrán obtener información necesaria, los datos vendrán falseados y distorsionados.

2. Confiabilidad de los instrumentos de medición

Es casi imposible que una medición sea perfecta, siempre existe un grado de error. El error debe ser el mínimo posible. La medición de cualquier fenómeno se puede conceptualizar con la siguiente fórmula:

$$O = T + \epsilon$$

Donde

O: Representa los valores observados (experimento).

T: Representa los valores verdadero o teórico.

ϵ : Representa el grado de error en la medición.

Si no hay error de medición ($\epsilon=0$), el valor observado y el verdadero son equivalentes. Esta situación es ideal.

$$O = T + \epsilon$$

Es importante que el error sea lo más reducido posible.

3. Construir un instrumento de medición

Existen diversos tipos de instrumentos de medición, cada uno con características diferentes. Sin embargo, el procedimiento general para construirlos es semejante.

Pasos para construir un instrumento de medición:

- Listar las variables que se pretenden medir u observar.
- Revisar su definición conceptual y comprender su significado.
- Revisar como han sido definidas operacionalmente las variables, esto es, como se ha medido cada variable.
- Asegurarnos de tener un número suficiente de ítems para medir las variables en toda su dimensión.

4. Tipos de instrumento de medición

4.1 Encuesta:

El diseño encuesta parte de la premisa de que si queremos conocer algo sobre el comportamiento de las personas, lo mejor, lo más directo y simple, es preguntárselo directamente a ellas. Se trata por tanto de requerir información a un grupo socialmente significativo de personas acerca del problema en estudio para luego, mediante un análisis de tipo cuantitativo, sacar las conclusiones que se correspondan con los datos.

Cuando se recaba información de toda la población involucrada en el problema en estudio se le llama Censo. Los censos, por las dificultades materiales que implica su realización, son casi siempre trabajos de envergadura, que sólo pueden ser acometidos por los Estados o por instituciones de muy amplios recursos.

Las encuestas por muestreo se escogen mediante procedimientos estadísticos en una parte significativa de la población. Las conclusiones que se obtienen para este grupo se proyectan luego a la totalidad del universo, teniendo en cuenta los errores muestrales que se calculen para el caso. De esta forma, los hallazgos obtenidos a partir de la muestra pueden generalizarse a todo el universo con un margen de error conocido y limitado previamente por el investigador.

Las principales ventajas de la encuesta son las siguientes:

- Su conocimiento de la realidad es primario, no mediado, y por lo tanto menos engañoso. Al acudir directamente a la gente para conocer su situación, conducta u opinión, nos precavemos contra una multiplicidad de distorsiones y nos ponemos a salvo de interpretaciones que pueden estar altamente teñidas de subjetividad.
- Es un método de trabajo relativamente económico y rápido. Si se cuenta con un equipo de entrevistadores y codificadores convenientemente entrenado, resulta fácil llegar rápidamente a una multitud de personas y obtener una gran cantidad de datos en poco tiempo.
- Las desventajas de la encuesta son:
- La encuesta recoge la visión que la gente tiene de sí misma; no puede dudarse de que ésta es siempre una imagen singular y muy subjetiva y que, para algunos temas, puede ser deliberadamente falsa e imprecisa. No es lo mismo lo que las personas hacen, sienten o creen, que lo que ellas mismas dicen que hacen, creen o sienten.
- La encuesta no relata los hechos sociales desde el punto de vista de sus actores; puede, en este sentido, llegar a una cierta profundidad y sistematicidad, pero resulta poco apta para reconocer las relaciones sociales ya sean interpersonales o institucionales.

4.2 Entrevista

Es una forma específica de interacción social que tiene por objeto recolectar datos para una indagación. El investigador formula preguntas a las personas capaces de aportarle datos de interés, estableciendo un diálogo peculiar, asimétrico, donde una de las partes busca recoger informaciones y la otra es la fuente de esas informaciones.

Ventaja de la entrevista:

- La ventaja esencial de la entrevista reside en que son los mismos actores sociales quienes proporcionan los datos relativos a sus conductas, opiniones, deseos, actitudes y expectativas, cosa que por su misma naturaleza es casi imposible de observar desde fuera. Nadie mejor que la misma persona involucrada para hablarnos acerca de todo aquello que piensa y siente, de lo que ha experimentado o proyecta hacer.

Desventaja de la entrevista:

- Cualquier persona entrevistada podrá hablarnos de aquello que le preguntemos pero siempre nos dará la imagen que tiene de las cosas, lo que cree que son, a través de toda su carga subjetiva de intereses, prejuicios y estereotipos. La propia imagen que el entrevistado tiene de sí mismo podrá ser radicalmente falsa y, en todo caso, estará siempre idealizada de algún modo, distorsionada, mejorada o retocada según factores que no es del caso analizar aquí, pero que nunca podemos prever en detalle.

Para tener éxito con una entrevista es recomendable prestar atención a una serie de factores.

La apariencia exterior del entrevistador debe ser adecuada al medio social donde habrá de formular sus preguntas, evitando innecesarias reacciones de temor, agresividad o desconfianza. El entrevistador deberá ser una persona de por lo menos una cultura media, que comprenda el valor y la importancia de cada dato recogido y la función que su trabajo desempeña en la investigación. Tendrá que ser mentalmente ágil, no tener prejuicios marcados frente a ninguna categoría de personas y, sobre todo, ser capaz de dejar hablar libremente a los demás, eliminando por completo todo intento de convencerlos, apresurarlos o agredirlos por sus opiniones. La entrevista deberá realizarse a las horas más apropiadas para las personas que responden, teniendo en cuenta que su posible duración no afecte la confiabilidad de los datos.

Clasificación de los diversos tipos de entrevistas:

Entrevistas no estructuradas

Una entrevista no estructurada o no formalizada es aquella en que existe un margen más o menos grande de libertad para formular las preguntas y las respuestas. No se guían por lo tanto por un cuestionario o modelo rígido, sino que discurren con cierto grado de espontaneidad, mayor o menor según el tipo concreto de entrevista que se realice. Entre estos tenemos los siguientes:

Entrevista informal: Es la modalidad menos estructurada posible de entrevista ya que la misma se reduce a una simple conversación sobre el tema de estudio. Lo importante no es aquí definir los límites de lo tratado ni ceñirse a algún esquema previo, sino “hacer hablar” al entrevistado, para obtener un panorama de los problemas más salientes, de los mecanismos lógicos y mentales del encuestado, de los temas que para él resultan de importancia.

Entrevista focalizada: Es prácticamente tan libre y espontánea como la anterior, pero tiene la particularidad de concentrarse en un único tema. El entrevistador deja hablar sin restricciones al entrevistado, proponiéndole apenas algunas orientaciones básicas pero, cuando éste se desvía del tema original y se desliza hacia otros distintos, el entrevistador vuelve a centrar la conversación sobre el primer asunto, y así repetidamente.

Entrevistas por pautas o guías: Se guían por una lista de puntos de interés que se van explorando en el curso de la entrevista. Los temas deben guardar una cierta relación entre sí.

El entrevistador, en este caso, hace muy pocas preguntas directas y deja hablar al encuestado siempre que vaya tocando alguno de los temas señalados en la pauta o guía. En el caso de que éste se aparte de ellos, o que no toque alguno de los puntos en cuestión, el investigador llamará la atención sobre ellos, aunque tratando siempre de preservar en lo posible la espontaneidad de la interacción.

Todas estas formas de entrevistas poseen la ventaja de permitir un diálogo más profundo y rico, de presentar los hechos en toda su complejidad, captando no sólo las respuestas a los temas elegidos sino también las actitudes, valores y formas de pensar de los entrevistados, a veces inaccesibles por otras vías. Su inconveniente radica en que es poco práctico sistematizar un gran número de entrevistas de este tipo, organizándolas estadísticamente, pues pueden tener muy pocos puntos de contacto entre sí. Otra dificultad que no debe omitirse es su costo, pues involucran la presencia de personal altamente especializado.

Entrevistas formalizadas

Estas se desarrollan en base a un listado fijo de preguntas cuyo orden y redacción permanece invariable. Comúnmente se administran a un gran número de entrevistados para su posterior tratamiento estadístico. Por este motivo, es la forma de recolección de datos más adecuada para el diseño de encuesta.

Entre sus ventajas principales mencionaremos su rapidez y el hecho de que pueden ser llevadas a cabo por personas con mediana preparación, lo cual redundará en su bajo costo. Otra ventaja evidente es su posibilidad de procesamiento estadístico, ya que al guardar una estricta homogeneidad sus respuestas resultan comparables y agrupables.

Su desventaja es que reducen grandemente el campo de información registrado, limitando los datos a los que surgen de una lista taxativa de preguntas. Esta lista de preguntas, que es el instrumento concreto de recolección empleado en este caso, recibe el nombre de cuestionario y puede ser administrado sin que necesariamente medie una entrevista.

Las preguntas se dividen en dos grandes tipos: de alternativas fijas, de final abierto.

Preguntas de alternativas fijas: Llamadas cerradas, formalizan más el cuestionario, pues en ellas sólo se otorga al entrevistado la posibilidad de escoger entre un número limitado de respuestas posibles. Por ejemplo ¿Cree Ud. en Dios?, sí o no; estas alternativas ofrecidas pueden ser dos, tres, etc., el encuestado solo puede elegir una respuesta que esté en la lista. Hay que tener sumo cuidado en la redacción de estas alternativas, procurando especialmente que ellas sean exhaustivas y mutuamente excluyentes, tal como ocurría para la confección en un sentido determinado.

Preguntas de final abierto: llamadas también simplemente abiertas, proporcionan una variedad más amplia de respuestas pues éstas pueden ser escogidas libremente por los encuestados. Su redacción debe ser muy cuidadosa para evitar respuestas confusas o erróneas, y para evitar además que ellas predispongan a los entrevistados en uno u otro sentido. Un ejemplo de pregunta abierta sería ¿Qué opina Ud. de Dios?; la respuesta aquí puede ser infinitamente variada, según lo opinión de cada persona.

4.3 Pruebas

Una variante de las encuestas son las pruebas o test, los cuales son de amplia utilización en el psicología y se emplean para medir una habilidad cognitiva, emocional, motriz o de otra naturaleza.

Tipos de Pruebas:

Prueba de referencia de norma: Mide diferencias individuales y se evalúa, por tanto, frente al desempeño del grupo. Ejemplo: el 80% está por debajo de Juan.

Prueba de referencia de criterio: Se evalúa cada individuo frente al dominio del conocimiento y no frente a otros individuos. Ejemplo: Juan respondió el 85% de las preguntas.

Pruebas individuales: Se aplican a un individuo a la vez. Muchas de ellas son orales.

Pruebas grupales: Se aplican a grupos de personas en forma simultánea. Son generalmente escritas.

Prueba objetiva: Es clara y no es ambigua en su puntuación, por lo que diferentes evaluadores deberían asignar igual puntuación. Ejemplo: selección múltiple, falso-verdadero, "matching".

Prueba subjetiva: La asignación de la puntuación está afectada por el juicio del calificador.

Pruebas de potencia: Tienen tiempo suficiente para responder cada pregunta.

Pruebas verbales: Enfatizan el uso del lenguaje como medio de responder.

Pruebas no verbales: No enfatizan el papel de la lectura en la prueba. Por ejemplo, en niños pequeños.

Pruebas sin lenguaje: Se administran por gestos o pantomima. Por ejemplo, en personas sordas o que hablan otro idioma.

Pruebas de signos: Distinguen un grupo de individuos de otro: *Son capaces de ... o no lo son.*

Pruebas de selección: Se emplean para admitir o rechazar a una persona. Por ejemplo, investigadores.

Pruebas de clasificación: Permiten ubicar en un determinado nivel o clase a los individuos.

Pruebas de diagnóstico: Evalúan fuerzas o debilidades en orden a mejorar el desempeño.

Tipos de Dominios:

Cognitivo: Enfatiza comprensión, retención y desarrollo de conocimiento e intelecto.

Afectivo: Enfatiza conductas como sentimientos, emociones o valores.

Sicomotor: Se refiere a comportamientos musculares o motores.

Otros: Comprensión literaria, comprensión inferencial, evaluación, apreciación, etc.

Tipos de preguntas:

a. Falso – verdadero:

Ventajas:

- Permiten hacer más preguntas por lo rápido.
- Son fáciles de construir.
- Son fáciles de evaluar.

Desventajas:

- No admite resultados intermedios.
- Permiten adivinar en alto grado (50%).

Sugerencias:

- Evitar dar pistas no intencionales.
- Hacer aproximadamente 50% falsas y 50% verdaderas.
- Cada pregunta debe ser inequívocamente falsa o verdadera.
- Incrementar el número de preguntas.
- Para reducir adivinanzas se sugiere corregir el puntaje así:
- Puntaje = Buenas – malas
- Puntaje = Buenas – {malas/ (opciones –1)}; cuando hay 2, 3 o más opciones de respuesta (selección múltiple).

b. Selección múltiple:

Ventajas:

- Es muy versátil para evaluar conocimiento.
- Permite evaluar gran cantidad de material en poco tiempo.
- La evaluación es objetiva.
- Permite aplicaciones de grado, no sólo de falso y verdadero.
- Reduce el efecto de adivinar.
- Tiene varias opciones: elegir, llenar, reagrupar, completar.

Sugerencias:

- Definir claramente el enunciado en forma positiva y evitar las preguntas en negativo.
- Evitar respuestas complejas. Ejemplo: *a y c son ciertas; a y b pero no c; ninguna de las anteriores; todas las anteriores.*
- Cambiar de lugar la opción correcta.
- Evitar usar 2 hojas, una con preguntas y otra para respuestas.

c. *Emparejamiento* (“*matching*”): Es una selección múltiple donde se deben asociar 2 columnas de ítem.

Ventajas y desventajas:

- Mide asociaciones.
- Reduce efecto de adivinar.
- Tiende a tener información trivial.

Sugerencias:

- Deben ser ítem homogéneos Ejemplo: *todos los ítem de una columna que sean eventos; todos los de la otra personas; evitar que al interior de una columna hayan personas, ciudades, fechas, etc.*
- Colocar más opciones que ítem. De este modo si se saben algunos no hay muchas opciones de adivinar los que quedan por marcar.
- Organizar ítem y opciones ordenadamente: A, B, C../ 1, 2, 3..
- Colocar 5 a 6 ítems por cada pregunta.

- Si los ítems son largos, usar opciones cortas.
- Los ítems y las opciones deben estar en la misma página.
- Evitar dar pistas.

d. *Test de completar*: tiene el propósito de completar frases donde falta alguna palabra o concepto.

Ventajas:

- Fácil de construir.
- No se puede adivinar.
- Sirve para evaluar respuestas precisas: nombres, fechas, definiciones, etc.

Sugerencias:

- Que haya sólo una respuesta.
- Usar sólo un espacio en blanco por pregunta.
- Tratar de ubicar el espacio en blanco al final de la frase.

e. Ensayos:

Ventajas:

- Permiten una respuesta libre.
- Elimina la posibilidad de adivinar.
- Permite medir pensamientos divergentes.

Desventajas:

- Son difíciles de evaluar.
- Consumen mucho tiempo.
- Tienen alto grado de subjetividad en la evaluación.
- Premian a los buenos escritores.

f. Otros:

Evaluación: se refiere a la calificación de las pruebas, encuestas o entrevistas.

Frecuencia: indica la regularidad de una prueba.

Las preguntas anteriores se pueden equiparar a las formuladas en el marco de las encuestas y las entrevistas, según los tipos definidos por Kerlinger (1983):

- *Reactivos de alternativas fijas* : que incluye las dicotómicas -falso-verdadero, sí-no, acuerdo-desacuerdo- y de alternativas múltiples donde el sujeto puede elegir entre varias posibilidades -elija la frase que mejor.., rangos de valores, frecuencias, etc.
- *Reactivos de escala*: se trata de alternativas fijas pero con niveles de grado -totalmente de acuerdo-parcialmente de acuerdo...; completa aprobación-aprobación alta...; califique de 1 a 5..., etc.
- *Reactivos de finalidad abierta*: no imponen restricciones a las respuestas.

4.4 Cuestionarios

Consiste en un conjunto de preguntas respecto a una o más variables a medir. El contenido de las preguntas de un cuestionario puede ser tan variado como los aspectos que mida. Y podemos hablar de dos tipos de preguntas: cerradas y abiertas.

Las preguntas cerradas contienen categorías o alternativas de respuestas que han sido delimitadas. Es decir, se presentan a los sujetos las posibilidades de respuestas y ellos deben circunscribirse a ellas. Pueden ser dicotómicas (dos alternativas de respuestas) o incluir varias alternativas de respuestas.

Las preguntas abiertas no delimitan de antemano las alternativas de respuesta. Por lo cual el número de categorías de respuesta es muy elevado; en teoría, es infinito.

Tamaño de un cuestionario:

No existe una regla al respecto, pero si es muy corto se pierde información y si es largo puede resultar tedioso. En este último caso, las personas pueden negarse a responder, o al menos, lo responderán en forma incompleta. El tamaño depende del número de variables y dimensiones a medir, el interés de los encuestados y la manera como es administrado. No hacer preguntas innecesarias o injustificadas.

Aplicar Cuestionarios:

Los cuestionarios pueden ser aplicados de diversas maneras:

Autoadministrado. El cuestionario se proporciona directamente a los encuestados, quienes lo contestan. No hay intermediarios y las respuestas las marcan ellos.

Entrevista personal. Un entrevistador aplica el cuestionario a los entrevistados.

Entrevista telefónica: Esta situación es similar a la anterior, sólo que la entrevista no es “cara a cara” sino a través del teléfono.

Por correo postal, electrónico o servicio de mensajería. Los encuestados contestan directamente el cuestionario, marcan o anotan las respuestas.

4.5 Observación

Consiste en el registro sistemático, válido y confiable de comportamiento o conducta manifiesta. Es un método más utilizado por quienes están orientados conductualmente.

Pasos para construir un sistema de observación:

- Definir con precisión el universo de aspectos, eventos o conductas a observar.
- Extraer una muestra representativa de los aspectos, eventos o conductas a observar.
- Un repertorio suficiente de conductas a observar.
- Establecer y definir las unidades de observación.
- Establecer y definir las categorías y subcategorías de observación.

Tipos de observación:

Observación participante: El observador interactúa con los sujetos observados.

Observación no participante: El observador no interactúa con los sujetos observados.

Ventajas de la observación:

Son técnicas de medición no obstructivas. En el sentido que el instrumento de medición no “estimula” el comportamiento de los sujetos (las escalas de actitud y los cuestionarios pretenden “estimular” una respuesta a cada ítem). Los métodos no obstructivos simplemente registran algo que fue estimulado por otros factores ajenos al instrumento de medición. Aceptan material no estructurado. Pueden trabajar con grandes volúmenes de datos.

5. Escala de medición

Todo problema de investigación, aún el más abstracto, implica de algún modo una tarea de medición de los conceptos que intervienen en el mismo.

La medida es intrínsecamente comparativa. En el caso más sencillo, es determinar cuántas veces una cierta unidad o patrón de medida cabe en el objeto a medir. Para medir la longitud de un objeto físico nosotros desplazamos una regla o cinta graduada sobre el mismo, observando cuántas unidades (centímetros) abarca el objeto en cuestión. Es decir que comparamos el objeto con nuestro patrón de medición para determinar cuántas unidades y fracciones del mismo incluye.

Si deseamos medir el peso de un objeto podremos expresar el valor del mismo en kilogramos, libras o cualquier unidad. En cambio para medir el grado de autoritarismo de un dirigente no existe una unidad ni una escala reconocida, por lo que el investigador se ve obligado a elegir alguna ya utilizada en otros trabajos o a construir una adaptada a sus necesidades.

Una escala puede concebirse como un continuo de valores ordenados correlativamente que admite un punto inicial y otro final.

Ejemplo.- Si evaluamos el rendimiento académico de estudiantes podemos asignar el valor cero al mínimo rendimiento imaginable; al mayor rendimiento posible podemos atribuirle un valor de 100 puntos. Con estos dos valores tendríamos ya marcados los límites de nuestra escala; para concluir de confeccionarla será necesario asignar a los rendimientos intermedios puntajes también intermedios. Con ello obtendremos una escala capaz de medir la variable rendimiento académico (trabajos, exámenes, etc.).

Para que una escala sea capaz de aportar información objetiva debe reunir los siguientes requisitos:

Confiability: Se refiere a la consistencia interior de la misma, a su capacidad para discriminar en forma constante entre un valor y otro.

Validez: Indica la capacidad de la escala para medir las cualidades para las cuales ha sido construida y no otras parecidas. Una escala tiene validez cuando verdaderamente mide lo que afirma medir.

Las escalas se clasifican en: escala nominal, ordinal, intervalos iguales y cociente o razón.

Escala nominal: Son aquellas en que sólo se manifiesta una equivalencia de categorías entre los diferentes puntos que asume la variable. Es como una simple lista de las diferentes posiciones que puede adoptar la variable, pero sin que en ella defina ningún tipo de orden o de relación.

Ejemplo en la variable cereal, los distintos valores que esa variable puede tomar serán: trigo, maíz, centeno, etc. Entre estos valores no cabe obviamente ninguna jerarquía, no se puede trazar ningún ordenamiento.

Escala ordinal: Distinguen los diferentes valores de la variable jerarquizándolos simplemente de acuerdo a un rango. Establecen que existe un orden entre uno y otro valor de la escala, de tal modo que cualquiera de ellos es mayor que el precedente y menor que el que le sigue a continuación. Sin embargo la distancia entre un valor y otro no queda definida sino que es indeterminada. En otras palabras, tales escalas nos esclarecen solamente el rango que las distintas posiciones guardan entre sí.

Ejemplo la variable “grado de escolaridad”: una persona que ha tenido 2 años de instrucción escolar ha recibido más instrucción que quien sólo tiene un año y menos que quien posee tres.

Escala de intervalos iguales además de poseer la equivalencia de categorías y el ordenamiento interno entre ellas, como en el caso de las ordinales, tienen las características de que la distancia entre sus intervalos está claramente determinada y que estos son iguales entre sí.

Ejemplo escalas termométricas. Entre 23 y 24 grados centígrados, existe la misma diferencia que hay entre 45 y 46 grados. La limitación que poseen es que no definen un cero absoluto, un valor límite que exprese realmente la ausencia completa de la cualidad medida.

Escala de cociente o razón. En ellas se conservan todas las propiedades de los casos anteriores pero además se añade la existencia de un valor cero, con lo que se hacen posibles ciertas operaciones matemáticas, tales como la obtención de proporciones y cocientes. Son escalas de cocientes las que miden la longitud, la masa, la intensidad de corriente eléctrica y otras variables del mundo físico.

5.1 Construcción de Escala

Escalas son un instrumento de medición o pruebas psicológicas que frecuentemente son utilizadas para la medición de actitudes.

El término actitud es la suma total de inclinaciones y sentimientos, prejuicios o distorsiones, nociones preconcebidas, ideas, temores, amenazas y convicciones de un individuo acerca de cualquier asunto específico. La actitud se expresa por medio de opiniones; por ejemplo, una persona que expresa su opinión sobre la caída del muro de Berlín refleja una actitud específica sobre los hechos referidos.

En una escala de medición de actitudes no interesa propiamente la opinión o el conjunto de palabras que expresa la persona. Lo que en realidad importa es la actitud de quién opina. La escala de medición de actitudes analiza los pensamientos y sentimientos de la persona hacia los hechos ya especificados.

Las actitudes pueden medirse a través de diversos tipos de escalas entre las que destacan la escala de actitudes tipo Likert y el escalograma de Guttman.

a) Escala de Likert

La escala de Likert mide actitudes o predisposiciones individuales en contextos sociales particulares. Se le conoce como escala sumada debido a que la puntuación de cada unidad de análisis se obtiene mediante la sumatoria de las respuestas obtenidas en cada ítem.

La escala se construye en función de una serie de ítems que reflejan una actitud positiva o negativa acerca de un estímulo o referente. Cada ítem está estructurado con cinco alternativas de respuesta:

- () Totalmente de acuerdo
- () De acuerdo
- () Indiferente
- () En desacuerdo
- () Totalmente en desacuerdo

La unidad de análisis que responde a la escala marcará su grado de aceptación o rechazo hacia la proposición expresada en el ítem, que por lo general tienen implícita una dirección positiva o negativa. Por ejemplo el ítem:

Los menonitas son un grupo étnico con excelentes valores hacia el trabajo.

- () Totalmente de acuerdo
- () De acuerdo
- () Indiferente
- () En desacuerdo
- () Totalmente en desacuerdo

Manifiesta una dirección positiva, en cambio si se expresará en la forma:

Los menonitas son un grupo étnico que tiene aversión al trabajo.

- () Totalmente de acuerdo
- () De acuerdo
- () Indiferente
- () En desacuerdo
- () Totalmente en desacuerdo

El ítem tiene una dirección negativa. La calificación o puntuación se asigna de acuerdo a la dirección del ítem; si tiene una dirección positiva la puntuación es:

- (+2) Totalmente de acuerdo
- (+1) De acuerdo
- (0) Indiferente
- (-1) En desacuerdo
- (-2) Totalmente en desacuerdo

En el caso de que el ítem posea una dirección negativa, la calificación se invierte. Los ítems se presentan en forma de enunciados cuyo grado de acuerdo o desacuerdo se solicita a la unidad de análisis. La cantidad de enunciados que integra una escala Likert varía de acuerdo a la naturaleza de la variable operacionalizada.

Los pasos a seguir para la construcción de la escala son:

1. Definición de la variable a medir.
2. Operacionalización de la variable, es decir, se determina como se habrá de medir y se señalan los indicadores.

3. Diseño de una cantidad suficiente de ítems favorables y desfavorables a la variable que se pretende medir. Se sugiere elaborar alrededor de 50 ítems, balanceando la escala con igual cantidad de enunciados favorables y desfavorables.
4. Depuración de la escala por medio de un estudio piloto con el propósito de seleccionar los ítems que habrán de integrarse a la versión final de la escala.
5. Administración de la versión final de la escala a las unidades de análisis que integran la unidad muestral del estudio.
6. Asignación de una puntuación a cada ítem de acuerdo al procedimiento descrito con anterioridad.
7. Obtención de la puntuación total de cada unidad muestral, reflejando la actitud global hacia la variable medida.

Es recomendable realizar un análisis de los ítems con el propósito de ser selectivos. Entre las técnicas de análisis se encuentran la correlación ítem-escala por medio del coeficiente de correlación de Pearson, el coeficiente gamma o el método de Edwards.

b) Escalograma de Guttman

Guttman desarrollo una técnica para la medición de actitudes en una dimensión única. Se le conoce como *Escalograma de Guttman*. Se caracteriza por medir la intensidad de la actitud a través de un conjunto de ítems.

La escala es unidimensional siempre y cuando sea de carácter acumulativo, es decir, que los ítems que la integran posean un escalamiento perfecto. Lo anterior se refiere a que el conjunto de ítems están encadenados entre sí de tal forma que si una unidad de análisis expresa estar de acuerdo con el primer ítem, deberá estar de acuerdo con el resto de ítems que constituyen el escalograma. Los ítems se ordenan de mayor a menor intensidad. Por ejemplo, ante los siguientes enunciados tiene que buscarse que cumplan con el principio de escalamiento:

1. A una excelente preparación y capacidad profesional corresponden excelentes ingresos económicos.
2. Los sueldos y salarios que devengan los empleados en las organizaciones deben asignarse de acuerdo a la preparación y capacidad profesional.
3. Un empleado nuevo de una organización con un alto nivel de preparación y capacidad y con un eficiente desempeño laboral debe obtener un salario más alto que un empleado de la misma organización con 10 o más años de experiencia pero con un bajo nivel de preparación y capacidad profesional, reflejado en un bajo desempeño laboral.

Si los enunciados anteriores constituyen una escala unidimensional, entonces deberán cumplir con el principio de escalamiento perfecto. La construcción del *Escalograma de Guttman* requiere de cumplir con las siguientes etapas:

1. Definir la variable de actitud a medir.
2. Operacionalizar la variable de actitud.
3. Desarrollar el conjunto de ítems con respecto al objeto de actitud o referente, especificando como alternativas de respuesta: () De acuerdo () En desacuerdo.
4. Desarrollar un estudio piloto con el propósito de verificar si la escala es unidimensional. Guttman sugiere administrar entre 10 y 12 ítems a un promedio de 100 personas para realizar la depuración de la escala con mayor confiabilidad.

5. Determinar si los ítems integran una escala acumulativa por medio del análisis de reproductividad de las respuestas, es decir, si los ítems cumplen con el principio de escalamiento. El análisis de reproductibilidad se determina mediante el coeficiente de reproductividad expresado como:

$$Cr = 1 - \frac{\text{Errores totales}}{(\text{Items})(\text{Sujetos})}$$

Donde:

Cr = Coeficiente de reproductividad

De acuerdo a Guttman un coeficiente de reproductividad adecuado debe tener un valor mínimo de 0.90 (90%) como criterio para aceptar que la escala es unidimensional.

6. Administrar el escalograma a la unidad muestral.
7. Asignar puntuaciones a cada ítem. El valor que se asigna a la alternativa de acuerdo es 1 y 0 a la alternativa en desacuerdo. Estas puntuaciones se utilizan como precedente para obtener el coeficiente de reproductividad, a la vez que se determinan los puntos de ruptura.
8. Determinar la actitud global de las unidades de análisis evaluadas con respecto al referente u objeto de actitud.

5.2 Construcción de índices

El índice es un instrumento de medición por medio del cual se asignan medidas a las unidades de análisis en función de la posesión de algún indicador social o económico.

En la investigación social y económica los índices tienen diversas aplicaciones, por ejemplo al utilizarse para el análisis de variables económicas es posible llegar a descubrir relaciones importantes en las variables estudiadas.

La medición de las variables se hace por medio de números índice expresados en términos de cantidad, precio o valor. Un número índice es un valor relativo, expresado como porcentaje, que mide un periodo dado contra un periodo base determinado.

Los números índice son un excelente medio para la toma de decisiones empresariales y para evaluar el efecto de programas de índole socioeconómica.

- a) **Índice de Precios Simple.** Se utiliza para medir la variación del precio de un bien o servicio en un tiempo que fluctúa entre el periodo base y el periodo actual o de referencia.

El método para establecer un índice simple consiste en dividir el precio del bien en el periodo actual por su precio en el periodo base multiplicado por 100; el resultado algebraicamente se expresa con la ecuación:

$$PI = \frac{PR \times 100}{PB}$$

Donde:

PI = Precio índice

PR = Precio del periodo de referencia

PB = Precio del periodo base

Por ejemplo, para calcular el índice de precios simple de un libro cuyo precio en 1997 es de 60 pesos y en 1998 es de 98 pesos, considerando el año 1997 como el periodo base, se tendría:

$$PI_{1997} = 60/60 \times 100 = 100$$

$$PI_{1998} = 98/60 \times 100 = 163.33$$

Los anteriores resultados reflejan que el índice de precios del libro aumentó de 100 a 163.33, por lo que es factible hacer la inferencia de que el precio del libro aumentó un 63.33 % de un año a otro. Para calcular el **incremento** del precio del libro se procede a obtener la diferencia entre ambos números índice dividido por el periodo base.

La expresión algebraica es:

$$\text{Incremento del Precio} = \frac{(PI - PIB) \times 100}{PIB}$$

Donde:

PIR = Precio índice del periodo de referencia

PIB = Precio índice del periodo base

Al aplicar la ecuación 3.2 al anterior ejemplo se tiene:

$$\text{Incremento del Precio} = \frac{(163.33 - 100)}{100} \times 100 = 63.33$$

En este tipo de índices el índice de precios del año base tiene invariablemente un valor de 100.

5.3 Ejemplos de Indicadores

INDICADORES JUDICIALES:

ESTÁNDARES DE PRODUCCIÓN.- Para poder definir e implementar los distintos estándares es necesario desarrollarlos con la participación de todos aquellos que de alguna manera forman parte del proceso; es decir, es conveniente la participación de jueces, personal judicial, conformando un mismo equipo.

Juzgado Civil: El estándar de producción en los juzgados civiles debe ser 38 procesos resueltos cada mes de acuerdo con datos estadísticos históricos (ver cuadro). Pero si somos más optimistas podemos fijar nuestro estándar en 54 procesos resueltos mensualmente similares al promedio del año 2005.

| Juzgados Civiles | 2005 | 2006 | 2007 | 2008-NOV | TOTAL |
|-------------------------|-------------|-------------|-------------|-----------------|--------------|
| 1° JUZGADO CMIL | 619 | 353 | 337 | 340 | 1649 |
| 2° JUZGADO CMIL | 450 | 323 | 334 | 292 | 1399 |
| 3° JUZGADO CMIL | 594 | 335 | 306 | 325 | 1560 |
| 4° JUZGADO CMIL | 706 | 262 | 272 | 351 | 1591 |
| 5° JUZGADO CMIL | 575 | 447 | 314 | 330 | 1666 |
| 6° JUZGADO CMIL | 649 | 386 | 367 | 288 | 1690 |
| 7° JUZGADO CMIL (03/06) | | 415 | 482 | 464 | 1361 |
| Total Producción | 3593 | 2521 | 2412 | 2390 | 10916 |
| Estandar Mensual | 54 | 34 | 31 | 34 | 38 |

CONGESTION: Congestión es un indicador que mide el incremento de la carga procesal respecto a la cantidad de escritos ingresados. Si el indicador de Congestión es cercano al 100% significa que el especialista legal está resolviendo una cantidad de escritos poco significativa.

En consecuencia, esta congestionándose al aumentar su carga procesal. Lo contrario sucede si el indicador es cercano al 0% ya que la congestión puede ser negativa y esto implicaría que el especialista está resolviendo sus escritos pendientes (Descongestionando),,,

La congestión Ideal debe ser 0%, es decir todo escrito que ingresa debe ser atendido en el plazo legal.

$$C = \left(1 - \frac{EA}{EI}\right) * 100\%$$

Donde: **EA** = Número total de escritos Atendidos
EI = Número total de escritos Ingresados

a) ESPECIALISTA LEGAL:

PRODUCTIVIDAD (P): Es un indicador que mide la productividad de cada Especialista Legal respecto a la cantidad de escritos ingresados. Si el indicador de productividad es Menor a 100% significa que el Especialista Legal está resolviendo una cantidad de escritos poco significativos. En consecuencia, se está congestionando de escritos por resolver. Lo contrario sucede si el indicador es Mayor a 100%.

La productividad Ideal debe ser el 100%, es decir todo escrito que ingresa debe ser atendido en el plazo legal.

$$P = \left(\frac{EA}{EI}\right) * 100\%$$

Donde: **EA** = Número total de escritos Atendidos
EI = Número total de escritos Ingresados

b) NOTIFICADOR:

EFICIENCIA (E): Es un indicador que mide la eficiencia de cada Notificador respecto a la cantidad de cédulas trabajadas (impresas). Si el indicador de eficiencia es Menor a 100% significa que el notificador no está notificando oportunamente. En consecuencia, no se cumplirá con el plazo legal. Lo contrario sucede si el indicador es Mayor a 100%.

La eficiencia Real debe ser el 100%, es decir toda notificación impresa debe ser entregada para su notificación en el plazo legal.

$$E = \left(\frac{NE}{NI}\right) * 100\%$$

Donde: **NE** = Número total de notificaciones entregadas para ser notificadas
NI = Número total de notificaciones impresas

c) ARCHIVO:

OPORTUNIDAD (O): Es un indicador que mide la oportunidad de cada Asistente de archivo respecto al movimiento de expedientes en el archivo. Si el indicador de eficiencia es Menor a 100% significa que el asistente no está notificando oportunamente. En consecuencia, no se cumplirá con el plazo legal. Lo contrario sucede si el indicador es Mayor a 100%.

La eficiencia Real debe ser el 100%, es decir toda notificación impresa debe ser entregada para su notificación en el plazo legal.

$$E = \left(\frac{NE}{NI} \right) * 100\%$$

Donde: **NE** = Número total de notificaciones entregadas para ser notificadas
NI = Número total de notificaciones impresas

d) SECRETARIOS:

PRODUCTIVIDAD (P): Es un indicador que mide la productividad de cada Secretario respecto a la cantidad de escritos ingresados. Si el indicador de productividad es Menor a 100% significa que el Secretario está resolviendo una cantidad de escritos poco significativos. En consecuencia, se está congestionando de escritos por resolver. Lo contrario sucede si el indicador es Mayor a 100%.

La productividad Ideal debe ser el 100%, es decir todo escrito que ingresa debe ser atendido en el plazo legal.

$$P = \left(\frac{EA}{EI} \right) * 100\%$$

Donde: **EA** = Número total de escritos Atendidos
EI = Número total de escritos Ingresados

e) TECNICO JUDICIAL

EFICIENCIA (E): Es un indicador que mide la eficiencia de cada Técnico Judicial respecto a la cantidad de cédulas trabajadas (impresas). Si el indicador de eficiencia es Menor a 100% significa que el Técnico Judicial no está notificando oportunamente. En consecuencia, no se cumplirá con el plazo legal. Lo contrario sucede si el indicador es Mayor a 100%.

La eficiencia Real debe ser el 100%, es decir toda notificación impresa debe ser entregada para su notificación en el plazo legal.

$$E = \left(\frac{NE}{NI} \right) * 100\%$$

Donde: **NE** = Número total de notificaciones entregadas para ser notificadas
NI = Número total de notificaciones impresas

6. Equipos de Medición

En estudios científicos se emplean múltiples equipos que permiten apoyar la medición de variables como por ejemplo:

El computador constituye hoy en día, el equipo básico en lo atinente al análisis e interpretación de imágenes satelitales y de fotografías aéreas y, además, constituye la herramienta básica para el análisis estadístico, el almacenamiento de información, la graficación o la digitación de los informes de investigación.

En el laboratorio se apoya de instrumentos tales como el estereoscopio, el microscopio, el espectrofotómetro, las cámaras de conteo y muchos más. Para los análisis físicos o químicos se utiliza también una serie de equipos de campo y laboratorio que incluyen medidores de gases, nutrientes, iones o elementos diversos, pH, temperatura, intensidad del sonido, etc.

El registro de poblaciones o comunidades bióticas en campo se vale entonces de cuadrantes, dragas, botellas muestreadoras, contadores manuales, balanzas, redes, trampas, cámaras de fotografía, binoculares, grabadoras, geoposicionadores, cuerdas metradas, papel milimetrado u otros.

7. Codificación

Consiste en agrupar numéricamente los datos que se expresen en forma verbal para poder luego operar con ellos como si se tratara, simplemente, de datos cuantitativos. Para lograrlo se habrá de partir de un cúmulo de informaciones que tengan una mínima homogeneidad, lo cual es necesario para poder integrarlas. Pueden tratarse de cientos de respuestas a una misma pregunta o de una variedad de posibles situaciones observadas mediante un mismo ítem de una pauta de observación: en ambos casos existirá una determinada variedad de respuesta o de observaciones que presenten las elecciones o los comportamientos de los objetos de estudio.

A cada categoría de respuestas habremos de darle un código particular, un número o letra diferente, que servirá para agrupar tras de sí a todas las respuestas u observaciones que sean idénticas o que, al menos, aparezcan como equivalentes. Luego procederemos a señalar a cada uno de los cuestionarios o pautas con el código que le corresponde en cada caso, con la que quedará sintetizada la respuesta que contiene. El proceso se facilita grandemente cuando se realiza mediante computadoras.

Supongamos que hemos preguntado la opinión que se tiene respecto a las Naciones Unidas. Si la pregunta ha sido abierta, cada encuestado habrá expuesto sus opiniones en algunas breves frases. La codificación nos permitirá agrupar sus respuestas, para poder evaluar cuáles son las opiniones más salientes al respecto. Nuestros códigos, por ejemplo, podrán ser:

Es una institución que garantiza (o protege) la paz mundial.

Es una institución que debería ser reformada.

Es útil por los servicios que presta a los países menos desarrollados.

Es inoperante, ineficiente, etc.

No sabe o no opina al respecto.

Por supuesto que podrían elaborarse otros códigos, categorizando más o menos detalladamente las respuestas. Cuando aparecen respuestas ambiguas, extrañas o simplemente anómalas, es conveniente agruparlas en el código "otras respuestas", para no abrir demasiadas categorías simultáneamente.

EJEMPLOS DE CUESTIONARIOS

Los cuestionarios electrónicos facilitan enormemente el trabajo de campo.

CUESTIONARIO 1

Formato de una Ficha de Trabajo:

| |
|---|
| <p>Ficha de Trabajo</p> <p>Fuente: Título: _____</p> <p>Autor: _____</p> <p>Año: _____ página(s) _____</p> <p>Edición _____ Editorial _____</p> <p>Tema: _____</p> <p>Subtema: _____</p> <p>Contenido:</p> <p>_____</p> <p>_____</p> <p>_____</p> |
|---|

CUESTIONARIO 2

Es Usted: Abogado o Litigante

SE ESTÁ ANALIZANDO LA CALIDAD DEL SERVICIO QUE PRESTA LA CORTE SUPERIOR DE LIMA NORTE.
(Marque con "x" la alternativa que crea correcta):

- En general ¿cómo califica el servicio que ha recibido?
Excelente () Bueno () Regular () Malo ()
- ¿Los escritos presentados, han sido resueltos dentro del plazo establecido?
Si () No ()
- ¿Cuál de estos trámites demoran más su atención?
a) Calificar demandas () b) Resolver escritos ()
c) Notificación Resolución () d) Lectura expediente ()
e) Entrevista con el Juez () e) Otros trámites.
- ¿Qué recomendaría usted, para mejorar el servicio de administrar Justicia en el Juzgado?
a) Rapidez en resolver () b) Calidad de resoluciones ()
c) Buen trato al público () d) Imparcialidad, procesos ()
e) Notificación rápida () f) Otras recomendaciones.

5. En general ¿cómo es el trato que ha recibido del personal en este juzgado?
 Excelente () Bueno () Regular () Malo ()
6. Cómo es el trato que se recibe en los juzgados:

| | Juez | Secretario o especialista | Mesa Partes | Vigilancia |
|-----------|------|---------------------------|-------------|------------|
| Excelente | | | | |
| Bueno | | | | |
| Regular | | | | |
| Malo | | | | |

CUESTIONARIO 3

Obtener información sobre la opinión de la población de Lima Metropolitana y Callao, de 18 a 70 años, acerca de la justicia en el Perú.

- ¿CUÁN INFORMADO ESTÁ SOBRE EL TEMA DE LA JUSTICIA EN EL PERÚ?
 - Informado
 - Poco informado
 - Nada informado
 - No sabe, no opina
- ¿CUÁNTO CONFÍA EN EL PODER JUDICIAL?
 - Mucho
 - Poco
 - Nada
 - No sabe, no contesta
- ¿CÓMO CALIFICA LA ADMINISTRACIÓN DE JUSTICIA EN EL PERÚ?
 - Buena
 - Regular
 - Mala
 - No sabe, no contesta
- ¿CÓMO CALIFICA EL NIVEL PROFESIONAL DE LOS JUECES EN EL PERÚ?
 - Buena
 - Regular
 - Mala
 - No sabe, no contesta

5. ¿CUÁL ES EL PRINCIPAL PROBLEMA DE LA ADMINISTRACIÓN DE LA JUSTICIA EN EL PERÚ?
Corrupción.
 - a. Lentitud de los procesos.
 - b. Desigualdad de acceso a la justicia.
 - c. Bajas remunerados.
 - d. Falta de autonomía.
 - e. Infraestructura deficiente.
 - f. No sabe, no contesta.

6. ¿LA SITUACIÓN DE LA JUSTICIA EN EL PERÚ HA MEJORADO, EN LOS ÚLTIMOS AÑOS?
 - a. Si
 - b. No
 - c. Sigue igual

7. HAY OCASIONES EN QUE LOS CIUDADANOS HACEN JUSTICIA POR PROPIA MANO. ¿USTED ESTÁ DE ACUERDO?
 - a. De acuerdo
 - b. Ni de acuerdo ni en desacuerdo
 - c. En desacuerdo
 - d. No sabe, no contesta

8. ¿HA LITIGADO ALGUNA VEZ EN EL PODER JUDICIAL?
 - a. Sí
 - b. No
 - c. No contesta

9. ¿EN ALGUNA OPORTUNIDAD LE HAN PEDIDO COIMA EN EL PODER JUDICIAL?
 - a. Sí
 - b. No
 - c. No contesta

10. ¿CUÁN EXCESIVO ES EL TIEMPO QUE DEMANDAN LOS PROCESOS JUDICIALES?
 - a. Mucho
 - b. Bastante
 - c. Poco
 - d. No sabe, no contesta

11. ¿LOS COSTOS DEL SERVICIO DE ADMINISTRACIÓN DE JUSTICIA EN EL PERÚ SON?
 - a. Caros
 - b. Moderados
 - c. Baratos
 - d. No sabe, no contesta

CUESTIONARIO 4



EMBOTELLADORA “NBQ” S.A.

PRODUCTO: CERVEZA

PRODUCCION N° _____

CANTIDAD (LITROS):

FECHA: DIA: _____ MES: _____ AÑO: 200__

| AGUA CALIENTE: | | <u>Cantidad</u> | | <u>Cantidad</u> |
|-----------------------|--|-----------------|------|-----------------------|
| AGUA (Litros) | | | Obs. | FINAL (HORA) |
| INICIO: (HORA) | | | | TEMPERATURA AGUA (°C) |
| TEMPERATURA AGUA (°C) | | | | PH |
| PH | | | | DENSIDAD |
| DENSIDAD | | | | |

| MACERACION: | | <u>Cantidad</u> | | <u>Cantidad</u> |
|-----------------------|--|-----------------|-----------------|------------------------|
| AGUA (Litros) | | | Obs. | PRUEBAN° 2 (HORA) |
| MALTA (Kg) | | | | TEMPERATURA AGUA (°C) |
| MALTA TOSTADA (Kg) | | | | ALMIDON (P. YODO) |
| INICIO (HORA): | | | | |
| TEMPERATURA AGUA (°C) | | | | FINAL (HORA) |
| PH | | | | TEMPERATURA AGUA (°C) |
| DENSIDAD | | | | ALMIDON (P. YODO) |
| | | | | PH |
| PRUEBAN° 1 (HORA) | | | | DENSIDAD |
| TEMPERATURA AGUA (°C) | | | | RECIRCULACION (Tiempo) |
| ALMIDON (P. YODO) | | | LAVADO (Tiempo) | |
| | | | AGUA (Litros) | |

| COCCION: | | <u>Cantidad</u> | | <u>Cantidad</u> |
|-----------------|--|-----------------|--|------------------|
| INICIO (HORA): | | | | |
| AGUA (Litros) | | | | FINAL HORA |
| HERVOR (HORA) | | | | REPOSO |
| LUPOLUS | | | | WIRPOOL (TIEMPO) |
| | | | | REPOSO (TIEMPO) |

CUESTIONARIO 6

CANTIDAD (LITROS):

FECHA INICIO: DIA: MES: AÑO: 200

| FILTRADO | | Cantidad | | Cantidad | |
|---------------------------------|--|----------|--|----------|--|
| | | | | | |
| | | | | | |
| | | | | | |
| CARBUATA | | Cantidad | | Cantidad | |
| | | | | | |
| | | | | | |
| | | | | | |
| EMBASADO | | Cantidad | | Cantidad | |
| | | | | | |
| | | | | | |
| | | | | | |
| PASTEURIZADO | | Cantidad | | Cantidad | |
| | | | | | |
| | | | | | |
| | | | | | |
| ETIQUETADO | | Cantidad | | Cantidad | |
| | | | | | |
| | | | | | |
| | | | | | |
| CONTROL DE CALIDAD FINAL | | Cantidad | | Cantidad | |
| | | | | | |
| | | | | | |
| | | | | | |

CUESTIONARIO 7

| N° | N° EXP. | PROCESADO | DELITO | CONFIRMADO | NULA | REVOCADO | PENA |
|----|---------|-----------|--------|------------|------|----------|------|
| 1 | | | | | | | |
| 2 | | | | | | | |
| 3 | | | | | | | |
| 4 | | | | | | | |
| 5 | | | | | | | |
| 6 | | | | | | | |
| 7 | | | | | | | |
| 8 | | | | | | | |
| 9 | | | | | | | |
| 10 | | | | | | | |
| 11 | | | | | | | |
| 12 | | | | | | | |
| 13 | | | | | | | |
| 14 | | | | | | | |
| 15 | | | | | | | |

Capítulo

6

Procesamiento de datos o información

El siguiente paso consiste en procesar los datos recolectados del cuestionario desarrollado con el fin de obtener información relevante para nuestra investigación.

Procesamiento de Datos: Consiste en ingresar los datos recolectados a un software. Si los datos son de un cuestionario desarrollado, llamaremos a este proceso *introducción directa de datos*. Una investigación también se realiza con información ya existente en un soporte electrónico de datos (archivos de Excel, SQL, Lotus, texto, etc.); a este proceso lo llamaremos *extrayendo un archivo de disco*.

Existen diversos software que ayudan en tal propósito mediante análisis cuantitativos de la información.

En este libro se utilizará el PASW llamado anteriormente SPSS.



PROCESO DE DATOS

Consiste en ingresar los datos recolectados a un software. Si los datos son de un cuestionario desarrollado. Llamaremos a este proceso *introducción directa de datos*. Una investigación también se realiza con información ya existente en un soporte electrónico de datos (archivos de Excel, SQL, Lotus, texto, etc.); a este proceso lo llamaremos *extrayendo un archivo de disco o introducción indirecta*.

Existen un sin número de software estadísticos para procesar datos; un software libre recomendable es "R". En este libro utilizaremos el software PASW, por ser un software fácil de entender.

INTRODUCCIÓN DIRECTA DE DATOS

Cuando recolectamos datos de un cuestionario compuesto de "j" preguntas (variables) y entrevistamos "n" sujetos (casos), se debe introducir los datos al software escogido.

Como es de su conocimiento, este libro trabaja con el software PASW. De aquí en adelante aprenderemos a utilizar dicho software.

Iniciar Pasw

Para ingresar a PASW se accede al botón Inicio y desde allí a Programas, luego PASW 18 for Windows.

Aparece la presentación del programa y al fondo el *Editor de datos*, luego hacemos clic en Cancelar y pasaremos directo a la ventana editor de datos.

El Editor de datos tiene dos pestañas, *Vista de datos (Data View)* y *Vista de variables (Variable View)*.

1. Vista de Datos (Data View): Ingreso de datos

Presenta una *matriz de n casos (sujetos) x j variables* (pregunta). Cada columna representa una variable única y cada fila un caso individual de cada variable.

Para introducir datos basta con pinchar sobre una casilla y teclear el dato. Si donde tecleamos el dato la variable no está definida, PASW asigna automáticamente un nombre de variable a la columna y la define con sus parámetros por defecto, como se muestra en el ejemplo siguiente.

En el ejemplo se ingresa en la primera columna (VAR0001) los valores 3, 2, 3, 2, así sucesivamente para las otras variables.

| | Variables | | | |
|---|-----------|----------|----------|-----|
| | VAR00001 | VAR00002 | VAR00003 | var |
| 1 | 3,00 | 1,00 | 1,00 | |
| 2 | 2,00 | 5,00 | | |
| 3 | 3,00 | 6,00 | . | |
| 4 | 2,00 | 7,00 | . | |
| 5 | | | | |
| 6 | | | | |
| 7 | | | | |

Si bien es cierto se puede empezar a introducir directamente los datos en la matriz de datos (*Vista de datos*), es recomendable primero la definición de las variables en *Vista de variables*.

2. Vista de Variables (Variable View): Definir variables

Pulsando en la pestaña *Vista de variables (Variable View)* accederemos a una matriz similar a la de datos, pero que en este caso contiene las variables definidas del modo siguiente: cada fila representa una variable (1,2,3....) y cada columna un parámetro o especificación de esa variable (Name, Type, etc.). Éstos se asignan por defecto, pero podemos modificarlos a nuestra conveniencia.

| | Nombre | Tipo | Anchura | Decimales | Etiqueta | Valores | Perdidos |
|----|--------|------|---------|-----------|----------|---------|----------|
| 1 | | | | | | | |
| 2 | | | | | | | |
| 3 | | | | | | | |
| 4 | | | | | | | |
| 5 | | | | | | | |
| 6 | | | | | | | |
| 7 | | | | | | | |
| 8 | | | | | | | |
| 9 | | | | | | | |
| 10 | | | | | | | |
| 11 | | | | | | | |

La lista de parámetros o especificación de las variables son:

Nombre (Name):

Debemos introducir el nombre de la variable, que pueden ser letras, números o el símbolo de subrayado (_). El primer carácter únicamente puede ser una letra.

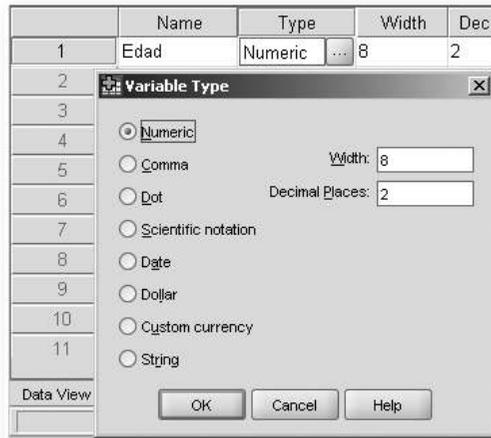
Por ejemplo en la fila "1" que representa la primera variable, escribimos edad para el parámetro Name (Nombre).

| | Name | Type | Width | Decimals |
|---|------|------|-------|----------|
| 1 | Edad | | | |

Tipo (Type):

Nos ubicados en el parámetro Type (Tipo). A través del botón  que aparece al lado del tipo (Numeric) se accede a un panel donde podemos seleccionar el tipo de variable a utilizar que puede ser: numérica, coma, fecha, moneda, cadena, etc., frecuentemente se trabaja con numérico.

Por ejemplo seleccionamos como tipo (Type) de variable numérico (Numeric).



Anchura (Width)

El parámetro Anchura (Width) especifica el número de caracteres o dígitos de la variable. El botón  que aparece al lado izquierdo permite aumentar o disminuir el número de dígitos o caracteres.

Por ejemplo seleccionamos como ancho el número “8” que salió por defecto.

| | Name | Type | Width |
|---|------|---------|-------|
| 1 | Edad | Numeric | 8 |

Decimales (Decimals)

El parámetro Decimales (Decimals) indica el número de decimales, si la variable es de tipo contable (numérico, moneda, etc.). El botón  que aparece al lado izquierdo permite aumentar o disminuir número de decimales.

Por ejemplo: el número de decimales se seleccionó el número “3”, es decir, 3 decimales.

| | Name | Type | Width | Decimals |
|---|------|---------|-------|----------|
| 1 | Edad | Numeric | 8 | 3 |

Etiqueta (Label)

Se usan etiquetas (Labels) para describir la variable de una forma más clara. También se utilizan en la presentación de resultados.

Por ejemplo: se escribió como etiqueta la palabra “Edad”.

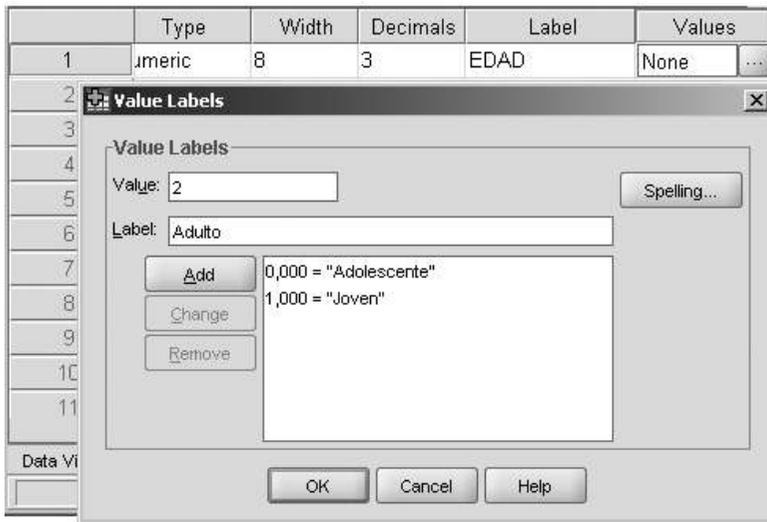
| | Name | Type | Width | Decimals | Label |
|---|------|---------|-------|----------|-------|
| 1 | Edad | Numeric | 8 | 3 | EDAD |

Valores (Values)

El parámetro Valores (Values) permite asignar etiquetas a determinados valores de las variables.

Por ejemplo, la variable numérica *edad* contiene valores: 0 si es adolescente, 1 si es joven y 2 si es adulto; entonces, para el valor (value) "0" su etiqueta de valor (value label) será *adolescente*, luego clic en Add. Para el valor "1" su etiqueta de valor será *joven* y para valor "2" su etiqueta de valor será *adulto* respectivamente. De esta forma, cuando trabajamos con la variable *edad* aparecerá *adolescente*, *joven* y *adulto* en lugar de 0, 1 y 2.

Aquí es donde es necesario recordar la codificación de variables.



Perdidos (Missing):

El parámetro Perdidos (Missing) permite definir los valores que se tratarán como perdidos. Hemos de distinguir entre los valores declarados por el sistema (ausencia de dato) y los valores declarados como perdidos por nosotros mismos.

Por ejemplo: Aquí se seleccionó ningún valor perdido None, preferible dejar siempre este valor (None).

| | Label | Values | Missing |
|---|-------|-----------------|---------|
| 1 | Edad | {,00, Adole ... | None |

Columnas (Columns)

El parámetro Columnas (Columns) indica la anchura de la columna en la *Vista de datos*. El botón  que aparece al lado izquierdo permite aumentar o disminuir la anchura.

Por ejemplo: El ancho de columna se selecciono de tamaño "8".

| Missing | Columns |
|---------|---------|
| None | 8 |

Alineación (Align)

El parámetro alineación indica la posición del dato en la casilla de *Vista de datos* que puede ser alineación Izquierda (Left) alineación Derecha (Right) o Centrado (Center).

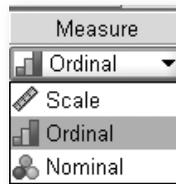
Por ejemplo: La alineación se seleccionó al lado derecho (Right)



Medida (Measure)

Este parámetro permite especificar el nivel de medida que puede ser: Escala (datos numéricos de una escala de intervalo o de razón), Ordinal (representan categorías ordenadas) y Nominal (representan categorías sin orden alguno). Esta especificación sólo se usa en procedimientos gráficos.

Por ejemplo: Se seleccionó la medida ordinal.



En general, es importante saber distinguir entre variable categórica y variable de escala. Una variable categórica es aquella que tiene un número limitado de valores o categorías distintas.

Todas las variables de cadena y las variables numéricas con etiquetas de valor definidas, o las variables numéricas definidas como nominales u ordinales, se tratan como categóricas. El resto se tratarán como de escala.

Ejemplo: En la encuesta realizada en el supermercado, definir las variables del cuestionario.

Encuesta sobre compras en el supermercado

Se quiere analizar el comportamiento de los consumidores que hicieron sus compras semanales en un Supermercado, realizando el siguiente cuestionario.

Marque con una "x" la alternativa correcta o escriba en las líneas Punteadas.

- 1) Edad:
 - a) Adolescente
 - b) Joven
 - c) Adulto
- 2) Sexo:
 - a) Varón
 - b) Mujer

3) Grado de Instrucción:

- a) Primaria b) Secundaria c) Superior d) No tiene

4) Estado Civil:

- a) Soltero b) Casado c) Divorciado d) Viudo

5) Horario de Compra

- a) Mañana b) Tarde c) Noche

6) Compras Semanales en Artículos de Aseo Personal (Nuevo Soles)

7) Compras Semanales en Verduras (Nuevo Soles)

8) Compras Semanales en Abarrotes (Nuevo Soles)

9) Compras Semanales en Bebidas (Nuevo Soles)

Codificando la encuesta: A partir de la variable X6 no es necesario codificar.

X1: Edad (0: Adolescente; 1: Joven; 2: Adulto).

X2: Sexo (0: Mujer; 1: Varón).

X3: Grado de Instrucción (1: No tiene; 2: Primaria; 3: Secundaria 4: Superior).

X4: Estado Civil (1: Soltero; 2: Casado; 3: Divorciado 4: Viudo).

X5: Turno (0: Mañana; 1: Tarde; 2: Noche).

X6: Venta de Artículos de Aseo Personal (Nuevo Soles).

X7: Venta de Verduras (Nuevo Soles).

X8: Venta de Abarrotes. (Nuevo Soles).

X9: Venta de Bebidas (Nuevo Soles).

Solución:

Las variables a definir serán nueve, como edad, sexo, grado de instrucción etc.

En cada fila se escribirá el **nombre** de la variable.

Además de otras propiedades como el **tipo** (numérico, cadena, etc.), **la anchura**, el número de **decimales**, la **etiqueta**, el **'valor'** y la 'etiqueta de valor' de los **valores** que admite, los valores **perdidos**, la **alineación** del texto, etc. Basta poseionar el ratón o el cursor en cada casilla y escribir la información deseada. En nuestro ejercicio la respuesta se muestra en el gráfico siguiente:

| | Name | Type | Width | Deci | Label | Values |
|---|----------|---------|-------|------|----------------------|------------------|
| 1 | edad | Numeric | 8 | 0 | Edad | {0, Adolescent |
| 2 | sexo | Numeric | 8 | 0 | Sexo | {0, Mujer}... |
| 3 | instrucc | Numeric | 8 | 0 | Grado de Instrucción | {1, No Tiene}... |
| 4 | civil | Numeric | 8 | 0 | Estado Civil | {1, Soltero}... |
| 5 | turno | Numeric | 8 | 0 | turno | {0, Mañana}... |
| 6 | aseo | Numeric | 8 | 2 | Venta de Artículos d | None |
| 7 | verduras | Numeric | 8 | 2 | Venta de Verduras | None |
| 8 | abarrote | Numeric | 8 | 2 | Venta de Abarrotes | None |
| 9 | bebidas | Numeric | 8 | 2 | Venta de Bebidas | None |

3. Introducir datos

Definida las variables, pasamos a *vista de datos* para introducir los datos en la matriz de datos.

Para desplazarnos por la matriz de datos podemos usar las flechas del cursor así como la tecla [Intro] para pasar al siguiente caso de la misma variable, y la tecla [Tab] para pasar a la siguiente variable.

| | edad | sexo | instruc | civil | turno | aseo | verduras | abarrotes |
|---|------|------|---------|-------|-------|------|----------|-----------|
| 1 | 0 | 0 | 3 | 2 | 0 | 2,00 | 17,00 | 34,50 |
| 2 | 2 | 0 | 4 | 1 | 2 | 3,00 | 20,00 | 40,00 |
| 3 | 2 | 0 | 2 | 3 | 2 | ,30 | 10,00 | 31,60 |
| 4 | 1 | 0 | 2 | 4 | 1 | 3,30 | 9,00 | 35,40 |
| 5 | 2 | 0 | 3 | 4 | 2 | 1,30 | 23,00 | 30,00 |
| 6 | 2 | 0 | 4 | 4 | 2 | ,40 | 13,00 | 32,90 |
| 7 | 1 | 0 | 2 | 2 | 0 | 1,50 | 12,00 | 33,20 |
| 8 | 0 | 0 | 3 | 2 | 0 | 4,50 | 19,00 | 33,10 |

El gráfico muestra los datos en forma de valor original.

Si queremos los datos definidos con etiquetas, debemos ir a opción del menú *Ver (View) / Etiquetas de valor (Value Labels)*.

En todo caso también con el botón *Value Labels (Etiquetas de valor)* podemos mostrar las etiquetas en lugar del valor original.



Luego los valores de datos definidos con etiquetas nos permiten introducirlos como tales o con el nombre de la etiqueta como se puede observar en la matriz de datos siguiente:

| | edad | sexo | instrucc | civil | turno | aseo | verduras | abarro |
|---|-------|-------|----------|--------|-------|------|----------|--------|
| 1 | Adole | Mujer | Secu | Casa | Mañ | 2,00 | 17,00 | 34,50 |
| 2 | Adult | Mujer | Super | Solter | Noc | 3,00 | 20,00 | 40,00 |
| 3 | Adult | Mujer | Prima | Divor | Noc | ,30 | 10,00 | 31,60 |
| 4 | Joven | Mujer | Prima | Viudo | Tard | 3,30 | 9,00 | 35,40 |
| 5 | Adult | Mujer | Secu | Viudo | Noc | 1,30 | 23,00 | 30,00 |
| 6 | Adult | Mujer | Super | Viudo | Noc | ,40 | 13,00 | 32,90 |
| 7 | Joven | Mujer | Prima | Casa | Mañ | 1,50 | 12,00 | 33,20 |
| 8 | Adole | Mujer | Secu | Casa | Mañ | 4,50 | 19,00 | 33,10 |

Al seleccionar una casilla aparece una lista desplegable con todas las opciones que tenemos para la variable.

Para modificar un dato es suficiente con pinchar su casilla y editarlo, bien en la misma casilla, bien en el *Editor de casillas* (situado entre la barra de herramientas y el *Editor de datos*). Debemos tener en cuenta que en el *Editor de casillas* no podemos introducir ninguna fórmula u operación.

| | edad | sexo | instrucc | civil | turno |
|---|--------|-------|-----------|------------|-------|
| 1 | Adoles | Mujer | Secundari | Casad ▾ | Mañan |
| | | | | Soltero | |
| | | | | Casado | |
| | | | | Divorciado | |
| | | | | Viudo | |

Lista desplegable

Es importante guardar la información introducida en el editor de datos en una unidad de disco a través del menú *Archivo* de la barra de herramientas o seleccionando el menú *Archivo*, luego se elige 'Guardar' (Save).

4. Uso del Editor de datos y variables

El *Editor de datos* nos permite realizar múltiples operaciones con la matriz de datos. Básicamente con las operaciones de cortar, copiar y pegar del menú *Edición* podemos realizar cualquier desplazamiento o copia de los datos que se nos antoje.

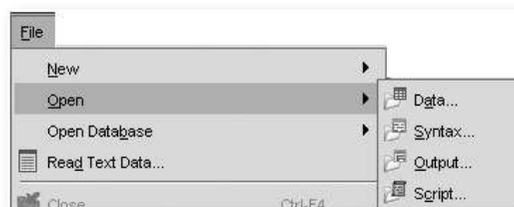
Podemos copiar un dato y pegarlo en otra o varias celdas, copiar uno o varios casos completos en otra u otras filas, copiar grupos de casos de una o varias variables, o variables completas sustituyendo a otras o creando otras nuevas. En resumen, casi cualquier operación en la matriz de datos es posible.

Los datos que no le proporcionemos directamente los completará el sistema con los valores definidos por defecto. En el caso de la ventana de variables el funcionamiento es el mismo, con la restricción de que no podemos modificar las especificaciones definidas de las variables. Asimismo, podemos insertar casos y variables mediante el menú *Datos* o los botones, o desplazarnos a un caso determinado con el comando *Ir a caso* o a una variable.

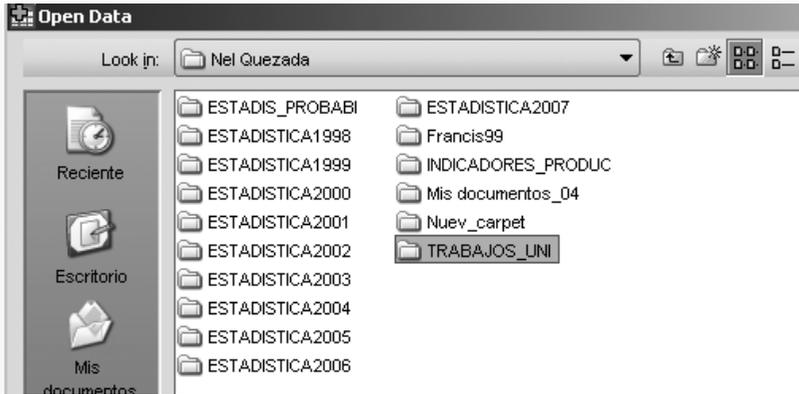
EXTRAYENDO UN ARCHIVO DE DISCO

Otra forma de obtener datos en PASW es extrayendo un archivo de disco llamado también Soporte Electrónico de Datos. Entenderemos por Soporte Electrónico de Datos a todo documento o elemento que conteniendo información es "leído" o interpretado por una unidad periférica de un ordenador (USB, CD, Disco Magnético Flexible, Disco Duro, PD, etc.). Así, la información recogida en un cuestionario o formato de recolección de datos debe ser registrado en un soporte para que la información pueda ser procesada por un ordenador.

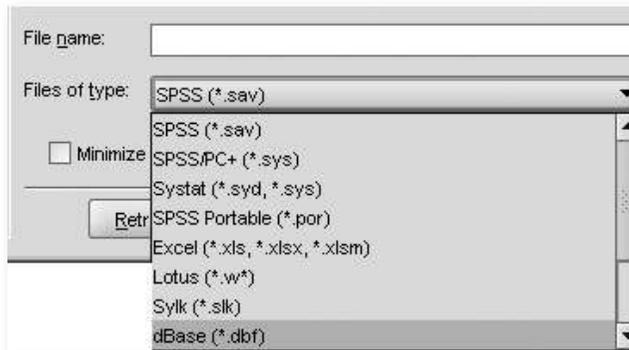
El procedimiento consiste en ir a la opción del menú *Archivo (File) / Abrir (Open) / Datos (Data)*.



Luego ubicar la unidad de disco o carpeta donde se encuentra el archivo que se desea extraer.



Este archivo puede ser de formato PASW (*.sav), Excel (*.xls), Lotus, Texto, etc.; ó de base de datos como Access, dBase, FoxPro, etc. Inclusive base de datos corporativos ó base de datos descargados de la web.



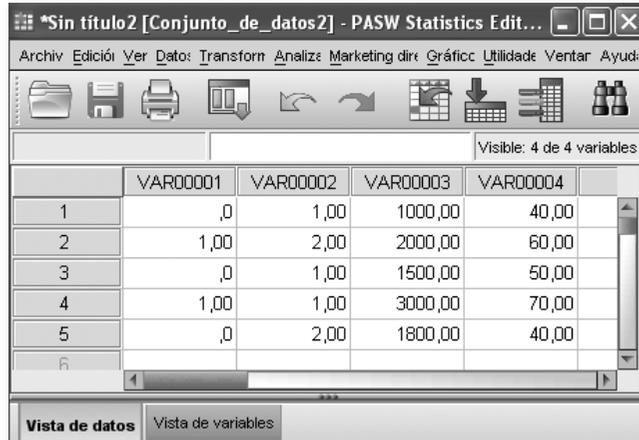
EJERCICIO 1: MANIPULACIÓN DE DATOS

Los datos corresponden a 5 sujetos (casos) medidos en cuatro variables: Sexo (0: mujer y 1: varón), Nivel cultural (1: bajo y 2: alto), Ingresos en nuevos soles y Peso en kilogramos.

| .Sujetos | Sexo | Nivel Cultural | Ingresos | Peso |
|----------|------|----------------|----------|------|
| 1 | 0 | 1 | 1000 | 40 |
| 2 | 1 | 2 | 2000 | 60 |
| 3 | 0 | 1 | 1500 | 50 |
| 4 | 1 | 1 | 3000 | 70 |
| 5 | 0 | 2 | 1800 | 40 |

Si bien es cierto es recomendable primero definir las variables, pero que sucede si primero ingresamos los datos a la matriz.

Los datos se introducen colocando un caso (sujeto) en cada fila y una variable en cada columna. Para ello, simplemente se escriben las puntuaciones correspondientes a cada caso. Para moverse de una celda a otra se pulsán las teclas marcadas con flechas. En nuestro ejemplo, los datos de los 5 primeros sujetos se colocan del siguiente modo:

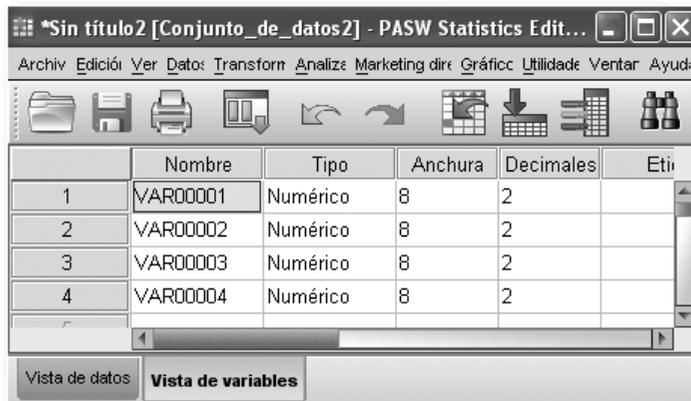


Visible: 4 de 4 variables

| | VAR00001 | VAR00002 | VAR00003 | VAR00004 |
|---|----------|----------|----------|----------|
| 1 | ,0 | 1,00 | 1000,00 | 40,00 |
| 2 | 1,00 | 2,00 | 2000,00 | 60,00 |
| 3 | ,0 | 1,00 | 1500,00 | 50,00 |
| 4 | 1,00 | 1,00 | 3000,00 | 70,00 |
| 5 | ,0 | 2,00 | 1800,00 | 40,00 |
| 6 | | | | |

Vista de datos Vista de variables

Se observa en la figura anterior que el PASW define por defecto los nombres de las variables como, *var00001*, *var00002*, *var00003* y *var00004*. Asimismo, ha definido los valores numéricos con dos decimales. Los datos representan una matriz de cinco filas por cuatro columnas.



| | Nombre | Tipo | Anchura | Decimales | Etiquetas |
|---|----------|----------|---------|-----------|-----------|
| 1 | VAR00001 | Numérico | 8 | 2 | |
| 2 | VAR00002 | Numérico | 8 | 2 | |
| 3 | VAR00003 | Numérico | 8 | 2 | |
| 4 | VAR00004 | Numérico | 8 | 2 | |

Vista de datos **Vista de variables**

Una vez introducidos la información en la matriz de datos, es necesario dar un nombre a cada una de las variables.

Si se desea cambiar el nombre de una variable y definir sus características (nombre, tipo, ancho de columna, decimales, etiquetas de valores, formato de columna, definición de valores perdidos, etc.) hay que pulsar con el ratón en la solapa Vista de variables (Variable View).

A continuación aparece una ventana que contiene tantas filas como variables definidas en el editor de datos. En nuestro caso son cuatro filas porque existen cuatro variables (Sexo, Nivel cultural, Ingresos y Peso).

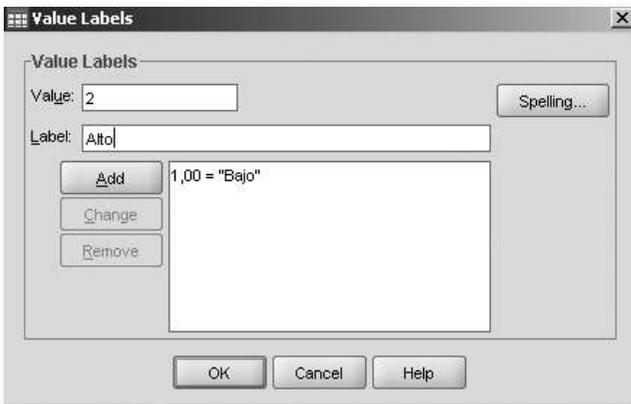
En cada fila se da la posibilidad de modificar los parámetros, como el **nombre (Name)** de la variable. Asimismo, de otras propiedades como el **tipo** (Type: numérico, cadena, etc.), la **anchura (Width)**, el número de **decimales (Decimals)**, la **etiqueta (Label)**, el significado de los **valores (Values)** que admite, los valores **perdidos**, la **alineación** del texto, etc. Basta con posesionar el cursor en cada casilla (Celda) y escribir la información deseada.

| | Name | Type | Width | Decimals | Label |
|---|----------|---------|-------|----------|----------------|
| 1 | sexo | Numeric | 8 | 2 | Sexo |
| 2 | nivel | Numeric | 8 | 2 | Nivel Cultural |
| 3 | ingresos | Numeric | 8 | 2 | Ingresos |
| 4 | peso | Numeric | 8 | 2 | Peso |
| 5 | | | | | |

En nuestro ejemplo, los nombres (names) de las variables son: *sexo, nivel cultural, ingresos y peso*, respectivamente.

Algunas de las casillas de la ventana ‘Vista de variables’ incorporan su propio cuadro diálogo que se activa al seleccionarlas.

Por ejemplo al pulsar en ‘Valores’, en la variable nivel cultural aparece lo siguiente:



Desde este cuadro de diálogo se puede dar etiqueta a los valores de una variable. Por ejemplo en nivel cultural el valor 1 significa nivel cultural bajo y el valor 2, nivel cultural alto. Cada vez que se introduzca la etiqueta de un valor se pulsa la tecla Añadir (add). Para terminar se pulsa en Aceptar (OK).

Una vez definidas todos los parámetros de las variables, podemos volver a visualizar los datos numéricos, para lo cual se pincha en la solapa “Vista de datos”.

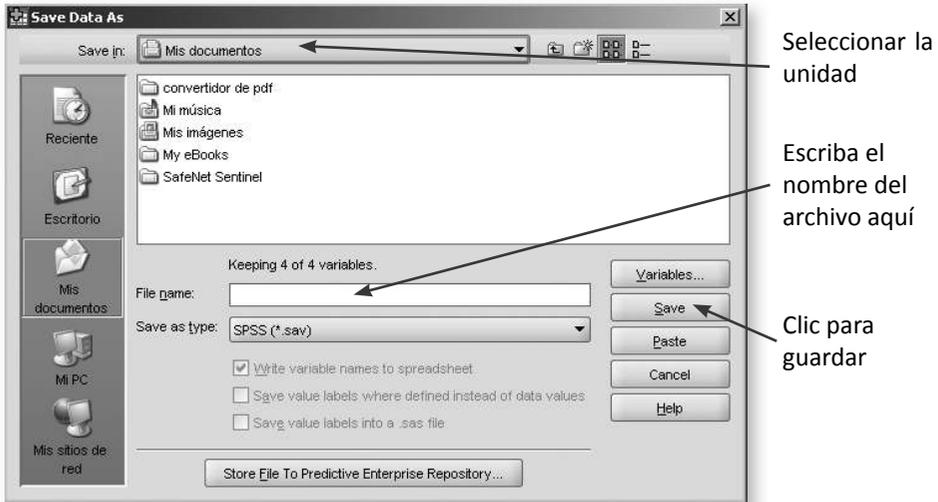
En nuestro ejemplo, en la matriz de datos que se muestra a continuación se puede observar que las variables ya están definidas:

| | sexo | nivel | ingresos | peso | var |
|---|------|-------|----------|-------|-----|
| 1 | 0,00 | Bajo | 1000,00 | 40,00 | |
| 2 | 1,00 | Alto | 2000,00 | 60,00 | |
| 3 | 0,00 | Bajo | 1500,00 | 50,00 | |
| 4 | 1,00 | Bajo | 3000,00 | 70,00 | |
| 5 | 0,00 | Alto | 1800,00 | 40,00 | |

Algo importante que debemos realizar antes de continuar es guardar la información en un archivo de disco.

Guardar

Para guardar la información introducida en el editor de datos se selecciona el menú *Archivo* de la barra de herramientas, se elige 'Guardar' y aparece el correspondiente cuadro de diálogo:

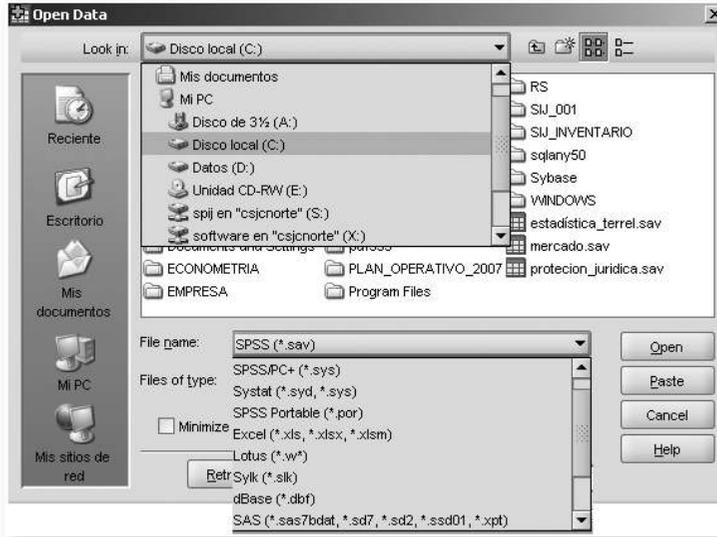


El PASW guarda archivos de datos con la extensión "**sav**", por tanto basta con dar un nombre a los datos (*ejercicio_1*) y por defecto se grabarán en un archivo con la extensión sav (aquí le hemos llamado: *g:\ejercicio_1.sav*).

También es posible guardar archivos de resultados. Para ello nos situamos en la ventana *Visor de resultados*, seleccionamos el menú 'Guardar' y aparece un cuadro de diálogo idéntico al de la figura anterior. El procedimiento a seguir es exacto al de archivar datos con la única diferencia de que la extensión de este tipo de archivos es "**spo**". Por tanto, los archivos PASW con la extensión "**sav**" contienen datos y los de la extensión "**spo**" resultados.

Abrir archivos de datos existentes

Cuando el archivo de datos se encuentra ya introducido en el ordenador o en un disquete, bastará simplemente con abrirlo. Para ello, se selecciona el menú 'Abrir' y la unidad donde se encuentra el archivo, el cuadro de diálogo correspondiente a 'Abrir'. Nuestros datos se encuentran grabados en el archivo *a:\practica.sav*. Para acceder a ellos los seleccionamos y hacemos clic en el botón *abrir*.



Una vez completada esta operación aparecen los datos en el *Editor de datos*. A continuación se muestra en vista datos de la matriz para los 5 sujetos con sus respectivas variables (sexo, nivel, ingresos y peso).

| | sexo | nivel | ingresos | peso |
|---|------|-------|----------|-------|
| 1 | .00 | 1.00 | 1000.00 | 40.00 |
| 2 | 1.00 | 2.00 | 2000.00 | 60.00 |
| 3 | .00 | 1.00 | 1500.00 | 50.00 |
| 4 | 1.00 | 1.00 | 3000.00 | 70.00 |
| 5 | .00 | 2.00 | 1800.00 | 40.00 |

Notas: Siempre que se desee iniciar una sesión, el PASW en la que se lleve a cabo uno o más análisis de datos, es necesario que el archivo donde se encuentran los datos esté abierto en el Editor de datos del PASW. Como veremos más adelante, los menús donde se encuentran las diferentes opciones de análisis no están activos a no ser que el editor de datos tenga algún tipo de contenido.

Para que el PASW realice algún tipo de análisis estadístico, no basta con que los datos estén grabados en un archivo, es necesario que dicho archivo esté abierto.

EJERCICIO 2: MANIPULACIÓN DE DATOS

Lenar los datos de la encuesta supermercado. Los datos de la matriz representan nueve preguntas realizados a 30 sujetos (nueve variables y 30 sujetos).

La codificación de la encuesta es la siguiente: A partir de la variable X6 no es necesario codificar.

X1: Edad (0: Adolescente; 1: Joven; 2: Adulto).

X2: Sexo (0: Mujer; 1: Varón).

X3: Grado de Instrucción (1: No tiene; 2: Primaria; 3: Secundaria 4: Superior).

X4: Estado Civil (1: Soltero; 2: Casado; 3: Divorciado 4: Viudo).

X5: Turno (0: Mañana; 1: Tarde; 2: Noche).

X6: Venta de Artículos de Aseo Personal (Nuevo Soles).

X7: Venta de Verduras (Nuevo Soles).

X8: Venta de Abarrotes. (Nuevo Soles).

X9: Venta de Bebidas (Nuevo Soles).

| .ni | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 |
|-----|----|----|----|----|----|-----|------|------|-----|
| 1 | 0 | 0 | 3 | 2 | 0 | 2,0 | 17,0 | 34,5 | 6,0 |
| 2 | 2 | 0 | 4 | 1 | 2 | 3,0 | 20,0 | 40,0 | 6,0 |
| 3 | 2 | 0 | 2 | 3 | 2 | 0,3 | 10,0 | 31,6 | 6,0 |
| 4 | 1 | 0 | 2 | 4 | 1 | 3,3 | 9,0 | 35,4 | 6,0 |
| 5 | 2 | 0 | 3 | 4 | 2 | 1,3 | 23,0 | 30,0 | 6,0 |
| 6 | 2 | 0 | 4 | 4 | 2 | 0,4 | 13,0 | 32,9 | 6,0 |
| 7 | 1 | 0 | 2 | 2 | 0 | 1,5 | 12,0 | 33,2 | 6,0 |
| 8 | 0 | 0 | 3 | 2 | 0 | 4,5 | 19,0 | 33,1 | 6,0 |
| 9 | 2 | 1 | 4 | 3 | 2 | 2,5 | 18,0 | 35,6 | 6,0 |
| 10 | 1 | 1 | 2 | 2 | 1 | 0,3 | 24,0 | 33,0 | 6,0 |
| 11 | 2 | 1 | 4 | 1 | 2 | 1,0 | 7,0 | 34,5 | 6,0 |
| 12 | 2 | 0 | 3 | 2 | 2 | 6,0 | 10,0 | 33,2 | 6,0 |
| 13 | 2 | 0 | 4 | 4 | 2 | 5,6 | 5,0 | 31,5 | 6,0 |
| 14 | 2 | 1 | 4 | 4 | 2 | 6,0 | 14,0 | 36,2 | 1,0 |
| 15 | 1 | 0 | 1 | 1 | 1 | 1,2 | 15,0 | 36,8 | 1,0 |
| 16 | 2 | 0 | 1 | 2 | 2 | 0,2 | 12,0 | 35,4 | 2,0 |
| 17 | 2 | 0 | 1 | 3 | 2 | 6,0 | 14,0 | 33,2 | 2,0 |
| 18 | 2 | 0 | 2 | 2 | 2 | 5,5 | 6,0 | 35,9 | 2,0 |
| 19 | 2 | 0 | 4 | 4 | 2 | 6,5 | 9,0 | 36,5 | 3,0 |
| 20 | 2 | 0 | 1 | 4 | 2 | 0,2 | 13,0 | 38,2 | 3,0 |
| 21 | 2 | 0 | 3 | 1 | 2 | 2,3 | 6,0 | 34,5 | 3,0 |
| 22 | 2 | 0 | 2 | 2 | 2 | 0,2 | 7,0 | 36,2 | 3,0 |
| 23 | 1 | 0 | 1 | 1 | 1 | 2,3 | 10,0 | 36,8 | 3,0 |
| 24 | 2 | 0 | 2 | 4 | 2 | 1,5 | 13,0 | 30,1 | 3,0 |
| 25 | 2 | 0 | 3 | 4 | 2 | 5,0 | 7,0 | 36,0 | 3,0 |
| 26 | 2 | 0 | 2 | 2 | 2 | 4,5 | 18,0 | 35,4 | 3,0 |
| 27 | 0 | 0 | 3 | 2 | 0 | 0,3 | 24,0 | 31,5 | 3,0 |
| 28 | 2 | 0 | 2 | 3 | 2 | 5,5 | 7,0 | 36,2 | 4,0 |
| 29 | 2 | 0 | 2 | 2 | 2 | 0,5 | 10,0 | 33,9 | 4,0 |
| 30 | 2 | 0 | 2 | 2 | 2 | 0,2 | 4,0 | 34,9 | 4,0 |

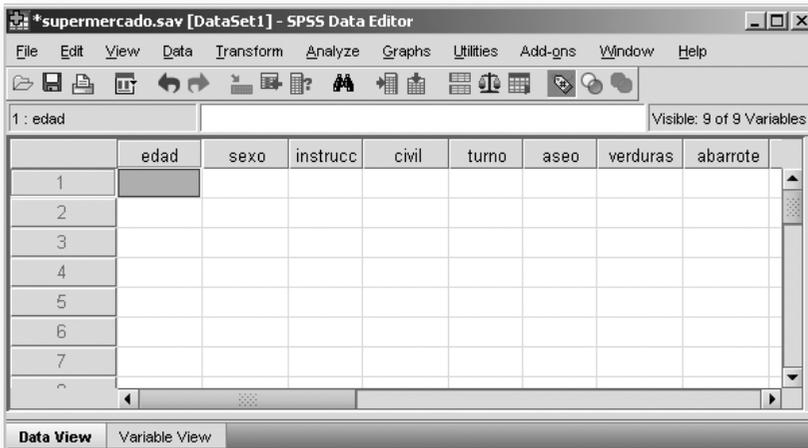
Primero definimos los parámetros para las variables *edad*, *sexo*, *grado de instrucción*, *estado civil*, *horario de compra* (*turno*, *venta de artículos de aseo*, *verduras* *abarrotes* y *bebida*).

En cada fila se escribirá el **nombre (name)** de las variables, el **tipo** (numérico), la **anchura** (8 caracteres), el número de **decimales** (2 decimales), la **etiqueta (Label)**, el significado de los **valores** que admite (tener presente que las variables *edad*, *sexo* *grado de instrucción* y *horario de compra* tienes valores), los valores **perdidos**, la **alineación** del texto, etc.

Para escribir la información de cada variable simplemente se posiciona el ratón o el cursor en cada casilla y escribimos la información deseada.

| | Name | Type | Width | Deci | Label | Values |
|---|-----------|---------|-------|------|----------------------|------------------|
| 1 | edad | Numeric | 8 | 0 | Edad | {0, Adolescent |
| 2 | sexo | Numeric | 8 | 0 | Sexo | {0, Mujer}... |
| 3 | instrucc | Numeric | 8 | 0 | Grado de Instrucción | {1, No Tiene}... |
| 4 | civil | Numeric | 8 | 0 | Estado Civil | {1, Soltero}... |
| 5 | turno | Numeric | 8 | 0 | turno | {0, Mañana}... |
| 6 | aseo | Numeric | 8 | 2 | Venta de Artículos d | None |
| 7 | verduras | Numeric | 8 | 2 | Venta de Verduras | None |
| 8 | abarrotes | Numeric | 8 | 2 | Venta de Abarrotes | None |
| 9 | bebidas | Numeric | 8 | 2 | Venta de Bebidas | None |

Una vez definida todas las variables de la encuesta supermercados, pasamos a la pestaña *editor de datos* cuyo aspecto es el siguiente:



Los datos se introducen colocando un caso (sujeto) en cada fila y una variable en cada columna. Para ello, simplemente se escriben las puntuaciones correspondientes a cada caso. Para moverse de una celda a otra se pulsan las teclas marcadas con flechas. En nuestro ejercicio, los datos están representados por 30 sujetos, cada sujeto tiene nueve variables. Es decir la matriz de datos será de 30 filas por 9 columnas, como se aprecia en el siguiente gráfico:

| | edad | sexo | instrucc | civil | turno | aseo | verduras | abarrotes | bebidas |
|---|------------|-------|----------|------------|--------|------|----------|-----------|---------|
| 1 | Adolesc... | Mujer | Secun... | Casado | Mañana | 2,00 | 17,00 | 34,50 | 6,00 |
| 2 | Adulto | Mujer | Superior | Soltero | Noche | 3,00 | 20,00 | 40,00 | 6,00 |
| 3 | Adulto | Mujer | Primaria | Divorciado | Noche | 0,30 | 10,00 | 31,60 | 6,00 |
| 4 | Joven | Mujer | Primaria | Viudo | Tarde | 3,30 | 9,00 | 35,40 | 6,00 |
| 5 | Adulto | Mujer | Secun... | Viudo | Noche | 1,30 | 23,00 | 30,00 | 6,00 |
| 6 | Adulto | Mujer | Superior | Viudo | Noche | 0,40 | 13,00 | 32,90 | 6,00 |
| 7 | Joven | Mujer | Primaria | Casado | Mañana | 1,50 | 12,00 | 33,20 | 6,00 |

Nota: Existen dos formas de introducir datos en el PASW. La más usual consiste en introducir los datos directamente en el editor de datos en el PASW. Aunque se puede grabar en otra aplicación y leerlos desde el PASW. La primera técnica es más usual ya que frecuentemente nos encontramos con encuestas que son necesarias procesar. La segunda técnica resulta útil cuando ya tenemos los datos en otro formato (Excel, Lotus, etc.) y queremos analizarlos con el PASW.

Capítulo

7

Estadísticas descriptivas

Análisis de Resultados y Conclusiones: Muestra los principales hallazgos de la investigación aplicando Estadísticas Descriptivas (tablas, cuadros, gráficas, etc.) y presenta una potente interpretación teórica que demuestra el dominio técnico del investigador. Las **conclusiones** extraen lo esencial de todo el proceso enfatizando especialmente la riqueza de la evidencia empírica aportada.



ESTADÍSTICA DESCRIPTIVA

Ciencia que analiza series de datos (por ejemplo, edad de una población, altura de los estudiantes de una escuela, temperatura en los meses de verano, etc.) y trata de extraer conclusiones sobre el comportamiento de estas variables.

Tipos de Variables: variables pueden ser de dos tipos:

Variables cualitativas o atributos: no se pueden medir numéricamente (por ejemplo: nacionalidad, color de la piel, sexo, etc.).

Variables cuantitativas: tienen valor numérico (edad, precio de un producto, ingresos anuales).

Las variables también se pueden clasificar en:

Variables unidimensionales: sólo recogen información sobre una característica (por ejemplo: edad de los alumnos de una clase).

Variables bidimensionales: recogen información sobre dos características de la población (por ejemplo: edad y altura de los alumnos de una clase).

Variables pluridimensionales: recogen información sobre tres o más características (por ejemplo: edad, altura y peso de los alumnos de una clase).

Las variables cuantitativas se pueden clasificar en discretas y continuas:

Discretas: sólo pueden tomar valores enteros (1, 2, 8, 4, etc.). Por ejemplo: número de hermanos (puede ser 1, 2, 3... etc., pero, por ejemplo, nunca podrá ser 3.45).

Continuas: pueden tomar cualquier valor real dentro de un intervalo. Por ejemplo, la velocidad de un vehículo puede ser 80,3 km/h, 94,57 km/h...etc.

DISTRIBUCION DE FRECUENCIAS

Distribución de Frecuencia: es la representación estructurada, en forma de tabla, de toda la información que se ha recogido sobre la variable que se estudia.

| Variable (Valor) | Frecuencias absolutas | | Frecuencias relativas | |
|---------------------|-----------------------|-------------------------------|-----------------------|-------------------------------|
| | Simple | Acumulada | Simple | Acumulada |
| X_1 | f_1 | f_1 | $h_1 = f_1/n$ | h_1 |
| X_2 | f_2 | $f_1 + f_2$ | $h_2 = f_2/n$ | $h_1 + h_2$ |
| ... | ... | ... | ... | ... |
| X_{k-1} | f_{k-1} | $f_1 + f_2 + \dots + f_{k-1}$ | $h_{k-1} = f_{k-1}/n$ | $h_1 + h_2 + \dots + h_{k-1}$ |
| X_k | f_k | $\sum f_i = n$ | $h_k = f_k/n$ | 1 |

Siendo X_i los distintos valores que puede tomar la variable. ($i=1, 2, 3 \dots k$).

Siendo f_i el número de veces que se repite cada valor.

Siendo h_i el porcentaje que la repetición de cada valor supone sobre el total.

Siendo n el número de observaciones realizadas.

EJEMPLO 1

Medimos la altura de los niños de una clase y obtenemos los siguientes resultados (cm):

| Alumno | Estatura | Alumno | Estatura | Alumno | Estatura |
|-----------|----------|-----------|----------|-----------|----------|
| Alumno 1 | 1,25 | Alumno 11 | 1,23 | Alumno 21 | 1,21 |
| Alumno 2 | 1,28 | Alumno 12 | 1,26 | Alumno 22 | 1,29 |
| Alumno 3 | 1,27 | Alumno 13 | 1,30 | Alumno 23 | 1,26 |
| Alumno 4 | 1,21 | Alumno 14 | 1,21 | Alumno 24 | 1,22 |
| Alumno 5 | 1,22 | Alumno 15 | 1,28 | Alumno 25 | 1,28 |
| Alumno 6 | 1,29 | Alumno 16 | 1,30 | Alumno 26 | 1,27 |
| Alumno 7 | 1,30 | Alumno 17 | 1,22 | Alumno 27 | 1,26 |
| Alumno 8 | 1,24 | Alumno 18 | 1,25 | Alumno 28 | 1,23 |
| Alumno 9 | 1,27 | Alumno 19 | 1,20 | Alumno 29 | 1,22 |
| Alumno 10 | 1,29 | Alumno 20 | 1,28 | Alumno 30 | 1,21 |

Si presentamos esta información estructurada obtendríamos la siguiente **tabla de frecuencias**:

| Variable (Valor) | Frecuencias absolutas | | Frecuencias relativas | |
|---------------------|-----------------------|-----------|-----------------------|-----------|
| | Simple | Acumulada | Simple | Acumulada |
| 1,20 | 1 | 1 | 3,3% | 3,3% |
| 1,21 | 4 | 5 | 13,3% | 16,6% |
| 1,22 | 4 | 9 | 13,3% | 30,0% |
| 1,23 | 2 | 11 | 6,6% | 36,6% |
| 1,24 | 1 | 12 | 3,3% | 40,0% |
| 1,25 | 2 | 14 | 6,6% | 46,6% |
| 1,26 | 3 | 17 | 10,0% | 56,6% |
| 1,27 | 3 | 20 | 10,0% | 66,6% |
| 1,28 | 4 | 24 | 13,3% | 80,0% |
| 1,29 | 3 | 27 | 10,0% | 90,0% |
| 1,30 | 3 | 30 | 10,0% | 100,0% |

Si los valores que toma la variable son diversos y cada uno de ellos se repite muy pocas veces, entonces conviene agruparlos por intervalos, ya que de otra manera obtendríamos una tabla de frecuencia muy extensa que aportaría poco valor a efectos de síntesis.

Resultados con el PASW.

| | | Frecuencia | Porcentaje | Porcentaje válido | Porcentaje acumulado |
|---------|-------|------------|------------|-------------------|----------------------|
| Válidos | 1,20 | 1 | 3,3 | 3,3 | 3,3 |
| | 1,21 | 4 | 13,3 | 13,3 | 16,7 |
| | 1,22 | 4 | 13,3 | 13,3 | 30,0 |
| | 1,23 | 2 | 6,7 | 6,7 | 36,7 |
| | 1,24 | 1 | 3,3 | 3,3 | 40,0 |
| | 1,25 | 2 | 6,7 | 6,7 | 46,7 |
| | 1,26 | 3 | 10,0 | 10,0 | 56,7 |
| | 1,27 | 3 | 10,0 | 10,0 | 66,7 |
| | 1,28 | 4 | 13,3 | 13,3 | 80,0 |
| | 1,29 | 3 | 10,0 | 10,0 | 90,0 |
| | 1,30 | 3 | 10,0 | 10,0 | 100,0 |
| | Total | 30 | 100,0 | 100,0 | |

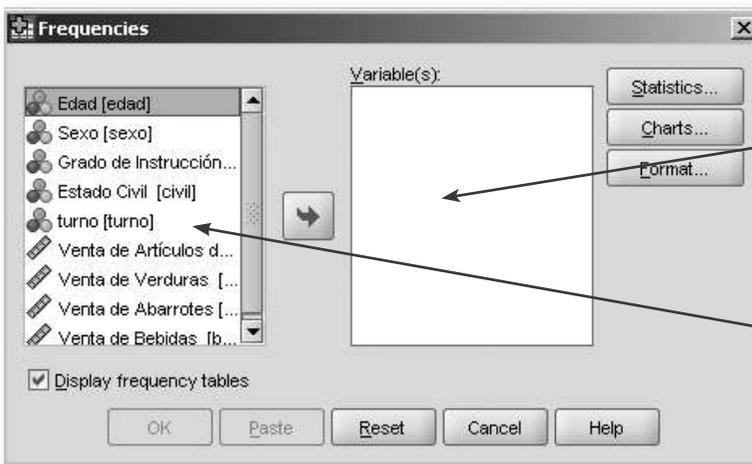
Ejercicio de distribución de frecuencia

PASW puede confeccionar distribuciones de frecuencias para organizar los datos correspondientes a una variable. En este apartado veremos cómo se construyen las tablas de frecuencias y cómo se realizan algunas de las representaciones gráficas.

Para la ilustración de los ejercicios utilizaremos la encuesta supermercado que ya fue ingresado anteriormente.

Las frecuencias se realizan desde el menú. *Analizar -> Estadísticas descriptivas: frecuencias*, haciendo clip sale el cuadro de dialogo.

El cuadro de diálogo de **frecuencias** ofrece el siguiente aspecto:

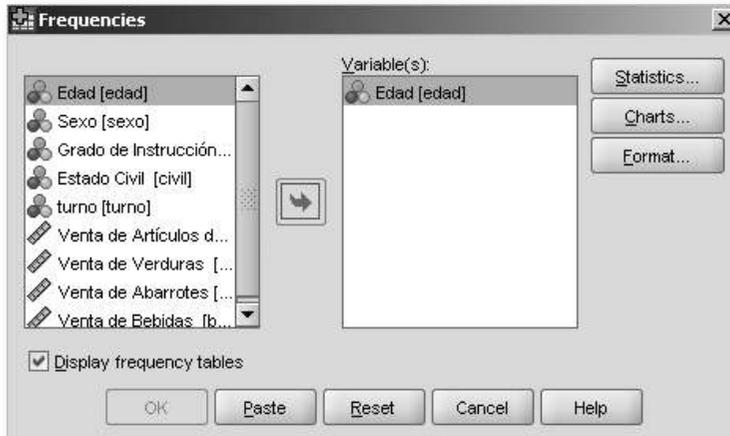


Trasladar aquí las variables sobre las que se quiera obtener la distribución de frecuencias.

Lista de variables del Archivo abierto en la ventana Editor de datos

El cuadro de la izquierda muestra el listado de las variables de la base de datos en estudio. El cuadro derecho está vacío, esperando que le indiquemos qué variables queremos analizar. Las variables se trasladan de un cuadro a otro mediante el botón . Por defecto, la opción “mostrar tablas de frecuencias (Display frequency tables)” está activada pero si no queremos que aparezca puede desactivarse.

Trabajemos con la variable “Edad”. Para ello trasladamos la variable al cuadro de la derecha del modo siguiente:



Luego, pulsar en el botón *Aceptar (OK)* para que se ejecute la orden. Los resultados obtenidos aparecen en el *Visor de resultados*.

Edad

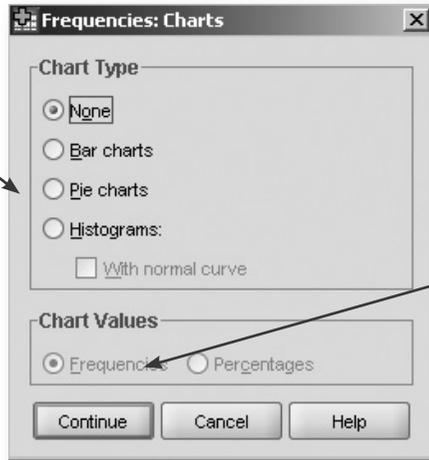
| | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------------------|-----------|---------|---------------|--------------------|
| Valid Adolescente | 3 | 10.0 | 10.0 | 10.0 |
| Joven | 5 | 16.7 | 16.7 | 26.7 |
| Adulto | 22 | 73.3 | 73.3 | 100.0 |
| Total | 30 | 100.0 | 100.0 | |

La tabla de frecuencias tiene el siguiente aspecto. En la primera columna aparecen los valores de la variable (Adolescente, Joven y Adulto); en la segunda, sus frecuencias absolutas en sentido ascendente (o descendente); en la tercera columna (Porcentaje) las frecuencias relativas en porcentaje; y en la última (porcentaje acumulado) las frecuencias relativas acumuladas en porcentaje.

REPRESENTACIONES GRÁFICAS

También es posible pedir al PASW que realice tres tipos de gráficos para observar las frecuencias de los datos. Desde el cuadro de diálogo anterior presionar Gráficos (Charts...). Se obtiene el siguiente cuadro de diálogo.

Sólo puede seleccionarse Uno. Una vez seleccionado. Pulsar en Continuar.



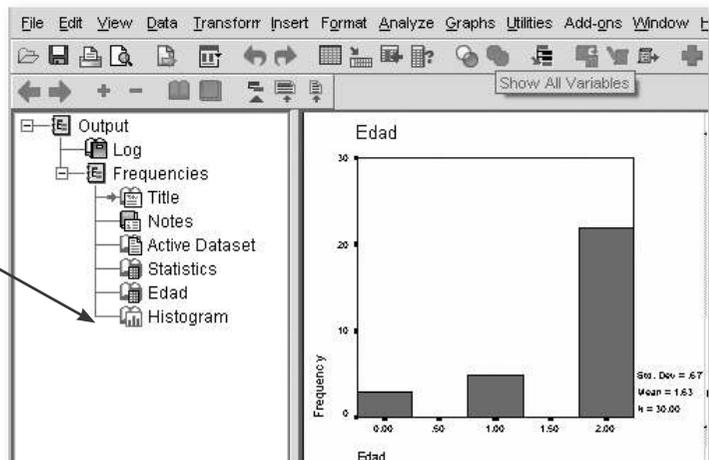
Permite confeccionar el gráfico con las frecuencias absolutas o con las relativas.

Como se observa, se puede obtener un histograma (con el ajuste a la curva normal), un diagrama de barras (Bar charts) y un diagrama de sectores o pictograma (Pie chats). Desde este cuadro de diálogo sólo es posible seleccionar uno de los gráficos. Si se quisiera obtener los tres hay que repetir la operación tres veces.

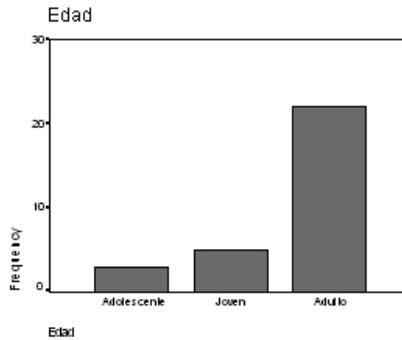
Seguimos con la variable 'Edad' como ejemplo. Ahora vamos a pedir al PASW que nos muestre los tres tipos de gráficos de la distribución de frecuencias para la edad. Si pulsamos en 'continuar' (OK) volvemos automáticamente al cuadro de diálogo anterior. Para ir de un resultado a otro puede utilizarse la parte izquierda de la pantalla, donde aparece el listado de resultados obtenidos (en el ejemplo: 'Edad' e 'Histograma'). Situándose con el ratón en cada uno de ellos puede accederse automáticamente al resultado correspondiente.

El gráfico siguiente muestra el histograma para la variable edad.

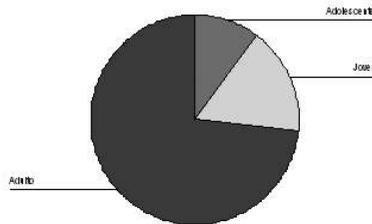
Para ir de un resultado a otro (por ejemplo a la tabla de frecuencias o al histograma) nos situamos con el ratón en el título correspondiente (en el ejemplo, Histograma).



Se puede pedir al PASW graficar el diagrama de barras para la variable 'Edad'.



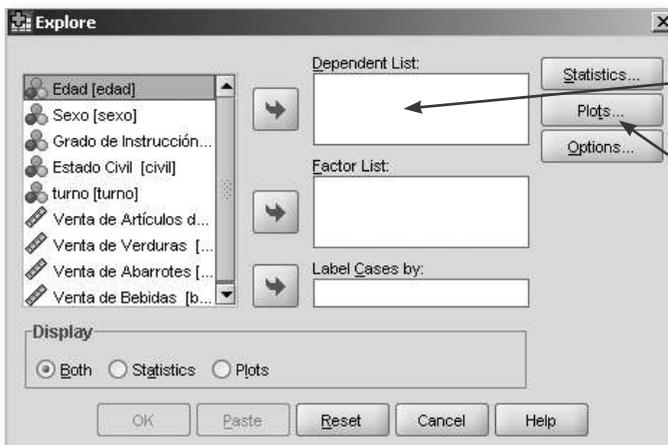
También es posible graficar el diagrama de sectores para la variable 'Edad'; el resultado obtenido es el siguiente:



En el gráfico se observa que los adultos son los que más compran en el supermercado, seguido por los jóvenes. Como *ejercicio* usted puede realizar gráficos (Histograma, Barras, Sectores), teniendo presente que el gráfico debe ser el que mejor describa a las variables.

DIAGRAMA DE TALLO Y HOJAS

Otra de las representaciones que sirven para reflejar distribuciones de frecuencias es el *diagrama de tallo y hojas*. Para confeccionarlo se utiliza el menú **Analizar** -> **Estadístico descriptivos** -> **Explorar**, cuyo cuadro de diálogo es el siguiente:

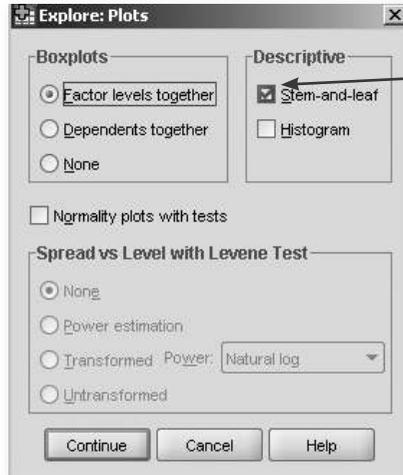


Trasladar aquí la variable o variables para las que se desee obtener un análisis.

Pulsar aquí para solicitar un diagrama de tallo y hojas.

El cuadro de diálogo ofrece diferentes posibilidades de análisis de datos. Por el momento nos centraremos tan sólo en las que nos interesan. En este caso, elaboramos un diagrama de tallo y hojas. Para ello, en primer lugar se seleccionan las variables para las que se desee este tipo de diagrama y se trasladan al cuadro 'Dependientes'.

Se selecciona el botón de la opción 'gráficos' y aparece el siguiente cuadro de diálogo:



Para diagramas de tallo y hojas, seleccionar esta opción.

Veamos un ejemplo con la variable 'Venta de Artículos de Aseo Personal'. Los resultados obtenidos son los siguientes:

| Venta de Artículos de Aseo Personal Stem-and-Leaf Plot | |
|--|---------------|
| Frequency | Stem & Leaf |
| 9.00 | 0 . 222233345 |
| 5.00 | 1 . 02355 |
| 4.00 | 2 . 0335 |
| 2.00 | 3 . 03 |
| 2.00 | 4 . 55 |
| 4.00 | 5 . 0556 |
| 4.00 | 6 . 0005 |
| Stem width: | 1.00 |
| Each leaf: | 1 case(s) |

MEDIDAS DE POSICIÓN Y DE TENDENCIA CENTRAL

Los estadísticos de posición y tendencia central para las variables definidas en el editor de datos pueden obtenerse desde diferentes cuadros de diálogo. A continuación veremos los procedimientos más empleados para cada una de ellas:

1. Medidas de posición central

Informan sobre los valores medios de la serie de datos. Las principales medidas de posición central son las siguientes:

- 1.- **Media:** Es el valor medio ponderado de la serie de datos. Se pueden calcular diversos tipos de media, siendo las más utilizadas:

Media aritmética: Se calcula multiplicando cada valor por el número de veces que se repite. La suma de todos estos productos se divide por el total de datos de la muestra:

$$\bar{X}_a = \frac{(X_1 * f_1) + (X_2 * f_2) + (X_3 * f_3) + \dots + (X_{n-1} * f_{n-1}) + (X_n * f_n)}{n}$$

Media geométrica: Se eleva cada valor al número de veces que se ha repetido. Se multiplican todos estos resultados y al producto final se le calcula la raíz "n" (siendo "n" el total de datos de la muestra).

$$\bar{X}_m = \left(X_1^{f_1} * X_2^{f_2} * X_3^{f_3} * \dots * X_k^{f_{k1}} \right)^{(1/n)}$$

Según el tipo de datos que se analice será más apropiado utilizar la media aritmética o la media geométrica.

La media geométrica se suele utilizar en series de datos como tipos de interés anuales, inflación, etc., donde el valor de cada año tiene un efecto multiplicativo sobre el de los años anteriores. En todo caso, la media aritmética es la medida de posición central más utilizada. Lo más positivo de la media es que en su cálculo se utilizan todos los valores de la serie, por lo que no se pierde ninguna información.

- 2.- **Mediana:** Es el valor de la serie de datos que se sitúa justamente en el centro de la muestra (un 50% de valores son inferiores y otro 50% son superiores).
- 3.- **Moda:** Es el valor que más se repite en la muestra.

Ejemplo: Vamos a utilizar la tabla de distribución de frecuencias con los datos de la estatura de los alumnos que vimos en la lección anterior.

| Variable (Valor) | Frecuencias absolutas | Frecuencias relativas | | |
|---------------------|-----------------------|-----------------------|--------|-----------|
| | Simple | Acumulada | Simple | Acumulada |
| 1,20 | 1 | 1 | 3,3% | 3,3% |
| 1,21 | 4 | 5 | 13,3% | 16,6% |
| 1,22 | 4 | 9 | 13,3% | 30,0% |
| 1,23 | 2 | 11 | 6,6% | 36,6% |
| 1,24 | 1 | 12 | 3,3% | 40,0% |
| 1,25 | 2 | 14 | 6,6% | 46,6% |
| 1,26 | 3 | 17 | 10,0% | 56,6% |
| 1,27 | 3 | 20 | 10,0% | 66,6% |
| 1,28 | 4 | 24 | 13,3% | 80,0% |
| 1,29 | 3 | 27 | 10,0% | 90,0% |
| 1,30 | 3 | 30 | 10,0% | 100,0% |

Vamos a calcular los valores de las distintas posiciones centrales:

- 1.- Media aritmética: 1.253 Por lo tanto, la estatura media de este grupo de alumnos es de 1,253 cm.
- 2.- Media geométrica: Luego: 1.253
En este ejemplo la media aritmética y la media geométrica coinciden, pero no tiene siempre por qué ser así.
- 3.- Mediana
La mediana de esta muestra es 1,26 cm, ya que por debajo está el 50% de los valores y por arriba el otro 50%. Esto se puede ver al analizar la columna de frecuencias relativas acumuladas.
En este ejemplo, como el valor 1,26 se repite en 3 ocasiones, la media se situaría exactamente entre el primer y el segundo valor de este grupo, ya que entre estos dos valores se encuentra la división entre el 50% inferior y el 50% superior.
- 4.- Moda
Hay 3 valores que se repiten en 4 ocasiones: el 1,21, el 1,22 y el 1,28, por lo tanto esta serie de datos cuenta con 3 modas.

Resultados con el PASW.

Estadísticos

Estatura

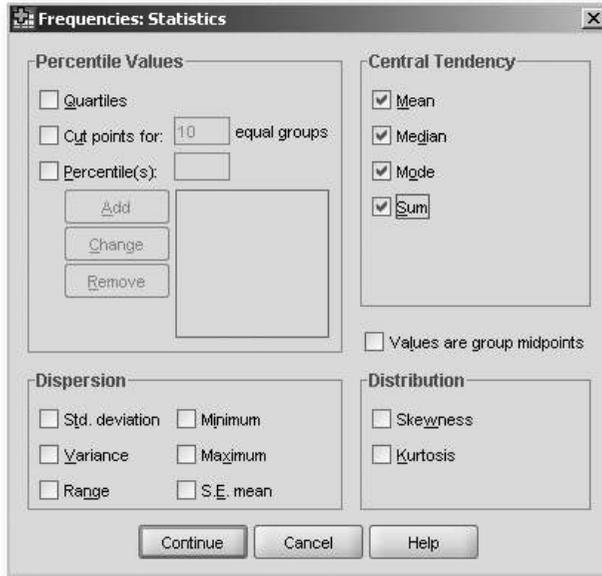
| | | |
|---------|----------|-------------------|
| N | Válidos | 30 |
| | Perdidos | 0 |
| Media | | 1,2533 |
| Mediana | | 1,2600 |
| Moda | | 1,21 ^a |
| Suma | | 37,60 |

a. Existen varias modas.
Se mostrará el menor de los valores.

Ejercicios de medidas de tendencia central: Media, mediana, moda y suma.

Estos estadísticos pueden obtenerse prácticamente desde cualquier cuadro de diálogo del menú *Analizar -> Estadísticas descriptivas*. Uno de los más empleados es el del cuadro de diálogo 'Frecuencias' ya que cuando se elabora una tabla de frecuencias normalmente también se desea obtener las medidas de posición.

Para ello se pulsa en el botón 'Estadísticas' y aparece el cuadro de diálogo, aquí se selecciona las estadísticas que se desean utilizar en el estudio. En nuestro ejemplo se pide calcular la 'media', 'mediana', 'moda' y 'suma'; esto se logra simplemente marcando cada opción.



Luego se presiona el botón continuar y después aceptar (OK). El PASW calcula estos cuatro valores estadísticos para el listado de variables seleccionadas.

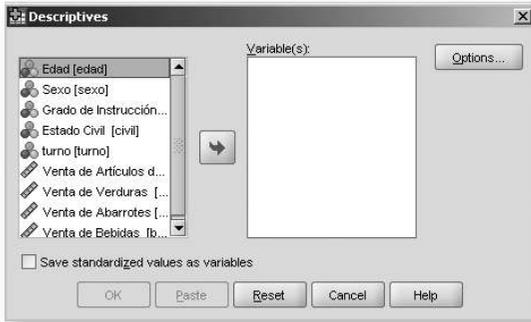
Statistics

Venta de Artículos de Aseo Personal

| | | |
|--------|---------|--------|
| N | Valid | 30 |
| | Missing | 0 |
| Mean | | 2.6533 |
| Median | | 2.1500 |
| Mode | | .20 |
| Sum | | 79.60 |

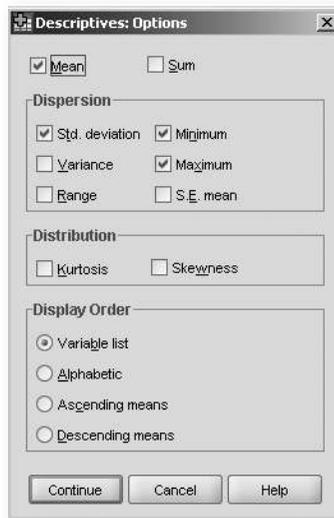
La media que se obtiene se corresponde con la media aritmética, la mediana con el valor central cuando los datos se ordenan en modo creciente y la moda con el valor que obtiene la frecuencia absoluta mayor.

Para conocer las estadísticas descriptivas de las variables también puede seleccionarse el procedimiento '*Descriptivas*'. Este procedimiento presenta un cuadro de diálogo muy parecido al del procedimiento '*Frecuencias*'. El cuadro de diálogo del procedimiento '*Descriptivas*' es el siguiente:



Pulsar para indicar qué estadísticos variados se desea obtener.

Para obtener las estadísticas descriptivas, primero se trasladan las variables correspondientes al cuadro 'variables' y luego se pulsa en *Opciones*, obteniéndose el cuadro de diálogo siguiente:



En este caso, solamente está disponible la estadística de tendencia central 'media' y 'suma' que, como antes, representa la media aritmética para las variables seleccionadas. Veamos un ejemplo para algunas variables del archivo.

Descriptive Statistics

| | N | Sum | Mean |
|-------------------------------------|----|---------|---------|
| Venta de Artículos de Aseo Personal | 30 | 79.60 | 2.6533 |
| Venta de Verduras | 30 | 376.00 | 12.5333 |
| Venta de Abarrotes | 30 | 1036.20 | 34.5400 |
| Venta de Bebidas | 30 | 125.00 | 4.1667 |
| Valid N (listwise) | 30 | | |

Se aprecia que las 30 personas gastaron en total S/.79.60 en artículos de aseo personal; en el supermercado la media de los gastos es S/. 2.6533. En forma similar, se aprecia en la venta de Verduras que las 30 personas gastaron S/. 356.00 en total con una media o promedio de gastos de S/. 12.533.

2. Medidas de posición no centrales

Informan de cómo se distribuyen el resto de los valores de la serie. Las medidas de posición no centrales permiten conocer otros puntos característicos de la distribución que no son los valores centrales.

Entre otros indicadores, se suelen utilizar una serie de valores que dividen la muestra en tramos iguales como:

Cuartiles: Son 3 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente.

Deciles: Son 9 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en diez tramos iguales, en los que cada uno de ellos concentra el 10% de los resultados.

Percentiles: Son 99 valores que distribuyen la serie de datos, ordenada de forma creciente o decreciente, en cien tramos iguales, en los que cada uno de ellos concentra el 1% de los resultados.

Ejemplo: Vamos a calcular los cuartiles de la serie de datos referidos a la estatura de un grupo de alumnos (lección 2ª). Los deciles y centiles se calculan de igual manera, aunque harían falta distribuciones con mayor número de datos.

| Variable (Valor) | Frecuencias absolutas | | Frecuencias relativas | |
|---------------------|-----------------------|-----------|-----------------------|-----------|
| | Simple | Acumulada | Simple | Acumulada |
| 1,20 | 1 | 1 | 3,3% | 3,3% |
| 1,21 | 4 | 5 | 13,3% | 16,6% |
| 1,22 | 4 | 9 | 13,3% | 30,0% |
| 1,23 | 2 | 11 | 6,6% | 36,6% |
| 1,24 | 1 | 12 | 3,3% | 40,0% |
| 1,25 | 2 | 14 | 6,6% | 46,6% |
| 1,26 | 3 | 17 | 10,0% | 56,6% |
| 1,27 | 3 | 20 | 10,0% | 66,6% |
| 1,28 | 4 | 24 | 13,3% | 80,0% |
| 1,29 | 3 | 27 | 10,0% | 90,0% |
| 1,30 | 3 | 30 | 10,0% | 100,0% |

1º cuartil: es el valor 1,22 cm, ya que por debajo suya se sitúa el 25% de la frecuencia (tal como se puede ver en la columna de la frecuencia relativa acumulada).

2º cuartil: es el valor 1,26 cm, ya que entre este valor y el 1º cuartil se sitúa otro 25% de la frecuencia.

3º cuartil: es el valor 1,28 cm, ya que entre este valor y el 2º cuartil se sitúa otro 25% de la frecuencia. Además, por encima suyo queda el restante 25% de la frecuencia.

Atención: cuando un cuartil cae en un valor que se ha repetido más de una vez (como ocurre en el ejemplo en los tres cuartiles), la medida de posición no central sería realmente una de las repeticiones.

Resultados con el PASW

Estadísticos

| Estatura | | |
|-------------|----------|--------|
| N | Válidos | 30 |
| | Perdidos | 0 |
| Percentiles | 25 | 1,2200 |
| | 50 | 1,2600 |
| | 75 | 1,2800 |

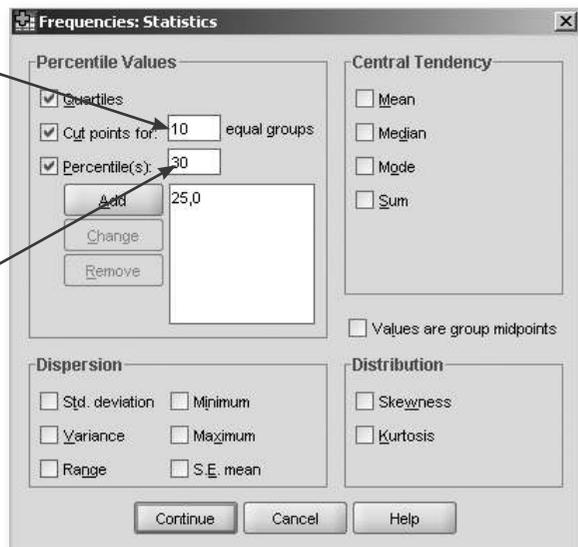
Ejercicios de medidas de tendencia no central

Los percentiles (Centiles), los deciles y los cuartiles:

Los centiles y otros cuantiles pueden obtenerse desde el procedimiento *Frecuencias*. En la parte inferior de este cuadro de diálogo hay un botón llamado 'Estadísticos'. Si se selecciona se obtiene el siguiente cuadro de diálogo:

Si se desea obtener los Cuarteles indicar grupos de 4, para los Deciles 10 y para los Centiles 100.

Esta opción es específica para los centiles y permite definir el centil (entre 1 y 99) que se desee obtener. Para indicar varios escribir el número y pulsar en Añadir (Add) sucesivamente



Desde aquí puede indicarse que se obtengan diferentes tipos de cuantiles o los percentiles. Se puede indicar el número exacto del percentil que se desee obtener e ir añadiendo a la lista inferior.

Continuemos con la variable artículos de aseo personal. Si le pedimos que nos dé los cuartiles, deciles y los percentiles 25 y 30, se obtiene:

Statistics

Venta de Artículos de Aseo Personal

| | | |
|-------------|---------|--------|
| N | Valid | 30 |
| | Missing | 0 |
| Percentiles | 10 | .2000 |
| | 20 | .3000 |
| | 25 | .3750 |
| | 30 | .6500 |
| | 40 | 1.3800 |
| | 50 | 2.1500 |
| | 60 | 2.8000 |
| | 70 | 4.5000 |
| | 75 | 5.1250 |
| | 80 | 5.5000 |
| | 90 | 6.0000 |

Como se observa, hay un 75% de sujetos que gastan S/. 5.1250 como máximo en las compras del supermercado y el 10% gasta S/. 0.20 como máximo en la compra de artículos de aseo personal.

3. Medidas de variación (dispersión)

Estudia la distribución de los valores de la serie, analizando si estos se encuentran más o menos concentrados o más o menos dispersos. Existen diversas **medidas de dispersión**; entre las más utilizadas podemos destacar las siguientes:

- 1.- **Rango:** Mide la amplitud de los valores de la muestra y se calcula por diferencia entre el valor más elevado y el valor más bajo.

$$R = X_{\max} - X_{\min}$$

- 2.- **Varianza:** Mide la distancia existente entre los valores de la serie y la media. Se calcula como sumatorio de las diferencias al cuadrado entre cada valor y la media, multiplicadas por el número de veces que se ha repetido cada valor. El sumatorio obtenido se divide por el tamaño de la muestra. La varianza siempre será mayor que cero. Mientras más se aproxima a cero, más concentrados están los valores de la serie alrededor de la media. Por el contrario, mientras mayor sea la varianza, más dispersos están.

$$S_x^2 = \frac{\sum_{i=1}^k (X_i - \bar{X})^2 f_i}{n}$$

- 3.- **Desviación típica:** Se calcula como raíz cuadrada de la varianza. S_x
- 4.- **Coefficiente de varización de Pearson:** Se calcula como cociente entre la desviación típica y la media.

$$C_v = S_x / \bar{x}$$

Ejemplo: Vamos a utilizar la serie de datos de la estatura de los alumnos de una clase (lección 2ª) y vamos a calcular sus medidas de dispersión.

| Variable (Valor) | Frecuencias absolutas | | Frecuencias relativas | |
|---------------------|-----------------------|-----------|-----------------------|-----------|
| | Simple | Acumulada | Simple | Acumulada |
| X | X | X | X | X |
| 1,20 | 1 | 1 | 3,3% | 3,3% |
| 1,21 | 4 | 5 | 13,3% | 16,6% |
| 1,22 | 4 | 9 | 13,3% | 30,0% |
| 1,23 | 2 | 11 | 6,6% | 36,6% |
| 1,24 | 1 | 12 | 3,3% | 40,0% |
| 1,25 | 2 | 14 | 6,6% | 46,6% |
| 1,26 | 3 | 17 | 10,0% | 56,6% |
| 1,27 | 3 | 20 | 10,0% | 66,6% |
| 1,28 | 4 | 24 | 13,3% | 80,0% |
| 1,29 | 3 | 27 | 10,0% | 90,0% |
| 1,30 | 3 | 30 | 10,0% | 100,0% |

Rango: Diferencia entre el mayor valor de la muestra (1,30) y el menor valor (1,20). Luego el rango de esta muestra es 10 cm.

Varianza: Recordemos que la media de esta muestra es 1,253. Luego aplicamos la fórmula, por lo tanto, la varianza es 0,0010

Desviación típica: Es la raíz cuadrada de la varianza. Luego: 0.0320

Coefficiente de variación de Pearson: Se calcula como cociente entre la desviación típica y la media de la muestra. Luego, $Cv = 0,0320 / 1,253$ $Cv = 0,0255$

Resultados con el PASW

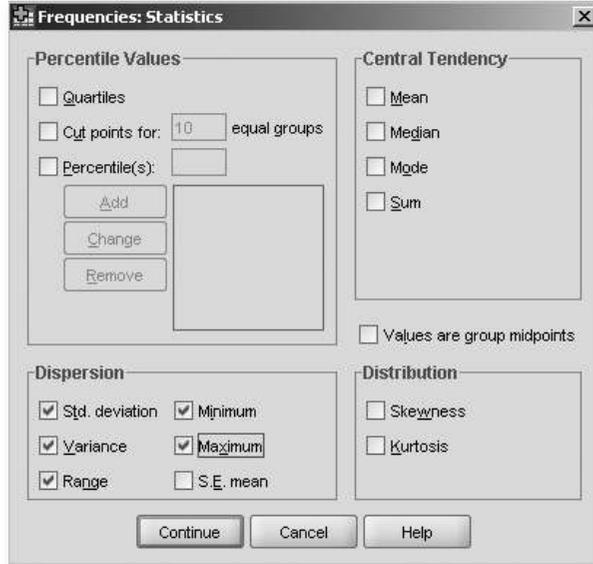
| Estadísticos | | |
|--------------|------------------------|--------|
| Estatura | | |
| N | Válidos | 30 |
| | Perdidos | 0 |
| | Error típ. de la media | ,00592 |
| | Desv. típ. | ,03241 |
| | Varianza | ,001 |
| | Rango | ,10 |
| | Mínimo | 1,20 |
| | Máximo | 1,30 |

Ejercicios de medidas de variación

Las medidas o estadísticas de variación se obtienen de forma muy similar y desde los mismos menús y los cuadros de diálogo que las de posición y tendencia central. Las más empleadas son: la varianza y la desviación típica, aparecen por defecto en muchos menús del PASW y en aquellas opciones que se denominan ‘descriptivas’ sin especificar exactamente a qué se refieren.

Varianza, desviación típica, rango, mínimo y máximo.

La varianza y desviación típica, rango, mínimo y máximo pueden obtenerse a partir del procedimiento 'Frecuencia' en el botón 'Estadísticas'. Desde aquí puede pedirse que se calcule: la desviación típica, la varianza y la amplitud total o rango de las puntuaciones.



Otra forma de proceder para obtener medidas de variación es desde el menú *Analizar -> Estadísticos descriptivos -> Descriptivos* o de desde el menú *Analizar -> Explorar (Estadísticas)*.

Ejemplo para la venta de aseo personal. Desde cualquiera de los cuadros de diálogo señalados se obtienen los siguientes resultados:

Statistics

Venta de Artículos de Aseo Personal

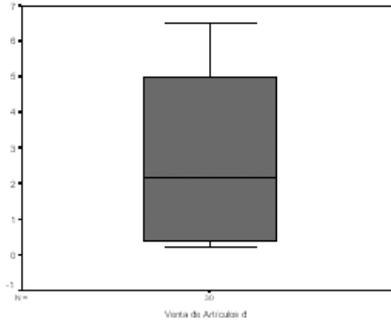
| | | |
|----------------|---------|--------|
| N | Valid | 30 |
| | Missing | 0 |
| Std. Deviation | | 2.2483 |
| Variance | | 5.0550 |
| Range | | 6.30 |
| Minimum | | .20 |
| Maximum | | 6.50 |

Nota Importante: Conviene tener en cuenta un aspecto importante. El programa PASW tiene la peculiaridad de que proporciona la cuasi varianza. Es decir, cuando se le pide que nos dé la varianza y la desviación típica, lo que da exactamente es el resultado de las fórmulas:

$$S_x^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} \quad S_x' = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

4. Gráficos de cajas y bigotes

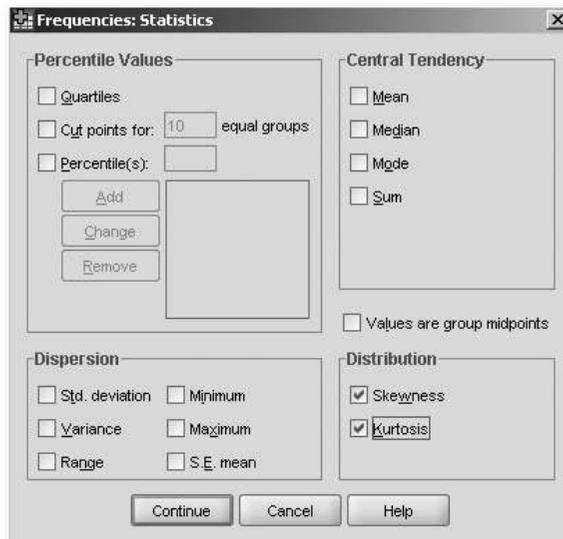
El PASW también ofrece la posibilidad de elaborar los gráficos que expresan la dispersión de los datos en una variable. En concreto, desde el menú **Analizar -> Explorar** (Gráficos). Puede definirse un diagrama de cajas y bigotes para una variable en estudio. Este tipo de diagramas ofrece información sobre el rango de la variable y los cuarteles, por ejemplo para la variable ‘artículos de aseo personal’ se obtiene el gráfico siguiente:



Se observa en la gráfica que los gastos oscilan entre 0,20 y 6,50 nuevos soles, estando la mediana en 2,15 nuevos soles.

5. Asimetría y Curtosis

La asimetría (Skewness) y la curtosis (Kurtosis) pueden obtenerse a partir del menú **Analizar -> Frecuencias** pulsando el botón ‘Estadísticas’, luego marcando las respectivas distribuciones.



También se obtiene el mismo resultado desde el menú **Analizar -> Estadísticas descriptivas -> Descriptivas**. Desde ambos cuadros de diálogo se ofrecen exactamente las mismas opciones: el índice de asimetría de Fisher para cada variable y el índice de curtosis. Además del valor de la estadística tanto para la asimetría como para la curtosis, el PASW proporciona información sobre el error típico en el cálculo de estos índices. Veamos un ejemplo para algunas variables de la encuesta en el supermercado.

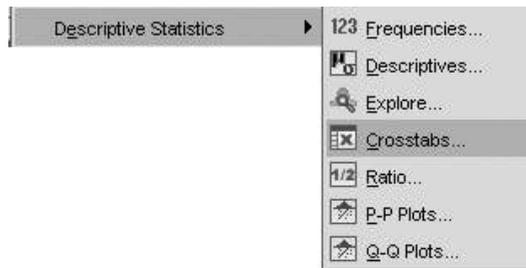
Statistics

| | | Venta de Artículos de Aseo Personal | Venta de Verduras | Venta de Abarrotes |
|------------------------|---------|-------------------------------------|-------------------|--------------------|
| N | Valid | 30 | 30 | 30 |
| | Missing | 0 | 0 | 0 |
| Skewness | | .435 | .562 | -.050 |
| Std. Error of Skewness | | .427 | .427 | .427 |
| Kurtosis | | -1.413 | -.565 | .082 |
| Std. Error of Kurtosis | | .833 | .833 | .833 |

6. Tablas de Contingencia

Veremos cómo se obtienen distribuciones conjuntas de frecuencias y su representación gráfica para dos o más variables de diversos tipos. Los datos utilizados son de la encuesta del supermercado.

La forma de proceder en el PASW es como ya vimos, todos los análisis se realizan desde el menú *Analizar -> Estadísticas descriptivas -> Tablas de contingencia (Crosstabs...)*.



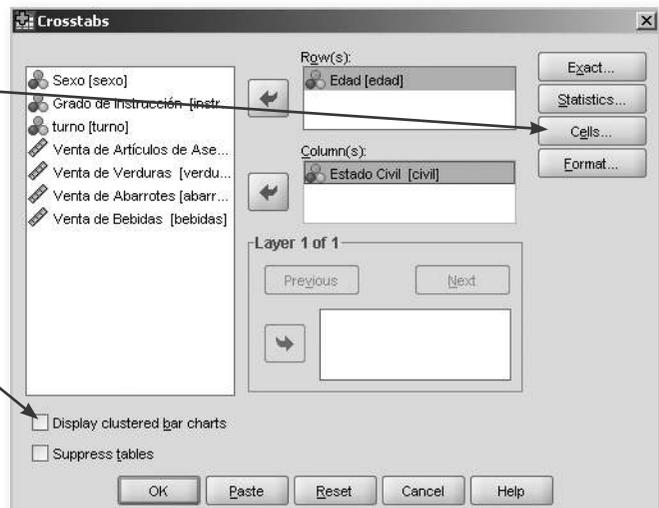
Dos variables cualitativas:

Para describir dos variables cualitativas primero debemos elaborar una *Tabla de contingencia* ((Crosstabs...)).

Si se selecciona el procedimiento '*Tablas de contingencia*' aparece el cuadro de diálogo que se muestra en la figura siguiente:

Pulsar para indicar si se desea obtener los datos en porcentajes.

Seleccionar si se desea obtener un Diagrama de barras para las variables

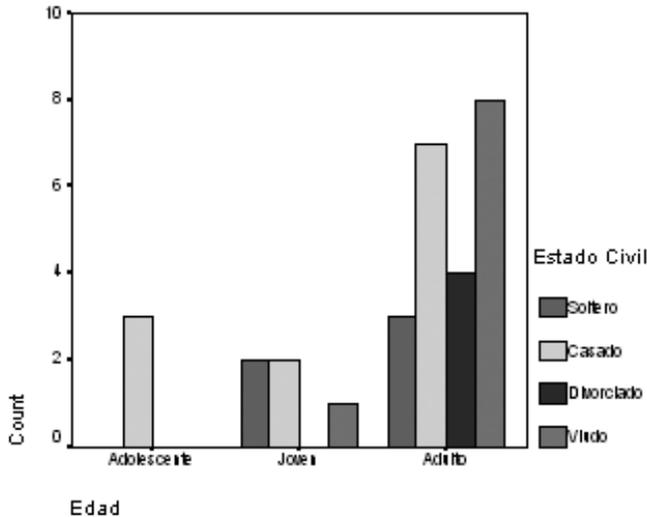


Supongamos que deseamos conocer la distribución conjunta de frecuencias de las variables 'Edad' y 'Estado Civil'. Para indicar al PASW esta instrucción, primero se seleccionan las variables. Por ejemplo, las categorías de la variable *Edad* pueden ser las filas (Row[s]) de la tabla de frecuencias y las de la variable *Estado Civil* las columnas (Column[s]) como se muestra en la figura anterior. Este cuadro de diálogo también permite obtener una representación gráfica para las dos variables.

Si se pulsa en el botón *Aceptar* el resultado obtenido es el siguiente:

Edad * Estado Civil Crosstabulation

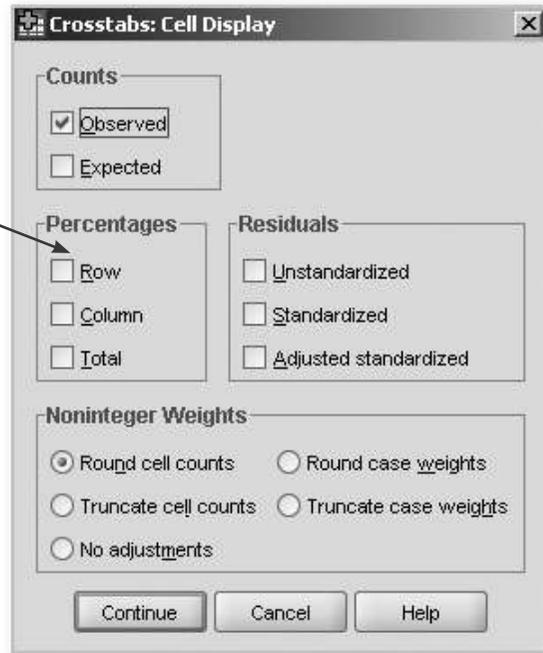
| Count | | Estado Civil | | | | Total |
|-------|-------------|--------------|--------|------------|-------|-------|
| | | Soltero | Casado | Divorciado | Viudo | |
| Edad | Adolescente | | 3 | | | 3 |
| | Joven | 2 | 2 | | 1 | 5 |
| | Adulto | 3 | 7 | 4 | 8 | 22 |
| Total | | 5 | 12 | 4 | 9 | 30 |



En la tabla de resultados se observa las *distribuciones conjuntas y marginales* de las variables y la representación gráfica del diagrama de barras conjunto. Como se observa, de los 30 sujetos de la muestra 3 son Adolescentes, 5 son Jóvenes y 22 son Adultos. De los Adultos la mayoría son viudos (8 son viudos) etc.

Las distribuciones conjuntas también pueden obtenerse en frecuencias relativas o porcentajes para ambas variables (filas y columnas). Para ello se selecciona el botón *Celdas (Cells)* del cuadro de diálogo anterior, cuyo aspecto es el siguiente:

Señalar si se desea obtener las frecuencias conjuntas relativas para las filas y las columnas



Seleccionando las opciones marcadas se obtienen los mismos resultados que antes pero expresados en porcentajes.

Edad * Estado Civil Crosstabulation

| | | | Estado Civil | | | | Total |
|--------|-----------------------|-----------------------|--------------|--------|------------|--------|--------|
| | | | Soltero | Casado | Divorciado | Viudo | |
| Edad | Adolescente | Count | | 3 | | | 3 |
| | | % within Edad | | 100.0% | | | 100.0% |
| | | % within Estado Civil | | 25.0% | | | 10.0% |
| Joven | Count | 2 | 2 | | 1 | 5 | |
| | % within Edad | 40.0% | 40.0% | | 20.0% | 100.0% | |
| | % within Estado Civil | 40.0% | 16.7% | | 11.1% | 16.7% | |
| Adulto | Count | 3 | 7 | 4 | 8 | 22 | |
| | % within Edad | 13.6% | 31.8% | 18.2% | 36.4% | 100.0% | |
| | % within Estado Civil | 60.0% | 58.3% | 100.0% | 88.9% | 73.3% | |
| Total | Count | 5 | 12 | 4 | 9 | 30 | |
| | % within Edad | 16.7% | 40.0% | 13.3% | 30.0% | 100.0% | |
| | % within Estado Civil | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | |

Veamos un ejemplo de interpretación para los Solteros. Del 100% de los Solteros se observa, que el 0% son Adolescentes, el 40% son Jóvenes y el 60% son Adultos (la lectura fue en columna).

Veamos un ejemplo de interpretación para los adultos. Del 100% de Adultos se observa, que el 13.6% son Solteros, el 31.8% son Casados, el 18.2% son Divorciados, y 36.4% son Viudos (la lectura fue en Fila).

Las tablas de contingencia sólo tienen sentido para variables de tipo cualitativo (nominal u ordinal). Si se desea representar la distribución conjunta de dos variables cuantitativas es necesario agrupar los valores de cada una de dichas variables en intervalos de la misma amplitud (para ello puede utilizarse el menú *Transformar -> Recodificar*). Una vez obtenidos los intervalos para dichas variables pueden elaborarse la tabla de contingencia siguiendo el mismo procedimiento que si las variables fueran cualitativas.

La gráfica más adecuada para dos variables cuantitativas es el diagrama de dispersión.

Una variable cualitativa y otra cuantitativa:

Si queremos cruzar una variable *cualitativa* con una *cuantitativa*, se procede igual que para variables cualitativas, excepto que para resumir los datos de la cuantitativa hay que emplear estadísticas (sumas, medias, etc.). También se puede representar gráficamente la relación entre estas dos variables. La presentación gráfica es más recomendable por que podemos análisis con mayor claridad las variables en estudio.

Una variable cuantitativa y dos cualitativas:

Es frecuente la representación de una variable dependiente cuantitativa en función de dos independientes de tipo cualitativo. En estos casos también se puede representar gráficamente la relación entre estas tres variables. En forma similar a la anterior la presentación gráfica es más recomendable por que podemos analizar con mayor claridad las variables en estudio, como veremos en el capítulo de Gráficos.

Prueba de independencia (χ^2)

EJERCICIO 1

Los datos aparecen recopilados en las variables **Día1**, **Día2**, **Día3** que indican el número de horas trabajadas en los tres días. Además se han recopilado la variable **Sexo** que toma valores 0 para mujeres y 1 para hombres. **Edad**, 1 para Adolescentes, 2 para Jóvenes y 3 para Adultos. Finalmente se incluyen tres datos relacionados con las terapias que ha recibido el sujeto. Todas ellas toman 1, si el sujeto ha recibido la terapia y 0 en caso contrario. Las variables de terapia son **Estrés**, **Ansiedad** y **Fobia**.

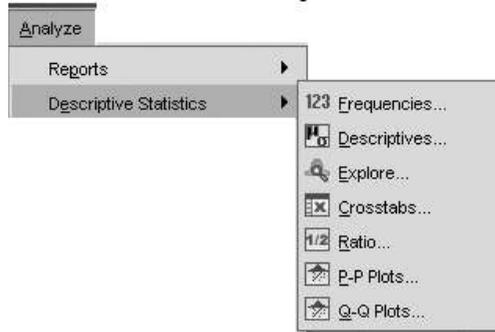
Los datos representan a 10 sujetos que participaron en el estudio, cuyos resultados obtenidos se muestra en la tabla siguiente:

| Sujeto | Día1 | Día2 | Día3 | Sexo | Edad | Estrés | Ansiedad | Fobia |
|--------|------|------|------|------|------|--------|----------|-------|
| 1 | 5.7 | 6.3 | 6.8 | 0 | 1 | 0 | 1 | 0 |
| 2 | 4.8 | 5.4 | 4.3 | 1 | 3 | 0 | 0 | 0 |
| 3 | 7.6 | 7.9 | 8.6 | 0 | 2 | 0 | 0 | 0 |
| 4 | 5.7 | 6.0 | 6.7 | 1 | 1 | 0 | 0 | 1 |
| 5 | 3.8 | 4.6 | 4.9 | 1 | 3 | 0 | 0 | 0 |
| 6 | 7.5 | 8.0 | 8.5 | 1 | 2 | 0 | 1 | 0 |
| 7 | 6.4 | 7.0 | 7.3 | 1 | 1 | 1 | 1 | 0 |
| 8 | 7.7 | 8.1 | 8.5 | 0 | 2 | 0 | 0 | 0 |
| 9 | 4.0 | 4.5 | 5.0 | 0 | 3 | 1 | 0 | 0 |
| 10 | 5.7 | 6.1 | 6.8 | 1 | 2 | 1 | 0 | 1 |

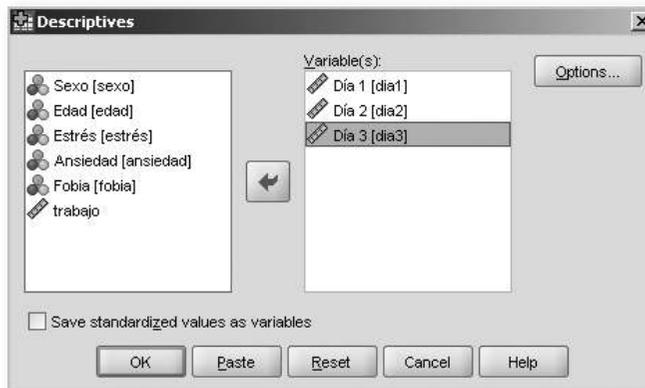
Estadísticas Descriptivas

Cuando realizamos análisis descriptivos de los datos se utiliza el procedimiento 'descriptivas' que proporcionan distintas estadísticas de tendencia central, dispersión, asimetría y curtosis. Además, permiten obtener puntuaciones típicas y guardarlos como nuevas variables.

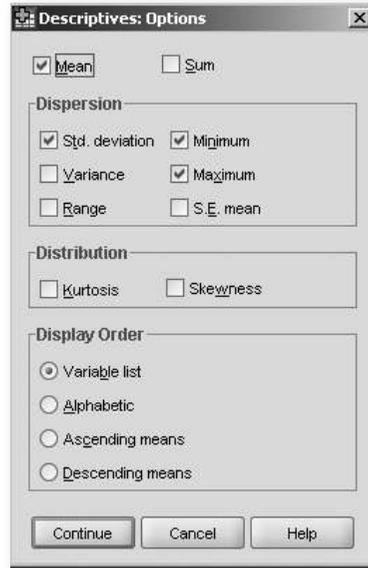
Para seleccionar el procedimiento descriptivo se debe escoger las opciones del menú: *Analizar>Estadísticas >Descriptivos*:



A continuación aparece un cuadro de diálogo en la que se listan las variables del archivo de datos. Para indicar al PASW cuáles son las variables que se desea analizar, se seleccionan estas variables de la lista pulsando sobre su nombre con el ratón. Luego, se pulsa sobre el botón flecha para trasladar las variables al cuadro de variables seleccionadas.



Siempre los cuadros de diálogo de procedimientos del PASW poseen varios botones. Los botones comunes son *Acertar (OK)*, *pegar*, *restablece*, *Cancelar* y *Ayuda*. El cuadro de diálogo del procedimiento Descriptivos contiene el botón específico denominado: *Opciones...* Los tres puntos significan que pulsando sobre este botón aparece un subcuadro de diálogo que permite seleccionar las estadísticas que se desean calcular.



Para poder realizar los cálculos estadísticos es necesario pulsar el botón **OK** del cuadro de diálogo del procedimiento *Descriptivos*. El PASW muestra los cálculos solicitados en la ventana visor de resultados.

Descriptive Statistics

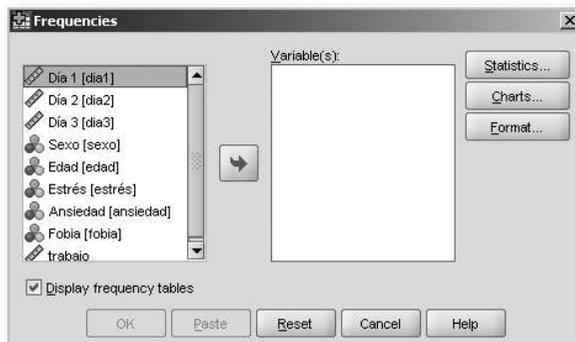
| | N | Minimum | Maximum | Mean | Std. Deviation |
|--------------------|----|---------|---------|--------|----------------|
| Día 1 | 10 | 3.80 | 7.70 | 5.8900 | 1.4255 |
| Día 2 | 10 | 4.50 | 8.10 | 6.3900 | 1.3404 |
| Día 3 | 10 | 4.30 | 8.60 | 6.7400 | 1.5756 |
| Valid N (listwise) | 10 | | | | |

Frecuencia:

Este procedimiento proporciona la distribución de frecuencias de una variable en forma de tabla y también en forma de gráfica, además de varias estadísticas descriptivas.

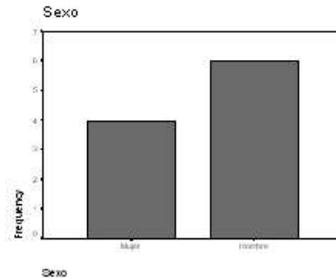
Para ejecutar este procedimiento es necesario seleccionar el menú **Análizar > Estadísticas descriptivas > frecuencias**.

Luego aparece un cuadro de diálogo en el cual se indican las variables del archivo de datos que se está trabajando. Al igual que en el procedimiento *Descriptivos*, el primer paso consiste en trasladar al recuadro de la derecha (Variables) aquellas variables que se desea analizar.



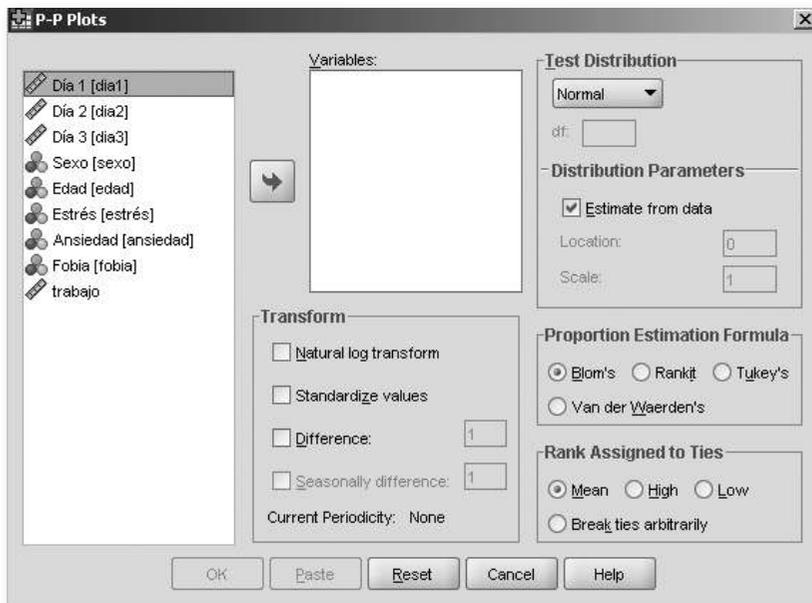
Si pulsamos sobre el botón Estadísticas pueden seleccionarse diversas estadísticas como pueden ser: de tendencia central y no central, dispersión, forma de distribución, percentiles, deciles etc.

En este cuadro de diálogo también se puede utilizar el botón gráficos (Charts...) para indicar al PASW que se desea realizar un gráfico de barras, de sectores o un histograma; por ejemplo si se quiere utilizar la gráfica de barras en la variable 'Sexo' el resultado se observa en el diagrama siguiente:



P-P Plots

Para crear un gráfico de P-P, el cuadro de diálogo del gráfico deseado es:

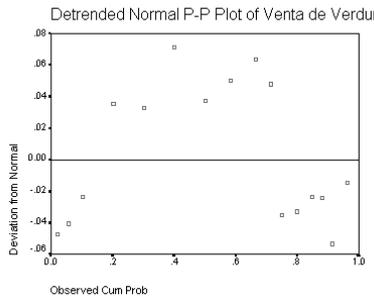
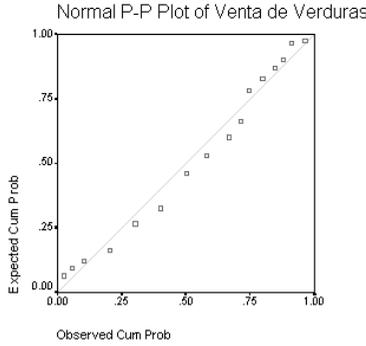


Crea un gráfico de las proporciones acumuladas de una variable respecto a las de una distribución cualquiera de prueba. Es decir, por cada variable especificada proporciona dos gráficos: en el primero se representan los valores de la función de distribución acumulativa esperada bajo el supuesto de normalidad (por defecto) frente a los observados. En el segundo se representan los residuos. Los gráficos de probabilidad suelen emplearse para determinar si la distribución de una variable coincide con una distribución dada.

Si la variable seleccionada coincide con la distribución de prueba, los puntos se concentran en torno a una línea recta. Entre las distribuciones de prueba disponibles se encuentran la beta, Chi-cuadrado,

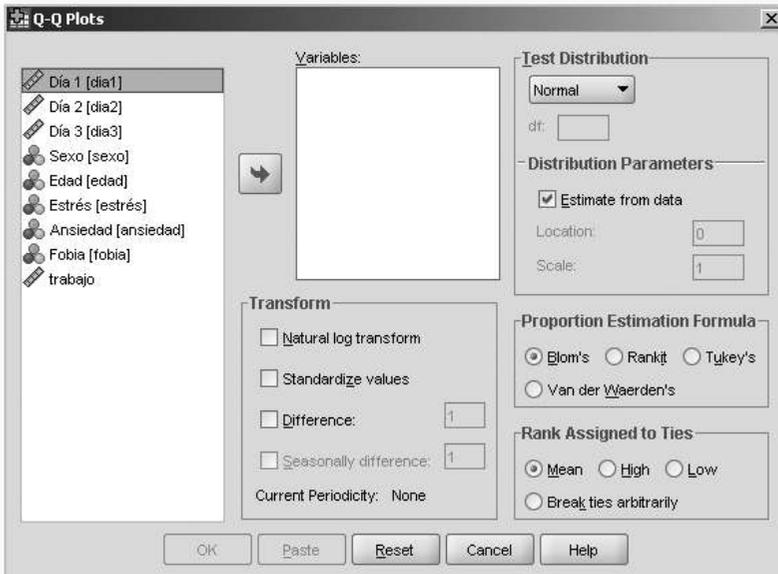
exponencial, gamma, semi-normal, Laplace, logístico, Lognormal, normal, Pareto, t de Student, Weibull y uniforme. Según la distribución elegida, pueden especificarse distintos grados de libertad y otros parámetros.

Ejemplo: Para la variable *venta de verduras* se puede decir que sí tiene una distribución normal.



Q-Q Plots

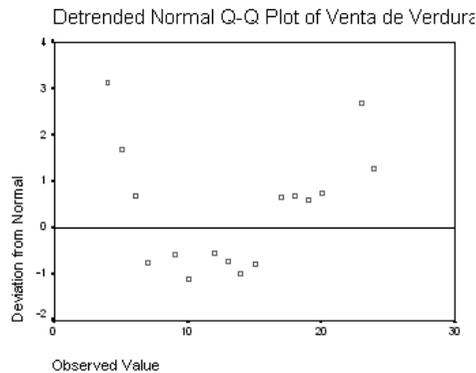
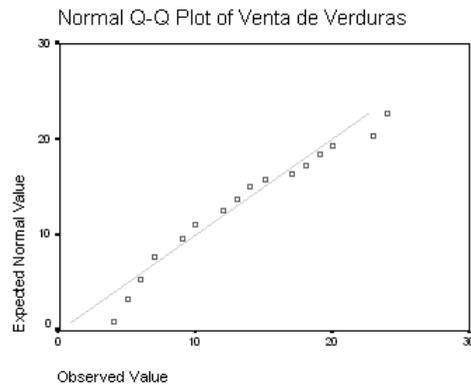
Para crear un gráfico de Q-Q, el cuadro de diálogo del gráfico deseado es:





Crea un gráfico con los cuantiles de distribución de una variable respecto a los cuantiles de una distribución cualquiera de prueba. Es decir, por cada variable especificada proporciona dos gráficos: en el primero se representan los valores de los cuantiles esperados bajo el supuesto de distribución normal (por defecto) de media cero y varianza uno frente a los observados. En el segundo, en lugar de los cuantiles esperados, se representan los residuos. Las distribuciones de prueba disponibles son las mismas que en P-P.

Ejemplo: Para la variable *venta de verduras* se puede decir que sí tiene una distribución normal.



Capítulo

8

Correlación

Análisis de Resultados y Conclusiones: Muestra los principales hallazgos de la investigación aplicando Estadísticas Descriptivas (tablas, cuadros, gráficas, etc.) y presenta una potente interpretación teórica que demuestra el dominio técnico del investigador. Las **conclusiones** extraen lo esencial de todo el proceso enfatizando especialmente la riqueza de la evidencia empírica aportada.



CORRELACIÓN

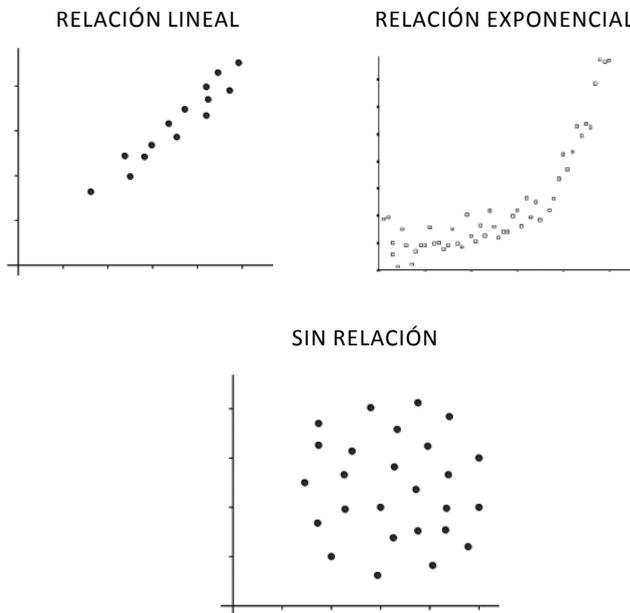
Permite saber si las variables en estudio de nuestro trabajo de investigación guardan algún tipo de relación entre ellas.

COEFICIENTE DE CORRELACIÓN LINEAL

En una distribución bidimensional puede ocurrir que las dos variables guarden algún tipo de relación entre sí.

Por ejemplo, si se analiza la estatura y el peso de los alumnos de una clase es muy posible que exista relación entre ambas variables: mientras más alto sea el alumno, mayor será su peso.

El coeficiente de correlación lineal mide el grado de intensidad de esta posible relación entre las variables. Este coeficiente se aplica cuando la relación que puede existir entre las variables es lineal (es decir, si representáramos en un gráfico los pares de valores de las dos variables la nube de puntos se aproximaría a una recta).



No obstante, puede que exista una relación que no sea lineal, sino exponencial, parabólica, etc. En estos casos, el coeficiente de correlación lineal mediría mal la intensidad de la relación las variables, por lo que convendría utilizar otro tipo de coeficiente más apropiado.

Para ver, por tanto, si se puede utilizar el coeficiente de correlación lineal, lo mejor es representar los pares de valores en un gráfico y ver la forma que describe.

El coeficiente de correlación lineal se calcula aplicando la siguiente fórmula:

$$r_{xy} = \frac{1/n * \sum (x_i - \bar{x}_m) * (y_i - \bar{y}_m)}{((1/n * \sum (x_i - \bar{x}_m)^2) * (1/n * \sum (y_i - \bar{y}_m)^2))^{1/2}}$$

Donde \bar{x}_m : Media de x

\bar{y}_m : Media de y

Es decir:

Numerador: se denomina covarianza y se calcula de la siguiente manera: en cada par de valores (x, y) se multiplica la "x" menos su media, por la "y" menos su media. Se suma el resultado obtenido de todos los pares de valores y este resultado se divide por el tamaño de la muestra.

Denominador se calcula el producto de las varianzas de "x" y de "y", y a este producto se le calcula la raíz cuadrada.

Los valores que puede tomar el coeficiente de correlación "r" son: $-1 < r < 1$

- a) Si "**r**" > 0, la correlación lineal es positiva (si sube el valor de una variable sube el de la otra). La correlación es tanto más fuerte cuanto más se aproxime a 1.

Ejemplo

La altura y peso: los alumnos más altos suelen pesar más.

- b) Si "**r**" < 0, la correlación lineal es negativa (si sube el valor de una variable disminuye el de la otra). La correlación negativa es tanto más fuerte cuanto más se aproxime a -1.

Ejemplo

Peso y velocidad: los alumnos más gordos suelen correr menos.

- c) Si "**r**" = 0, no existe correlación lineal entre las variables. Aunque podría existir otro tipo de correlación (parabólica, exponencial, etc.).

De todos modos, aunque el valor de "r" fuera próximo a 1 o -1, tampoco esto quiere decir obligatoriamente que existe una relación de causa-efecto entre las dos variables, ya que este resultado podría haberse debido al azar.

Ejercicio 1

Vamos a calcular el coeficiente de correlación de la siguiente serie de datos de altura y peso de los alumnos de una clase:

| Alumno | Estatura | Peso | Alumno | Estatura | Peso | Alumno | Estatura | Peso |
|----------|----------|------|-----------|----------|------|-----------|----------|------|
| Alumno 1 | 1,25 | 32 | Alumno 11 | 1,25 | 33 | Alumno 21 | 1,25 | 33 |
| Alumno 2 | 1,28 | 33 | Alumno 12 | 1,28 | 35 | Alumno 22 | 1,28 | 34 |
| Alumno 3 | 1,27 | 34 | Alumno 13 | 1,27 | 34 | Alumno 23 | 1,27 | 34 |
| Alumno 4 | 1,21 | 30 | Alumno 14 | 1,21 | 30 | Alumno 24 | 1,21 | 31 |
| Alumno 5 | 1,22 | 32 | Alumno 15 | 1,22 | 33 | Alumno 25 | 1,22 | 32 |
| Alumno 6 | 1,29 | 35 | Alumno 16 | 1,29 | 34 | Alumno 26 | 1,29 | 34 |
| Alumno 7 | 1,30 | 34 | Alumno 17 | 1,30 | 35 | Alumno 27 | 1,30 | 34 |
| Alumno 8 | 1,24 | 32 | Alumno 18 | 1,24 | 32 | Alumno 28 | 1,24 | 31 |
| Alumno 9 | 1,27 | 32 | Alumno 19 | 1,27 | 33 | Alumno 29 | 1,27 | 35 |
| Alumno10 | 1,29 | 35 | Alumno 20 | 1,29 | 33 | Alumno 30 | 1,29 | 34 |

Aplicamos la fórmula:

$$R = \frac{(1/30) * (0,826)}{\sqrt{((1/30) * (0,02568)) * ((1/30) * (51,366))}}^{(1/2)}$$

Luego, R=0.8281. Por lo tanto, la correlación existente entre estas dos variables es elevada (0,82) y de signo positivo.

Resultados con aplicando el software PASW:

Correlaciones

| | | Estatura | Peso |
|----------|------------------------|----------|--------|
| Estatura | Correlación de Pearson | 1 | ,828** |
| | Sig. (unilateral) | | ,000 |
| | N | 30 | 30 |
| Peso | Correlación de Pearson | ,828** | 1 |
| | Sig. (unilateral) | ,000 | |
| | N | 30 | 30 |

** . La correlación es significativa al nivel 0,01 (unilateral).

Ejercicio 2

Aprendiendo a realizar análisis de Correlación. Desde el menú *Analizar -> Correlación (Correlate)*.



Para realizar análisis con los procedimientos de Correlación (*Bivariadas, Parciales y Distancias*), se deben seguir los cuadros de diálogo.

A continuación, se presenta una explicación detallada del cuadro de diálogo correspondiente a los principales procedimientos que nos tocará analizar en este tema, como es la herramienta '**Correlación**'.

Los datos utilizados serán los recopilados en la encuesta del supermercado.

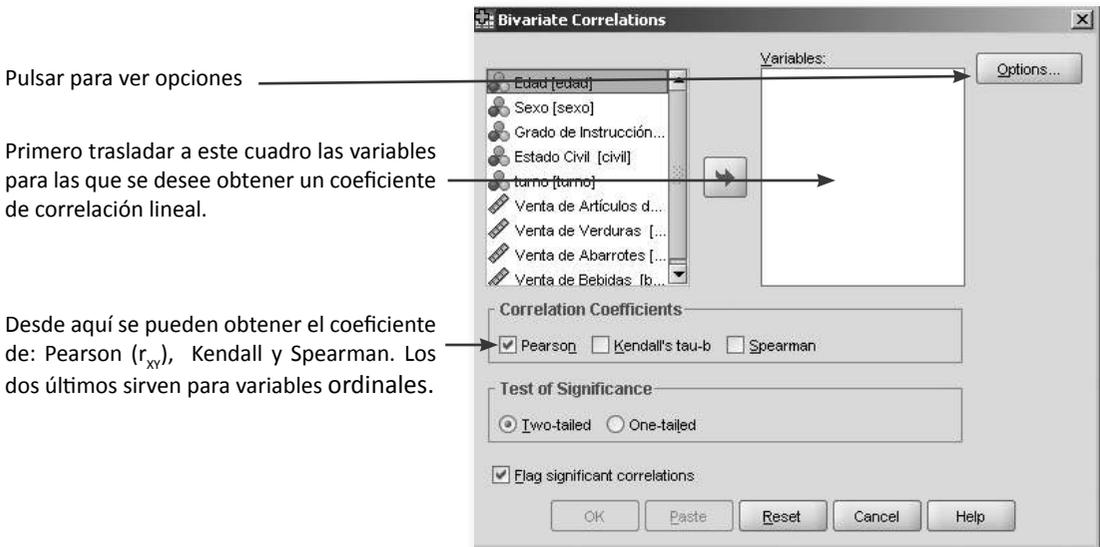
Bivariadas

Índice de asociación lineal

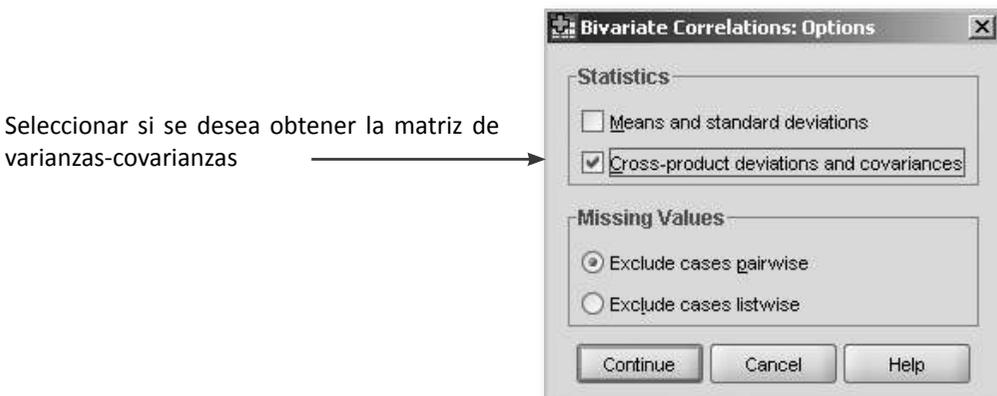
El índice de asociación lineal es medido por el procedimiento Bivariada.

Para obtener los índices de asociación lineal con variables (la covarianza, S_{xy} y el coeficiente de correlación de Pearson, r_{xy}), se utiliza el procedimiento bivariadas y se selecciona el menú **Analizar -> Correlaciones -> Bivariadas**.

El cuadro de diálogo para el procedimiento *Correlaciones bivariadas* tiene el siguiente aspecto:



Si pulsamos el botón '**Opciones**' aparece el cuadro de diálogo de la figura siguiente:



Desde aquí puede indicarse que se muestren los estadísticos descriptivos (media y desviación típica) para cada una de las variables seleccionadas y también la matriz de varianzas-covarianzas (selección productos cruzados y covarianzas).

La matriz de varianzas-covarianzas

Este procedimiento se consigue pulsando opciones, luego marcado varianzas-covarianzas (cross-product deviations and covariances).

En la covarianza pasa como en la varianza: el estadístico ofrecido es el insesgado. En este caso:

$$S'_{xy} = \frac{\sum_{i=1}^n x_i y_i}{n - 1}$$

La matriz de correlaciones:

Desde el cuadro de diálogo de correlaciones bivariadas se selecciona las variables, para las que se desea tener una correlación lineal.

Luego se pulsa en *Aceptar* y aparece la matriz de correlaciones de Pearson de todas las variables seleccionadas.

Veamos cómo se resuelve en el PASW alguno de los ejemplos.

En este ejemplo parecen los datos de 30 sujetos para dos variables; X: Venta de Verduras e Y: Venta de Abarrotes. Finalmente se pulsa en el botón *Aceptar*. Los resultados obtenidos son los siguientes:

Correlations

| | | Venta de Verduras | Venta de Abarrotes |
|--------------------|-----------------------------------|-------------------|--------------------|
| Venta de Verduras | Pearson Correlation | 1.000 | -.187 |
| | Sig. (2-tailed) | . | .323 |
| | Sum of Squares and Cross-products | 949.467 | -71.340 |
| | Covariance | 32.740 | -2.460 |
| | N | 30 | 30 |
| Venta de Abarrotes | Pearson Correlation | -.187 | 1.000 |
| | Sig. (2-tailed) | .323 | . |
| | Sum of Squares and Cross-products | -71.340 | 153.372 |
| | Covariance | -2.460 | 5.289 |
| | N | 30 | 30 |

Se ha señalado el coeficiente de Pearson con un círculo. Como se observa, se obtienen $r_{xy} = -0.187$.

Entonces la correlación entre x e y es igual a -0.187 (187%). Esto implica que la venta de verduras y la venta de abarroses tienen una baja correlación negativa.

La Media y la Desviación estándar se muestran en la tabla siguiente:

Descriptive Statistics

| | Mean | Std. Deviation | N |
|--------------------|---------|----------------|----|
| Venta de Verduras | 12.5333 | 5.7219 | 30 |
| Venta de Abarrotes | 34.5400 | 2.2997 | 30 |

Veamos otro ejemplo a partir de los datos de la encuesta. Tomemos las variables 'Venta de aseo personal', 'Venta de Verduras' y 'Venta de Bebidas' y obtengamos sus coeficientes de correlación. Los resultados obtenidos son los siguientes:

Correlations

| | | Venta de Artículos de Aseo Personal | Venta de Verduras | Venta de Bebidas |
|-------------------------------------|-----------------------------------|-------------------------------------|-------------------|------------------|
| Venta de Artículos de Aseo Personal | Pearson Correlation | 1.000 | -.208 | -.154 |
| | Sig. (2-tailed) | . | .271 | .418 |
| | Sum of Squares and Cross-products | 146.595 | -77.453 | -17.667 |
| | Covariance | 5.055 | -2.671 | -.609 |
| | N | 30 | 30 | 30 |
| Venta de Verduras | Pearson Correlation | -.208 | 1.000 | .182 |
| | Sig. (2-tailed) | .271 | . | .335 |
| | Sum of Squares and Cross-products | -77.453 | 949.467 | 53.333 |
| | Covariance | -2.671 | 32.740 | 1.839 |
| | N | 30 | 30 | 30 |
| Venta de Bebidas | Pearson Correlation | -.154 | .182 | 1.000 |
| | Sig. (2-tailed) | .418 | .335 | . |
| | Sum of Squares and Cross-products | -17.667 | 53.333 | 90.167 |
| | Covariance | -.609 | 1.839 | 3.109 |
| | N | 30 | 30 | 30 |

- El coeficiente de correlación de venta de artículos de aseo personal y la venta de verduras es -0.208
- El coeficiente de correlación de venta de artículos de aseo personal y la venta de bebidas es -0.154
- El coeficiente de correlación de la venta de verduras y la venta de bebidas es 0.182

Se observa que existe una baja correlación entre las variables.

Se han sombreado los tres coeficientes de Pearson. Como se observa, no se ha obtenido un valor razonable de correlación lineal. Es decir las correlaciones son muy bajas.

El cuadro siguiente muestra la media y la desviación estándar.

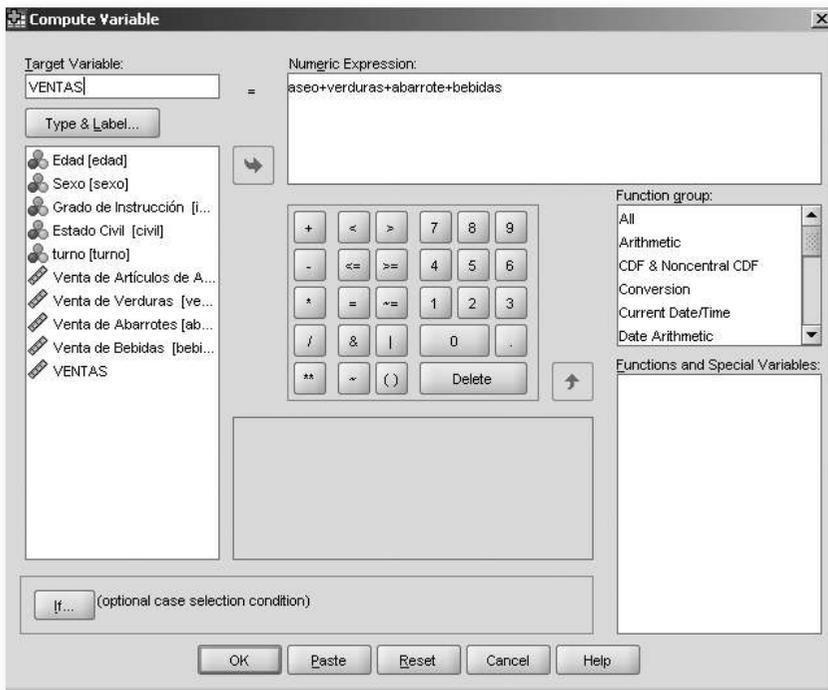
Descriptive Statistics

| | Mean | Std. Deviation | N |
|-------------------------------------|---------|----------------|----|
| Venta de Artículos de Aseo Personal | 2.6533 | 2.2483 | 30 |
| Venta de Verduras | 12.5333 | 5.7219 | 30 |
| Venta de Bebidas | 4.1667 | 1.7633 | 30 |

COMBINACIÓN LINEAL DE VARIABLES

En este apartado veremos cómo se realizan transformaciones lineales con los datos de dos o más variables para obtener nuevas puntuaciones (por ejemplo, $T_i = X_i + Y_i$; $T_i = AX_i + BY_i$; etc) y cómo se cumplen las propiedades sobre la media y la varianza en estas nuevas variables.

El primer paso que hay que dar es definir la nueva variable: T_i . Para ello se utiliza el menú *Transformar -> Calcular*. Lo primero es dar un nombre a la nueva variable y a continuación se indica la expresión numérica que la define. Con los datos de la encuesta del supermercado crearemos una combinación lineal con las cuatro variables de ventas (Útiles de aseo personal, Verduras, Abarrotes y Bebidas). A la variable resultante la llamaremos *Ventas*. El cuadro de diálogo de la *Transformar -> Calcular* queda del siguiente modo:



La nueva variable es del tipo $T_i = X_i + Y_i + V_i + W_i$. Obtenemos ahora sus estadísticos descriptivos univariados desde el menú *Analizar -> Estadísticos descriptivos -> Descriptivos* y las covarianzas desde *Analizar -> Correlaciones -> Bivariadas (Opciones)*. El resultado es el siguiente:

Descriptive Statistics

| | N | Mean | Variance |
|-------------------------------------|----|---------|----------|
| Venta de Artículos de Aseo Personal | 30 | 2.6533 | 5.055 |
| Venta de Verduras | 30 | 12.5333 | 32.740 |
| Venta de Abarrotes | 30 | 34.5400 | 5.289 |
| Venta de Bebidas | 30 | 4.1667 | 3.109 |
| VENTAS | 30 | 53.8933 | 37.361 |
| Valid N (listwise) | 30 | | |

Correlations

| | | Venta de Artículos de Aseo Personal | Venta de Verduras | Venta de Abarrotes | Venta de Bebidas |
|-------------------------------------|-----------------------------------|-------------------------------------|-------------------|--------------------|------------------|
| Venta de Artículos de Aseo Personal | Pearson Correlation | 1.000 | -.208 | .160 | -.154 |
| | Sig. (2-tailed) | . | .271 | .399 | .418 |
| | Sum of Squares and Cross-products | 146.595 | -77.453 | 23.966 | -17.667 |
| | Covariance | 5.055 | -2.671 | .826 | -.609 |
| | N | 30 | 30 | 30 | 30 |
| Venta de Verduras | Pearson Correlation | -.208 | 1.000 | -.187 | .182 |
| | Sig. (2-tailed) | .271 | . | .323 | .335 |
| | Sum of Squares and Cross-products | -77.453 | 949.467 | -71.340 | 53.333 |
| | Covariance | -2.671 | 32.740 | -2.460 | 1.839 |
| | N | 30 | 30 | 30 | 30 |
| Venta de Abarrotes | Pearson Correlation | .160 | -.187 | 1.000 | -.331 |
| | Sig. (2-tailed) | .399 | .323 | . | .074 |
| | Sum of Squares and Cross-products | 23.966 | -71.340 | 153.372 | -38.900 |
| | Covariance | .826 | -2.460 | 5.289 | -1.341 |
| | N | 30 | 30 | 30 | 30 |
| Venta de Bebidas | Pearson Correlation | -.154 | .182 | -.331 | 1.000 |
| | Sig. (2-tailed) | .418 | .335 | .074 | . |
| | Sum of Squares and Cross-products | -17.667 | 53.333 | -38.900 | 90.167 |
| | Covariance | -.609 | 1.839 | -1.341 | 3.109 |
| | N | 30 | 30 | 30 | 30 |

Como se observa la media en la nueva variable es 53.8933 valor que se corresponde exactamente con el resultado de la suma de las medias en las cuatro variables originales. En cuanto a la varianza el resultado es 37.361.

Ejercicio 1

Los datos para este ejercicio aparecen recopilados en las variables **Día1**, **Día2**, **Día3**, que indican el número de horas trabajadas en los tres días.

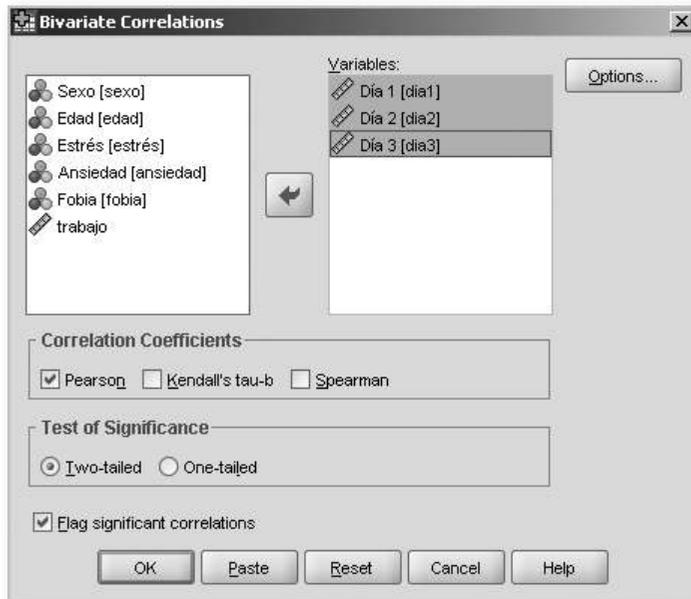
Además se han recopilado la variable **Sexo** que toma valores 0 para mujeres y 1 para hombres. **Edad**, 1 para Adolescentes, 2 para Jóvenes y 3 para Adultos.

Finalmente se incluyen tres datos relacionados con las terapias que ha recibido el sujeto. Todas ellas toman 1, si el sujeto ha recibido la terapia y 0 en caso contrario las variables son **Estrés**, **Ansiedad** y **Fobia**.

| Sujeto | Día1 | Día2 | Día3 | Sexo | Edad | Estrés | Ansiedad | Fobia |
|--------|------|------|------|------|------|--------|----------|-------|
| 1 | 5.7 | 6.3 | 6.8 | 0 | 1 | 0 | 1 | 0 |
| 2 | 4.8 | 5.4 | 4.3 | 1 | 3 | 0 | 0 | 0 |
| 3 | 7.6 | 7.9 | 8.6 | 0 | 2 | 0 | 0 | 0 |
| 4 | 5.7 | 6.0 | 6.7 | 1 | 1 | 0 | 0 | 1 |
| 5 | 3.8 | 4.6 | 4.9 | 1 | 3 | 0 | 0 | 0 |
| 6 | 7.5 | 8.0 | 8.5 | 1 | 2 | 0 | 1 | 0 |
| 7 | 6.4 | 7.0 | 7.3 | 1 | 1 | 1 | 1 | 0 |
| 8 | 7.7 | 8.1 | 8.5 | 0 | 2 | 0 | 0 | 0 |
| 9 | 4.0 | 4.5 | 5.0 | 0 | 3 | 1 | 0 | 0 |
| 10 | 5.7 | 6.1 | 6.8 | 1 | 2 | 1 | 0 | 1 |

Correlaciones

El procedimiento **Correlaciones** permite calcular la matriz de correlaciones de un conjunto de variables. Para calcular la correlación de Pearson se ejecuta el procedimiento: **Analizar > Correlaciones > Bivariadas**. Por ejemplo, para calcular las correlaciones entre los tres días se indica:



Pulsando en el botón *Opciones* pueden obtenerse los estadísticos descriptivos y la matriz de covarianzas.

Según puede verse en el cuadro de diálogo, este procedimiento también permite calcular la correlación de tau de Kendall y la de Spearman, que utilizan únicamente las propiedades ordinales de los datos.

Al pulsar en *Aceptar*, el PASW muestra la matriz de correlaciones entre las tres variables:

Descriptive Statistics

| | Mean | Std. Deviation | N |
|-------|--------|----------------|----|
| Día 1 | 5.8900 | 1.4255 | 10 |
| Día 2 | 6.3900 | 1.3404 | 10 |
| Día 3 | 6.7400 | 1.5756 | 10 |

Correlations

| | | Día 1 | Día 2 | Día 3 |
|-------|-----------------------------------|--------|--------|--------|
| Día 1 | Pearson Correlation | 1.000 | .996** | .953** |
| | Sig. (2-tailed) | . | .000 | .000 |
| | Sum of Squares and Cross-products | 18.289 | 17.119 | 19.264 |
| | Covariance | 2.032 | 1.902 | 2.140 |
| | N | 10 | 10 | 10 |
| Día 2 | Pearson Correlation | .996** | 1.000 | .946** |
| | Sig. (2-tailed) | .000 | . | .000 |
| | Sum of Squares and Cross-products | 17.119 | 16.169 | 17.984 |
| | Covariance | 1.902 | 1.797 | 1.998 |
| | N | 10 | 10 | 10 |
| Día 3 | Pearson Correlation | .953** | .946** | 1.000 |
| | Sig. (2-tailed) | .000 | .000 | . |
| | Sum of Squares and Cross-products | 19.264 | 17.984 | 22.344 |
| | Covariance | 2.140 | 1.998 | 2.483 |
| | N | 10 | 10 | 10 |

** . Correlation is significant at the 0.01 level (2-tailed).

El coeficiente de correlación de Pearson de Día 1 y Día 1 es = 1

El coeficiente de correlación de Pearson de Día 1 y Día 2 es = 0.996

El coeficiente de correlación de Pearson de Día 1 y Día 3 es = 0.953

El coeficiente de correlación de Pearson de Día 2 y Día 2 es = 1

El coeficiente de correlación de Pearson de Día 2 y Día 3 es = 0.946

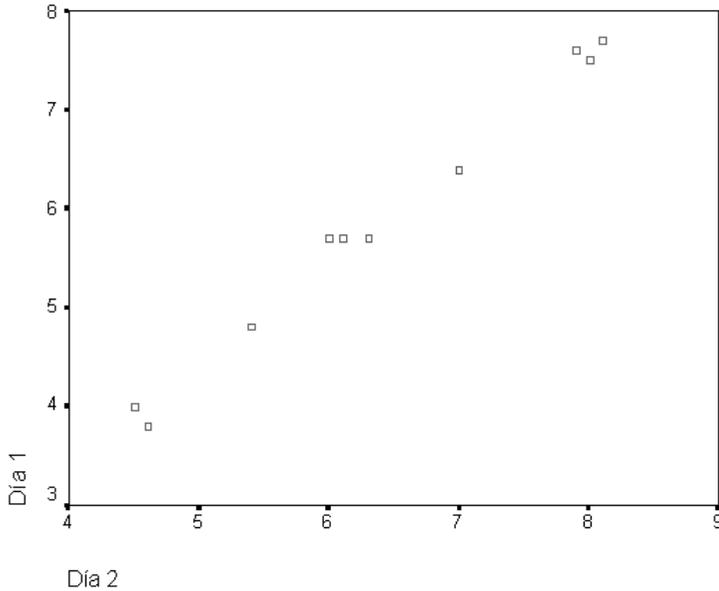
El coeficiente de correlación de Pearson de Día 3 y Día 3 es = 1

Además de los coeficientes de correlación de Pearson (r_{ij}) para cada combinación de las variables incluidas, la tabla incluye el nivel crítico que permite tomar una decisión respecto a cada hipótesis nula : $H_0 : \rho_{ij} = 0.000$. La última fila de la tabla contiene el número de casos (10).

Correlación en Gráficos

Las correlaciones lineales permiten evaluar la existencia de relaciones lineales entre variables. Sin necesidad de calcular ningún estadístico, es posible elaborar un gráfico de dispersión que informe sobre la posibilidad de que se dé este tipo de relación.

Por ejemplo, mediante el procedimiento: **Gráficos > Dispersión > Simples** es posible elaborar el siguiente gráfico de dispersión de las variables número de horas trabajadas durante los **Día1** y **Día2**.



Cada punto del gráfico se corresponde con las puntuaciones de un sujeto en ambas variables. El gráfico sugiere que existe una relación lineal directa (o positiva) entre ambas, de modo que los sujetos que trabajan más el primer día también lo hacen el segundo día. Para comprobar si esta relación es estadísticamente significativa se utilizan los procedimientos **Correlaciones**.

Capítulo

9

Regresión

Análisis de Resultados y Conclusiones: Muestra los principales hallazgos de la investigación aplicando Regresión (lineal simple y múltiple), y presenta una potente interpretación teórica que demuestra el dominio técnico del investigador. Las **conclusiones** extraen lo esencial de todo el proceso enfatizando especialmente la riqueza de la evidencia empírica aportada.



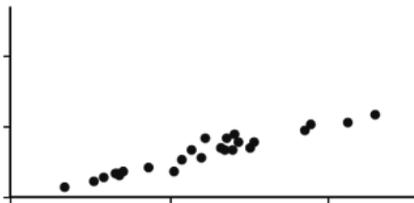
REGRESIÓN LINEAL SIMPLE

El análisis de regresión lineal es una técnica estadística que se utiliza para estudiar la relación entre variables.

En la investigación estadística suele emplearse para pronosticar valores de una variable criterio (*Y: Dependiente*) desde las puntuaciones de una variable predictora (*X: Independiente*).

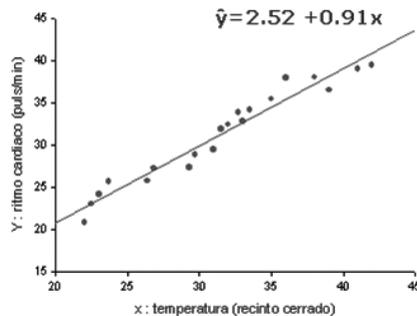
Por ejemplo, representamos en un gráfico los pares de valores de una distribución bidimensional: la variable “x” en el eje horizontal o eje de abscisa, y la variable “y” en el eje vertical o eje de ordenada.

Vemos que la nube de puntos sigue una tendencia lineal:



El **coeficiente de correlación lineal** nos permite determinar si, efectivamente, existe relación entre las dos variables.

Una vez que se concluye que sí existe relación, la **regresión** nos permite definir la recta que mejor se ajusta a esta nube de puntos.



Una recta viene definida por la siguiente fórmula: $Y = a + bX$

Donde:

“y” sería la variable dependiente, es decir, aquella que viene definida a partir de la otra variable.

“x” (variable independiente).

Para definir la recta hay que determinar los valores de los parámetros “a” y “b”:

El **parámetro “a”** es el valor que toma la variable dependiente “y” cuando la variable independiente “x” vale 0, y es el punto donde la recta cruza el eje vertical.

El **parámetro "b"** determina la pendiente de la recta, su grado de inclinación.

La regresión lineal nos permite calcular el valor de estos dos parámetros, definiendo la recta que mejor se ajusta a esta nube de puntos.

El **parámetro "b"** viene determinado por la siguiente fórmula:

$$b = \frac{1/n * \sum (x_i - \bar{x}_m) * (y_i - \bar{y}_m)}{1/n * \sum (x_i - \bar{x}_m)^2}$$

Es la covarianza de las dos variables, dividida por la varianza de la variable "x".

El **parámetro "a"** viene determinado por: $a = \bar{y}_m - (b * \bar{x}_m)$. Es la media de la variable "Y", menos la media de la variable "X" multiplicada por el parámetro "b" que hemos calculado.

Ejemplo 1

Vamos a calcular la recta de regresión de la siguiente serie de datos de altura y peso de los alumnos de una clase. Vamos a considerar que la altura es la variable independiente "X" y que el peso es la variable dependiente "Y" (podríamos hacerlo también al contrario):

| Alumno | Estatura | Peso | Alumno | Estatura | Peso | Alumno | Estatura | Peso |
|-----------|----------|------|-----------|----------|------|-----------|----------|------|
| Alumno 1 | 1,25 | 32 | Alumno 11 | 1,25 | 33 | Alumno 21 | 1,25 | 33 |
| Alumno 2 | 1,28 | 33 | Alumno 12 | 1,28 | 35 | Alumno 22 | 1,28 | 34 |
| Alumno 3 | 1,27 | 34 | Alumno 13 | 1,27 | 34 | Alumno 23 | 1,27 | 34 |
| Alumno 4 | 1,21 | 30 | Alumno 14 | 1,21 | 30 | Alumno 24 | 1,21 | 31 |
| Alumno 5 | 1,22 | 32 | Alumno 15 | 1,22 | 33 | Alumno 25 | 1,22 | 32 |
| Alumno 6 | 1,29 | 35 | Alumno 16 | 1,29 | 34 | Alumno 26 | 1,29 | 34 |
| Alumno 7 | 1,30 | 34 | Alumno 17 | 1,30 | 35 | Alumno 27 | 1,30 | 34 |
| Alumno 8 | 1,24 | 32 | Alumno 18 | 1,24 | 32 | Alumno 28 | 1,24 | 31 |
| Alumno 9 | 1,27 | 32 | Alumno 19 | 1,27 | 33 | Alumno 29 | 1,27 | 35 |
| Alumno 10 | 1,29 | 35 | Alumno 20 | 1,29 | 33 | Alumno 30 | 1,29 | 34 |

El parámetro "b" viene determinado por:

$$b = (1/30) * 1,034 / ((1/30) * 0,00856) = 40,265$$

Y el parámetro "a" por: $a = 33,1 - (40,265 * 1,262) = -17,714$

Por lo tanto, la **recta** que mejor se ajusta a esta serie de datos es:

$$Y = -17,714 + (40,265 * X)$$

Esta recta define un valor de la variable dependiente (peso), para cada valor de la variable independiente (estatura):

| Estatura | Peso |
|----------|------|
| X | Y |
| 1,20 | 30,6 |
| 1,21 | 31,0 |
| 1,22 | 31,4 |
| 1,23 | 31,8 |
| 1,24 | 32,2 |
| 1,25 | 32,6 |
| 1,26 | 33,0 |
| 1,27 | 33,4 |
| 1,28 | 33,8 |
| 1,29 | 34,2 |
| 1,30 | 34,6 |

Resultados con el software PASW

Resumen del modelo

| Modelo | R | R cuadrado | R cuadrado corregida | Error típ. de la estimación |
|--------|-------------------|------------|----------------------|-----------------------------|
| 1 | ,828 ^a | ,686 | ,675 | ,82519 |

a. Variables predictoras: (Constante), Estatura

Coefficientes^a

| Modelo | | Coefficients no estandarizados | | Coefficients tipificados | t | Sig. |
|--------|-------------|--------------------------------|------------|--------------------------|--------|------|
| | | B | Error típ. | Beta | | |
| 1 | (Constante) | -17,714 | 6,500 | | -2,725 | ,011 |
| | Estatura | 40,265 | 5,149 | ,828 | 7,819 | ,000 |

a. Variable dependiente: Peso

ANOVA^b

| Modelo | | Suma de cuadrados | gl | Media cuadrática | F | Sig. |
|--------|-----------|-------------------|----|------------------|--------|-------------------|
| 1 | Regresión | 41,634 | 1 | 41,634 | 61,142 | ,000 ^a |
| | Residual | 19,066 | 28 | ,681 | | |
| | Total | 60,700 | 29 | | | |

a. Variables predictoras: (Constante), Estatura

De los cuadros de resultados se tiene:

El **parámetro “b”** viene determinado por estatura y su valor es:
b = 40,265

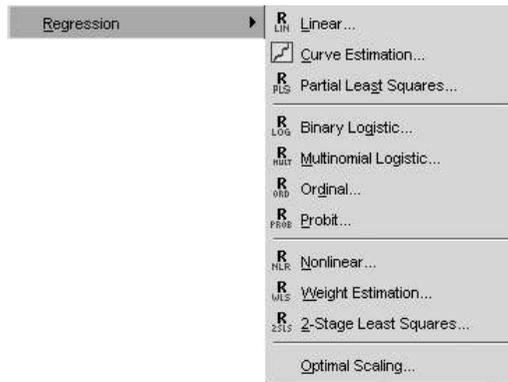
Y el **parámetro “a”** es la constante y su valor es: **a = -17,714**

Por lo tanto, la **recta** que mejor se ajusta a esta serie de datos es:

$$Y = -17,714 + (40,265 * X)$$

Esta recta define un valor de la variable dependiente (peso), para cada valor de la variable independiente (estatura):

Aprendiendo a realizar análisis de Regresión con el software PASW. Desde el menú *Analizar -> Regresión (Regression)*.



ANÁLISIS DE REGRESIÓN LINEAL SIMPLE

La regresión puede ser lineal con una sola variable predictora (regresión simple), o múltiple ya que puede efectuarse a partir de 2 o más variables predictoras (X1, X2, etcétera).

DIAGRAMA DE DISPERSIÓN

Herramienta que permite explicar con claridad un modelo.

La primera aproximación a la relación entre dos variables (X e Y) puede hacerse a partir de un diagrama de dispersión.

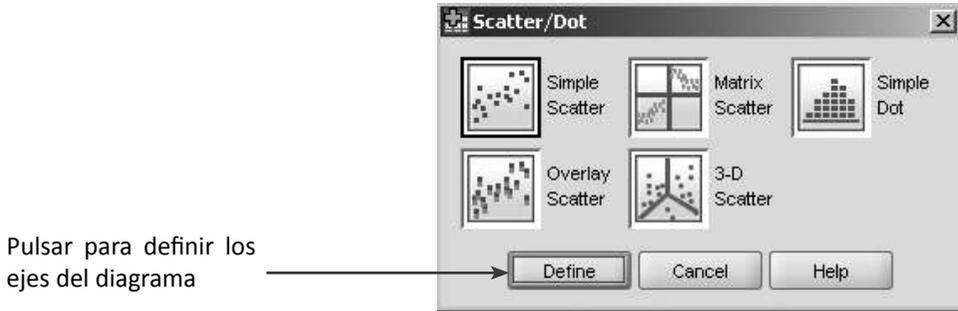
Por ejemplo, supongamos que tenemos los siguientes datos:

| Sujetos | Edad | Peso |
|---------|------|------|
| 1 | 39 | 58 |
| 2 | 30 | 55 |
| 3 | 50 | 65 |
| 4 | 52 | 70 |
| 5 | 62 | 69 |

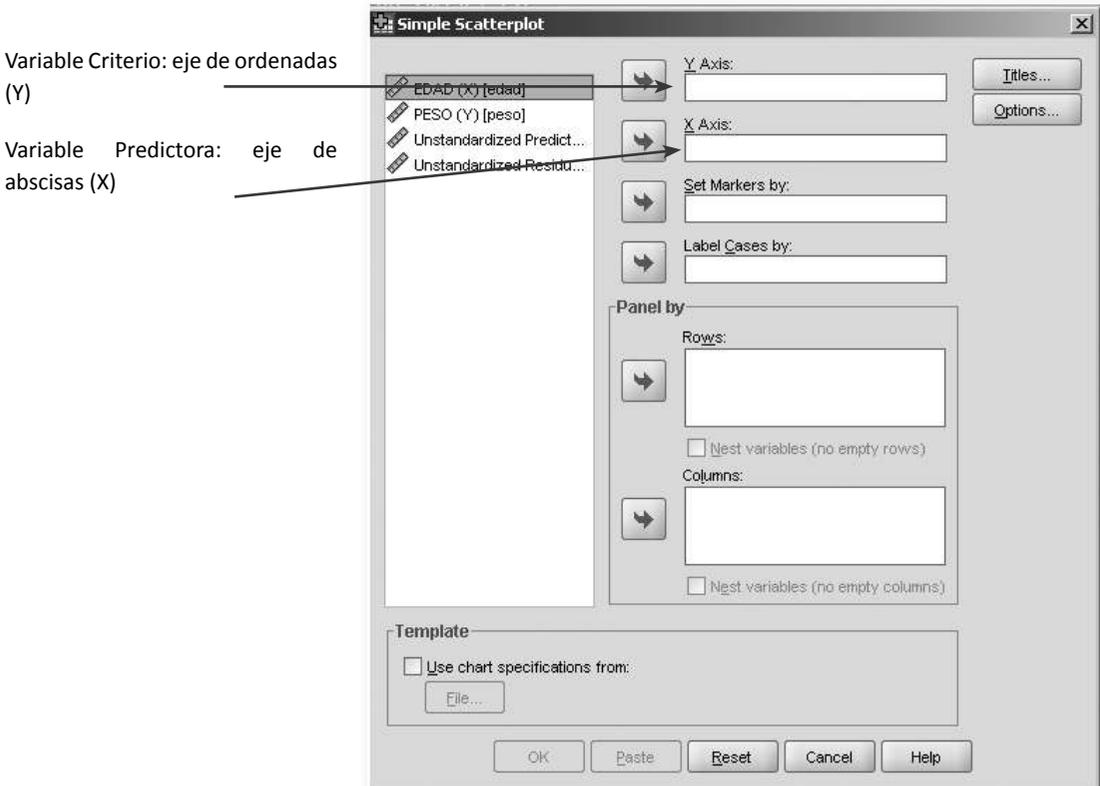
Los datos corresponden a 5 sujetos medidos en dos variables: Edad y Peso.

Como en otros ejemplos previos, tendríamos que introducir los datos correspondientes a los 5 sujetos en las dos variables en un archivo nuevo del editor de datos. Una vez introducidos los datos guardar como *regresion.sav*.

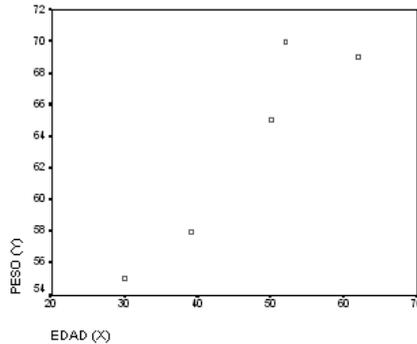
Para realizar el diagrama de dispersión se selecciona el menú *Gráficos* de la barra de menús del *Editor de datos* y se elige el procedimiento 'Dispersión' (Scatter...). El cuadro de diálogo correspondiente es el siguiente:



Luego pulsar definir para definir los ejes del diagrama de las variables, obteniendo el cuadro de diálogo siguiente:



Con esta definición se obtiene el siguiente diagrama de dispersión simple en el visor de resultados:



Como se observa, la forma de este diagrama indica que los puntos en él no están perfectamente alineados pero se acercan a una hipotética línea recta.

ECUACIÓN DE REGRESIÓN

Podrían trazarse diferentes rectas para realizar pronósticos de una variable a partir de la otra (por ejemplo de Y a partir de X en una regresión de Y sobre X). Las rectas de regresión tienen una fórmula muy simple:

$$Y'_i = A_{YX} + B_{YX}X_i$$

El objetivo es encontrar aquella recta que minimice la distancia entre lo encontrado (Y) y lo pronosticado (Y'). Es decir, que minimice la expresión:

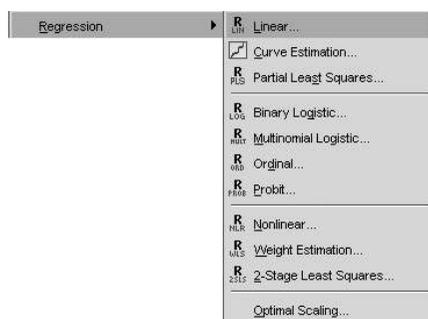
$$\sum \frac{(Y_i - Y'_i)^2}{n}$$

Para ello, calculamos los coeficientes del modelo mediante:

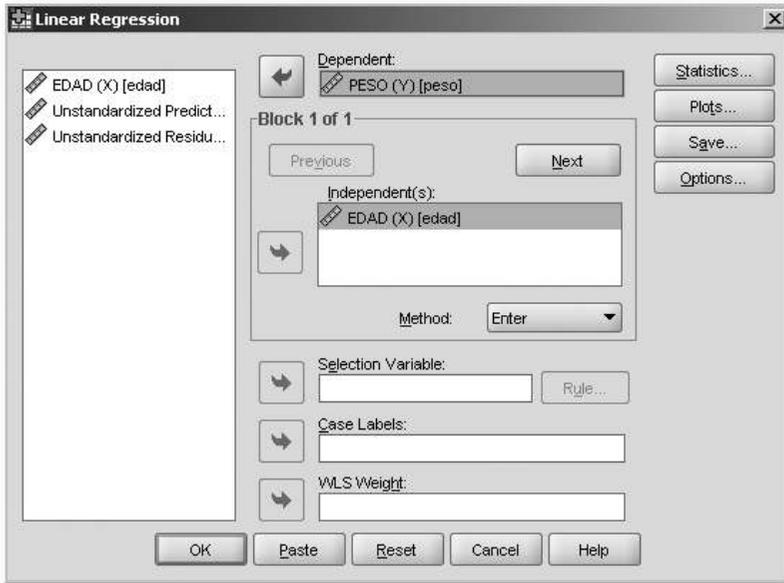
$$B_{YX} = \frac{n \cdot \sum X_i \cdot Y_i - \sum X_i \cdot \sum Y_i}{n \cdot \sum X_i^2 - (\sum X_i)^2} \quad A_{YX} = \bar{Y} - B_{YX} \bar{X}$$

Ahora veamos cómo se procede en el PASW para calcular dichos coeficientes, obtener la recta de regresión y valorar la bondad del modelo.

Se selecciona el menú **Analizar -> Regresión -> Lineal**: como se observa en la figura siguiente:

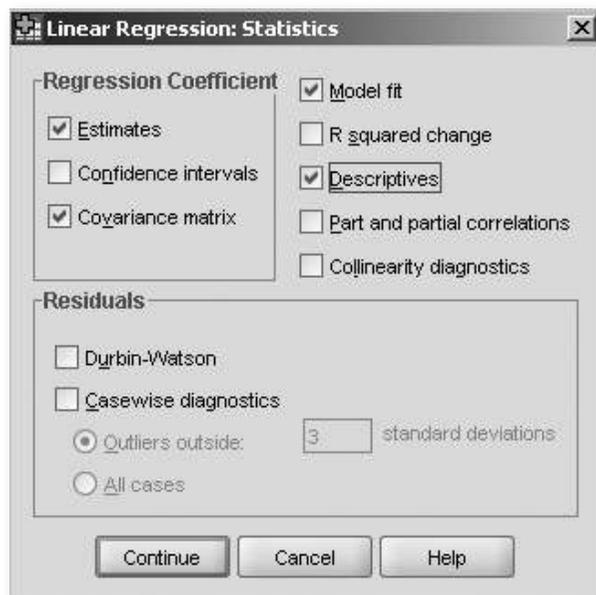


El cuadro de diálogo correspondiente al procedimiento *Regresión lineal* aparece en la figura siguiente:



Como en otros cuadros de diálogo del PASW, lo primero es seleccionar las variables. En este caso hay que distinguir entre 'Dependiente' e 'Independiente' (o independientes si se trata de una regresión múltiple).

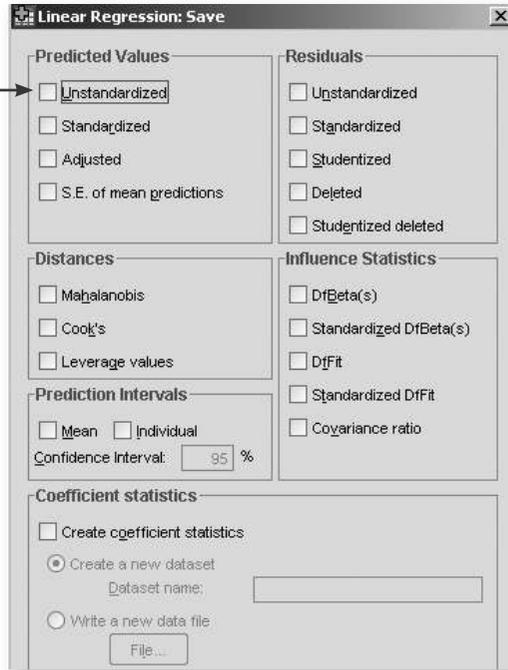
Dentro de este menú hay otras opciones. Por el momento, la que nos interesa es la que se encuentra en el botón '*Estadísticos*', cuyo cuadro de diálogo aparece en la figura:



Desde aquí podemos solicitar que se ofrezcan las estimaciones de la pendiente y el origen de la recta de regresión (según el criterio de mínimos cuadrados), la matriz de covarianzas para las variables, el ajuste del modelo (coeficiente de Pearson al cuadrado o *coeficiente de determinación*) y los estadísticos descriptivos (media y varianza):

Otra opción relevante del menú regresión lineal es la de 'Guardar'. Desde su correspondiente cuadro de diálogo es posible indicar que se guarden los valores pronosticados por el modelo (las Y_i) y los residuos (las $Y_i - Y'_i$) en el editor de datos:

Seleccionar si se desea guardar los valores pronosticados por el modelo (las Y_i) para cada uno de los sujetos.



Veamos cuál es el resultado para el ejemplo. Si indicamos todas estas instrucciones, el resultado que ofrece el PASW es el siguiente (nótese que se ofrecen muchas tablas de resultados. Aquí explicaremos sólo las más importantes.

Regression

Descriptive Statistics

| | Mean | Std. Deviation | N |
|----------|---------|----------------|---|
| PESO (Y) | 63.4000 | 6.6558 | 5 |
| EDAD (X) | 46.6000 | 12.3612 | 5 |

Correlations

| | | PESO (Y) | EDAD (X) |
|---------------------|----------|----------|----------|
| Pearson Correlation | PESO (Y) | 1.000 | .935 |
| | EDAD (X) | .935 | 1.000 |
| Sig. (1-tailed) | PESO (Y) | . | .010 |
| | EDAD (X) | .010 | . |
| N | PESO (Y) | 5 | 5 |
| | EDAD (X) | 5 | 5 |

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .935 ^a | .875 | .833 | 2.7198 |

Esto es el coeficiente de determinación, r^{2XY} o la proporción de varianza en común entre X e Y.

a. Predictors: (Constant), EDAD (X)

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|--------|-------------------|
| 1 | Regression | 155.008 | 1 | 155.008 | 20.954 | .020 ^a |
| | Residual | 22.192 | 3 | 7.397 | | |
| | Total | 177.200 | 4 | | | |

a. Predictors: (Constant), EDAD (X)

b. Dependent Variable: PESO (Y)

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 39.932 | 5.269 | | 7.579 | .005 |
| | EDAD (X) | .504 | .110 | .935 | 4.578 | .020 |

a. Dependent Variable: PESO (Y)

En la Columna B. Los coeficientes a y b del modelo pronosticado en directas se ven en esta columna: siendo a (Constant) = 39,932 y b (Edad x)= 0,504.

Luego la recta hallada es $Y' = 39.932 + 0,504 X$

En la columna Beta. Aparece el modelo en típicas: $Z_y = r_{xy} Z_x$ (en el ejemplo: $Z_y = 0.935 Z_x$).

Como se observa, el PASW ofrece muchas tablas de datos como resultado de la regresión. Por el momento solamente nos fijaremos en dos: La primera se refiere a los coeficientes del modelo y la segunda a su bondad.

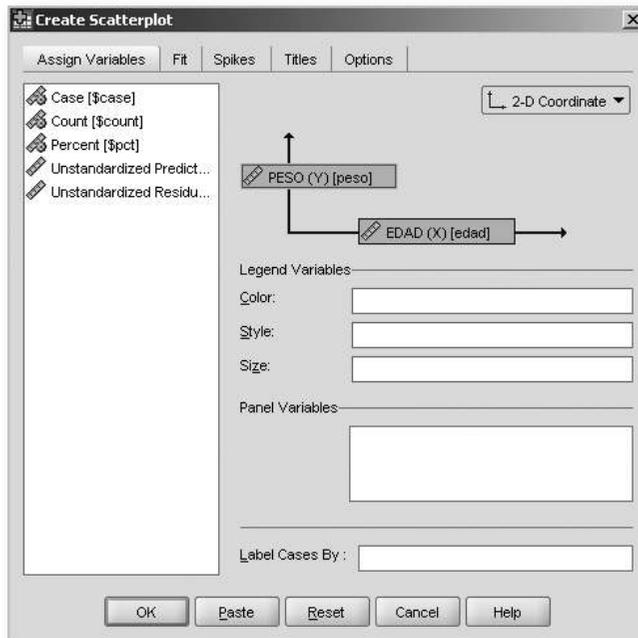
En cuanto a la primera, se toman los coeficientes no estandarizados. En este caso el mejor modelo para pronosticar Y_i a partir de X_i es $Y_i = 39.932 + 0,504 X_i$

BONDAD DE AJUSTE DEL MODELO

Además de la fórmula de la recta de regresión, resulta necesario disponer de información sobre el grado en que el modelo se ajusta a los datos observados (nube de puntos).

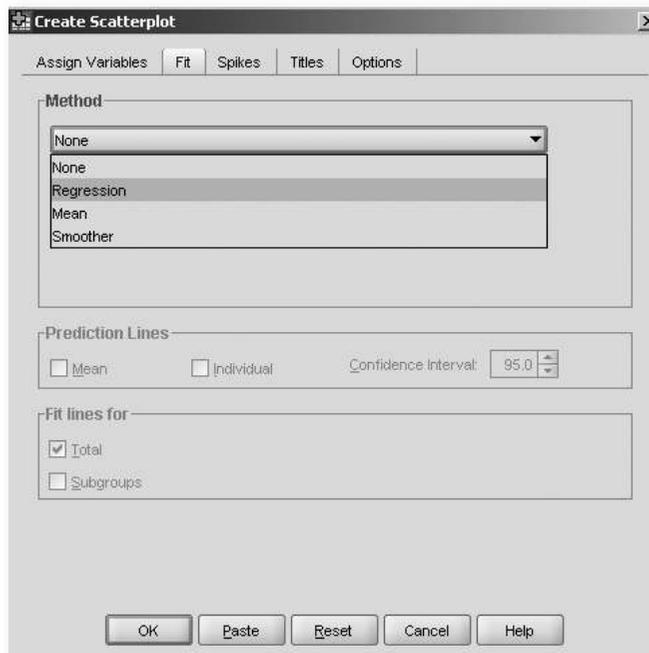
Una primera aproximación es la interpretación gráfica del problema. Para elaborar la gráfica del ajuste de la recta a los datos observados se pulsa el menú Gráficos -> Interactivos -> Diagramas de dispersión.

Luego el cuadro de diálogo de Diagramas de dispersión interactivos es el siguiente:

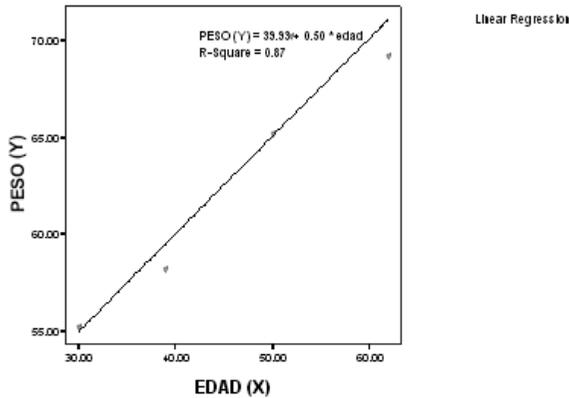


Lo primero es definir la variable del criterio (en el eje de ordenadas) y la variable predictora (en el eje de abscisas) desde la solapa 'Asignar variables'.

A continuación se selecciona el método 'regresión' desde la solapa 'Ajuste'(Fit). El cuadro de diálogo desde el que se hacen estas selecciones tiene el siguiente aspecto:



El gráfico obtenido es el siguiente:



Como se observa, los puntos no se alejan bastante de la recta, luego el ajuste es aceptable. Además de la interpretación gráfica del problema, la forma de cuantificar la bondad del modelo es mediante el coeficiente de determinación, r^2_{XY} .

Se trata de una medida estandarizada que toma valores entre 0 y 1 y cuya interpretación es muy sencilla: representa la proporción de varianza explicada de la variable del criterio a partir de la predictora. En nuestro ejemplo, $r^2_{XY} = 0,875$ por lo que el modelo **SI** es adecuado para hacer pronósticos de Y a partir de X.

La bondad del modelo también puede valorarse a partir del análisis de los errores en los pronósticos, frecuentemente llamados residuos ($Y_i - Y'_i$). Nótese que ejecutando las órdenes indicadas en el anterior apartado han resultado dos nuevas variables en el editor de datos (la variable 'pre_1' que se corresponde con las Y'_i y lleva la etiqueta 'Unstandardized predicted value' y la variable 'res_1' que se refiere a las $Y_i - Y'_i$ y lleva la etiqueta 'Unstandardized residual').

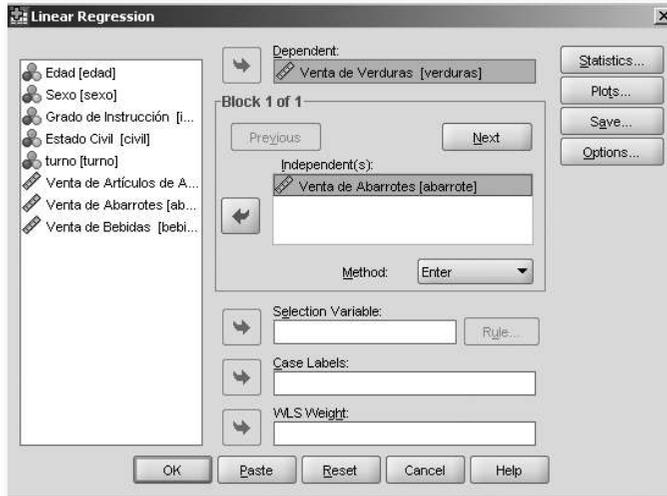
Teniendo los datos para Y_i , Y'_i e $(Y_i - Y'_i)$ podemos evaluar la bondad del modelo a partir de la descomposición de la varianza del criterio ($S^2_Y = S^2_{Y'} + S^2_{Y-Y'}$). Para ello entramos en el menú Analizar -> Estadísticos descriptivos -> Descriptivos: Opciones e indicamos en el cuadro de diálogo que se calcule la varianza para estas tres variables (Y, pre_1 y res_1). Los resultados obtenidos en el visor son los siguientes:

Descriptive Statistics

| | N | Variance |
|--------------------------------|---|----------|
| PESO (Y) | 5 | 44.300 |
| Unstandardized Predicted Value | 5 | 38.752 |
| Unstandardized Residual | 5 | 5.548 |
| Valid N (listwise) | 5 | |

Como se comprueba, la varianza del criterio ($S^2_Y = 44.3$) se descompone en la varianza de los pronósticos ($S^2_{Y'} = 38.752$) y la de los errores ($S^2_{Y-Y'} = 5.548$).

Veamos ahora un ejemplo a partir de los datos de la encuesta. Supóngase que queremos predecir la variable 'Venta de verduras (V)' a partir de la variable 'Venta de abarrotes (A)'. Para ello, construimos la recta de regresión $V_i = a + b A_i$. La definición del modelo en el PASW es la siguiente:



Los resultados obtenidos para el modelo planteado son los siguientes:

Variables Entered/Removed^a

| Model | Variables Entered | Variables Removed | Method |
|-------|---------------------------------|-------------------|--------|
| 1 | Venta de Abarrotes ^a | . | Enter |

a. All requested variables entered.

b. Dependent Variable: Venta de Verduras

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .187 ^a | .035 | .000 | 5.7205 |

a. Predictors: (Constant), Venta de Abarrotes

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|-------|-------------------|
| 1 | Regression | 33.183 | 1 | 33.183 | 1.014 | .323 ^a |
| | Residual | 916.283 | 28 | 32.724 | | |
| | Total | 949.467 | 29 | | | |

a. Predictors: (Constant), Venta de Abarrotes

b. Dependent Variable: Venta de Verduras

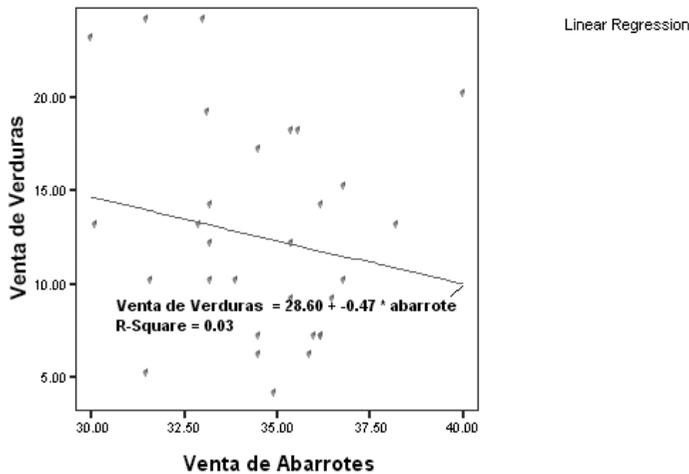
Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|--------------------|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 28.599 | 15.989 | | 1.789 | .084 |
| | Venta de Abarrotes | -.465 | .462 | -.187 | -1.007 | .323 |

a. Dependent Variable: Venta de Verduras

El modelo resultante es: $V_i = 28.599 - 0.465A_i$. Como se observa, el coeficiente de determinación (R al cuadrado) es 0,035 por lo que el modelo no es adecuado para explicar la relación entre la variable ‘Venta de Verduras’ y la variable ‘Venta de Abarrotes’. Es decir, la variable ‘Venta de Abarrotes’ no tiene una capacidad predictiva para explicar la variable del criterio (Venta de Verduras).

El gráfico obtenido definiendo las opciones del diagrama de dispersión interactivo de la figura es el siguiente:



ANÁLISIS DE REGRESIÓN LINEAL MÚLTIPLE

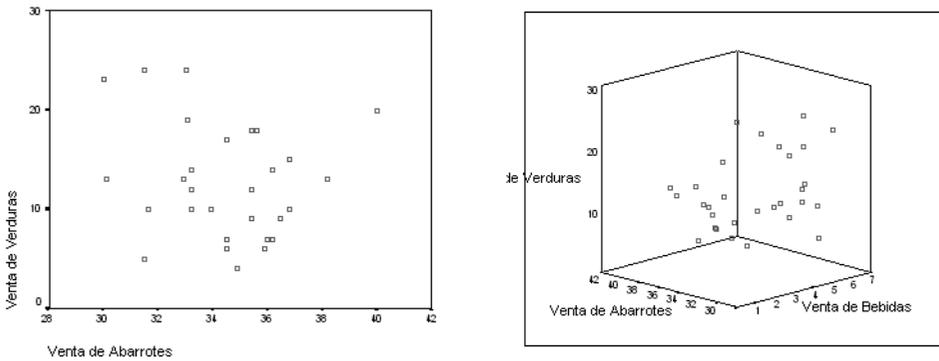
Una alternativa a esta situación de poco ajuste es incluir una variable predictora adicional (o más de una) y observar si su inclusión produce algún cambio en R cuadrado. Si consideramos dos variables predictoras estamos planteando el siguiente modelo:

$$Y'_i = a + b_1 X_{i1} + b_2 X_{i2}$$

Retomemos el ejemplo de la ‘Venta de Abarrotes’ y ‘Venta de Verduras’, considerando ahora la ‘Venta de Bebidas (B)’ como segunda variable predictora. El modelo planteado es:

$$V_i = a + b_1 A_i + b_2 B_i$$

La primera aproximación al problema puede hacerse a partir del gráfico de dispersión. En este caso puede elaborarse el gráfico simple para observar la relación entre la edad y el peso, así como también un gráfico 3-D para observar la relación entre las dos predictoras y el criterio. El resultado obtenido es:



Para obtener los coeficientes del modelo y valorar el ajuste del modelo se selecciona el menú *Analizar -> Regresión -> Lineal*. Lo primero es trasladar las variables 'Venta de Abarrotes' y 'Venta de Bebidas' a la casilla *Independientes*, así como la variable 'Venta de Verduras' a la casilla *Dependiente*. Los resultados encontrados son muy similares a los de la regresión simple, aunque ahora hay un coeficiente más que interpretar:

Variables Entered/Removed^a

| Model | Variables Entered | Variables Removed | Method |
|-------|---|-------------------|--------|
| 1 | Venta de Bebidas , Venta de Abarrotes ^a | | Enter |

a. All requested variables entered.

b. Dependent Variable: Venta de Verduras

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .226 ^a | .051 | -.019 | 5.7761 |

a. Predictors: (Constant), Venta de Bebidas , Venta de Abarrotes

ANOVA^a

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|------|-------------------|
| 1 | Regression | 48.648 | 2 | 24.324 | .729 | .492 ^a |
| | Residual | 900.819 | 27 | 33.364 | | |
| | Total | 949.467 | 29 | | | |

a. Predictors: (Constant), Venta de Bebidas , Venta de Abarrotes

b. Dependent Variable: Venta de Verduras

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|--------------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 22.926 | 18.168 | | 1.262 | .218 |
| | Venta de Abarrotes | -.354 | .494 | -.142 | -.716 | .480 |
| | Venta de Bebidas | .439 | .645 | .135 | .681 | .502 |

a. Dependent Variable: Venta de Verduras

Como se observa, el modelo resultante

$$V_i = 22.926 - 0.354 A_i + 0.439 B_i$$

No aporta casi nada a la explicación de la variable del criterio.

Ejercicio 1

El presente ejercicio se trabaja con la siguiente información:

Los datos aparecen recopilados en las variables **Día1**, **Día2**, **Día3**, que indican el número de horas trabajadas en los tres días. Además se han recopilado la variable **Sexo** que toma valores 0 para mujeres y 1 para hombres. **Edad**, 1 para Adolescentes, 2 para Jóvenes y 3 para Adultos.

Finalmente se incluyen tres datos relacionados con las terapias que ha recibido el sujeto. Todas ellas toman 1, si el sujeto ha recibido la terapia y 0 en caso contrario; las variables son **Estrés**, **Ansiedad** y **Fobia**.

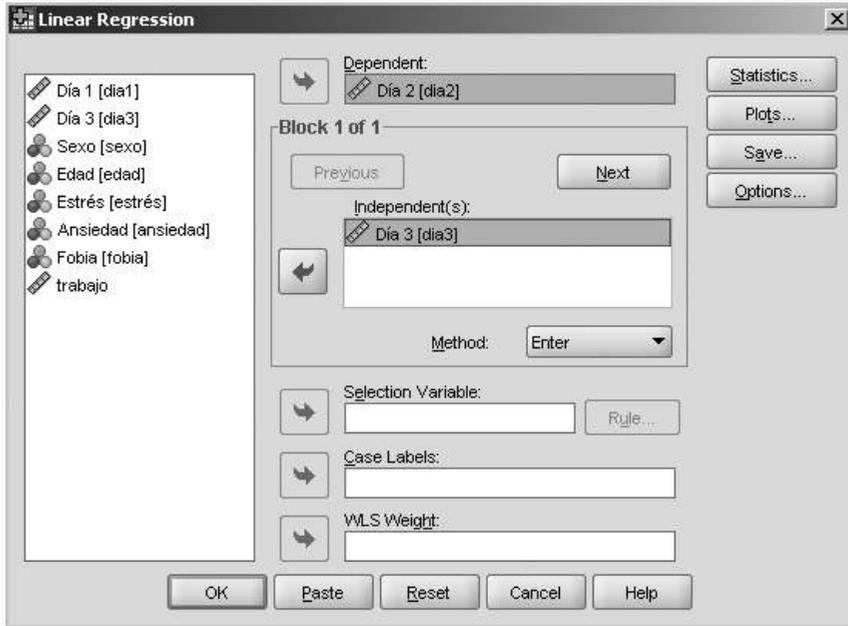
| Sujeto | Día1 | Día2 | Día3 | Sexo | Edad | Estrés | Ansiedad | Fobia |
|--------|------|------|------|------|------|--------|----------|-------|
| 1 | 5.7 | 6.3 | 6.8 | 0 | 1 | 0 | 1 | 0 |
| 2 | 4.8 | 5.4 | 4.3 | 1 | 3 | 0 | 0 | 0 |
| 3 | 7.6 | 7.9 | 8.6 | 0 | 2 | 0 | 0 | 0 |
| 4 | 5.7 | 6.0 | 6.7 | 1 | 1 | 0 | 0 | 1 |
| 5 | 3.8 | 4.6 | 4.9 | 1 | 3 | 0 | 0 | 0 |
| 6 | 7.5 | 8.0 | 8.5 | 1 | 2 | 0 | 1 | 0 |
| 7 | 6.4 | 7.0 | 7.3 | 1 | 1 | 1 | 1 | 0 |
| 8 | 7.7 | 8.1 | 8.5 | 0 | 2 | 0 | 0 | 0 |
| 9 | 4.0 | 4.5 | 5.0 | 0 | 3 | 1 | 0 | 0 |
| 10 | 5.7 | 6.1 | 6.8 | 1 | 2 | 1 | 0 | 1 |

Regresión

El procedimiento **Analizar > Regresión lineal** permite obtener la regresión lineal simple y múltiple. Por ejemplo, para estimar los parámetros del modelo:

$$Día2_i = a + b Día3_i + E_i$$

Se indica:



Los resultados muestran el coeficiente de determinación (*R cuadrado*) en la tabla *Resumen del modelo*. Se observa que el modelo estimado para *día 2* explica el 89.50% de la varianza. La tabla de regresión en formato ANOVA muestra la descomposición de la suma de cuadrados total en la parte debida a la regresión y el error:

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .946 ^a | .895 | .882 | .4602 |

a. Predictors: (Constant), Día 3

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|--------|-------------------|
| 1 | Regression | 14.475 | 1 | 14.475 | 68.349 | .000 ^a |
| | Residual | 1.694 | 8 | .212 | | |
| | Total | 16.169 | 9 | | | |

a. Predictors: (Constant), Día 3

b. Dependent Variable: Día 2

En el ejemplo, el nivel crítico asociado a la F del ANOVA permite rechazar la hipótesis nula de que no existe relación lineal entre las variables ($H_0 : \beta = 0$).

A continuación aparecen los valores estimados para los parámetros:

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|------------|-----------------------------|------------|---------------------------|-------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | .965 | .672 | | 1.436 | .189 |
| | Día 3 | .805 | .097 | .946 | 8.267 | .000 |

a. Dependent Variable: Día 2

Estadístico de contraste T y nivel crítico para $H_0 : \beta = 0$

La tabla de coeficientes muestra los términos de la ecuación de regresión.

El término denominado Constante es el origen A de la ecuación (estimador de α). La pendiente B (estimador de β) aparece en la fila de la tabla indicada por día 3.

Los valores de a y b se muestran en la primera columna:

$$\text{Día2} = 0.965 + 0.805 \text{ Día3},$$

A continuación el error típico de estimación y los valores de a y b en caso de que se utilicen puntuaciones típicas.

Por último, aparece el valor del estadístico “t” de contraste de las hipótesis nulas $\beta = 0$ y $\alpha = 0$. La última columna contiene el nivel crítico.

Capítulo 10

Prueba de hipótesis

Análisis de Resultados y Conclusiones: Muestra los principales hallazgos de la investigación aplicando prueba de hipótesis (*Medias, Prueba T para una muestra, Prueba T para muestras independientes, Prueba T para muestras relacionadas, ANOVA de un factor*), y presenta una potente interpretación teórica que demuestra el dominio técnico del investigador. Las **conclusiones** extraen lo esencial de todo el proceso enfatizando especialmente la riqueza de la evidencia empírica aportada.



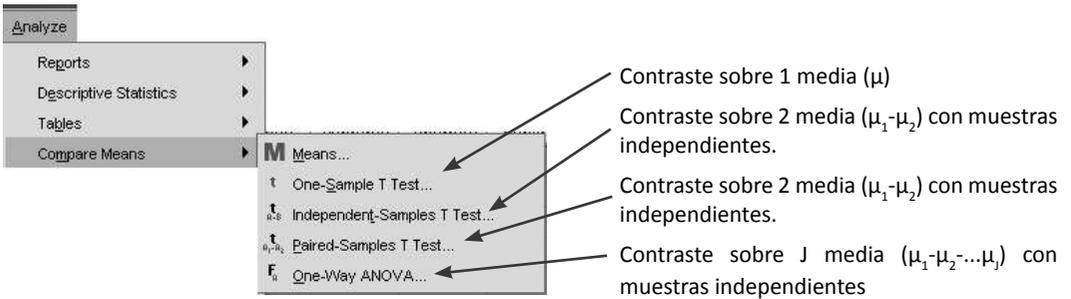
PRUEBA DE HIPÓTESIS

Prueba de hipótesis llamado en software PASW comparar medias, es un procedimiento estadístico que permite realizar distintos tipos de **contrastes sobre una, dos y hasta “n” medias**.

Los procedimientos que estudiaremos son: *Medias, Prueba T para una muestra, Prueba T para muestras independientes, Prueba T para muestras relacionadas, ANOVA de un factor.*

Los datos utilizados serán los recopilados en la ‘encuesta supermercado’.

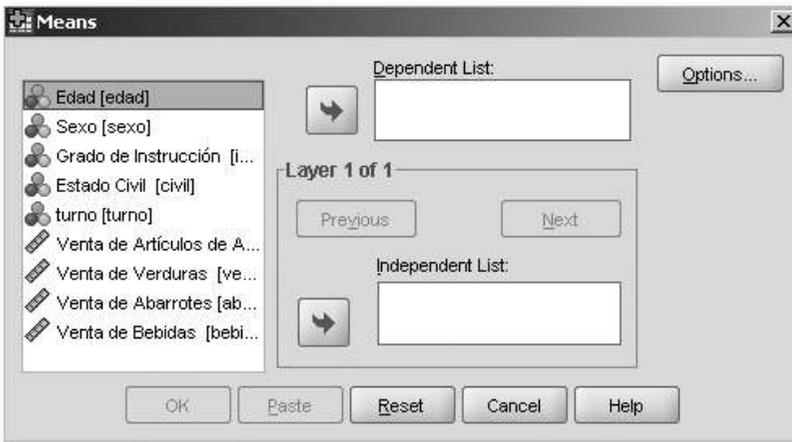
En el menú *Analyze* se encuentra el procedimiento **Comparar medias**, que permite realizar diferentes tipos de contrastes sobre una, dos y hasta “n” medias. El aspecto de dicho menú es el siguiente: *Analyze -> Comparación de Medias (Compare Means)*.



A continuación se presenta una explicación detallada del cuadro de diálogo correspondiente a los principales procedimientos.

Medias:

Utilizando el procedimiento Medias del PASW, se pueden calcular Medias y variancias. El cuadro de diálogo de la herramienta estadística ‘**Medias**’ ofrece el siguiente aspecto:



En el cuadro de la izquierda se muestra el listado de las variables de la que consta el archivo de datos que está abierto. Los cuadros de la derecha están vacíos pues están esperando a que le indiquemos qué variable queremos analizar. Dichas variables se trasladan de un cuadro a otro mediante el botón .

Veamos un ejemplo con las variables de la encuesta del supermercado. Para ello, primero debemos trasladar las variables a quienes se desea analizar a los cuadros de la derecha del modo siguiente: (en nuestro caso las variables son: Venta de artículos de aseo y edad).



Luego pulsar en el botón *Aceptar (OK)* para que se ejecute la orden. Los resultados obtenidos aparecen en la tabla siguiente:

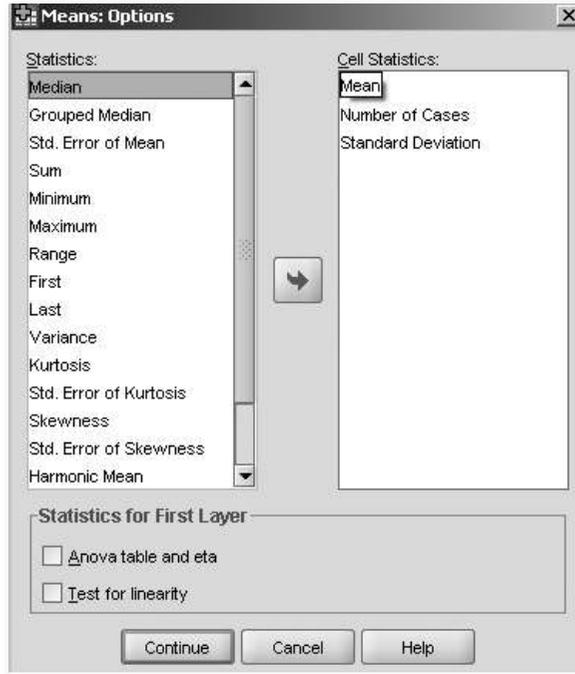
Report

Venta de Artículos de Aseo Personal

| Edad | Mean | N | Std. Deviation |
|-------------|--------|----|----------------|
| Adolescente | 2.2667 | 3 | 2.1127 |
| Joven | 1.7200 | 5 | 1.1367 |
| Adulto | 2.9182 | 22 | 2.4510 |
| Total | 2.6533 | 30 | 2.2483 |

La tabla muestra una comparación de las medias de la venta de artículos de Aseo Personal en cada subgrupo de la variable Edad. Como es el caso que la media de los adolescentes en la compra de artículos de aseo personal es 2.2667. La media de los jóvenes en la compra de artículos de aseo personal es 1.7200.

Además, si pulsamos en el botón ‘Opciones’ obtenemos como resultado el cuadro de diálogo siguiente:



Donde se observa un listado de estadística como son: media, mediana, valor mínimo, valor máximo, es decir todas las estadísticas que se aprecian en el cuadro izquierdo.

Para realizar los cálculos simplemente debemos pasar las estadísticas elegidas al cuadro de la derecha (celda de estadísticas). Dichas estadísticas se trasladan de un cuadro a otro mediante el botón .

Ejemplo: Para ello hay que trasladar al cuadro de la derecha las estadísticas media, desviación estándar, mínimo, máximo, obteniendo la tabla de resultados siguiente:

Report

Venta de Artículos de Aseo Personal

| Edad | Mean | N | Std. Deviation | Minimum | Maximum |
|-------------|--------|----|----------------|---------|---------|
| Adolescente | 2.2667 | 3 | 2.1127 | .30 | 4.50 |
| Joven | 1.7200 | 5 | 1.1367 | .30 | 3.30 |
| Adulto | 2.9182 | 22 | 2.4510 | .20 | 6.50 |
| Total | 2.6533 | 30 | 2.2483 | .20 | 6.50 |

Se observa en la tabla de resultados los valores de la media, desviación estándar, el valor mínimo, el valor máximo de los grupos de edades en las ventas de artículos de aseo personal.

También se pueden comparar medias como son: media aritmética, media armónica y media geométrica. Como se observa en el cuadro de resultados siguiente:

Report

Venta de Abarrotes

| Estado Civil | Mean | Harmonic Mean | Geometric Mean |
|--------------|---------|---------------|----------------|
| Soltero | 36.5200 | 36.4116 | 36.4652 |
| Casado | 34.1833 | 34.1282 | 34.1559 |
| Divorciado | 34.1500 | 34.0483 | 34.0993 |
| Viudo | 34.0889 | 33.8444 | 33.9674 |
| Total | 34.5400 | 34.3904 | 34.4655 |

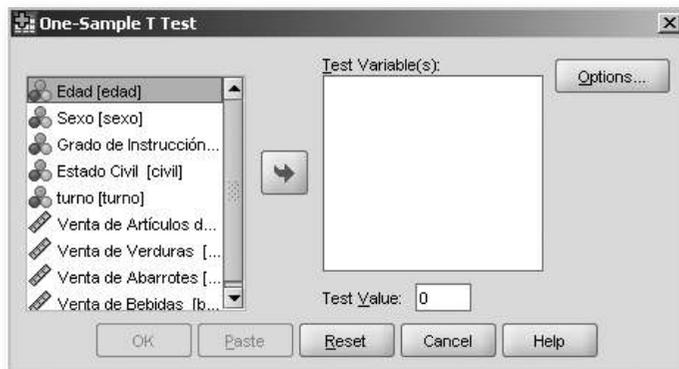
Se observa en la tabla de resultados la comparación de medias de los grupos de edades en las ventas de abarrotes.

Prueba T

Para comparar medias utilizando la prueba T hay tres opciones diferentes que permiten realizar contrastes de hipótesis sobre una media, dos medias con muestras independientes y dos medias con muestras relacionadas.

Prueba T para una muestra

Contraste sobre una media: Al seleccionar la opción menú **Analizar -> Comparar medias -> Prueba T para una muestra (one-sample T Test)** aparece el cuadro de diálogo siguiente:

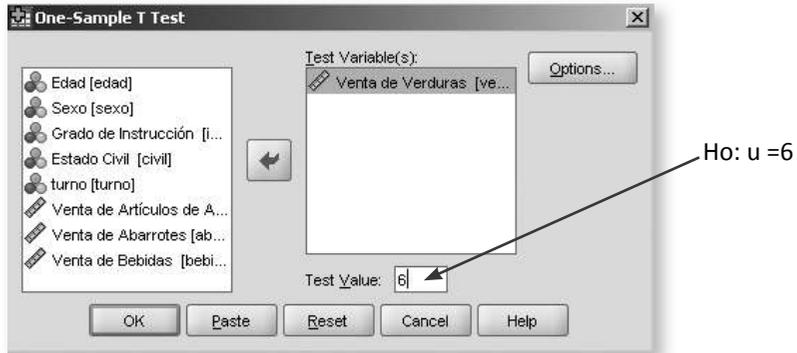


En el cuadro de la izquierda se muestra el listado de las variables de la que consta el archivo de datos que está abierto. Los cuadros de la derecha están vacíos pues están esperando a que le indiquemos qué variable queremos analizar. Dichas variables se trasladan de un cuadro a otro mediante el botón.

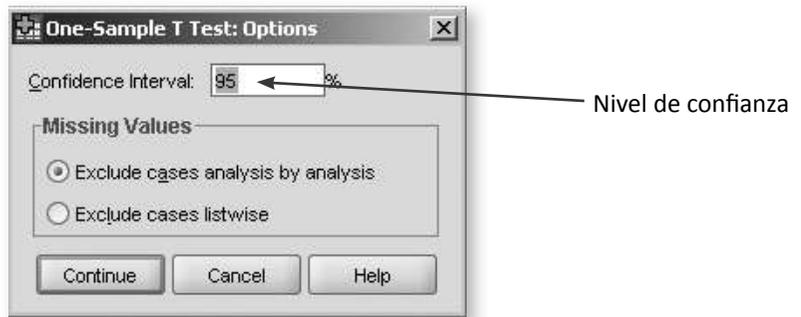


Para realizar el contraste hay que trasladar al cuadro **Contrastar variables** (Test Variable(s)) la variable cuya media se desea contrastar, indicar en **Valor de prueba** (Test Value.) el valor de la media poblacional en la hipótesis nula y luego pulsar el botón **Aceptar (OK)**.

Por ejemplo, si se quiere comprobar que la media poblacional de la Venta de Verduras es igual a “6” ($H_0: \mu=6$ y $\alpha = 0,05$), se seleccionan las opciones mostradas en cuadro de diálogo de la Prueba T para una muestra (One Sample T Test).



Además, si pulsamos sobre el botón *Opciones* puede modificarse el nivel de confianza que por defecto es 0,95. El cuadro de diálogo de Opciones es el siguiente:



El resultado del análisis se muestra en el visor de resultados: el PASW proporciona dos tablas de resultados.

En la primera aparecen el número de casos, la media, la desviación típica y el error típico de la media.

$$(\sigma_{\bar{x}} = S_{n-1} / \sqrt{n})$$

One-Sample Statistics

| | N | Mean | Std. Deviation | Std. Error Mean |
|-------------------|----|---------|----------------|-----------------|
| Venta de Verduras | 30 | 12.5333 | 5.7219 | 1.0447 |

En la segunda tabla aparecen los resultados de la prueba T. En primer lugar el valor de la estadística de contraste. Los grados de libertad (df), el nivel crítico bilateral, la diferencia entre la media muestral y la media poblacional (μ) de la hipótesis nula y, por último, los límites inferior y superior del intervalo de confianza para la diferencia (Media Muestral – μ).

| One-Sample Test | | | | | | |
|-------------------|-------|----|-----------------|-----------------|---|--------|
| Test Value = 6 | | | | | | |
| | t | df | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
| | | | | | Lower | Upper |
| Venta de Verduras | 6.254 | 29 | .000 | 6.5333 | 4.3967 | 8.6699 |

Cuando se realizan contrastes de hipótesis con el PASW, la decisión sobre H_0 debe tomarse a partir del nivel crítico, que es la probabilidad asociada al estadístico de contraste.

El PASW suele ofrecer el nivel crítico bilateral. En el ejemplo, el valor de T ha resultado ser 6,254 y nivel crítico bilateral 0,000. Esto significa que $P(T \leq -6.254) + P(T \geq 6.254) = 0,000$, lo que nos permite rechazar H_0 con $\alpha = 0,05$ ($\alpha = 0,05 > 0,000$). Esta decisión también puede tomarse a partir del intervalo de confianza para la diferencia de medias, el cual no incluye el valor 0.

Para poder visualizar mejor se presenta el gráfico de la **distribución T Bilateral (dos Colas)**.

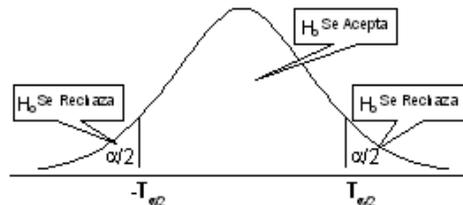
Hipótesis Planteada: $H_0: \mu = \mu_0$
 $H_1: \mu \neq \mu_0$

Nivel de Significancia: $\alpha \%$

Estadístico de Prueba:

$$t_{calc} = \frac{\hat{\theta} - E(\hat{\theta})}{\sqrt{Var(\hat{\theta})}} \approx t_{\alpha/2, n-1}$$

Regla de decisión



$T < -T_{\alpha/2}$ y $T > T_{\alpha/2}$ Se Rechaza la Hipótesis H_0 .

$-T_{\alpha/2} < T < T_{\alpha/2}$ Se Acepta la Hipótesis H_0 .

En caso de haber deseado realizar un contraste unilateral, es necesario dividir por dos el nivel crítico bilateral. Continuando con el ejemplo, para contrastar la hipótesis:

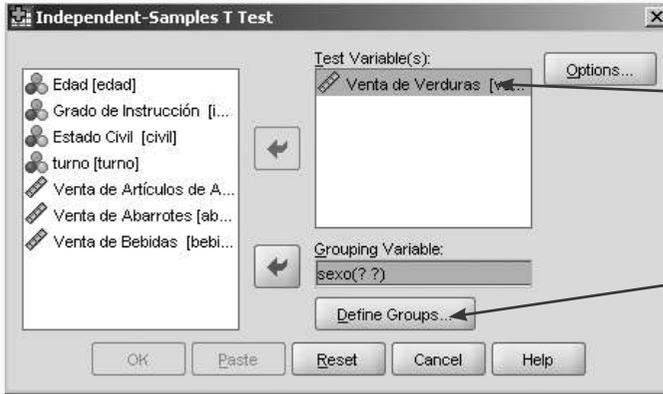
$H_0: \mu \leq 6$

$H_1: \mu > 6$

Hay que dividir por dos el nivel crítico bilateral: $0,000 / 2 = 0,00$, que es la probabilidad correspondiente al suceso: $T \geq 6.264$. En este caso también se rechazaría H_0 con $\alpha = 0,05$.

Prueba T para muestras independientes

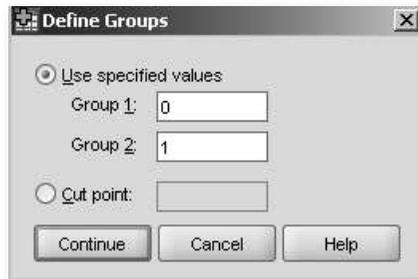
Contraste sobre dos medias independientes. Al seleccionar la opción menú **Analizar->Comparar medias-> Prueba T para muestras independientes (Independent-Samples T Test)** aparece el cuadro de diálogo siguiente:



Trasladar aquí la variable sobre la que se desea realizar el contraste.

Para realizar el contraste es necesario indicar cómo están definidos los dos grupos en el editor de datos.

En el cuadro de diálogo se ha indicado un contraste de las medias de las variables Venta de Verduras, en los dos grupos de la variable **Sexo**: Además es necesario indicar cuál es valor de la variable **Sexo** en cada grupo, para lo cual se pulsa el botón *Definir grupos (Define Groups)*, con lo que aparece un subcuadro de diálogo.



De este modo, se indica que se van a comparar los dos grupos, en los que **Sexo** toma los valores 0 (mujer) y 1 (varón):

La prueba T sobre dos medias independientes tiene dos versiones, dependiendo de sí se asumen varianzas iguales o distintas.

El PASW ejecuta en primer lugar la prueba de Levene para contrastar la hipótesis nula de que las varianzas poblacionales de los dos grupos son iguales. En la siguiente línea presenta la prueba de contrastar la hipótesis nula de que las varianzas poblacionales de los dos grupos son diferentes.

A continuación proporciona el resultado de las dos versiones de la prueba T.

Independent Samples Test

| | | Levene's Test for Equality of Variances | | t-test for Equality of Means | | | | | | |
|-------------------|-----------------------------|---|------|------------------------------|-------|-----------------|-----------------|-----------------------|---|--------|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
| | | | | | | | | | Lower | Upper |
| Venta de Verduras | Equal variances assumed | .262 | .613 | -1.218 | 28 | .233 | -3.7115 | 3.0479 | -9.9549 | 2.5318 |
| | Equal variances not assumed | | | -.996 | 3.564 | .382 | -3.7115 | 3.7258 | -14.5744 | 7.1513 |

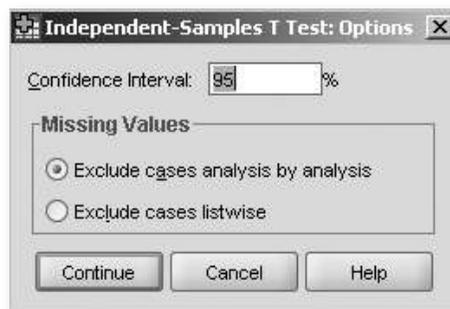
Es responsabilidad del usuario escoger el resultado de una versión de la prueba T, en función del resultado del contraste sobre varianzas. El nivel crítico de cada contraste aparece en la columna *Sig.*

La hipótesis nula de la prueba de Levene es $H_0: \sigma_1^2 = \sigma_2^2$, es decir, que la varianza poblacional es igual en ambos grupos. En este ejemplo, el estadístico de contraste (con valor $F = 0,262$) no ha resultado significativo, según indica el nivel crítico igual a 0,613. Por tanto, al mantenerse H_0 puede asumirse que las varianzas son iguales.

Se puede observar que en la tabla de resultados aparecen dos valores de T. En este caso se utiliza el primero de ellos, correspondiente al supuesto de igualdad de varianzas ($T = -1.218$).

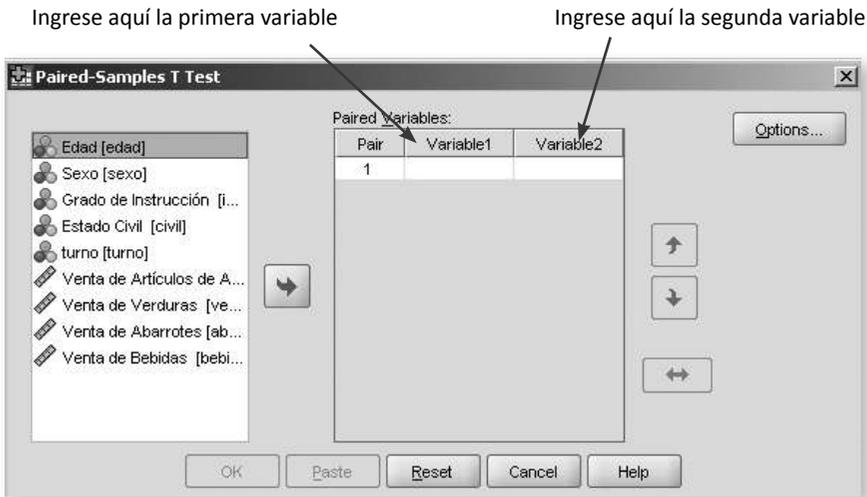
De nuevo, la decisión sobre la hipótesis nula debe tomarse a partir del nivel crítico (igual a 0,233), que en este caso indica que el valor de T no es significativo y por tanto, que las medias de las *Venta de Verduras* son iguales en varones y mujeres.

Además si pulsamos sobre el botón *Opciones* puede modificarse el nivel de confianza que por defecto es 0,95. El cuadro de diálogo de Opciones es el siguiente:



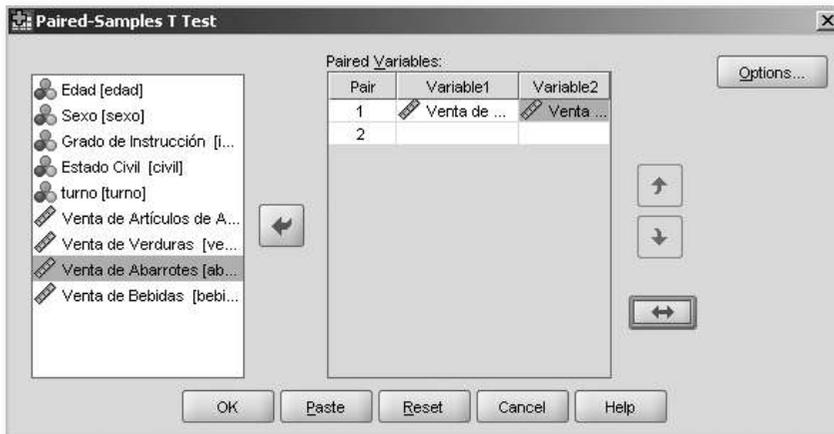
Prueba T para muestras relacionadas

Contraste sobre dos medias relacionadas: Al seleccionar la opción menú **Analizar->Comparar medias -> Prueba T para muestras relacionadas (Paired Samples T Test)** aparece el cuadro de diálogo siguiente:



Para comparar las medias con dos muestras relacionadas es necesario seleccionar dos variables de la lista, tal y como se observa en el cuadro de diálogo.

Luego para contrastar la hipótesis nula de que la diferencias de medias de **Venta de Verduras (Variable 1)** y **Venta de Abarrotes (Variable 2)** es significativamente diferente de cero. Se pulsa sobre el botón flecha para trasladar las variables al cuadro de variables seleccionadas, así como se muestra en la figura siguiente:



Luego pulsar en el botón *Aceptar (OK)* para que se ejecute la orden contraste. Los resultados obtenidos aparecen en la ventana *Visor de resultados*. Los resultados obtenidos son los siguientes:

Paired Samples Statistics

| | | Mean | N | Std. Deviation | Std. Error Mean |
|------|--------------------|---------|----|----------------|-----------------|
| Pair | Venta de Verduras | 12.5333 | 30 | 5.7219 | 1.0447 |
| 1 | Venta de Abarrotes | 34.5400 | 30 | 2.2997 | .4199 |

Paired Samples Correlations

| | | N | Correlation | Sig. |
|------|--|----|-------------|------|
| Pair | Venta de Verduras & Venta de Abarrotes | 30 | -.187 | .323 |

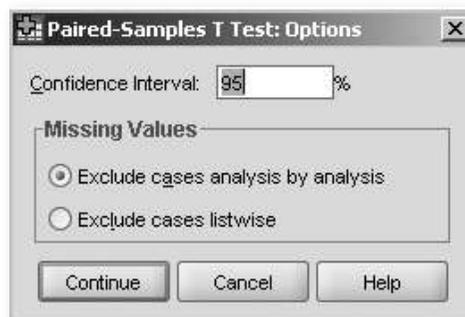
Paired Samples Test

| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|------|--|--------------------|----------------|-----------------|---|--------|-------|----|-----------------|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference | | | | |
| | | | | | Lower | Upper | | | |
| Pair | Venta de Verduras - Venta de Abarrotes | -22.01 | 6.5535 | 1.1965 | -24.45 | -19.56 | -18.4 | 29 | .000 |

Para realizar este contraste se calcula una nueva variable que es la diferencia entre las originales $D = \text{Venta de verduras} - \text{Venta de abarrotes}$. La salida de resultados muestra la media de la variable D , así como su desviación típica, el error típico de la media y el intervalo de confianza. También puede verse el estadístico de contraste T , los grados de libertad (df) y el nivel crítico bilateral.

Puesto que el nivel crítico vale $p = 0,000$ se rechaza H_0 con un nivel de confianza de 0,95 y se concluye que la media en la población de diferencia es significativamente distinta de cero. Es decir, existen diferencias significativas entre el promedio de Ventas tanto Verduras como de Abarrotes.

Además, si pulsamos sobre el botón *Opciones* puede modificarse el nivel de confianza que por defecto es 0,95. El cuadro de diálogo de Opciones es el siguiente:



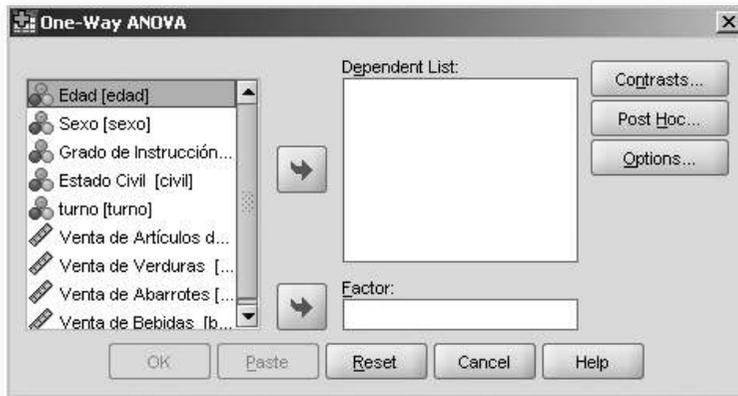
ANOVA de un factor

Análisis de varianza de un factor en un diseño completamente aleatorizado

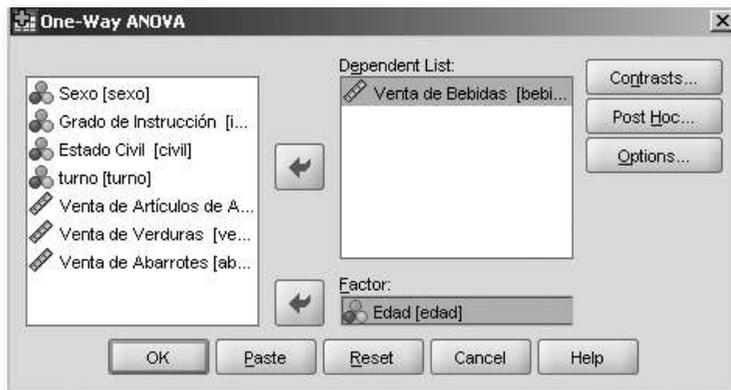
ANOVA A – EF – CA

Existen distintos procedimientos en PASW para realizar un análisis de varianza de un factor en un diseño completamente aleatorizado.

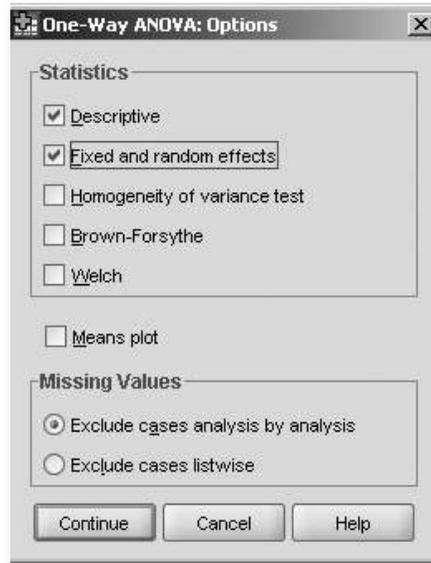
Uno de los más directos y sencillos se encuentra dentro del menú: **Analizar ->Comparar medias -> ANOVA de un factor (One-Way ANOVA)**. Se obtiene el siguiente cuadro de diálogo:



Por ejemplo, para contrastar la hipótesis de igualdad de medias en la variable venta de **bebidas** en los tres grupos de **Edad** (Adolescente, Joven y Adulto). Se ingresa al menú Analizar ->Comparar medias -> ANOVA de un factor y se procede llenar el cuadro de diálogo de la manera siguiente:



Dentro del cuadro de diálogo ANOVA de un factor hay varios sub-menús. Por ejemplo, si pulsando el botón *Opciones...* puede indicarse que PASW calcule estadísticos descriptivos para las variables en los “n” grupos y también que compruebe el supuesto de homocedasticidad (homogeneidad de varianzas). El cuadro de diálogo de Opciones es el siguiente:



Si se pulsa *Aceptar (OK)*, el resultado obtenido es el siguiente:

Descriptives

Venta de Bebidas

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|-------------|----|--------|----------------|------------|----------------------------------|-------------|---------|---------|
| | | | | | Lower Bound | Upper Bound | | |
| Adolescente | 3 | 5.0000 | 1.7321 | 1.0000 | .6973 | 9.3027 | 3.00 | 6.00 |
| Joven | 5 | 4.4000 | 2.3022 | 1.0296 | 1.5415 | 7.2585 | 1.00 | 6.00 |
| Adulto | 22 | 4.0000 | 1.6903 | .3604 | 3.2506 | 4.7494 | 1.00 | 6.00 |
| Total | 30 | 4.1667 | 1.7633 | .3219 | 3.5082 | 4.8251 | 1.00 | 6.00 |

Lo primero que aparece en el visor de resultados son los estadísticos descriptivos para los “n” (3) grupos. Como se observa, el grupo que más gasta en bebidas es el de Adolescente y el que menos gasta en bebidas es el de Adultos.

A continuación se muestra el resultado del test de Levene sobre homogeneidad de varianzas, cuya hipótesis nula es $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$.

Test of Homogeneity of Variances

Venta de Bebidas

| Levene Statistic | df1 | df2 | Sig. |
|------------------|-----|-----|------|
| .798 | 2 | 27 | .460 |

Puesto que el nivel crítico ($p = 0,460$) es mayor que α se mantiene la hipótesis nula sobre igualdad de varianzas. Por tanto, podemos asumir que se cumple el supuesto de homocedasticidad (homogeneidad de varianzas).

A continuación PASW muestra la tabla resumen el ANOVA:

ANOVA

Venta de Bebidas

| | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|----|-------------|------|------|
| Between Groups | 2.967 | 2 | 1.483 | .459 | .637 |
| Within Groups | 87.200 | 27 | 3.230 | | |
| Total | 90.167 | 29 | | | |

En la tabla aparecen las sumas de cuadrados inter- grupos, intra grupos (error) y total. La tabla también contiene los grados de libertad, medias cuadráticas, estadísticos de contraste F y su nivel crítico (*Sig.*). En este ejemplo, puesto que el nivel crítico (0.6379) es mayor de 0,05 se acepta la hipótesis nula del ANOVA ($H_0: \mu_1 = \mu_2 = \mu_3$) y puede concluirse que no existen diferencias significativas en el gasto medio de bebidas por los sujetos de cada edad (Adolescente, Joven y Adulto).

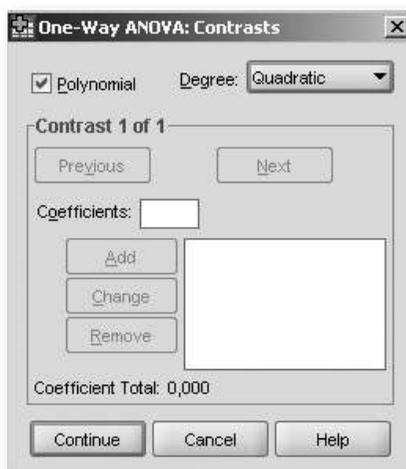
Comparaciones múltiples entre medias

Comparaciones de tendencia y F planeadas

Es necesario pulsar el botón **Contrastes**, que se encuentra situado en el cuadro de diálogo del procedimiento **ANOVA de un factor**.

En este ejemplo, como el número de media es $n = 3$ se podría realizar el contraste de tendencia lineal o cuadrática etc.

Si se quiere realizar la comparación de tendencia sobre el componente cuadrático se indicaría del siguiente modo:



El PASW proporciona los resultados del componente indicado y también de los de orden inferior, en este ejemplo, puesto que $n = 3$, son el cuadrático y el lineal.

ANOVA

Venta de Bebidas

| | | | Sum of Squares | df | Mean Square | F | Sig. |
|----------------|----------------|------------|----------------|----|-------------|------|------|
| Between Groups | (Combined) | | 2.967 | 2 | 1.483 | .459 | .637 |
| | Linear Term | Unweighted | 2.640 | 1 | 2.640 | .817 | .374 |
| | | Weighted | 2.933 | 1 | 2.933 | .908 | .349 |
| | | Deviation | 3.393E-02 | 1 | 3.393E-02 | .011 | .919 |
| | Quadratic Term | Unweighted | 3.393E-02 | 1 | 3.393E-02 | .011 | .919 |
| | | Weighted | 3.393E-02 | 1 | 3.393E-02 | .011 | .919 |
| Within Groups | | | 87.200 | 27 | 3.230 | | |
| Total | | | 90.167 | 29 | | | |

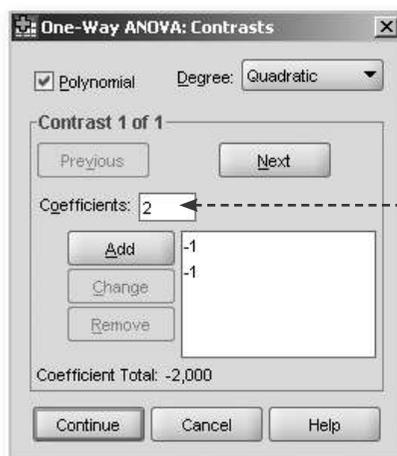
La tabla de resultados proporciona la F del ANOVA. Puede verse que su valor es 0.459 en este ejemplo (el mismo que en la salida de resultados comentada anteriormente). Además proporciona la F del componente lineal en la fila denominada No ponderado, que se toma el valor 0.817 y la del componente cuadrático, que ha resultado ser 0.011. La columna *Sig* se refiere a la hipótesis sobre cada tendencia, que se acepta en ambos casos con $p < 0.05$ en la tendencia lineal y $p < 0.01$ en la cuadrática; por tanto la relación entre ambas variables es Lineal.

Para llevar a cabo *comparaciones planeadas* es necesario introducir los coeficientes directamente. Por ejemplo, para comparar la media en **venta de bebidas** del grupo 3 (Adulto) de **Edad** con la media de los grupos 1 (Adolescente) y 2 (Joven) tomados juntos, hipótesis planteada sería:

$$H_0: (-1)\mu_1 + (-1)\mu_2 + (2)\mu_3 = 0$$

$$H_1: (-1)\mu_1 + (-1)\mu_2 + (2)\mu_3 \neq 0$$

Y en PASW se indicaría mediante:



Escriba aquí el valor de cada coeficiente y pulsar en Añadir.

La salida de los resultados incluye los coeficientes que se han utilizado en la comparación, lo cual permite comprobar si han sido asignados correctamente a los grupos. También incluye el estadístico T del contraste de medias, en sus dos versiones: asumiendo varianzas iguales y distintas.

Puesto que la prueba de Levene indicó igualdad de varianzas, el estadístico de prueba T vale - 0.921 y su nivel crítico 0,365 por lo que se mantiene H_0 y no puede concluirse que existan diferencias significativas entre los sujetos del grupo Adolescentes Jóvenes de frente a los Adultos.

Como se aprecia en los siguientes cuadros:

Contrast Coefficients

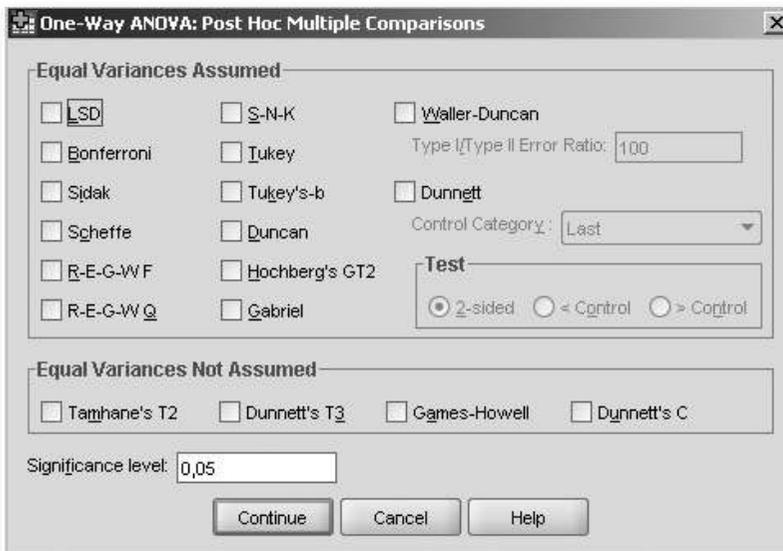
| Contrast | Edad | | |
|----------|-------------|-------|--------|
| | Adolescente | Joven | Adulto |
| 1 | -1 | -1 | 2 |

Contrast Tests

| | Contrast | Value of Contrast | Std. Error | t | df | Sig. (2-tailed) | |
|------------------|------------------------|-------------------|------------|--------|-------|-----------------|------|
| Venta de Bebidas | Assume equal variances | 1 | -1.4000 | 1.5198 | -.921 | 27 | .365 |
| | Does not assume equal | 1 | -1.4000 | 1.6061 | -.872 | 8.383 | .408 |

Comparaciones a posteriori

Al pulsar el botón *Post Hoc* aparece el cuadro de diálogo correspondiente a las comparaciones a posteriori:



Puede verse que permite seleccionar distintas pruebas: Tukey, Dunnett, Scheffé, etc. Continuando con el ejemplo, al seleccionar la prueba de Tukey, la salida PASW muestra la tabla de resultados de dicha prueba.

En la tabla puede verse la diferencia entre las medias de la venta de **Bebidas** en cada uno de los pares de grupos definidos por la variable **Edad**. También se incluye el error típico de la diferencia de medias, el nivel crítico asociado a dicha diferencia (*Sig.*) y los límites del intervalo de confianza. Como se observa, los resultados indican que No existen diferencias significativas entre el grupo Adolescentes y Joven ($p < 0,05$) y Adolescentes y Adulto ($p < 0,05$).

Multiple Comparisons

Dependent Variable: Venta de Bebidas

Tukey HSD

| (I) Edad | (J) Edad | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|-------------|-------------|-----------------------|------------|------|-------------------------|-------------|
| | | | | | Lower Bound | Upper Bound |
| Adolescente | Joven | .6000 | 1.3124 | .892 | -2.6541 | 3.8541 |
| | Adulto | 1.0000 | 1.1060 | .642 | -1.7424 | 3.7424 |
| Joven | Adolescente | -.6000 | 1.3124 | .892 | -3.8541 | 2.6541 |
| | Adulto | .4000 | .8904 | .895 | -1.8076 | 2.6076 |
| Adulto | Adolescente | -1.0000 | 1.1060 | .642 | -3.7424 | 1.7424 |
| | Joven | -.4000 | .8904 | .895 | -2.6076 | 1.8076 |

Capítulo

11

Modelo lineal general - ANOVA

Análisis de Resultados y Conclusiones: Muestra los principales hallazgos de la investigación aplicando el Modelo lineal general (Los Procedimiento utilizados son ANOVA A-EF-CA 'Univariante' y ANOVA de 2 factores en un diseño completamente aleatorizado), y presenta una potente interpretación teórica que demuestra el dominio técnico del investigador. Las **conclusiones** extraen lo esencial de todo el proceso enfatizando especialmente la riqueza de la evidencia empírica aportada.



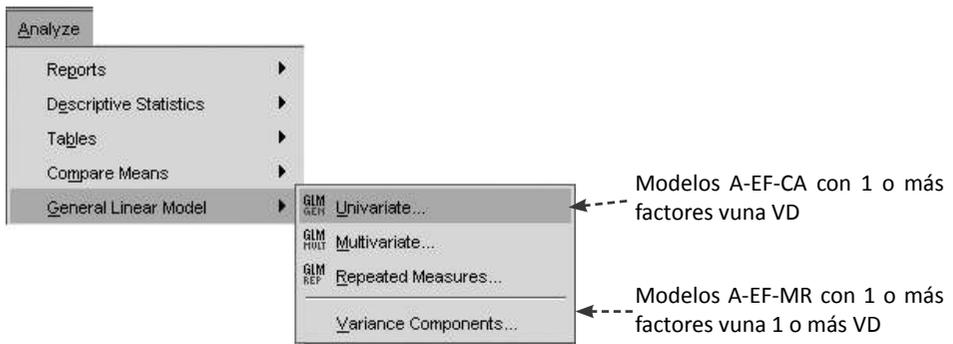
MODELO LINEAL GENERAL

Modelo lineal general permite realizar ANOVA de un factor, o cualquier otro tipo de ANOVA.

Los Procedimiento utilizados son: ANOVA A-EF-CA (Univariante) y ANOVA de 2 factores en un diseño completamente aleatorizado.

Para los ejercicios utilizaremos los datos de la encuesta supermercado.

El procedimiento consiste en hacer clic en el menú **Analizar -> Modelo lineal general**, cuyo aspecto es el siguiente:



A continuación, se presenta una explicación detallada del cuadro de diálogo correspondiente a los principales procedimientos.

Análisis de varianza de dos factores

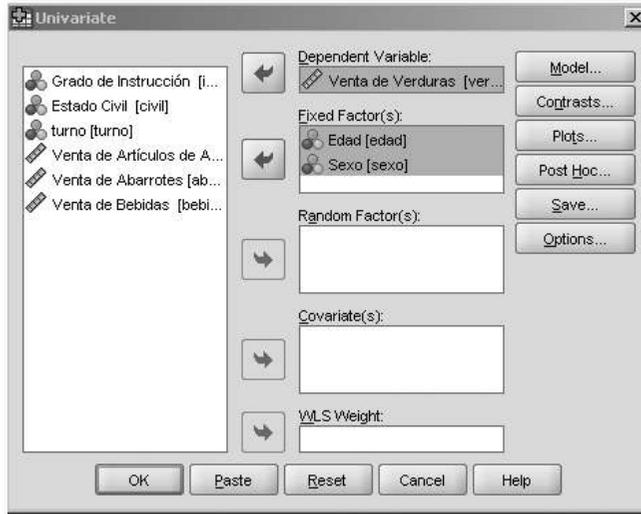
El software estadístico PASW ofrece buenas posibilidades para realizar análisis de varianza en diseños factoriales, con efectos fijos, aleatorios y mixtos. El procedimiento se encuentra ubicado en la opción del menú: **Analizar -> Modelo lineal general**.

ANOVA AB – EF – CA

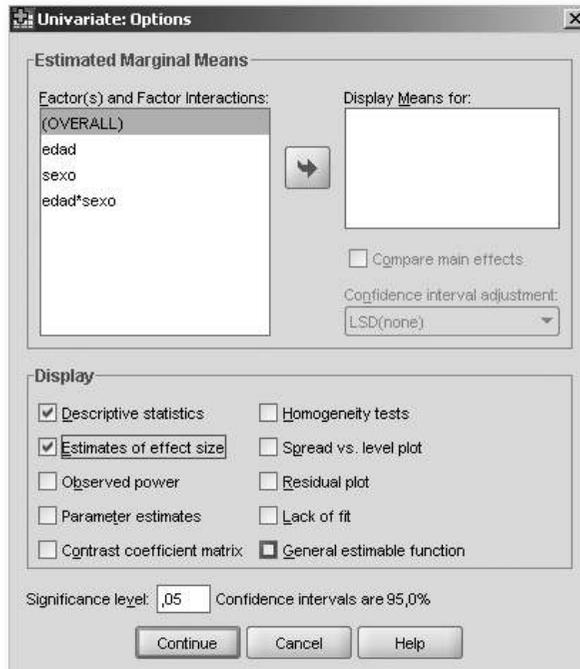
Este procedimiento estadístico permite realizar análisis de varianza con más de un factor, incluyendo factores de efectos fijos y aleatorios, además de análisis de covarianza (ANCOVA).

Supongamos que quisiera realizarse un ANOVA factorial de la variable **Venta de Verduras**, utilizando como variable independiente el sexo del sujeto y su edad. Esto se especificaría del siguiente modo: Ingresamos al menú: **Analizar -> Modelo lineal general. ->Univariante (univariate)**.

El procedimiento **Univariante** presenta el cuadro de diálogo siguiente:



El cuadro de diálogo del botón Opciones del procedimiento Univariante es el siguiente:



Pulsando *continuar* luego pulsando en el botón *Aceptar*, PASW realiza el ANOVA indicado y ofrece los resultados:

Descriptive Statistics

Dependent Variable: Venta de Verduras

| Edad | Sexo | Mean | Std. Deviation | N |
|-------------|-------|---------|----------------|----|
| Adolescente | Mujer | 20.0000 | 3.6056 | 3 |
| | Total | 20.0000 | 3.6056 | 3 |
| Joven | Mujer | 11.5000 | 2.6458 | 4 |
| | Varón | 24.0000 | . | 1 |
| | Total | 14.0000 | 6.0415 | 5 |
| Adulto | Mujer | 10.8947 | 5.1950 | 19 |
| | Varón | 13.0000 | 5.5678 | 3 |
| | Total | 11.1818 | 5.1606 | 22 |
| Total | Mujer | 12.0385 | 5.4734 | 26 |
| | Varón | 15.7500 | 7.1356 | 4 |
| | Total | 12.5333 | 5.7219 | 30 |

En Opciones se han solicitado los estadísticos descriptivos, la prueba de homogeneidad y las estimaciones de tamaño del efecto.

El PASW ofrece las medias y desviaciones típicas para las JK combinaciones de variables. Como se observa, quienes gastan más son los varones jóvenes.

Levene's Test of Equality of Error Variance

Dependent Variable: Venta de Verduras

| F | df1 | df2 | Sig. |
|------|-----|-----|------|
| .999 | 4 | 25 | .427 |

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a. Design: Intercept+EDAD+SEXO+EDAD * SEXO

A continuación la prueba de homocedasticidad o igualdad de varianzas, supuesto que se cumple al mantener H_0

La salida de resultados incluye la tabla – resumen del ANOVA y las medidas del tamaño del efecto:

Tests of Between-Subjects Effects

Dependent Variable: Venta de Verduras

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
|-----------------|-------------------------|----|-------------|---------|------|---------------------|
| Corrected Model | 354.677 ^a | 4 | 88.669 | 3.727 | .016 | .374 |
| Intercept | 2883.366 | 1 | 2883.366 | 121.193 | .000 | .829 |
| EDAD | 306.376 | 2 | 153.188 | 6.439 | .006 | .340 |
| SEXO | 130.390 | 1 | 130.390 | 5.481 | .028 | .180 |
| EDAD * SEXO | 66.047 | 1 | 66.047 | 2.776 | .108 | .100 |
| Error | 594.789 | 25 | 23.792 | | | |
| Total | 5662.000 | 30 | | | | |
| Corrected Total | 949.467 | 29 | | | | |

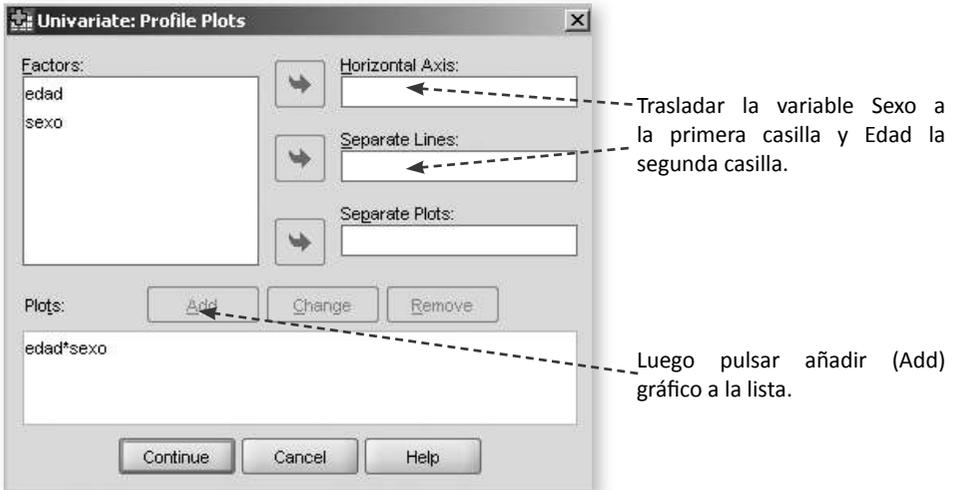
a. R Squared = .374 (Adjusted R Squared = .273)

$$\sum_i \sum_j \sum_k Y_{ijk}^2$$

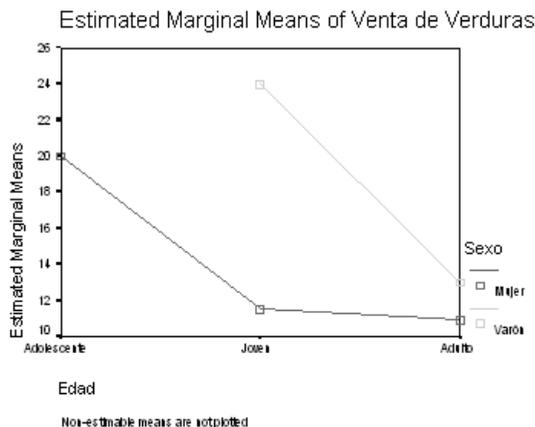
Las estadísticas de contraste F correspondiente a los efectos de las variables **Edad** y **Sexo**, así como a la interacción entre ambas (**Edad *Sexo**), aparecen en la fila etiquetadas con el nombre de la variable. La decisión sobre H_0 se toma a partir del punto crítico. Puede verse que se rechaza la hipótesis nula correspondiente a la variable **Edad**, también se rechaza la variable **Sexo** y se mantiene la correspondiente a la interacción **Edad* Sexo**.

Para realizar un gráfico con los efectos de los factores hay que pulsar el botón *Gráficos* del cuadro de diálogo del procedimiento **Univariante**.

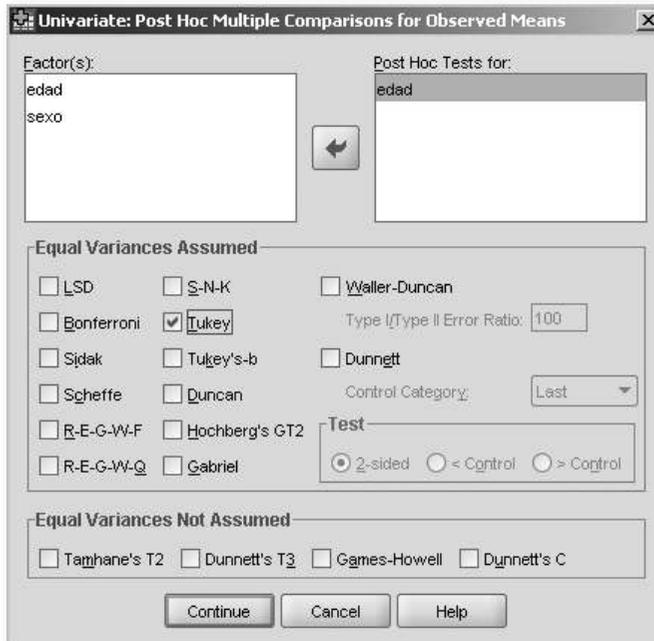
De este modo, aparece un subcuadro de diálogo en el que se especifica la forma en que se realizan uno o más gráficos. Por ejemplo.



De este modo, se ha especificado un gráfico de los valores de venta verduras por **Edad**, con líneas distintas para cada nivel de **Sexo**. A continuación, hay que pulsar el botón *Añadir* para que este gráfico se incluya en la lista de gráficos que se van a realizar. La salida de resultados del ANOVA incluye todos los gráficos que se hayan especificado. En nuestro caso:



El cuadro de diálogo del procedimiento **Univariante** incluye también el botón *Post Hoc*, que se utiliza para realizar contrastes a posteriori sobre los factores de efectos fijos. Para realizar la prueba de Tukey sobre los niveles de **Edad** se indicaría:



Los resultados de la prueba de Tukey aparecen junto con los de ANOVA.

La tabla muestra las diferencias entre cada par de medias de la variable **Venta de verduras** en los grupos de sujetos definidos por **Edad** y **Sexo**. Además, se incluye el error típico de cada diferencia, el nivel crítico y el intervalo de confianza para la diferencia de medias.

Multiple Comparisons

Dependent Variable: Venta de Verduras
Tukey HSD

| (I) Edad | (J) Edad | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
|-------------|-------------|-----------------------|------------|------|-------------------------|-------------|
| | | | | | Lower Bound | Upper Bound |
| Adolescente | Joven | 6.0000 | 3.5621 | .231 | -2.8727 | 14.8727 |
| | Adulto | 8.8182* | 3.0020 | .019 | 1.3407 | 16.2957 |
| Joven | Adolescente | -6.0000 | 3.5621 | .231 | -14.8727 | 2.8727 |
| | Adulto | 2.8182 | 2.4166 | .484 | -3.2011 | 8.8374 |
| Adulto | Adolescente | -8.8182* | 3.0020 | .019 | -16.2957 | -1.3407 |
| | Joven | -2.8182 | 2.4166 | .484 | -8.8374 | 3.2011 |

Based on observed means.

*. The mean difference is significant at the .05 level.

MEDIDAS REPETIDAS

Análisis de varianza de un factor con medidas repetidas

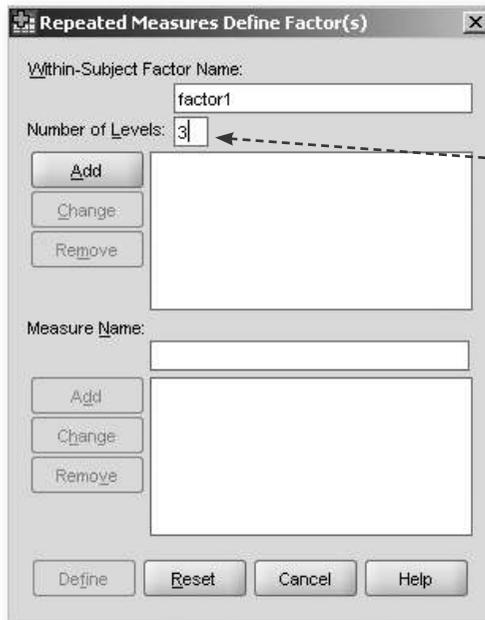
Anteriormente hemos visto el menú 'Comparar medias' para llevar a cabo ANOVAS de un factor. En PASW hay otro menú que puede utilizarse para llevar a cabo cualquier tipo de ANOVA. Se trata del menú que estamos estudiando **Analizar -> Modelo lineal general**.

ANOVA A - EF- MR

Para realizar un ANOVA de un factor de diseño de medidas repetidas se ejecuta la opción del menú: **Analizar -> Modelo lineal general -> Medidas repetidas (Repeated Measures Define Factor[s])**.

El objetivo es comparar las medias de distintas variables medidas sobre los mismos sujetos. Por ejemplo, supongamos que se desea comparar las medias de la **Venta de Verduras**, **Venta de Abarrotes** y **Venta de Bebidas** con el objeto de determinar la evolución de las ventas en cada tipo de artículo. Por tanto, tenemos un factor intra-sujetos con $J = 3$ niveles.

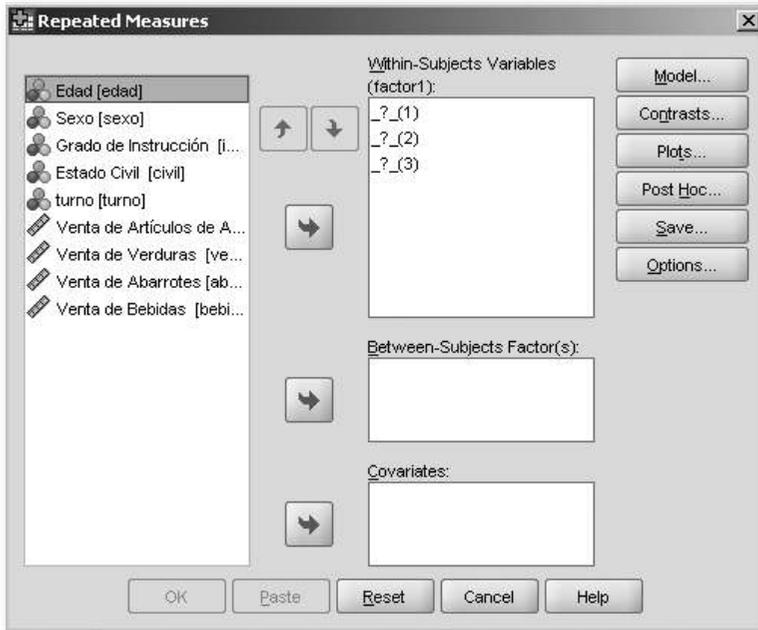
Cuando se selecciona el procedimiento **Medidas repetidas** aparece un cuadro de diálogo en el que se indica cuántas variables se van a comparar. En nuestro caso, son tres variables, luego:



Definir el número de niveles y pulsar en *Añadir (Add)*. A continuación pulsar en *Definir*.

A continuación, se pulsa sobre el botón *Añadir* para que el PASW interprete que se ha definido un factor *intra-sujetos* (denominado en el ejemplo **Factor 1**) con tres niveles diferentes. Después de *Añadir* se pulsa el botón *Definir* para indicar al PASW cuáles son las variables que forman cada uno de los niveles de Factor 1, en este caso tres niveles.

Al igual que en otros procedimientos, el PASW muestra un cuadro de diálogo con todas las variables del archivo de datos. En este cuadro se seleccionan las variables que forman los grupos del factor intra-sujetos, que en el ejemplo son, **Venta de Verduras**, **Venta de Abarrotes** y **Venta de Bebidas**.



A continuación, pulsando sobre el botón flecha se trasladan estas variables al cuadro denominado *Variables intra-sujetos*. De este modo se ha indicado cuáles son los tres niveles del factor intra-sujetos **Factor 1**.

El PASW proporciona distintas tablas de resultados. La tabla correspondiente a la F del factor intra-sujetos se denomina **Pruebas de efectos intra - sujetos**:

Tests of Within-Subjects Effects

Measure: MEASURE_1

| Source | | Type III Sum of Squares | df | Mean Square | F | Sig. |
|----------------|--------------------|-------------------------|--------|-------------|---------|------|
| FACTOR1 | Sphericity Assumed | 14768.339 | 2 | 7384.169 | 513.974 | .000 |
| | Greenhouse-Geisser | 14768.339 | 1.404 | 10517.592 | 513.974 | .000 |
| | Huynh-Feldt | 14768.339 | 1.454 | 10156.961 | 513.974 | .000 |
| | Lower-bound | 14768.339 | 1.000 | 14768.339 | 513.974 | .000 |
| Error(FACTOR1) | Sphericity Assumed | 833.275 | 58 | 14.367 | | |
| | Greenhouse-Geisser | 833.275 | 40.721 | 20.463 | | |
| | Huynh-Feldt | 833.275 | 42.166 | 19.762 | | |
| | Lower-bound | 833.275 | 29.000 | 28.734 | | |

La tabla contiene la suma de cuadrados Inter-grupos y Error con los correspondientes grados de libertad, medias cuadráticas, el estadístico de contraste F y el nivel crítico. El PASW ofrece los resultados bajo varios estadísticos según se cumple o no el supuesto de esfericidad.

Como puede verse en el cuadro de diálogo y en el visor de resultados, este procedimiento tiene muchas más opciones además de las indicadas. Por ejemplo, pueden realizarse pruebas de tendencia, comparaciones entre medias, gráficos, etc. Asimismo, desde este menú también puede realizar ANOVA con medidas repetidas en un más de un factor y con medidas repetidas sólo en algunos de ellos (diseños mixtos).

Capítulo 12

Pruebas no paramétricas

Análisis de Resultados y Conclusiones: Muestra los principales hallazgos de la investigación aplicando Pruebas no paramétricas, y presenta una potente interpretación teórica que demuestra el dominio técnico del investigador. Las **conclusiones** extraen lo esencial de todo el proceso enfatizando especialmente la riqueza de la evidencia empírica aportada.



PRUEBAS NO PARAMÉTRICAS

Los temas que se verán son los siguientes: Pruebas de los signos (binomial), Dos muestras independientes (prueba de Mann-Whitney), Dos muestras relacionadas (prueba de Wilcoxon), Más de dos muestras independientes (prueba de Kruskal Wallis), Más de dos muestras relacionadas (prueba de Friedman), K Muestras relacionadas, Contrastes sobre proporciones, Contraste sobre una proporción, Dos proporciones relacionadas, Más de dos proporciones relacionadas, Dos proporciones independientes, Prueba X^2 de Pearson, X^2 sobre Bondad de ajuste, Tablas de contingencia.

Los ejercicios se desarrollarán con los datos que aparecen recopilados en las variables **Día1**, **Día2**, **Día3** que indican el número de horas trabajadas en los tres días. Además se han recopilado las variables **Sexo** que toma valores 0 para mujeres y 1 para hombres, así como **Edad**, 1 para Adolescentes, 2 para Jóvenes y 3 para Adultos.

Finalmente, se incluyen tres datos relacionados con las terapias que ha recibido el sujeto. Todas ellas toman 1, si el sujeto ha recibido la terapia y 0 en caso contrario; las variables son **Estrés**, **Ansiedad** y **Fobia**.

Los datos representan 10 sujetos que participaron en el estudio, cuyo resultado se muestra en la tabla siguiente:

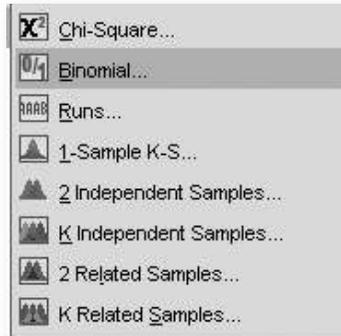
| Sujeto | Día1 | Día2 | Día3 | Sexo | Edad | Estrés | Ansiedad | Fobia |
|--------|------|------|------|------|------|--------|----------|-------|
| 1 | 5.7 | 6.3 | 6.8 | 0 | 1 | 0 | 1 | 0 |
| 2 | 4.8 | 5.4 | 4.3 | 1 | 3 | 0 | 0 | 0 |
| 3 | 7.6 | 7.9 | 8.6 | 0 | 2 | 0 | 0 | 0 |
| 4 | 5.7 | 6.0 | 6.7 | 1 | 1 | 0 | 0 | 1 |
| 5 | 3.8 | 4.6 | 4.9 | 1 | 3 | 0 | 0 | 0 |
| 6 | 7.5 | 8.0 | 8.5 | 1 | 2 | 0 | 1 | 0 |
| 7 | 6.4 | 7.0 | 7.3 | 1 | 1 | 1 | 1 | 0 |
| 8 | 7.7 | 8.1 | 8.5 | 0 | 2 | 0 | 0 | 0 |
| 9 | 4.0 | 4.5 | 5.0 | 0 | 3 | 1 | 0 | 0 |
| 10 | 5.7 | 6.1 | 6.8 | 1 | 2 | 1 | 0 | 1 |

CONTRASTES NO PARAMÉTRICOS

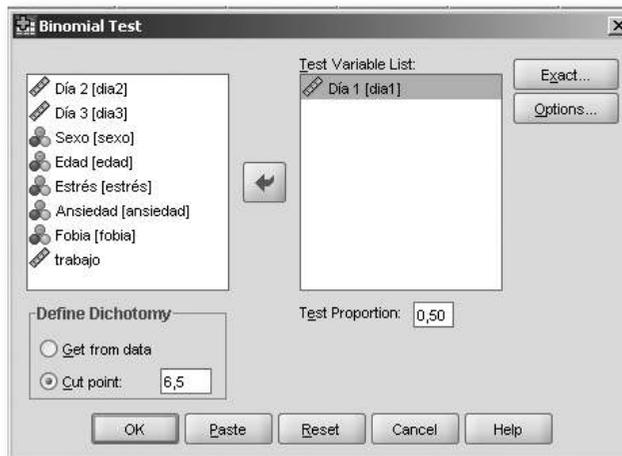
Los contrastes no paramétricos se ejecutan mediante la opción del menú: **Analizar > Pruebas no paramétricas**.

PRUEBAS DE LOS SIGNOS (BINOMIAL)

En el PASW se ejecuta mediante la opción **Prueba Binomial**.



Por ejemplo, supongamos que se desea contrastar la hipótesis de que la mediana de la variable Día1 es 6,5. Esto se indica del siguiente modo en el cuadro de diálogo de la prueba binomial:



En la casilla denominada *Punto de corte* se ha introducido el valor para la mediana en la hipótesis nula, en *Contrastar proporción* el valor 0,5, que es la proporción de sujetos con puntuaciones menores o iguales que 6,5, según la hipótesis nula.

La salida de resultados que se basa en el estadístico de contraste es:

Binomial Test

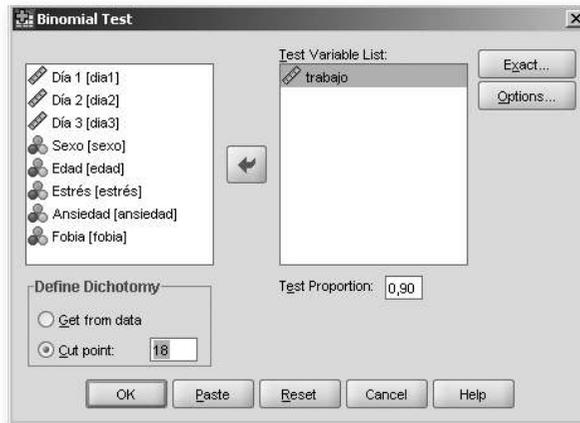
| | Category | N | Observed Prop. | Test Prop. | Exact Sig. (2-tailed) | |
|-------|----------|--------|----------------|------------|-----------------------|------|
| Día 1 | Group 1 | <= 6.5 | 7 | .70 | .50 | .344 |
| | Group 2 | > 6.5 | 3 | .30 | | |
| Total | | | 10 | 1.00 | | |

En el ejemplo, la tabla indica que 7 sujetos obtuvieron puntuaciones iguales o menores que 6,5. Esto corresponde al 70% de los sujetos de la muestra, mientras que el porcentaje esperado bajo H_0 es el 50%.

El valor del nivel crítico bilateral es 0,344, y se utiliza para tomar una decisión sobre H_0 .

La prueba binomial también permite realizar contrastes sobre cualquier otro centil, basta con cambiar la proporción introducida en la casilla *Contrastar proporción*.

Por ejemplo, para contrastar la hipótesis de que el centil 90 de la variable Trabajo (Día1+ Día2 + Día3) es el valor 18, se indica:



La salida de resultados basada en el estadístico de contraste es:

Binomial Test

| | Category | N | Observed Prop. | Test Prop. | Exact Sig. (1-tailed) |
|---------|----------|-------|----------------|------------|-----------------------|
| TRABAJO | Group 1 | <= 18 | 3 | .3 | .9 |
| | Group 2 | > 18 | 7 | .7 | |
| Total | | | 10 | 1.0 | |

a. Alternative hypothesis states that the proportion of cases in the first group < .9.

En el ejemplo, la tabla indica que 3 sujetos obtuvieron puntuaciones iguales o menores que 18. Esto corresponde al 30% de los sujetos de la muestra, mientras que el porcentaje esperado bajo H_0 es el 90%.

El valor del nivel crítico bilateral es 0,00, y se utiliza para tomar una decisión sobre H_0 .

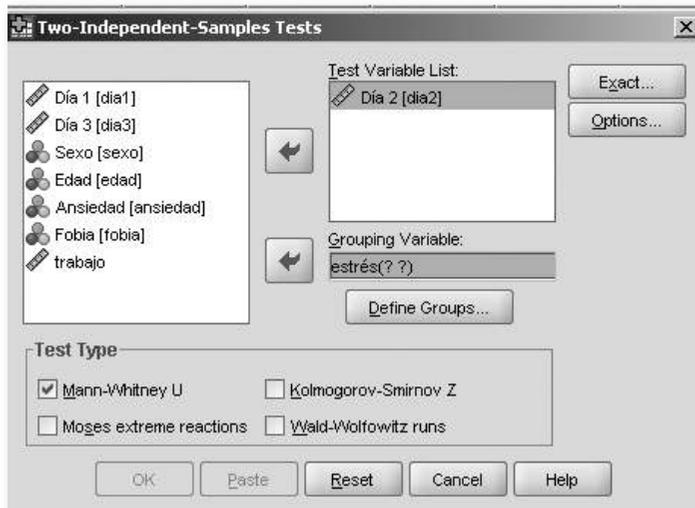
La opción *Definir la dicotomía* especifica que se va a comprobar si el 90% de los sujetos tienen valores iguales o menores que 18.

Contrastes para dos muestras

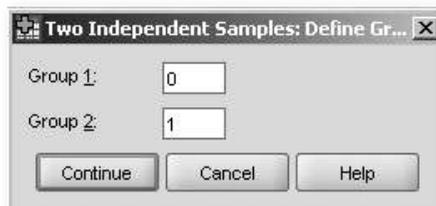
Dos muestras independientes, prueba de Mann-Whitney



Seleccionando los contrastes **2 Muestras independientes** aparece un cuadro de diálogo en el que se especifican las variables y el tipo de contraste. Para contrastar la hipótesis de que el valor esperado de **Día2** es igual en los dos grupos de la variable **Estrés**, se indica:



Mediante el botón *Definir grupos* es necesario indicar que los dos grupos de la variable **Estrés** son los correspondientes a los valores 0 y 1.



La salida de resultados muestra distintas tablas. Una de ellas contiene la suma de los rangos asignados a las puntuaciones en **Día2** en los dos grupos de **Estrés**.

Ranks

| | Estrés | N | Mean Rank | Sum of Ranks |
|-------|--------|----|-----------|--------------|
| Día 2 | No | 7 | 6.00 | 42.00 |
| | Si | 3 | 4.33 | 13.00 |
| | Total | 10 | | |

En la tabla separada aparece el estadístico de contraste U de Mann – Whitney y el nivel crítico.

Test Statistics^b

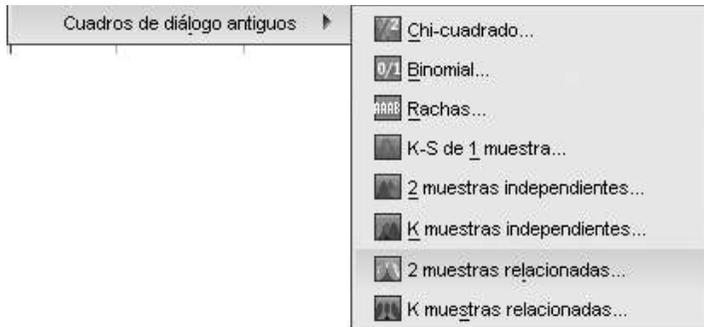
| | Día 2 |
|--------------------------------|-------------------|
| Mann-Whitney U | 7.000 |
| Wilcoxon W | 13.000 |
| Z | -.798 |
| Asymp. Sig. (2-tailed) | .425 |
| Exact Sig. [2*(1-tailed Sig.)] | .517 ^a |

a. Not corrected for ties.

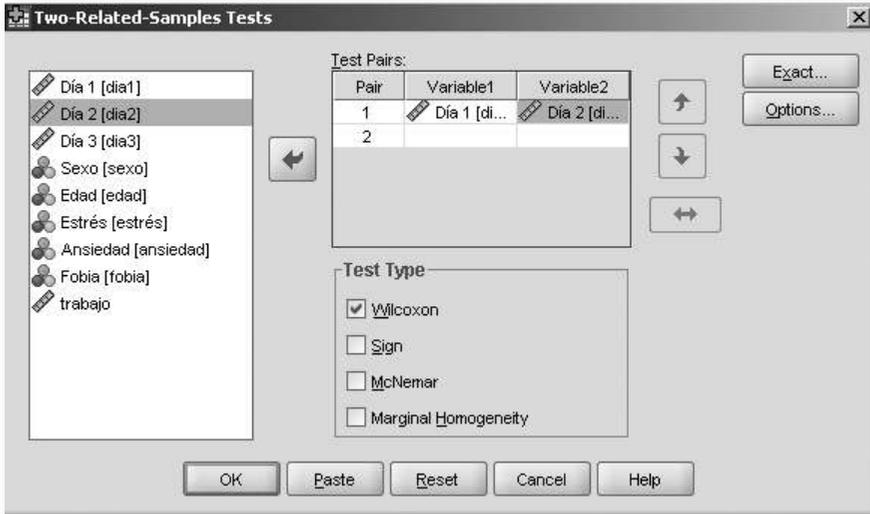
b. Grouping Variable: Estrés

Dos muestras relacionadas, prueba de Wilcoxon

Se encuentran en la opción **2 Muestras relacionadas**.



Por ejemplo, los sujetos del grupo **Día1** son los mismos que los de **Día2**, por lo que ambas son muestras relacionadas. Para comprobar que sus medianas son iguales se indica:



En las tablas siguientes aparece el estadístico de contraste:

Ranks

| | | N | Mean Rank | Sum of Ranks |
|---------------|----------------|-----------------|-----------|--------------|
| Día 2 - Día 1 | Negative Ranks | 0 ^a | .00 | .00 |
| | Positive Ranks | 10 ^b | 5.50 | 55.00 |
| | Ties | 0 ^c | | |
| | Total | 10 | | |

a. Día 2 < Día 1

b. Día 2 > Día 1

c. Día 1 = Día 2

Test Statistics^b

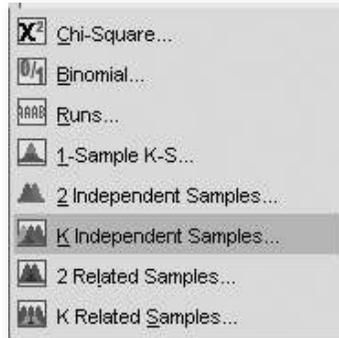
| | Día 2 - Día 1 |
|------------------------|---------------------|
| Z | -2.816 ^a |
| Asymp. Sig. (2-tailed) | .005 |

a. Based on negative ranks.

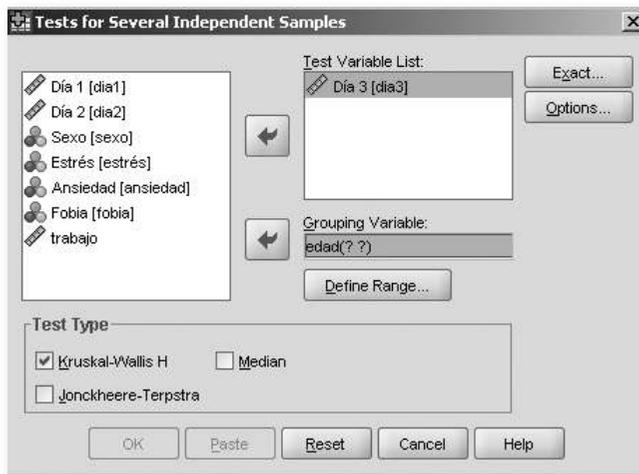
b. Wilcoxon Signed Ranks Test

Contrastes para más de dos muestras

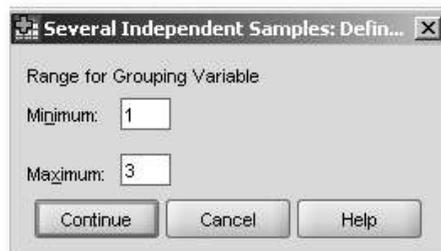
Más de dos muestras independientes, prueba de Kruskal Wallis: Se especifica mediante la opción **K Muestras independientes**.



A continuación, puede verse cómo contrastar la hipótesis de que la mediana de **Día3** es la misma en los tres grupos de edad.



Mediante *Definir rango* se ha indicado que los grupos de **Edad** vienen definidos por los valores 1 a 3.



La salida de resultados muestra la suma de rangos en cada grupo de **Edad** y el valor del estadístico χ^2 .

Ranks

| Edad | N | Mean Rank |
|-------------------|----|-----------|
| Día 3 Adolescente | 3 | 5.50 |
| Jóven | 4 | 8.13 |
| Adulto | 3 | 2.00 |
| Total | 10 | |

Test Statistics^{a,b}

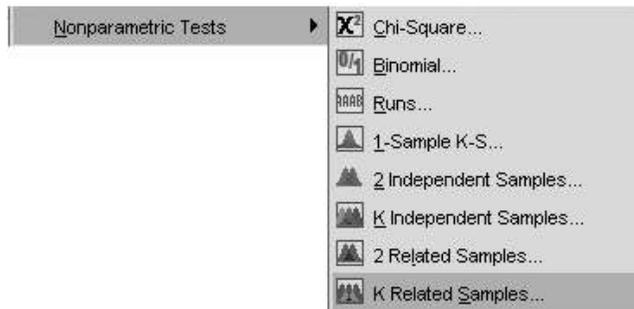
| | Día 3 |
|-------------|-------|
| Chi-Square | 7.102 |
| df | 2 |
| Asymp. Sig. | .029 |

a. Kruskal Wallis Test

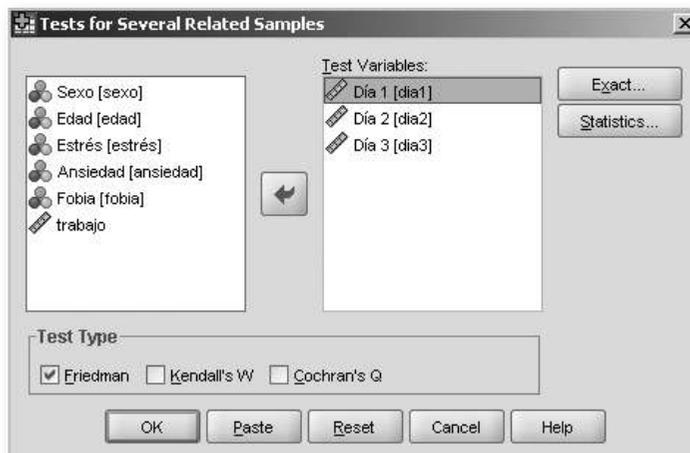
b. Grouping Variable: Edad

Más de dos muestras relacionadas, prueba de Friedman

Se especifica mediante la opción **K Muestras relacionadas**.



En el ejemplo, las variables de los tres Días están formadas por los mismos sujetos, por lo que son relacionadas. Del siguiente modo se comprueba si sus medianas son iguales:



La salida de resultados muestra el número de casos, estadístico de contraste X, los grados de libertad y el nivel crítico.

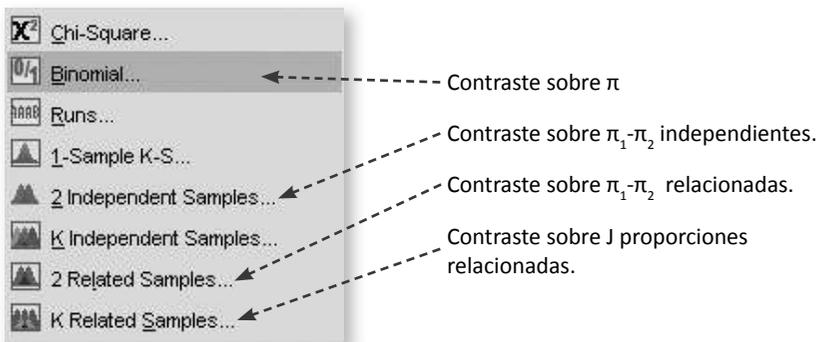
| Ranks | | Test Statistics ^a | |
|-------|-----------|------------------------------|--------|
| | Mean Rank | N | 10 |
| Día 1 | 1.10 | Chi-Square | 14.600 |
| Día 2 | 2.10 | df | 2 |
| Día 3 | 2.80 | Asymp. Sig. | .001 |

a. Friedman Test

El nivel crítico asociado al estadístico X^2 permite rechazar la hipótesis nula de igualdad de medianas en las tres variables.

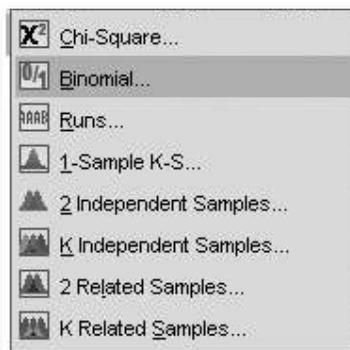
Contrastes sobre proporciones

Los procedimientos para realizar contrastes sobre proporciones se encuentran dentro del menú **Analizar > Pruebas no paramétricas**. El procedimiento particular varía dependiendo del número de grupos y de si están relacionados.

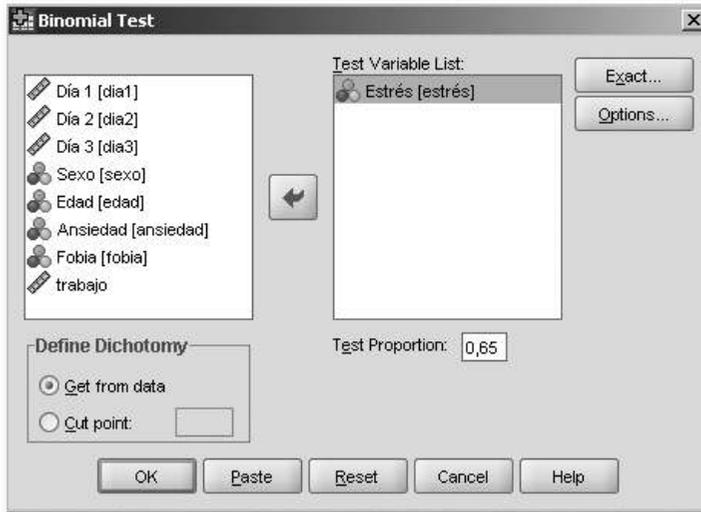


Contraste sobre una proporción

Este contraste es equivalente a la prueba no paramétrica binomial, vista en la práctica anterior.



Por ejemplo, para contrastar la hipótesis de que la proporción de sujetos que reciben la terapia de estrés es 0,65 se indica:



Luego se pulsa aceptar (OK) y se obtiene el resultado de la tabla:

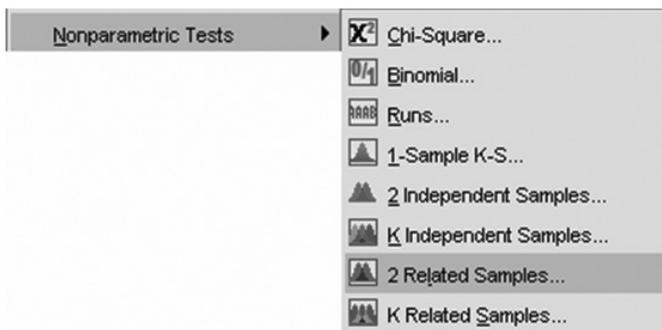
Binomial Test

| | Category | N | Observed Prop. | Test Prop. | Exact Sig. (1-tailed) |
|--------|------------|----|----------------|------------|-----------------------|
| Estrés | Group 1 No | 7 | .70 | .65 | .514 |
| | Group 2 Si | 3 | .30 | | |
| Total | | 10 | 1.00 | | |

El nivel crítico unilateral aparece en la última columna de la tabla de resultados. Utilizando $\alpha = 0,05$ no puede rechazarse la hipótesis de que la verdadera proporción de sujetos que recibe la terapia de estrés es 0,65.

Dos proporciones relacionadas

La prueba de McNemar se encuentra disponible en el procedimiento: **Analizar >Pruebas no paramétricas >2 Muestras relacionadas.**



Por ejemplo, sabemos que los sujetos de esta muestra pueden haber recibido terapia contra el estrés, así como otras dos terapias contra situaciones de ansiedad y fobias. Se desea comprobar si los sujetos que reciben la segunda terapia contra el estrés, así como otras dos terapias contra situaciones de ansiedad y fobias.

Se desea comprobar si los sujetos que reciben la segunda terapia es más probable que reciban también la tercera. Esto indicaría que los individuos que presentan un trastorno tienen mayor probabilidad de presentar el otro.

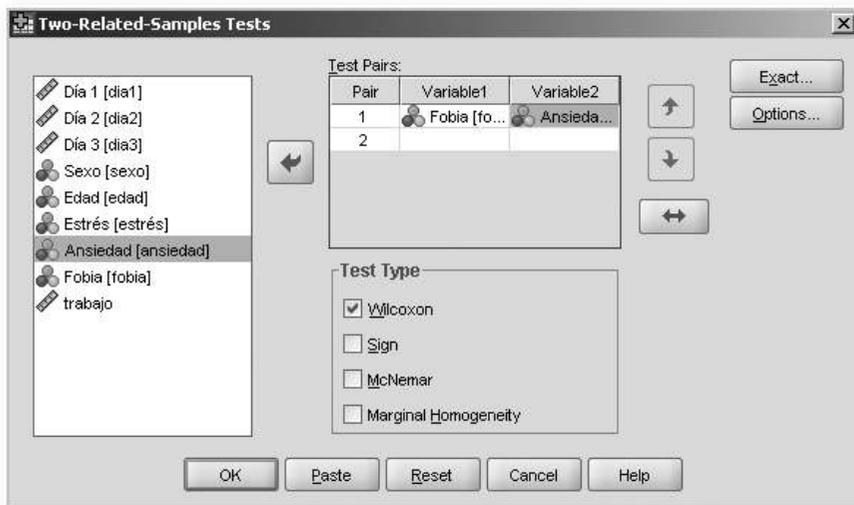
El contraste se realiza mediante el procedimiento **Analizar > Pruebas no paramétricas > 2 Muestras relacionadas**.

El cuadro de diálogo de este procedimiento permite escoger distintas pruebas. En concreto, para realizar la prueba de McNemar sobre las variables **ansiedad** y **fobia** se indica si seleccionamos las dos variables **ansiedad** y **fobia**. Se desea estudiar si la proporción de sujetos en ambas es igual o no. Entonces la prueba de hipótesis es la siguiente a:

$$H_0 : \pi_1 - \pi_2 = 0$$

$$H_1 : \pi_1 - \pi_2 \neq 0$$

Señalar la prueba de Mc Nemar. Esta prueba es la que contrasta hipótesis sobre 2 proporciones relacionadas.



El visor de resultados muestra la tabla de contingencia de ambas variables y los estadísticos de contraste:

Ansiedad & Fobia

| Ansiedad | Fobia | |
|----------|-------|---|
| | 0 | 1 |
| 0 | 5 | 2 |
| 1 | 3 | 0 |

Test Statistics^b

| | |
|-----------------------|--------------------|
| | Ansidad & Fobia |
| N | 10 |
| Exact Sig. (2-tailed) | 1.000 ^a |

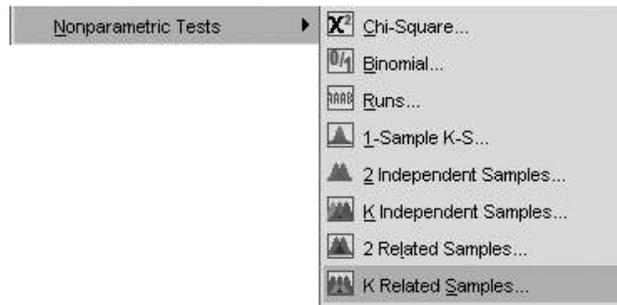
a. Binomial distribution used.

b. McNemar Test

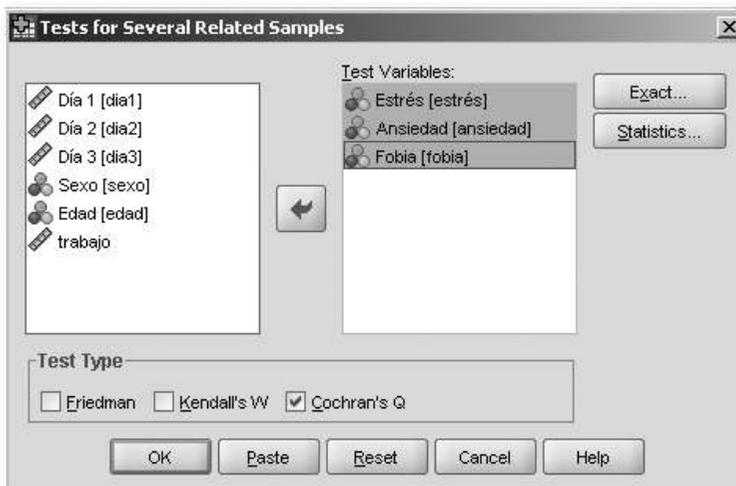
Resultado del test de la hipótesis. Como $p=1.000^a > \alpha$ se mantiene H_0 : Es decir no existe diferencia significativa entre la proporción de sujetos que reciben las terapias de **ansiedad y fobia**.

Más de dos proporciones relacionadas

Para realizar la prueba de Cochran con el PASW se escoge la opción: **Analizar > Pruebas no paramétricas > K Muestras relacionadas**.



Por ejemplo, si quisiera contrastarse la hipótesis de que la proporción de sujetos que reciben alguna de las tres terapias es la misma, se indica:



La salida de resultados muestra la distribución de frecuencias de cada variable y el estadístico de contraste Q :

Frecuencias

| | Value | |
|----------|-------|---|
| | 0 | 1 |
| Estrés | 7 | 3 |
| Ansiedad | 7 | 3 |
| Fobia | 8 | 2 |

Test Statistics

| | |
|-------------|-------------------|
| N | 10 |
| Cochran's Q | .333 ^a |
| df | 2 |
| Asymp. Sig. | .846 |

a. 0 is treated as a success.

Resultado del test de la hipótesis. Como $p=0.846 > \alpha$ se mantiene H_0 : Es decir no existe diferencia significativa entre la proporción de sujetos que reciben las terapias de **estrés**, **ansiedad** y **fobia**.

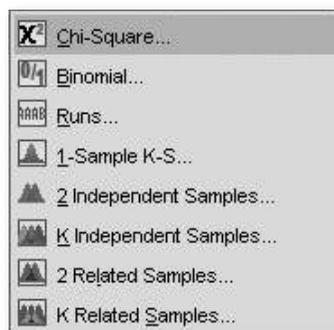
Dos proporciones independientes

Este tipo de contrastes se realizan mediante la prueba X^2 sobre igualdad de proporciones, que se comenta en la siguiente práctica.

Prueba X^2 de Pearson

X^2 sobre Bondad de ajuste

Para comprobar el ajuste de la distribución de frecuencias de una variable discreta a una distribución teórica se utiliza la prueba X que se encuentra disponible en **Analizar > Pruebas no paramétricas > Prueba Chi-cuadrado**.



A modo de ejemplo, se va a comprobar el ajuste de la variable **Edad** a una distribución uniforme. Utilizando el procedimiento **Frecuencias** puede obtenerse la distribución de frecuencias de esta variable ($H_0 : \pi_1 = \pi_2 = \pi_3$).

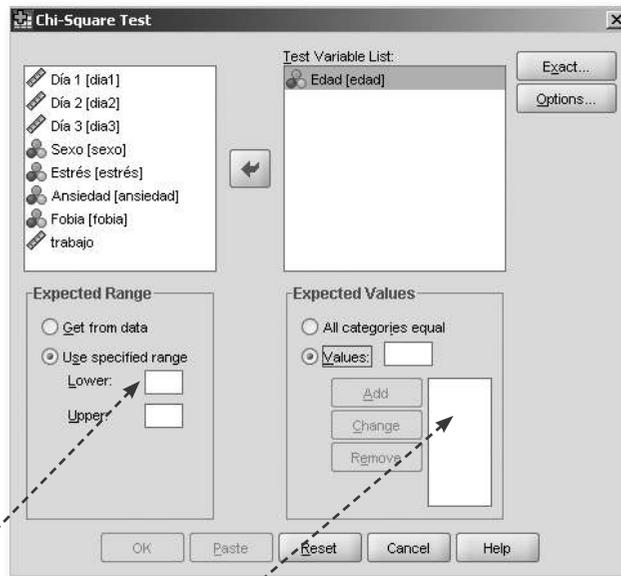
| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------------|-----------|---------|---------------|--------------------|
| Valid | Adolescente | 3 | 30.0 | 30.0 | 30.0 |
| | Jóven | 4 | 40.0 | 40.0 | 70.0 |
| | Adulto | 3 | 30.0 | 30.0 | 100.0 |
| | Total | 10 | 100.0 | 100.0 | |

Vemos que hay 3 sujetos adolescentes y adultos, 4 sujetos jóvenes de la variable **Edad**. La distribución uniforme asume que la frecuencia de sujetos es la misma en todos los niveles de edad.

Para comprobar si la desviación respecto a este supuesto es estadísticamente significativa, se utiliza el procedimiento de chi-cuadrado.

Al escoger el procedimiento **Analizar > Pruebas no paramétricas > Prueba Chi-cuadrado** aparece el cuadro de diálogo de la prueba χ^2 .

En el ejemplo se utiliza la variable **Edad**, luego:



Aquí se indica la categoría de la variable. Por ejemplo 0 y 1

Aquí se indica la frecuencia: $m_i - n\pi_i$ Para cada categoría. Por ejemplo. Adolescentes = $10 \cdot 0.3 = 3$; adultos = $10 \cdot 0.2 = 2$ y jóvenes = $10 \cdot 0.5 = 5$

En nuestro ejemplo el modelo a contrastar es el uniforme; se seleccionan todas las categorías iguales.

En el ejemplo, se selecciona la variable **Edad** y la opción *Todas las categorías iguales*, que indica que la frecuencia esperada de todos los valores de **Edad** es la misma, y por tanto que se está comprobando el ajuste a una distribución uniforme: $H_0 f(x) = M$ ($n = 10$ y $\pi = 1/3$).

La salida de resultados muestra una tabla con las frecuencias esperadas, observadas y las residuales (la diferencia entre ambas) y otra con el valor del estadístico de contraste X^2 , los grados de libertad y el nivel crítico.

Edad

| | Observed N | Expected N | Residual |
|-------------|------------|------------|----------|
| Adolescente | 3 | 3.3 | -.3 |
| Jóven | 4 | 3.3 | .7 |
| Adulto | 3 | 3.3 | -.3 |
| Total | 10 | | |

Test Statistics

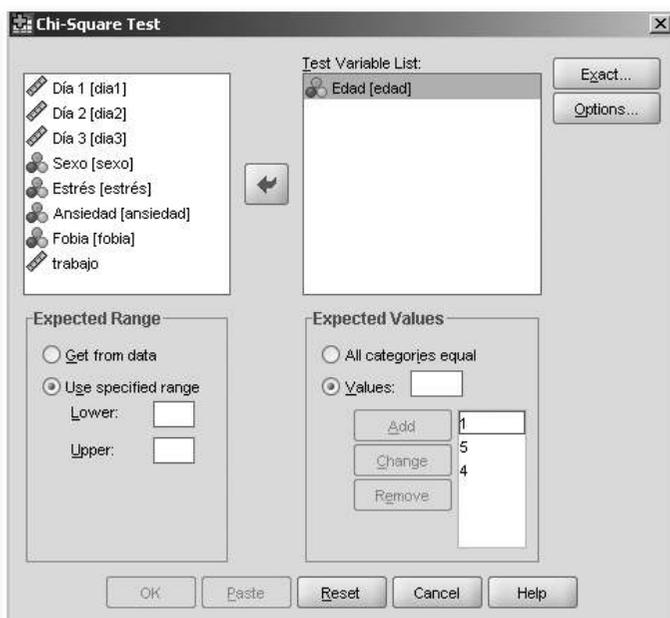
| | Edad |
|-------------------------|------|
| Chi-Square ^a | .200 |
| df | 2 |
| Asymp. Sig. | .905 |

a. 3 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 3.3.

Este procedimiento también permite comprobar el ajuste a cualquier otra distribución discreta. Para ello hay que indicar directamente la frecuencia esperada para cada valor.

Por ejemplo, si se quiere comprobar que la frecuencia esperada de los niveles 1, 2 y 3 de **Edad** es 4, 5 y 1 respectivamente, es decir:

$$H_0 f(x) = M(n=10, \pi_1 = 0.40, \pi_2 = 0.50, \pi_3 = 0.10) :$$



Obteniendo los siguientes resultados:

Edad

| | Observed N | Expected N | Residual |
|-------------|------------|------------|----------|
| Adolescente | 3 | 4.0 | -1.0 |
| Jóven | 4 | 5.0 | -1.0 |
| Adulto | 3 | 1.0 | 2.0 |
| Total | 10 | | |

Test Statistics

| | Edad |
|-------------------------|-------|
| Chi-Square ^a | 4.450 |
| df | 2 |
| Asymp. Sig. | .108 |

a. 2 cells (66.7%) have expected frequencies less than 5. The minimum expected cell frequency is 1.0.

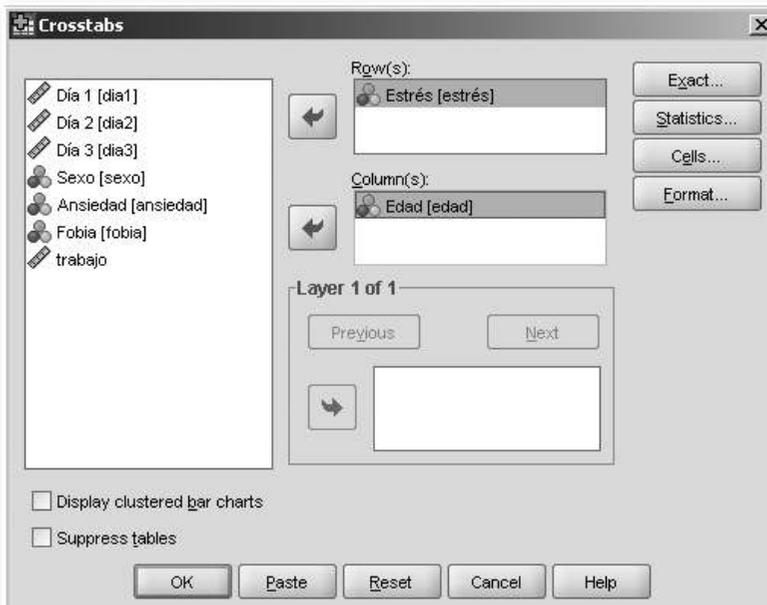
Tablas de contingencia

Este tipo de contrastes también se pueden realizar mediante la prueba χ^2 de las tablas de contingencia.

Las tablas de contingencia contienen la distribución conjunta de dos o más variables, y se han estudiado en relación con los contrastes sobre independencia e igualdad de proporciones.

En el PASW se pueden llevar a cabo mediante el procedimiento **Analizar > Estadísticos descriptivos > Tablas de contingencia**.

Vamos a obtener la tabla de contingencia de las variables **Estrés y Edad**. Esto se indica:



La salida de resultados del PASW muestra la tabla:

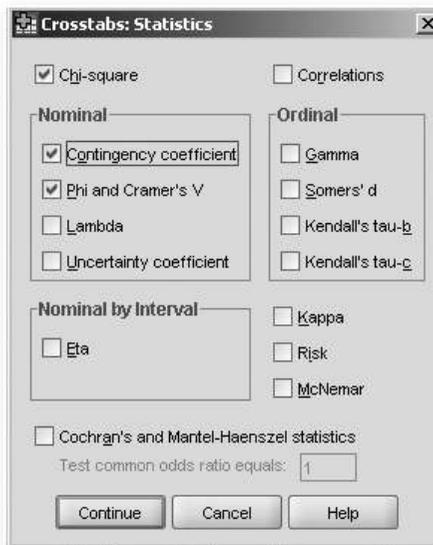
Estrés * Edad Crosstabulation

Count

| | | Edad | | | Total |
|--------|----|-------------|-------|--------|-------|
| | | Adolescente | Jóven | Adulto | |
| Estrés | No | 2 | 3 | 2 | 7 |
| | Si | 1 | 1 | 1 | 3 |
| Total | | 3 | 4 | 3 | 10 |

Cada casilla contiene el número observado de sujetos que obtienen un determinado par de valores en **Estrés** y **Edad**. Por ejemplo, hay 2 sujetos, el grupo adolescentes y que reciben la Terapia de Estrés, el número total de sujetos en la categoría Adulto es 3, etc.

Para calcular el valor de X^2 (y contrastar la hipótesis H_0 : X e Y son independientes), se pulsa en el botón *Estadísticos* del cuadro de diálogo del procedimiento tablas de contingencia, cuyo aspecto es el siguiente:



Este ejemplo se puede interpretar también como un contraste de igualdad de proporciones, en el que se compara la distribución de la variable **Estrés** (variable dicotómica) en cada grupo de Edad ($J=3$). El valor del estadístico aparece del siguiente modo en los resultados:

Chi-Square Tests

| | Value | df | Asymp. Sig. (2-sided) |
|------------------------------|-------------------|----|-----------------------|
| Pearson Chi-Square | .079 ^a | 2 | .961 |
| Likelihood Ratio | .080 | 2 | .961 |
| Linear-by-Linear Association | .000 | 1 | 1.000 |
| N of Valid Cases | 10 | | |

a. 6 cells (100.0%) have expected count less than 5. The minimum expected count is .90.

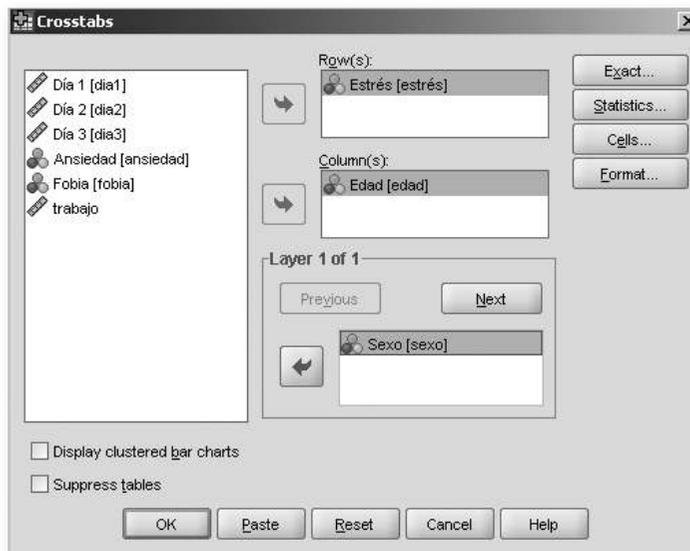
El valor de X está en la primera línea del cuadro, junto con los grados de libertad y el nivel crítico ($P(X \geq 0,833) = 0,361$). Los índices de asociación basados en chi-cuadrado aparecen en otra tabla diferente. En este caso, puesto que las variables son independientes, estos índices adoptan valores muy pequeños.

Symmetric Measures

| | | Value | Approx. Sig. |
|-----------------------|-------------------------|-------|--------------|
| Nominal by Nominal | Phi | .089 | .961 |
| | Cramer's V | .089 | .961 |
| | Contingency Coefficient | .089 | .961 |
| N of Valid Cases | | 10 | |

- a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

Es posible obtener tablas de contingencia con más de dos variables utilizando la casilla inferior del cuadro de diálogo. Por ejemplo, del siguiente modo se especifica una tabla de contingencia con las variables **Edad**, **Estrés** y **Sexo**:



La salida de resultados muestra la tabla de contingencia de las tres variables:

Estrés * Edad * Sexo Crosstabulation

| Count | | | Edad | | | Total |
|--------|--------|----|-------------|-------|--------|-------|
| Sexo | | | Adolescente | Jóven | Adulto | |
| Mujer | Estrés | No | 1 | 2 | | 3 |
| | | Si | | | 1 | 1 |
| | Total | | 1 | 2 | 1 | 4 |
| Hombre | Estrés | No | 1 | 1 | 2 | 4 |
| | | Si | 1 | 1 | | 2 |
| | Total | | 2 | 2 | 2 | 6 |

Lo que no permite el procedimiento Tablas de contingencia es realizar el análisis estadístico de más de dos variables cualitativas. Dicho análisis se lleva a cabo mediante los *modelos loglineales* y *modelos logit*.

Capítulo
13

Conglomerados de K medias

Conglomerados de K media - Conglomerados jerárquicos

Análisis Discriminante -Análisis Factorial

En estos temas sólo se realizará una pequeña introducción con ejemplos sencillos, simplemente para que usted tenga una idea de los temas tratados.



CONGLOMERADOS DE K MEDIAS

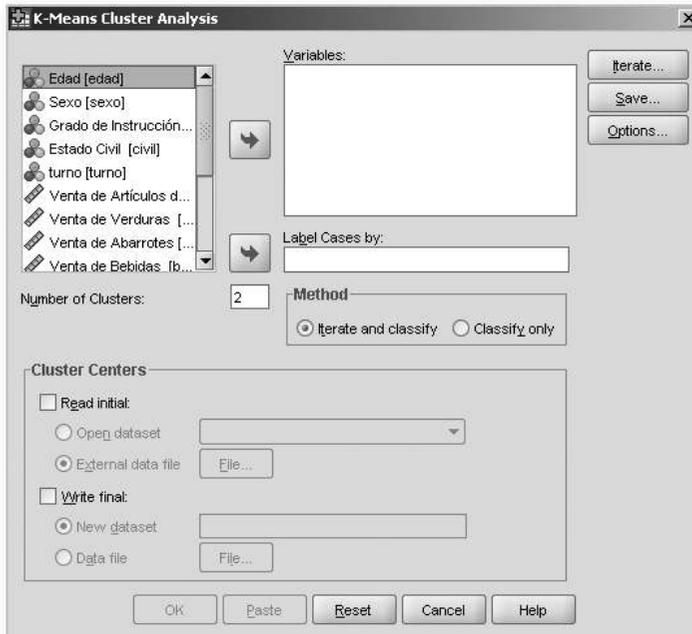
Intenta identificar grupos de casos relativamente homogéneos basándose en las características seleccionadas y utilizando un algoritmo que puede gestionar un gran número de casos. Sin embargo, el algoritmo requiere que el usuario especifique el número de conglomerados.

Se pueden especificar los centros iniciales de los conglomerados si se conoce de antemano dicha información. Se pueden elegir uno de los dos métodos disponibles para clasificar los casos: la actualización de los centros de los conglomerados de forma iterativa o sólo la clasificación.

Asimismo, se puede guardar la pertenencia a los conglomerados, información de la distancia y los centros de los conglomerados finales. Permite especificar una variable cuyos valores sean utilizados para etiquetar los resultados por casos.

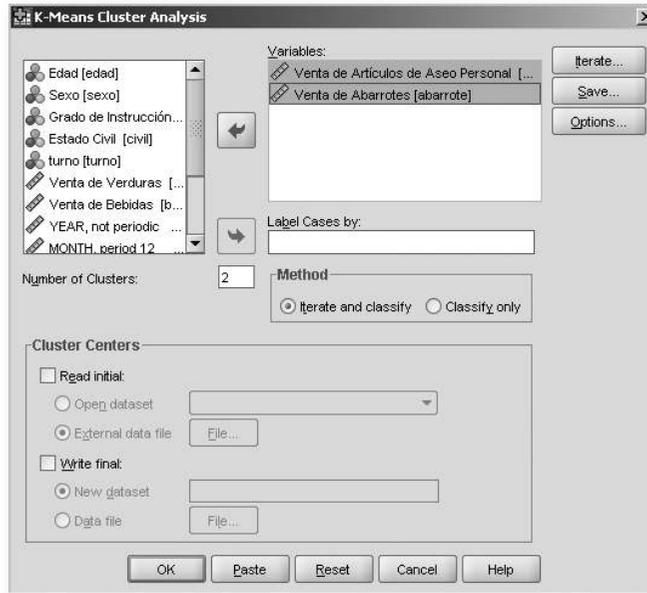
También se pueden solicitar los estadísticos F de los análisis de varianza. Aunque estos estadísticos son oportunistas (ya que el procedimiento trata de formar grupos que de hecho difieran), el tamaño relativo de los estadísticos proporciona información acerca de la contribución de cada variable a la separación de los grupos.

Para llevar a cabo un análisis de conglomerados de K medias, seleccionamos la opción **Analizar -> Clasificar->Conglomerados de K medias**, obteniéndose el cuadro de diálogo siguiente:



El cuadro de la izquierda muestra la lista de variables del archivo (numérico y cadena). Pero las variables de cadena sólo sirven para etiquetar casos.

Para obtener un análisis de conglomerados de K Medias se seleccionan las variables numéricas, y se traslada al cuadro del lado derecho las variables con los que se quiere formar conglomerados. En nuestro caso las variables son: Venta de artículos de aseo personal y Venta de abarrotes.



Opcionalmente, se selecciona una variable para identificar los casos en las tablas de resultados y en los gráficos. Luego, trasladar a **Etiquetar casos Mediante (Label Cases by:)**.

El **Número de conglomerados (Number of Clusters:)** Se encuentra seleccionado por defecto dos conglomerados. Para solicitar un número mayor de conglomerados introducir el número deseado en el cuadro.

Existen dos **Método (Method)**. El primero **Iterar y clasificar (Iterate and classify)** Este procedimiento se encarga de estimar los centros iterativamente y de clasificar a los sujetos con arreglo a los centros estimados. En el segundo, **Solo Clasificar (classify only)**, se clasifican a los sujetos según los centros iniciales (sin actualizar sus valores iterativamente). Este botón suele utilizarse junto con el botón **Centros>>**

Luego de realizar estas selecciones presionamos aceptar (OK). El visor presenta los siguientes resultados:

El primer cuadro muestra los centros iniciales, es decir, los valores que corresponden a los dos casos que han sido elegidos como centros respectivos de los conglomerados solicitados, en las dos variables de clasificación utilizadas.

Initial Cluster Centers

| | Cluster | |
|-------------------------------------|---------|-------|
| | 1 | 2 |
| Venta de Artículos de Aseo Personal | 1.30 | 3.00 |
| Venta de Abarrotes | 30.00 | 40.00 |

Por ejemplo, 1.30 primer centro (1) de la variable Venta de artículos de aseo personal y 3.00 segundo centro (2) de la variable Venta de artículos de aseo personal. Además 30.00 primer centro (1) de la variable Venta de abarrotes y 40.00 segundo centro (2) de la variable Venta de abarrotes.

Una vez seleccionados los centros de los conglomerados, cada caso es asignado al conglomerado de cuyo centro se encuentra más próximo y comienza un proceso de ubicación iterativa de los centros.

En la primera iteración se reasigna los casos por su distancia al nuevo centro y, tras la reasignación, se vuelve a actualizar el valor del centro. En la siguiente iteración se vuelven a reasignar los casos y a actualizar el valor del centro.

El cuadro siguiente resume el historial de iteraciones (3 en nuestro ejemplo) con indicación del cambio (desplazamiento) experimentado por cada centro en cada iteración. Puede observarse que, conforme avanzan las iteraciones, el desplazamiento de los centros se va haciendo más y más pequeño hasta llegar a la tercera iteración, en la que ya no existe desplazamiento alguno.

Iteration History

| Iteration | Change in Cluster Centers | |
|-----------|---------------------------|-------|
| | 1 | 2 |
| 1 | 2.962 | 3.540 |
| 2 | .187 | .251 |
| 3 | .000 | .000 |

a. Convergence achieved due to no or small distance change. The maximum distance by which any center has changed is .000. The current iteration is 3. The minimum distance between initial centers is 10.143.

En el PASW el proceso de iteración se detiene por defecto cuando alcanza 10 (u otro número ingresado) o cuando de una iteración a otra no se produce ningún cambio. En nuestro ejemplo se detiene en la tercera iteración porque no se produce ningún cambio.

El cuadro siguiente ofrece los centros de los conglomerados finales, es decir, los centros de los conglomerados tras el proceso de actualización iterativa. Comparando los centros finales (tras la iteración) de esta tabla con los centros iniciales (antes de la iteración):

Final Cluster Centers

| | Cluster | |
|-------------------------------------|---------|-------|
| | 1 | 2 |
| Venta de Artículos de Aseo Personal | 1.99 | 3.52 |
| Venta de Abarrotes | 33.00 | 36.55 |

En ejemplo, se puede apreciar que el primer (1) conglomerado está constituido por las ventas en menor cantidad en nuevos soles, tanto en artículos de aseo personal y de abarrotes, y el segundo conglomerado (2) por las ventas en mayor cantidad en nuevos soles de aseo personal y de abarrotes.

Por último, el cuadro siguiente muestra el número de casos asignados a cada conglomerado:

Number of Cases in each Cluster

| | | |
|---------|---|--------|
| Cluster | 1 | 17.000 |
| | 2 | 13.000 |
| Valid | | 30.000 |
| Missing | | .000 |

En nuestro ejemplo, los tamaños de los conglomerados son: primer conglomerado 17 y segundo conglomerado 13.

El cuadro siguiente es la tabla de ANOVA que muestra un resumen del análisis de varianza con una estadística F univariada para cada una de las variables incluidas en el análisis.

ANOVA

| | Cluster | | Error | | F | Sig. |
|-------------------------------------|-------------|----|-------------|----|--------|------|
| | Mean Square | df | Mean Square | df | | |
| Venta de Artículos de Aseo Personal | 17.048 | 1 | 4.627 | 28 | 3.685 | .065 |
| Venta de Abarrotes | 93.040 | 1 | 2.155 | 28 | 43.179 | .000 |

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

El cuadro siguiente muestra un listado de todos los casos utilizados en el análisis con indicación del conglomerado al que ha sido asignado cada caso y la distancia euclídea existente entre cada caso y el centro de su conglomerado.

Cluster Membership

| Case Number | Cluster | Distance |
|-------------|---------|----------|
| 1 | 1 | 1.500 |
| 2 | 2 | 3.484 |
| 3 | 1 | 2.198 |
| 4 | 2 | 1.174 |
| 5 | 1 | 3.079 |
| 6 | 1 | 1.597 |
| 7 | 1 | .533 |
| 8 | 1 | 2.508 |
| 9 | 2 | 1.393 |
| 10 | 1 | 1.694 |
| 11 | 1 | 1.800 |
| 12 | 1 | 4.011 |
| 13 | 1 | 3.905 |
| 14 | 2 | 2.510 |
| 15 | 2 | 2.328 |
| 16 | 1 | 2.996 |
| 17 | 1 | 4.011 |
| 18 | 2 | 2.090 |
| 19 | 2 | 2.985 |
| 20 | 2 | 3.702 |
| 21 | 1 | 1.531 |
| 22 | 2 | 3.334 |
| 23 | 2 | 1.240 |
| 24 | 1 | 2.942 |
| 25 | 2 | 1.585 |
| 26 | 2 | 1.517 |
| 27 | 1 | 2.263 |
| 28 | 2 | 2.016 |
| 29 | 1 | 1.744 |
| 30 | 1 | 2.613 |

El siguiente cuadro muestra la distancia euclídea existente entre los centros de los conglomerados finales:

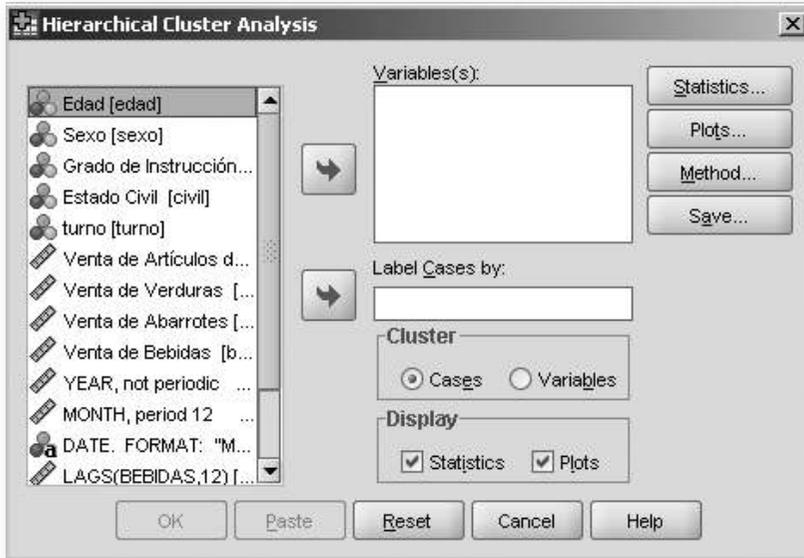
| | | |
|---------|---|--------|
| Cluster | 1 | 17.000 |
| | 2 | 13.000 |
| Valid | | 30.000 |
| Missing | | .000 |

CONGLOMERADOS JERÁRQUICOS

Intenta identificar grupos relativamente homogéneos de casos (o de variables) basándose en las características seleccionadas, mediante un algoritmo que comienza con cada caso (o cada variable) en un conglomerado diferente y combina los conglomerados hasta que sólo queda uno.

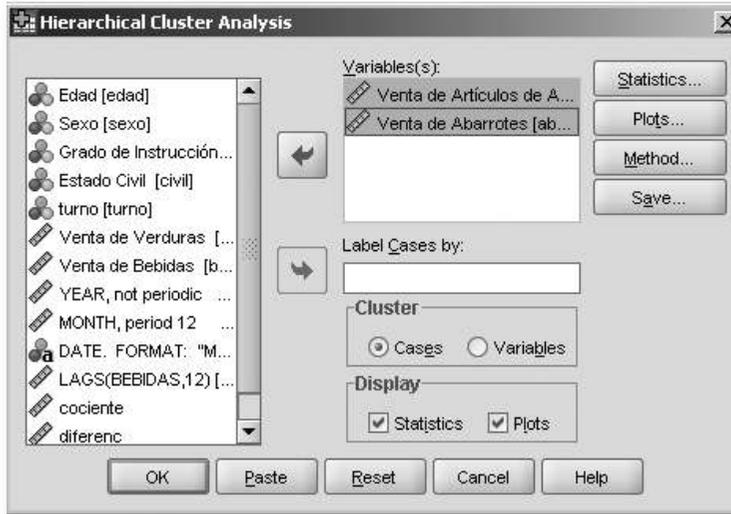
Es posible analizar las variables brutas o elegir de entre una variedad de transformaciones de estandarización. Las medidas de distancia o similitud se generan mediante el procedimiento *Proximidades*. Los estadísticos se muestran en cada etapa para ayudar a seleccionar la mejor solución.

Para llevar a cabo un análisis de conglomerados de K medias, seleccionamos la opción **Analizar -> Clasificar->Conglomerados jerárquicos**, obteniéndose el cuadro de diálogo siguiente:



El cuadro de la izquierda muestra la lista de variables del archivo (numérico y cadena). Pero las variables de cadena sólo sirven para etiquetar casos.

Para obtener un análisis de conglomerados jerárquicos se seleccionan las variables numéricas que se desea utilizar para diferenciar los casos y formar conglomerados, y se traslada al cuadro del lado derecho. En nuestro caso las variables son: Venta de artículos de aseo personal y Venta de abarrotes.



Opcionalmente, se selecciona una variable para identificar los casos en las tablas de resultado y en los gráficos. Luego, trasladamos a **Etiquetar casos Mediante (Label Cases by:)**.

Conglomerar (Clusters): Permiten decidir qué elementos del archivo de datos se desea agrupar.

Casos (Cases): Se agrupan los casos a partir de sus puntuaciones en las variables seleccionadas y es la opción por defecto.

Variables (Variables): Se agrupan las variables seleccionadas en la lista Variables a partir de las puntuaciones de los casos válidos del archivo de datos. Esta opción exige incluir en el análisis al menos tres variables.

Mostrar (Display): Las opciones permiten controlar el tipo de resultado que mostrará el visor de resultados (seleccionadas por defecto).

Estadísticas (Statistics): El visor de resultados sólo muestra las tablas de resultados. Desactivando esta opción se anula el botón Estadísticas.

Gráficos (Plots): El visor de resultados sólo muestra los gráficos.

Luego de realizar estas selecciones presionamos aceptar (OK); el visor presenta los siguientes resultados:

El primer cuadro muestra un resumen de los casos procesados, el número y porcentaje de casos válidos analizados, el número y porcentaje con valores perdidos, el tamaño total de la muestra.

Case Processing Summary^{a,b}

| Cases | | | | | |
|-------|---------|---------|---------|-------|---------|
| Valid | | Missing | | Total | |
| N | Percent | N | Percent | N | Percent |
| 30 | 100.0 | 0 | .0 | 30 | 100.0 |

a. Squared Euclidean Distance used

b. Average Linkage (Between Groups)

La nota “a.” Indica el nombre de la medida utilizada para obtener la matriz de distancia *Squared Euclidean Distance used (Distancia euclídea al cuadrado)*. Y “b.” el método de conglomerados utilizado *Vinculación promedio (Inter-Grupos)*.

El cuadro siguiente muestra el historial del proceso de conglomerados, etapa por etapa. En cada etapa se unen dos elementos. Como la muestra analizada tiene 30 casos, sólo se realizan 29 etapas de fusión.

- La columna *Conglomerados que se combina (Cluster Combinet)* informa sobre los conglomerados (o casos) fundidos en cada etapa. En nuestro ejemplo en la primera etapa se han fundido 12 y 17.

Agglomeration Schedule

| Stage | Cluster Combined | | Coefficients | Stage Cluster First Appears | | Next Stage |
|-------|------------------|-----------|--------------|-----------------------------|-----------|------------|
| | Cluster 1 | Cluster 2 | | Cluster 1 | Cluster 2 | |
| 1 | 12 | 17 | .000 | 0 | 0 | 18 |
| 2 | 3 | 27 | 1.000E-02 | 0 | 0 | 21 |
| 3 | 6 | 10 | 2.000E-02 | 0 | 0 | 15 |
| 4 | 5 | 24 | 5.000E-02 | 0 | 0 | 21 |
| 5 | 18 | 28 | 9.000E-02 | 0 | 0 | 8 |
| 6 | 1 | 21 | 9.000E-02 | 0 | 0 | 16 |
| 7 | 16 | 30 | .250 | 0 | 0 | 13 |
| 8 | 18 | 25 | .275 | 5 | 0 | 12 |
| 9 | 14 | 19 | .340 | 0 | 0 | 12 |
| 10 | 11 | 29 | .610 | 0 | 0 | 17 |
| 11 | 4 | 9 | .680 | 0 | 0 | 16 |
| 12 | 14 | 18 | 1.097 | 9 | 8 | 19 |
| 13 | 16 | 22 | 1.165 | 7 | 0 | 23 |
| 14 | 15 | 23 | 1.210 | 0 | 0 | 22 |
| 15 | 6 | 7 | 1.390 | 3 | 0 | 17 |
| 16 | 1 | 4 | 1.755 | 6 | 11 | 22 |
| 17 | 6 | 11 | 1.825 | 15 | 10 | 23 |
| 18 | 8 | 12 | 2.260 | 0 | 1 | 20 |
| 19 | 14 | 26 | 2.320 | 12 | 0 | 26 |
| 20 | 8 | 13 | 3.290 | 18 | 0 | 26 |
| 21 | 3 | 5 | 3.475 | 2 | 4 | 27 |
| 22 | 1 | 15 | 4.630 | 16 | 14 | 24 |
| 23 | 6 | 16 | 5.153 | 17 | 13 | 24 |
| 24 | 1 | 6 | 7.564 | 22 | 23 | 27 |
| 25 | 2 | 20 | 11.080 | 0 | 0 | 29 |
| 26 | 8 | 14 | 12.213 | 20 | 19 | 28 |
| 27 | 1 | 3 | 19.854 | 24 | 21 | 28 |
| 28 | 1 | 8 | 27.633 | 27 | 26 | 29 |
| 29 | 1 | 2 | 36.753 | 28 | 25 | 0 |

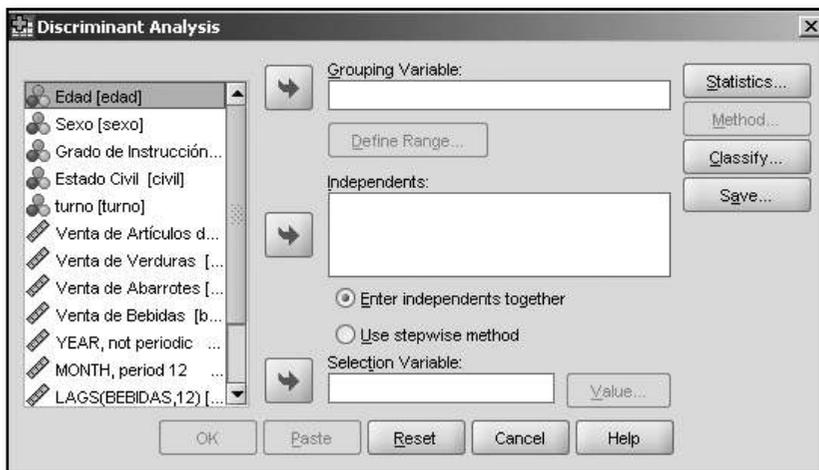
ANÁLISIS DISCRIMINANTE

Resulta útil para las situaciones en las que se desea construir un modelo predictivo para pronosticar el grupo de pertenencia de un caso a partir de las características observadas de cada caso.

El procedimiento genera una función discriminante (o, para más de dos grupos, un conjunto de funciones discriminantes), basada en combinaciones lineales de las variables predictoras que proporcionan la mejor discriminación posible entre los grupos.

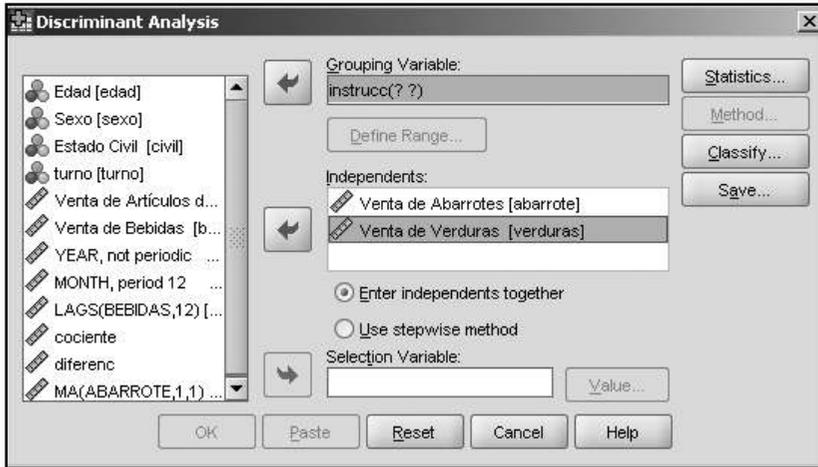
Las funciones se generan a partir de una muestra de casos para los que se conoce el grupo de pertenencia; posteriormente, las funciones pueden ser aplicadas a nuevos casos que dispongan de medidas para las variables predictoras pero de los que se desconozca el grupo de pertenencia. La variable de agrupación puede tener más de dos valores. Los códigos de la variable de agrupación han de ser números enteros y es necesario especificar sus valores máximo y mínimo. Los casos con valores fuera de estos límites se excluyen del análisis.

Para llevar a cabo un análisis discriminante, seleccionamos la opción **Analizar -> Clasificar-> Análisis Discriminante**, obteniéndose el cuadro de diálogo siguiente:

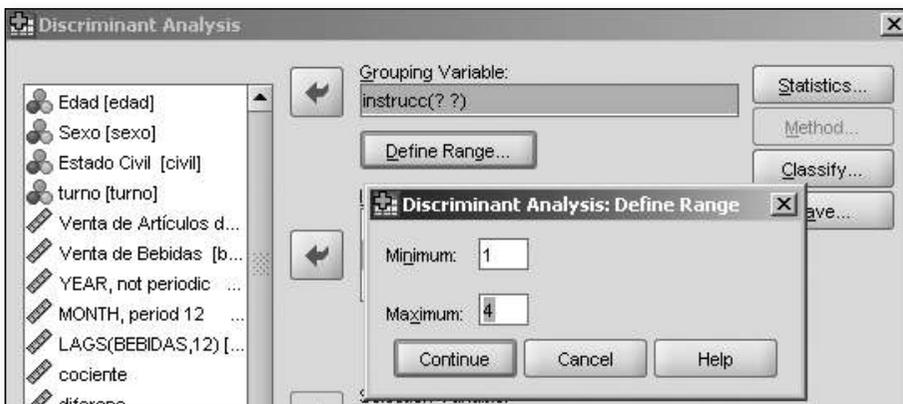


El cuadro de la izquierda muestra la lista de variables del archivo (excepto las que tienen formato cadena). Para obtener un análisis discriminante con las especificaciones que el PASW tiene establecidas por defecto, seleccionamos una variable categórica (nominal u ordinal), y lo trasladamos al cuadro del lado derecho, variable de agrupación (Grouping Variable). La variable de agrupación es aquella que define los grupos que se desea comparar. En nuestro caso será la variable Estado Civil.

Luego, seleccionamos al menos una variable cuantitativa (intervalo o razón) y trasladamos a la lista independiente (independents). Las variables independientes son aquellas en las que se desea comparar los grupos. En nuestro caso será la Venta de abarrotes y Venta de verduras, como se muestra en el siguiente gráfico:



Luego debemos pulsar en el botón *Definir rango...* (*Define Range...*) para acceder al cuadro de diálogo siguiente. En nuestro ejemplo la variable Estado Civil tiene rango mínimo 1 y rango máximo 4.



Una vez ingresados los rangos, pulsar el botón *continuar* para volver al cuadro de diálogo principal.

Luego de realizar estas selecciones presionamos aceptar (OK); el visor presenta los siguientes resultados:

El primer cuadro muestra un resumen con el total de los casos procesados, el número de casos válidos para el análisis y el número de casos excluidos. Dentro de los casos excluidos se distinguen entre los que son excluidos porque su código en la variable de agrupación no está dentro del rango seleccionado, los que son excluidos por tener un valor perdido en al menos una variable discriminante, y los que cumplen las dos condiciones anteriores.

Analysis Case Processing Summary

| Unweighted Cases | | N | Percent |
|------------------|---|----|---------|
| Valid | | 30 | 100.0 |
| Excluded | Missing or out-of-range group codes | 0 | .0 |
| | At least one missing discriminating variable | 0 | .0 |
| | Both missing or out-of-range group codes and at least one missing discriminating variable | 0 | .0 |
| | Total | 0 | .0 |
| Total | | 30 | 100.0 |

El cuadro siguiente muestra un resumen del número de casos válidos en cada variable discriminante. La información de esta tabla posee un interés especial, pues un número desigual de casos en cada uno de los grupos puede afectar la clasificación.

Group Statistics

| Estado Civil | | Valid N (listwise) | |
|--------------|--------------------|--------------------|----------|
| | | Unweighted | Weighted |
| Soltero | Venta de Abarrotes | 5 | 5.000 |
| | Venta de Verduras | 5 | 5.000 |
| Casado | Venta de Abarrotes | 12 | 12.000 |
| | Venta de Verduras | 12 | 12.000 |
| Divorciado | Venta de Abarrotes | 4 | 4.000 |
| | Venta de Verduras | 4 | 4.000 |
| Viudo | Venta de Abarrotes | 9 | 9.000 |
| | Venta de Verduras | 9 | 9.000 |
| Total | Venta de Abarrotes | 30 | 30.000 |
| | Venta de Verduras | 30 | 30.000 |

El cuadro siguiente contiene los autovalores y algunas estadísticas descriptivas Multivariantes. En la tabla aparece una fila numerada por cada función discriminante, en nuestro ejemplo hay dos funciones. La primera función explica el 89.8% de las diferencias existentes entre los sujetos de los grupos. Y la segunda función explica 10.2% de las diferencias existentes entre los sujetos de los grupos.

Eigenvalues

| Function | Eigenvalue | % of Variance | Cumulative % | Canonical Correlation |
|----------|-------------------|---------------|--------------|-----------------------|
| 1 | .182 ^a | 89.8 | 89.8 | .392 |
| 2 | .021 ^a | 10.2 | 100.0 | .142 |

a. First 2 canonical discriminant functions were used in the analysis.

[% de variancia (% of Variance) % acumulado (Cumulative %)]

- **El autovalor** (Eigenvalue) permite comparar cómo se distribuye la dispersión Inter-grupos cuando existe más de una función. Un autovalor tiene un valor mínimo de cero pero no tiene un máximo, lo cual lo hace difícilmente interpretable por sí sólo. Por esta razón, se utiliza la estadística lambda de Wilks que se encuentra estrechamente relacionada con los autovalores.

- **La correlación canónica** (canonical Correlation) es la correlación entre la combinación lineal de las variables independientes (función discriminante) y una combinación lineal de variables indicador que recogen la pertenencia de los sujetos a los grupos. Una correlación canónica alta indica que las variables discriminantes permiten diferenciar entre los grupos.

En el caso de nuestro ejemplo el autovalor es próximo a cero y la correlación canónica es moderada por lo que debemos suponer que las variables discriminantes utilizadas (Venta de abarrotes y Venta de verduras) no permiten distinguir demasiado bien entre los grupos en ambas funciones.

El cuadro siguiente muestra el estadístico *lambda de Wilks* que expresa la proporción de variabilidad total no debida a las diferencias entre los grupos; permite contrastar la hipótesis nula de que las medias multivariantes de los grupos (los centroides) son iguales.

Wilks' Lambda

| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
|---------------------|---------------|------------|----|------|
| 1 through 2 | .829 | 4.872 | 6 | .560 |
| 2 | .980 | .533 | 2 | .766 |

Por tanto, valores de *lambda de Wilks* cercanos a **1** indican un gran parecido entre grupo, mientras que valores próximo a cero indican una gran diferencia entre ellos. En nuestro ejemplo, los valores cercanos a 1 indican un gran parecido entre grupos o existe bastante solapamiento entre los grupos. Sin embargo, el valor transformado de *lambda* (*Chi-Square*) es de 4.872 con 6 grados de libertad y un nivel crítico (Sig) de 0.560, por lo que podemos aceptar la hipótesis nula de que los grupos comparados tienen promedios iguales en las dos variables discriminantes en la primera función. Para la segunda función el *Chi-Square* = 0.533 con 2 grados de libertad y un nivel crítico (Sig) de 0.766, por lo que podemos aceptar la hipótesis nula de que los grupos comparados tienen promedios iguales en las dos variables discriminantes.

El cuadro siguiente muestra la tabla de *coeficientes estandarizados* que contiene una versión estandarizada de los coeficientes de la función canónica discriminante:

Standardized Canonical Discriminant Function Coefficients

| | Function | |
|--------------------|----------|-------|
| | 1 | 2 |
| Venta de Abarrotes | 1.003 | .163 |
| Venta de Verduras | .016 | 1.016 |

En la función 1 la Venta de abarrotes tiene mayor importancia que las ventas de verduras y en la función 2 la Venta de verduras tiene mayor importancia que las ventas de abarrotes.

El cuadro siguiente muestra la *Matriz de estructura* que contiene las correlaciones entre las variables discriminantes y la función discriminante estandarizada.

Structure Matrix

| | Function | |
|--------------------|----------|-------|
| | 1 | 2 |
| Venta de Abarrotes | 1.000* | -.016 |
| Venta de Verduras | -.161 | .987* |

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions
 Variables ordered by absolute size of correlation within function

*. Largest absolute correlation between each variable and any discriminant function

Podemos apreciar que la Venta de abarrotes correlaciona alto con la función discriminante uno. Que la Venta de verduras correlaciona alto con la función discriminante Dos.

El cuadro siguiente muestra la ubicación de *los centroides* en la función discriminante. Esta tabla es de gran utilidad para interpretar la función discriminante.

Functions at Group Centroids

| Estado Civil | Function | |
|--------------|----------|-----------|
| | 1 | 2 |
| Soltero | .886 | -1.43E-02 |
| Casado | -.157 | .153 |
| Divorciado | -.176 | -7.67E-02 |
| Viudo | -.204 | -.162 |

Unstandardized canonical discriminant functions evaluated at group means

Podemos observar que el grupo de Solteros se encuentra localizado en promedio, en las puntuaciones (ventas) positivas de la primera función, mientras que los Casados, Divorciados y Viudos, se encuentran en las puntuaciones (ventas) negativas respecto de la primera función.

ANÁLISIS FACTORIAL

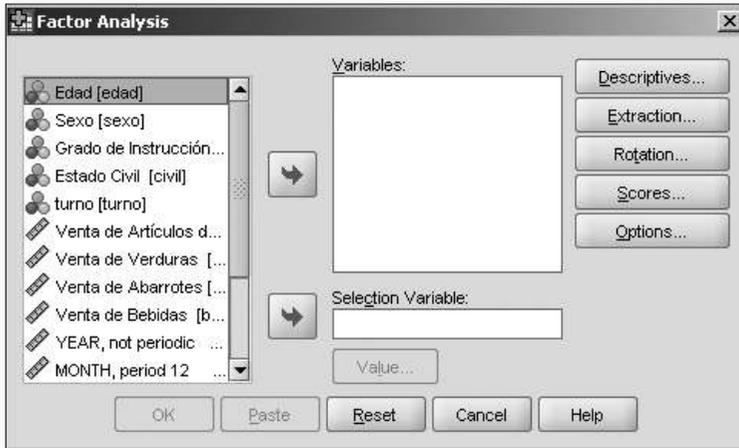
Intenta identificar variables subyacentes o factores que expliquen la configuración de las correlaciones dentro de un conjunto de variables observadas. Se suele utilizar en la reducción de los datos para identificar un pequeño número de factores que expliquen la mayoría de la varianza observada en un número mayor de variables manifiestas.

También puede utilizarse para generar hipótesis relacionadas con los mecanismos causales o para inspeccionar las variables para análisis subsiguientes (por ejemplo, para identificar la colinealidad antes de realizar un análisis de regresión lineal). Este procedimiento dispone de 7 métodos de extracción factorial; 5 métodos de rotación, entre ellos el oblimin directo y el promax para rotaciones no ortogonales; y 3 métodos para calcular las puntuaciones factoriales, que pueden guardarse como variables para análisis adicionales.

El análisis de factorial consta de cuatro fases características:

- El cálculo de una matriz capaz de expresar la variabilidad conjunta de todas las variables.
- La extracción del número óptimo de factores.
- La rotación de la solución para facilitar su interpretación.
- La estimación de las puntuaciones de los sujetos en las nuevas dimensiones.

Para llevar a cabo un análisis de factorial, seleccionamos la opción **Analizar** -> **Reducción de Datos** -> **Análisis factorial**, obteniéndose el cuadro de diálogo siguiente:



El cuadro de la izquierda muestra la lista de todas las variables del archivo (también las que tienen formato cadena aunque éstas sólo pueden utilizarse como variables de selección).

Para llevar a cabo Análisis de factorial con PASW seleccionamos el conjunto de variables que se desea analizar, y se trasladamos al cuadro del lado derecho lista de **Variables** (*Variables*):

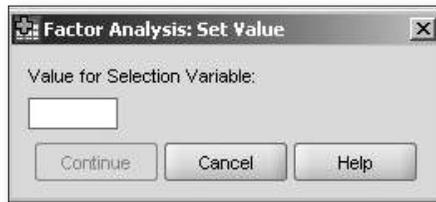




En nuestro ejemplo las variables Edad, Grado de instrucción, Estado Civil, Turno, Venta de verduras, Venta de abarroses y Venta de bebidas, fueron seleccionadas para realizar Análisis Factorial.

La **Variable de Selección** (*Selection Variable*): Este cuadro permite seleccionar una de las variables del archivo de datos como variable *filtro*: para definir una submuestra de sujetos que cumplen una determinada condición.

Si trasladamos la variable al cuadro Variable de selección y pulsamos en el botón **Valor** (*Value...*), se obtiene el cuadro de diálogo siguiente:



Introducir en el cuadro el valor de la variable de selección que identifica a los casos que se desea incluir en el análisis.

Sigamos con nuestro ejemplo. Luego de seleccionar las variables mencionadas, aceptamos (OK); el visor de resultados muestra los siguientes cuadros:

El cuadro siguiente contiene las comunalidades (Communalities) asignadas inicialmente a las variables (initial) y las comunalidades reproducidas por la solución factorial (Extraction). La comunalidad de una variable es la proporción de su varianza que puede ser explicada por el modelo factorial obtenido.

Communalities

| | Initial | Extraction |
|----------------------|---------|------------|
| Edad | 1.000 | .911 |
| Grado de Instrucción | 1.000 | .802 |
| Estado Civil | 1.000 | .702 |
| turno | 1.000 | .903 |
| Venta de Verduras | 1.000 | .452 |
| Venta de Abarrotes | 1.000 | .739 |
| Venta de Bebidas | 1.000 | .722 |

Extraction Method: Principal Component Analysis.

En nuestro ejemplo, la variable Venta de verduras es la peor explicada: el modelo sólo es capaz de reproducir el 45.2% de su variabilidad original. Para la solución factorial se ha utilizado el método de extracción de componentes principales. También podemos plantear en este momento, si dando por bueno el número de factores extraídos, alguna de las variables incluidas podría quedar fuera del análisis.

El cuadro siguiente muestra la tabla de *porcentajes de varianza explicada*. Se ofrece un listado de los autovalores (Eigenvalues) de la matriz varianza-covarianza y del porcentaje de varianza que representa cada uno de ellos. Los autovalores expresan la cantidad de la varianza total que está explicada por cada factor. Por defecto, se extraen tantos factores como autovalores mayores que 1 tiene la matriz analizada.

Total Variance Explained

| Component | Initial Eigenvalues | | | Extraction Sums of Squared Loadings | | |
|-----------|---------------------|---------------|--------------|-------------------------------------|---------------|--------------|
| | Total | % of Variance | Cumulative % | Total | % of Variance | Cumulative % |
| 1 | 2.513 | 35.894 | 35.894 | 2.513 | 35.894 | 35.894 |
| 2 | 1.716 | 24.519 | 60.413 | 1.716 | 24.519 | 60.413 |
| 3 | 1.001 | 14.297 | 74.710 | 1.001 | 14.297 | 74.710 |
| 4 | .744 | 10.622 | 85.332 | | | |
| 5 | .584 | 8.346 | 93.678 | | | |
| 6 | .414 | 5.912 | 99.591 | | | |
| 7 | 2.864E-02 | .409 | 100.000 | | | |

Extraction Method: Principal Component Analysis.

En nuestro ejemplo hay tres autovalores mayores que uno, por lo que el procedimiento extrae 3 factores que consiguen explicar el 74.71% de la varianza de los datos originales. La tabla también muestra, para cada factor con autovalor mayor que 1, la suma de las saturaciones al cuadrado.

La información de esta tabla puede utilizarse para tomar una decisión sobre el número idóneo de factores que deben extraerse. Por ejemplo, para un mínimo del 85% de la variabilidad contenida en los datos sería necesario obtener cuatro factores.

El cuadro siguiente muestra la tabla donde se encuentra la solución factorial propiamente dicha. Contiene las correlaciones entre las variables originales (o saturaciones) y cada uno de los factores.

Esta matriz se llama matriz de componentes porque el método utiliza componentes principales.

Component Matrix

| | Component | | |
|----------------------|-----------|-------|-------|
| | 1 | 2 | 3 |
| Edad | .938 | .175 | .031 |
| Grado de Instrucción | .010 | .713 | .541 |
| Estado Civil | .325 | .585 | -.504 |
| turno | .935 | .166 | .032 |
| Venta de Verduras | -.650 | .140 | -.098 |
| Venta de Abarrotes | .297 | -.576 | .565 |
| Venta de Bebidas | -.377 | .675 | .351 |

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

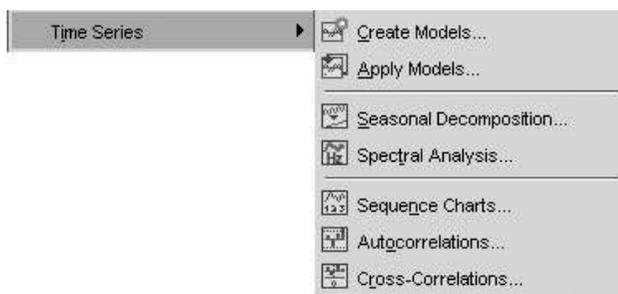
En el ejemplo, comparando las saturaciones relativas de cada variable en cada uno de los tres factores, se observa que el primero está constituido por las variables Edad (0.938) y Turno (0.935); el segundo factor está constituido por las variables Grado de instrucción, Estado civil, Venta de verduras y Venta de bebidas, y tercer factor está constituido por la variable Venta de abarroses.

Capítulo 14

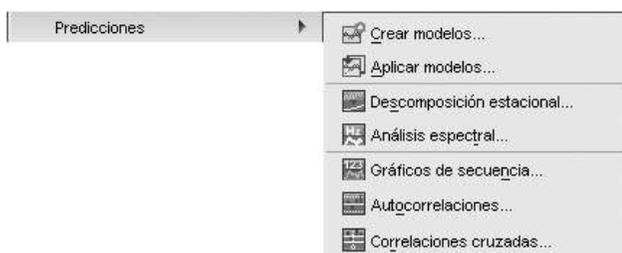
Series temporales

Podemos definir una serie temporal como una sucesión de observaciones cuantitativas de un fenómeno ordenadas en el tiempo. Para la realización de un análisis clásico descriptivo de una serie temporal con PASW se tiene los siguientes procedimientos estadísticos:

Inglés



Español





Se define los submenús más importantes:

SUAVIZADO EXPONENCIAL

Suaviza componentes irregulares de datos de series temporales; para ello hace uso de una variedad de modelos que incorporan diferentes supuestos acerca de la tendencia y la estacionalidad. Seleccionando un modelo estacional y definiendo la periodicidad mediante *Definir fechas*, se puede seleccionar una variable que represente los factores estacionales.

Los modelos difieren en los componentes estacionales y de tendencia. Se dispone de los modelos Simple, Holt, Winters y Personalizado.

AUTORREGRESIÓN

Estima un modelo de regresión lineal con errores autorregresivos de primer orden. Los métodos disponibles son: Máxima verosimilitud exacta, Cochrane-Orcutt y Prais-Winsten.

Se debe desactivar *Incluir constante en el modelo* si no se desea estimar un término constante en la ecuación de regresión.

SEQUENCE CHARTS

Crea un gráfico de casos en secuencia. Este procedimiento requiere datos de serie temporal u otros en los que se ordenen los casos con un criterio significativo. Representa, en dos dimensiones, los valores de cada una de las variables especificadas frente al número de secuencia en el archivo de datos. Como hemos dicho, si la variable es una serie temporal a cuyos valores se han asignado fechas (opción *Definir fechas*), la representación se realiza frente al instante de tiempo correspondiente.

ARIMA

Estima modelos Arima (Modelo Autorregresivo Integrado de Media Móvil) univariados estacionales y no estacionales (también conocidos como modelos “Box-Jenkins”).

Se pueden transformar las series antes de la estimación mediante una de las opciones de la lista *Transformar*:

- *Ninguna*. No se lleva a cabo ninguna transformación.
- *Log natural*. Transforma las series antes de la estimación con el logaritmo en base e.
- *Log base 10*. Transforma las series antes de la estimación con el logaritmo en base 10.

El grupo *Modelo* permite especificar los tres parámetros del modelo Arima:

Autorregresivo, *Diferenciación* y *Media móvil*. Estos parámetros se conocen de forma más común como p, d y q, respectivamente.

También se pueden definir los parámetros estacionales correspondientes, si se tiene un modelo estacional y se ha definido la estacionalidad de los datos en *Definir fechas*. Además, si se ha definido la periodicidad (que se muestra en la parte inferior del cuadro), se pueden especificar los parámetros estacionales correspondientes introduciendo los valores en los cuadros sp, sd y sq estacionales.

DESCOMPOSICIÓN ESTACIONAL

Estima factores estacionales multiplicativos o aditivos para las series temporales. Precisa que esté definida la periodicidad mediante *Definir fechas*. Permite seleccionar el tipo de modelo utilizado para la descomposición estacional: Multiplicativo o Aditivo.

Las opciones de *Ponderación de la media móvil* permiten especificar la manera de tratar las series en el cálculo de medias móviles. Estas opciones sólo están disponibles si la periodicidad de la serie es par (si es impar, todos los puntos son ponderados por igual):

- *Todos los puntos por igual* calcula las medias móviles con una amplitud igual a la periodicidad y todos los puntos ponderados por igual.
- *Puntos finales ponderados por 5* calcula las medias móviles con una amplitud igual a la periodicidad más uno y con los puntos finales ponderados por 0,5. Mediante *Mostrar el listado por casos* se imprime un listado por casos que contiene un resumen en una línea en cada iteración, así como los estadísticos finales.

Ejercicio N° 1

Se trabajará con el archivo de la encuesta del supermercado.

Serie temporal

Se puede definir una serie temporal como una sucesión de observaciones cuantitativas de un fenómeno ordenadas en el tiempo. Toda serie temporal, desde el enfoque del análisis clásico, puede ser descompuesta en cuatro componentes principales:

Ciclo:

Representa las oscilaciones de la serie que se producen con un período superior al año.

Tendencia:

Representa el movimiento general a largo plazo de una serie temporal.

Estacionalidad:

Representa oscilaciones que se producen con una periodicidad igual o inferior al año, y que se reproducen de manera reconocible cada año.

Residuo:

Movimiento que no muestra un carácter periódico reconocible.

El objeto del análisis de series temporales será la determinación de cada una de estas componentes con la finalidad de realizar predicciones o conocer de manera descriptiva el comportamiento de la serie temporal.

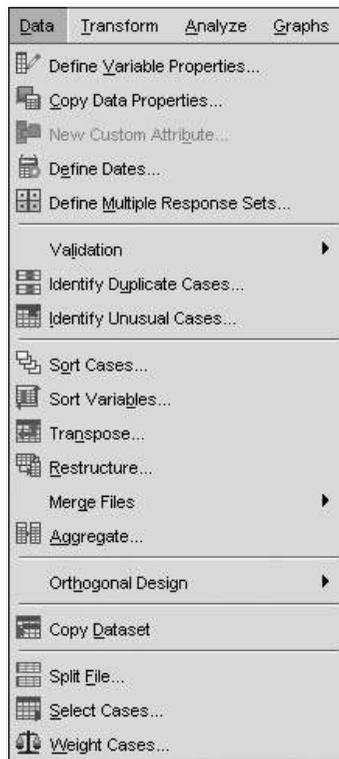
SERIES TEMPORALES CON PASW

Si queremos realizar un análisis clásico descriptivo de una serie temporal con PASW habrá que seguir los pasos siguientes:

- Definición de fechas.
- Representación gráfica de una serie temporal.
- Distinción entre los esquemas aditivos y multiplicativos de una serie temporal.
- Determinación de la tendencia.
- Determinación de la estacionalidad.

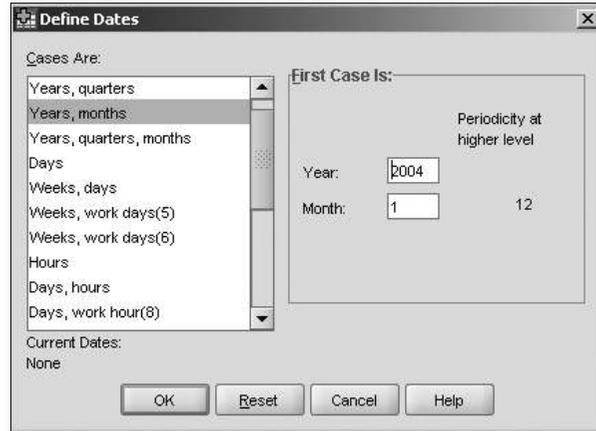
Definición de fechas

Una serie temporal exige que las observaciones del fenómeno objeto de estudio estén ordenadas en el tiempo, esto debido a la definición de serie temporal. El primer paso que hay que dar para la realización de un análisis clásico descriptivo de una serie temporal pasa por la necesidad de la definición de las fechas en el PASW. El procedimiento a seguir para la definición de las fechas de la serie temporal se definirá utilizando el comando **Definir fechas**, el cual se ejecuta al hacer un solo clic con el botón izquierdo del ratón cuando el puntero del mismo se encuentre sobre la opción del mismo nombre que aparece al desplegarse el menú **Datos (Datos / Definir fechas)**.



El procedimiento *Definir fechas* de PASW nos permite generar variables de fecha que se utilizarán para establecer la periodicidad de una serie temporal y para etiquetar los resultados del análisis de series temporales

Luego de la ejecución de este procedimiento se obtendrá el cuadro de diálogo representado en la siguiente figura (Cuadro de diálogo *Definir fechas*):



En el recuadro izquierdo (recuadro con encabezamiento *Los casos son*) de este cuadro de diálogo seleccionamos el formato de fecha deseado.

Una vez realizada esta operación habrá que indicar la fecha de la primera observación de la serie en el recuadro. *El primer caso es (First Case Is)*.

Finalmente, se hará clic con el botón izquierdo del ratón cuando el puntero se sitúe sobre *Aceptar (OK)*.

Una vez realizada la operación descrita en el párrafo anterior, en el visor de resultados aparecerá una anotación en la que se indica el nombre y la etiqueta de las nuevas variables creadas a consecuencia de la operación realizada (*year*, *month* y *date*), como se muestra en el resultado siguiente:

The following new variables are being created:

| Name | Label |
|--------|--------------------------|
| YEAR_ | YEAR, not periodic |
| MONTH_ | MONTH, period 12 |
| DATE_ | DATE. FORMAT: "MMM YYYY" |

En el editor de datos aparecerán los respectivos valores de estas variables, como se muestra en la siguiente tabla:

| | aseo | verduras | abarrotes | bebidas | year | month | date_ |
|----|------|----------|-----------|---------|------|-------|----------|
| 1 | 2.00 | 17.00 | 34.50 | 6.00 | 2004 | 1 | JAN 2004 |
| 2 | 3.00 | 20.00 | 40.00 | 6.00 | 2004 | 2 | FEB 2004 |
| 3 | .30 | 10.00 | 31.60 | 6.00 | 2004 | 3 | MAR 2004 |
| 4 | 3.30 | 9.00 | 35.40 | 6.00 | 2004 | 4 | APR 2004 |
| 5 | 1.30 | 23.00 | 30.00 | 6.00 | 2004 | 5 | MAY 2004 |
| 6 | .40 | 13.00 | 32.90 | 6.00 | 2004 | 6 | JUN 2004 |
| 7 | 1.50 | 12.00 | 33.20 | 6.00 | 2004 | 7 | JUL 2004 |
| 8 | 4.50 | 19.00 | 33.10 | 6.00 | 2004 | 8 | AUG 2004 |
| 9 | 2.50 | 18.00 | 35.60 | 6.00 | 2004 | 9 | SEP 2004 |
| 10 | .30 | 24.00 | 33.00 | 6.00 | 2004 | 10 | OCT 2004 |

Las variables creadas en el editor de datos al definir fechas (datos mensuales) son: *year*, *month* y *date*.

Si se realiza un análisis clásico descriptivo de una serie temporal resultaría incompleto si no se lleva a cabo la representación gráfica de la serie temporal, la cual persigue identificar patrones de comportamiento regulares que permitan plantear hipótesis sobre la estacionalidad y el ciclo, el esquema según el cual combinan las componentes de la serie temporal o la existencia de outliers o valores atípicos en la serie temporal que adviertan de errores en la recolección de los datos o de los fenómenos imprevistos de un impacto excepcionalmente grande en los valores de la variable.

Representación gráfica de una serie temporal

Para la representación gráfica de una serie temporal en PASW se debe seleccionar el comando **Secuencia** de PASW.

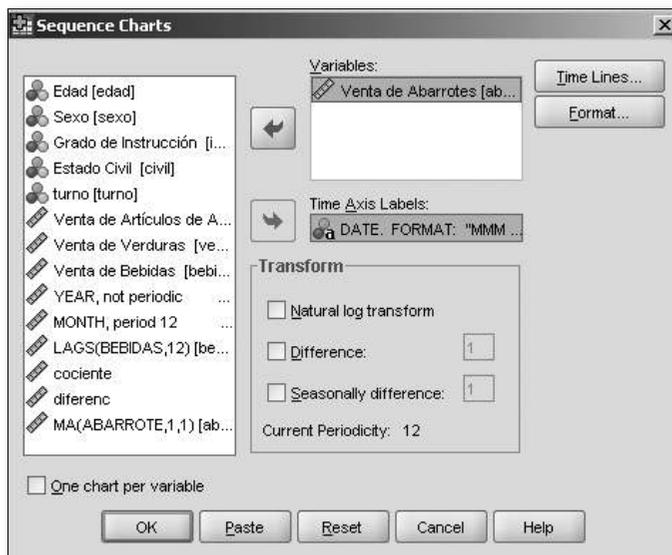
Dicho comando se ejecuta al hacer un solo clic con el botón izquierdo del ratón cuando el puntero del mismo se encuentre sobre la opción del mismo nombre que aparece al desplegarse el menú **Análisis** de la barra de menú de PASW (**Serie de tiempo / Gráfico de Secuencia**).

La ejecución de este procedimiento requiere disponer de datos de serie temporal o bien de otro tipo de datos en los cuales los casos aparezcan ordenados según un criterio significativo.

Una vez ejecutado el comando *Secuencia*, aparece en la pantalla el cuadro de diálogo *Gráficos de secuencia*.

En este cuadro de diálogo se introducirá en el recuadro *Variables* la variable o variables cuantitativas a representar y en el recuadro *Etiquetas del eje de tiempo* una variable categórica o numérica la cual se empleará para etiquetar los ejes temporales de la representación o representaciones gráficas.

El cuadro de dialogo de **Gráficos de secuencia** se presenta en la siguiente figura:



Cuadro de diálogo Gráficos de secuencia

Finalmente, sin modificar el resto de opciones que aparecen por defecto en el cuadro de diálogo, se hará clic con el botón izquierdo del ratón cuando el puntero se encuentre sobre el botón *Aceptar* para ejecutar el comando **Secuencia** que nos mostrará en el visor de resultados el gráfico de la serie temporal.

Distinción entre el esquema aditivo y esquema multiplicativo de una serie temporal

En lo referente al esquema según el cual se combinan las componentes de la serie temporal, pese a la existencia de métodos analíticos para la determinación del mismo, se va a optar por el método gráfico, el cual consiste en identificar el esquema de la serie a través de la observación de la representación gráfica de la misma. De esta forma, una serie posee un esquema aditivo si la variabilidad de la misma permanece constante a lo largo del tiempo tal y como se muestra en la figura:

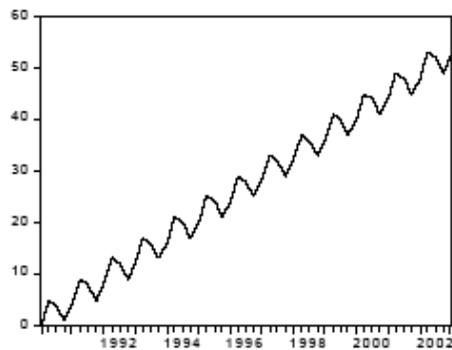


Figura: Gráfica de una serie aditiva

Y se dice que la serie posee un esquema multiplicativo cuando esta variabilidad aumenta con el tiempo tal y como se muestra en la figura:

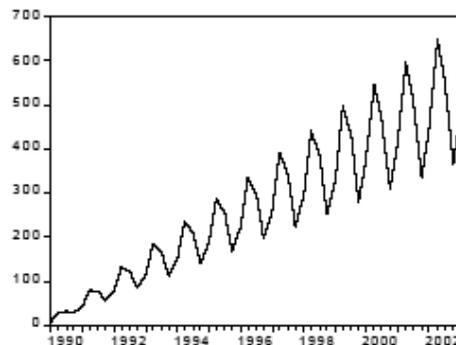


Figura: Gráfica de una serie multiplicativa

En la práctica, para diferenciar los modelos de series aditivos y multiplicativos se utilizan las técnicas del Método Analítico de las Diferencias y Cocientes Estacionales y el Método Gráfico.

Método Analítico de las Diferencias y Cocientes Estacionales

Esta técnica permite determinar el tipo de esquema según el cual se combinan las componentes del análisis clásico de series temporales mediante la comparación de los coeficientes de variación de Pearson de las series de las diferencias ($d = y_t - y_{tp}$) y de los cocientes [$C = y_t / y_{tp}$] estacionales.

$$CV_d = \frac{S_d}{\bar{Y}_d} \quad \text{ó}$$

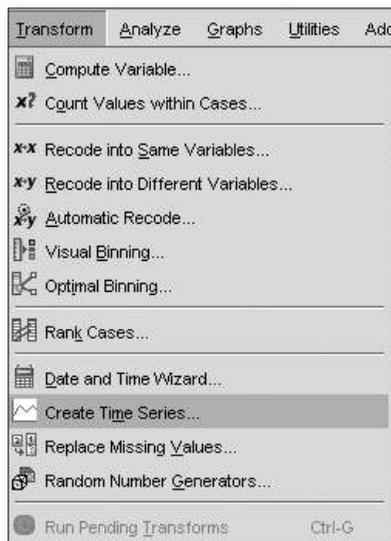
$$CV_d = \frac{S_d}{\bar{Y}_d}$$

De esta forma, el esquema será aditivo o multiplicativo de acuerdo a la siguiente regla de decisión:

- Si $CV_d > CV_c$ entonces el esquema es multiplicativo.
- Si $CV_c > CV_d$ entonces el esquema es aditivo.

Los pasos a seguir en PASW son los que se indican a continuación:

- Seleccionar como valor de p aquel que represente el período de oscilaciones más importante de la serie (12 si la serie es mensual, 4 si es trimestral, 3 si es cuatrimestral...).
- Crear la variable de retardos y_{tp} .
Para su obtención habrá que hacer un solo clic con el botón izquierdo del ratón sobre la opción **Crear serie temporal** que se despliega cuando su puntero se encuentra situado sobre el menú Transformar de la barra de menú de PASW (**Transformar / Crear serie temporal**). Seleccionar la función retardo y elegir como orden el valor de p fijado previamente en el paso 1.



A continuación, añadir a la ventana *nuevas variables* la variable $y = \text{Bebida}$. Tras realizar esta operación se activa el cuadro de diálogo que se representa en la figura:

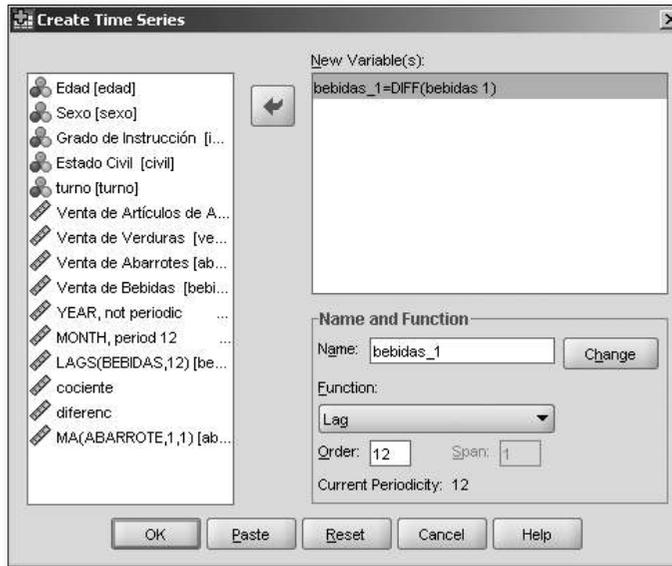


Figura: Obtención de la serie de los retardos

3. Crear la serie de las diferencias y de los cocientes.

Para su obtención habrá que hacer un solo clic con el botón izquierdo del ratón sobre la opción **Calcular** que se despliega cuando su puntero se encuentra situado sobre el menú **Transformar** de la barra de menú de PASW (**Transformar/ Calcular**).

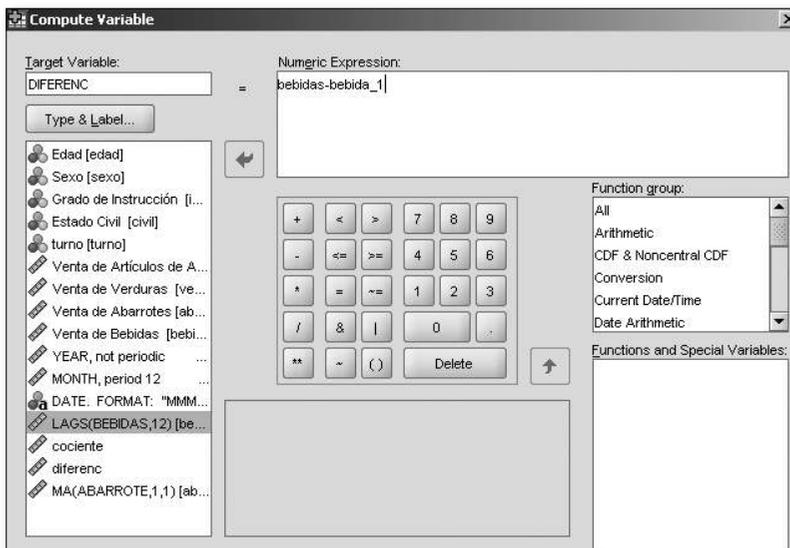


Figura: Obtención de la serie de diferencias

A continuación, se crean dos variables nuevas a partir de las variables ya existentes $y = \text{Bebida}$ e $y = \text{Bebida}_1$ llamadas *diferencia*, la cual se obtiene como diferencia de las variables $y_t - y_{tp}$ y *cociente*, obtenida con el cociente y_t / y_{tp} .

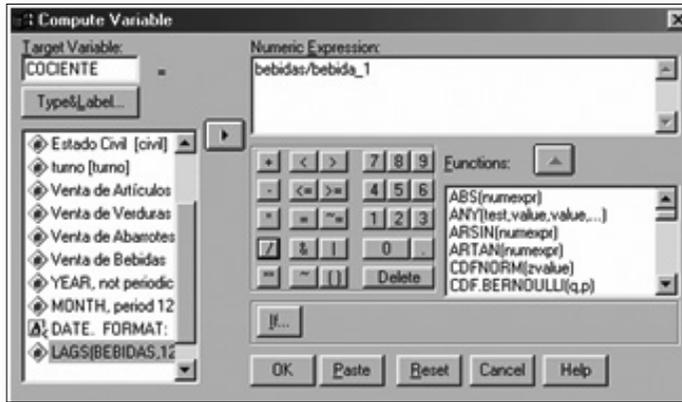


Figura: Obtención de la serie de los cocientes

4. Calcular el coeficiente de variación de Pearson a partir de la media y desviación típica de las variables *diferencia* y *cociente* ya obtenidas en el paso 3.

Para su obtención habrá que hacer un solo clic con el botón izquierdo del ratón sobre la opción **Estadísticos Descriptivos** que se despliega cuando su puntero se encuentra situado sobre el menú **Analizar** de la barra de menú de PASW (**Analizar/ Estadísticos Descriptivos/ Descriptivos**). Tras realizar esta operación se activa el cuadro de diálogos que se representa en la figura:

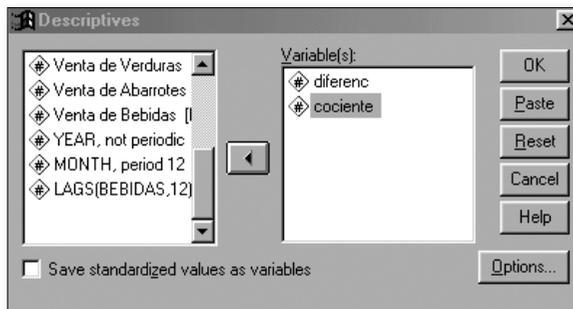


Figura: Obtención de los estadísticos media y desviación típica.

Finalmente, seleccionar con el botón *Opciones* los estadísticos *media* y *desviación* típica para las variables *diferencia* y *cociente*.

Descriptive Statistics

| | N | Minimum | Maximum | Mean | Std. Deviation |
|--------------------|----|---------|---------|---------|----------------|
| DIFERENC | 18 | -5.00 | 2.00 | -1.8333 | 2.6624 |
| COCIENTE | 18 | .17 | 3.00 | .9907 | .9540 |
| Valid N (listwise) | 18 | | | | |

Método Gráfico de la Media y de la Desviación Típica

Esta técnica consiste en representar gráficamente la media aritmética y la desviación típica de grupos de observaciones de la serie temporal.

Si en el gráfico se observa (no se observa) una clara tendencia creciente o decreciente, entonces el esquema según el cual se combinan las componentes de la serie será multiplicativo (aditivo).

Determinación de la tendencia de una serie temporal

Para la determinación de la tendencia de la serie temporal se pueden usar alternativamente dos métodos: El Método del Ajuste Analítico y el Método de la Media Móvil.

Método de Ajuste Analítico:

La idea fundamental de este método consiste en ajustar una curva a las medias aritméticas anuales de las observaciones que componen la serie temporal por el procedimiento de los Mínimos Cuadrados Ordinarios (MCO). Para la aplicación de esta metodología en PASW habrá que seguir los pasos:

1. Con el objeto de eliminar el efecto producido por la componente estacional que distorsiona la tendencia, se hace necesario tomar medias anuales, para lo que tendremos que recurrir al comando **Agregar**.

Para ejecutar el comando **Agregar** habrá que hacer un solo clic con el botón izquierdo del ratón cuando el puntero del mismo se encuentra sobre el menú Datos de la barra de menú (**Datos Agregar**).

Tras ejecutar el comando *Agregar*, aparece el cuadro de diálogo representado en la figura:

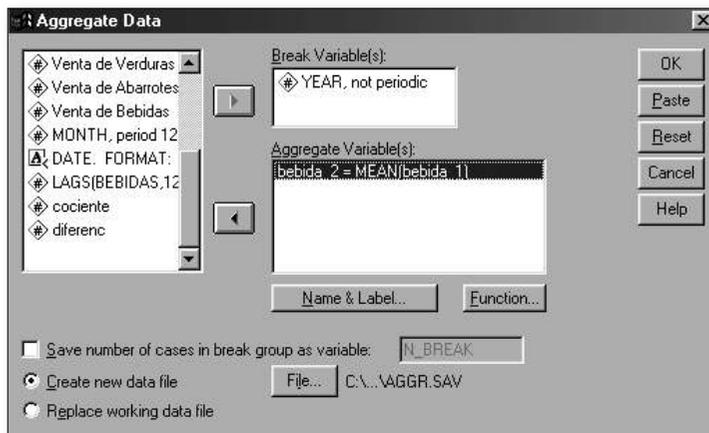


Figura: Cuadro de diálogo Agregar datos

En el cuadro de diálogo *Agregar datos* se introducirá la/s variable/s que van a definir los grupos a agregar en el recuadro *Variables de segmentación*, mientras que en el recuadro *Agregar variables* se introducirá la/s variable /s que van a ser agregadas indicando el nombre y etiquetas y la función de agregación (en este caso la media, la cual aparece seleccionada por defecto) de las mismas cuando estén seleccionadas, haciendo clic con el botón izquierdo del ratón con su puntero sobre el botón *Nombre y etiqueta* y *Función*, respectivamente.

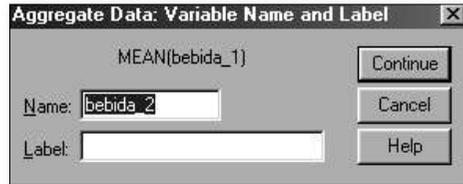


Figura: Nombre y etiquetas de variables del cuadro de diálogo *Agregar datos*.

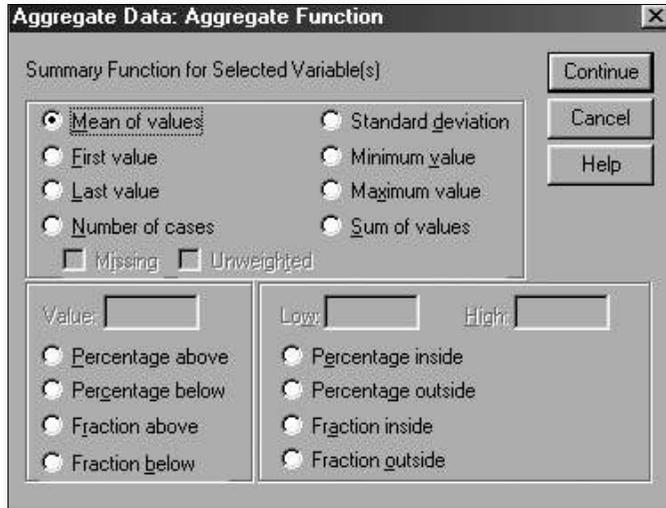


Figura : Función de agregación del cuadro de diálogo *Agregar datos*

Una vez definido el nombre y etiqueta de la nueva variable creada y la función de agregación de la misma se ejecutará el comando haciendo un clic con el botón izquierdo del ratón cuando su puntero se encuentre situado sobre el botón **Aceptar** del cuadro de diálogo **Agregar**.

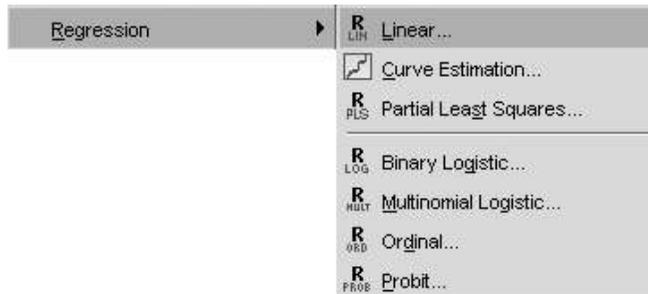
El resultado de la operación descrita en el párrafo anterior será guardado en la ubicación del disco duro que se indique antes de hacer clic en Aceptar haciendo clic con el botón izquierdo del ratón cuando su puntero se encuentre situado sobre el botón Archivo de este cuadro de diálogo.

Tras activar este botón aparecerá en la pantalla el cuadro de diálogo representado en la figura, en el que se indicará la ubicación en la que se desea guardar el resultado de la operación y el nombre que recibirá el archivo que recoge esta operación.

Por ejemplo, al nuevo archivo se le puede denominar *Nombre Supermerdo_agregado.sav*.

2. Tras abrir el nuevo archivo creado (**Supermerdo_agregado**), se procederá a aplicar un análisis de la regresión en el que la variable dependiente sea la variable que recoge las medias anuales (Bebidas) y la variable independiente la que representa el tiempo (year _ 1).

Con la opción **Regresión lineal** que se encuentra situada sobre la opción **Analizar** del menú PASW (*Analizar Regresión lineal*).



Tras realizar esta operación se activa el cuadro de diálogo que se representa en la figura:



Figura: Cuadro de diálogo Regresión lineal

Model Summary

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|-------------------|----------|-------------------|----------------------------|
| 1 | .668 ^a | .446 | .426 | 1.3354 |

a. Predictors: (Constant), YEAR, not periodic

ANOVA^b

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|-------|------------|----------------|----|-------------|--------|-------------------|
| 1 | Regression | 40.238 | 1 | 40.238 | 22.566 | .000 ^a |
| | Residual | 49.929 | 28 | 1.783 | | |
| | Total | 90.167 | 29 | | | |

a. Predictors: (Constant), YEAR, not periodic

b. Dependent Variable: Venta de Bebidas

Coefficients^a

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|-------|--------------------|-----------------------------|------------|---------------------------|--------|------|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 3106.833 | 653.149 | | 4.757 | .000 |
| | YEAR, not periodic | -1.548 | .326 | -.668 | -4.750 | .000 |

a. Dependent Variable: Venta de Bebidas

3. Representar gráficamente e interpretar los valores pronosticados en el análisis de regresión (la tendencia) y los valores agregados anuales de la serie utilizando el comando gráfico Secuencia de PASW.

Analizar /Regresión lineal / Guardar valores pronosticados no tipificados Gráfico / Secuencia

Observación: Los valores pronosticados no tipificados de este análisis de regresión son los valores de la tendencia de cada año.

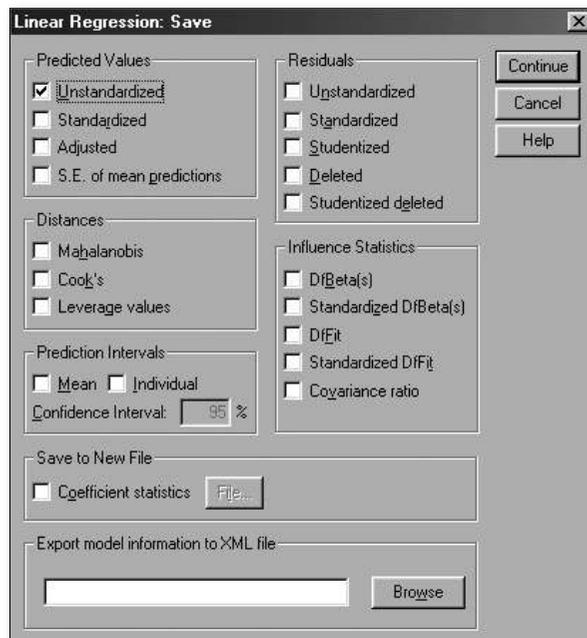


Figura: Cuadro de diálogo Guardar nuevas variables.

Método de la Media Móvil para determinar la tendencia

Previamente al método de la media móvil hay que utilizar el comando **Definir fechas** para definir el tiempo de una serie temporal.

Este método consiste en determinar la tendencia de la serie por medio del uso de medias móviles de orden p . Para la aplicación de esta metodología habrá que dar los siguientes pasos:

1. Elegir un valor de p que represente el período de oscilaciones más importante que posee la serie (normalmente, es el número de observaciones existentes dentro de cada año en ausencia de ciclo).

2. Calcular las medias móviles centradas mediante el comando **Crear serie temporal** de PASW. Para ejecutar este comando habrá que hacer un solo clic con el botón izquierdo del ratón sobre la opción del mismo nombre que se despliega cuando su puntero se encuentra situado sobre el menú Transformar de la barra de menú de PASW (**Transformar / Crear serie temporal**). Tras realizar esta operación se activa el cuadro de diálogo que se representa en la figura:

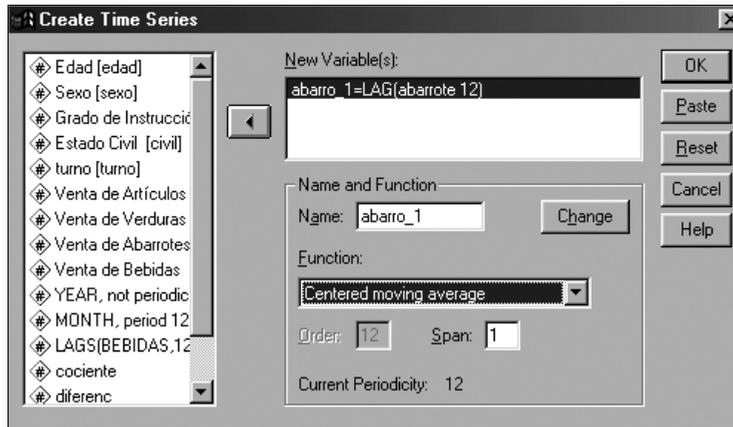


Figura: Cuadro de diálogo Crear serie temporal.

En el cuadro de diálogo *Crear serie temporal* se introducirá en el recuadro *Nuevas variables* la variable cuya tendencia se quiere determinar y en el recuadro *Nombre y función* se indicará el Nombre de la nueva variable a crear que recogerá la tendencia de la serie (MMC), la *Función* de la serie que recoge la tendencia de la serie (Media móvil centrada) y la *Amplitud* ($p = 12$ para series mensuales, 4 para series trimestrales...).

Tras completar el cuadro de diálogo con las especificaciones indicadas en el párrafo anterior, se ejecutará el comando haciendo clic con el botón izquierdo del ratón cuando su puntero se encuentre situado sobre el botón *Aceptar*.

Como resultado de esta última operación se creará una nueva variable, con el nombre que se haya indicado, que recogerá la tendencia de la serie temporal.

3. Representar gráficamente los valores de la tendencia recogidos en la variable MMC y los valores que toma la serie temporal en los distintos momentos del tiempo utilizando el comando **Secuencia** de PASW (**Gráfico / Secuencia**).

Determinación de la componente estacional de una serie temporal

De entre todos los métodos estudiados en teoría para la descomposición estacional de una serie temporal, el único que va a ser expuesto para su resolución con PASW va a ser el Método de la Media Móvil.

Método de la Media Móvil para determinar la estacionalidad

Este método requiere, en primer lugar obtener la componente extraestacional de la serie (ciclo + tendencia) utilizando el método de las Medias Móviles para la tendencia estudiado en el apartado anterior y, en segundo lugar, calcular los índices de variación estacional (IVE) según el esquema de acuerdo con el que combinan las componentes de la serie. La aplicación de este método con PASW implica dar los siguientes pasos:

1. Ejecutar el comando **Descomposición estacional** de PASW haciendo clic con el botón izquierdo del ratón en la opción del mismo nombre que se despliega al desplazar el puntero del ratón sobre la opción **Series temporales** de la barra de menú **Analizar** de PASW (**Analizar / Series temporales / Descomposición estacional**).

“El cálculo de los IVE en PASW se lleva a cabo utilizando el promedio medial definido como la media de todos los valores menos el máximo y el mínimo”.

Tras realizar la operación descrita en el párrafo anterior aparecerá en la pantalla el cuadro de diálogo que se representa en la figura:

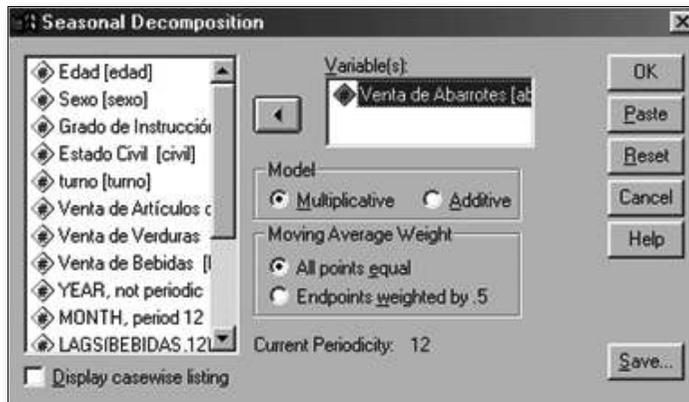


Figura: Cuadro de diálogo Descomposición estacional

En el cuadro de diálogo *Descomposición estacional* se introducirá en el recuadro *Variables*, la/s variable/s a las que se les va a realizar la descomposición estacional y se indicará en la opción *Modelo* el esquema que sigue la serie (Multiplicativo o Aditivo).

Observación: Si alguna de las variables introducidas en la casilla *Variables* tiene valores perdidos, el procedimiento de *Descomposición estacional* no se ejecuta.

A continuación, en el recuadro **Ponderación de la media móvil** se seleccionará una de las dos opciones siguientes:

- *Todos los puntos por igual* cuando el número de subdivisiones hechas dentro del año (p) sea un número impar.
- *Puntos finales ponderados por 5* en caso de que p sea un número par.

Finalmente, pulsando el botón Guardar del cuadro de diálogo representado anteriormente se indicará cómo se van a tratar las nuevas variables creadas como consecuencia de la descomposición estacional de la serie pudiéndose optar por:

- *Añadir al archivo.* Añade las nuevas variables creadas mediante este procedimiento al archivo activo.
- *Reemplazar las existentes.* Guarda sólo las variables del procedimiento actual Si en el archivo existen ya estas variables, se sustituirían.
- *No crear.* Indica que no se deberá guardar ninguna variable.

2. Interpretación de los resultados de la descomposición estacional.

Como consecuencia de la descomposición estacional se crean 4 nuevas variables que se añaden al archivo (en el caso de que se haya seleccionado la opción *Añadir al archivo*) en el cuadro de diálogo que aparece al pulsar el botón **Guardar** del cuadro de diálogo que aparece al pulsar el botón Guardar del cuadro de diálogo *Descomposición estacional* tal y como se muestra a continuación en la figura.

El contenido de cada una de estas nuevas variables es el que denota a continuación:

- **ERR.** Contiene la componente residual o errática de la serie. Es posible su cálculo utilizando el resto de las componentes si se tiene en cuenta el esquema de composición de la misma.
- **SAS.** Recoge los valores de la serie desestacionalizada y se calcula por la diferencia entre los valores de estacionalidad en el caso de que el esquema sea aditivo y por el cociente en el caso de que el esquema sea multiplicativo.
- **SAF.** Recoge los factores de estacionalidad de la serie los cuales se repiten cada 3, 4, 12... veces según sea la serie cuatrimestral, trimestral, anual.

Su interpretación depende de si en la descomposición estacional se ha considerado un esquema aditivo o multiplicativo:

- **Aditivo.** Cuando toma el valor igual a 0 no existe estacionalidad en el período concreto en el que lo toma; cuando éste es positivo en un período concreto el valor de la variable toma valores superiores a los de la media en ese período, mientras que cuando es negativo ocurre lo contrario.
- **Multiplicativo.** Cuando toma el valor igual a 1 no existe estacionalidad en el período concreto en el que lo toma; cuando éste es mayor que 1 en un período concreto el valor de la variable toma valores superiores a los de la media en ese período, mientras que cuando es menor que 1 ocurre lo contrario.
- **STC.** Recoge la componente tendencia – ciclo resultante de aplicar un análisis de la tendencia utilizando el método de la media móvil. El comando *Descomposición estacional* permite, por lo tanto, la determinación de la tendencia a través del método de la media móvil cuando el valor de p es igual al número de subdivisiones que se realicen dentro del año.

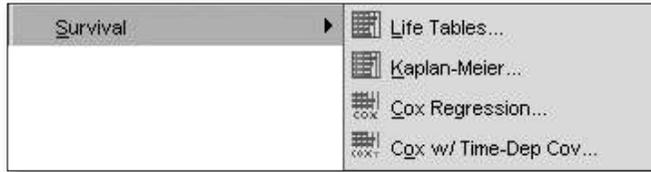
Capítulo
15

Supervivencia y respuestas múltiples

Presentamos una reseña simple sobre supervivencia y respuestas múltiples, simplemente para conocimientos del lector.



SUPERVIVENCIA



Tablas de mortalidad

Existen muchas situaciones en las se desea examinar la distribución de un período entre dos eventos, como la duración del empleo (tiempo transcurrido entre el contrato y el abandono de la empresa). Sin embargo, este tipo de datos suele incluir algunos casos para los que no se registra el segundo evento (censurados); por ejemplo, la gente que todavía trabaja en la empresa al final del estudio.

Las razones para que no se verifique el segundo evento pueden ser muy variadas: en algunos casos, el evento simplemente no tiene lugar antes de que finalice el estudio; en otros, el investigador puede haber perdido el seguimiento de su estado en algún momento anterior a que finalice el estudio; y existen además casos que no pueden continuar por razones ajenas al estudio (como el caso en que un empleado se enferme y se acoja a una baja laboral).

Estos casos se conocen globalmente como casos censurados y hacen que el uso de técnicas tradicionales como las pruebas *t* o la regresión lineal sea inapropiado para este tipo de estudio. Existe una técnica estadística útil para este tipo de datos llamada tabla de mortalidad de 'seguimiento'. La idea básica de la tabla de mortalidad es subdividir el período de observación en intervalos de tiempo más pequeños.

En cada intervalo, se utiliza toda la gente que se ha observado como mínimo durante ese periodo de tiempo para calcular la probabilidad de que un evento terminal tenga lugar dentro de ese intervalo.

Las probabilidades estimadas para cada intervalo se utilizan para estimar la probabilidad global de que el evento tenga lugar en diferentes puntos temporales.

Kaplan-Meier

Es un método de estimación de modelos hasta el evento en presencia de casos censurados. Se basa en la estimación de las probabilidades condicionales en cada punto temporal cuando tiene lugar un evento y en tomar el límite del producto de esas probabilidades para estimar la tasa de supervivencia en cada punto temporal.

Regresión de Cox

Del mismo modo que los anteriores, la *Regresión de Cox* es un método para crear modelos para datos de tiempos de espera hasta un evento con casos censurados presentes. Sin embargo, la Regresión de Cox permite incluir en los modelos variables predictoras (covariables).

Por ejemplo, permite construir un modelo de la duración en el empleo como función del nivel educativo y de la categoría laboral. Gestiona los casos censurados correctamente y proporciona las estimaciones de los coeficientes para cada una de las covariables, permitiendo evaluar el impacto de múltiples covariables en el mismo modelo. Además, es posible utilizar este método para examinar el efecto de covariables continuas.

Cox con covariable dep. del tiempo

Existen ciertas situaciones en las que interesa calcular un modelo de Regresión de Cox, pero no se cumple el supuesto de tasas de impacto proporcionales. Es decir, que las tasas de impacto cambian con el tiempo: los valores de una (o de varias) de las covariables son diferentes en los distintos puntos del tiempo. En esos casos, es necesario utilizar un modelo de regresión de Cox extendido, que permita especificar covariables dependientes del tiempo (T_COV_).

Con el fin de analizar dicho modelo, se debe definir primero una covariable dependiente del tiempo. Para facilitar esta tarea contamos con una variable del sistema, llamada T_, que representa el tiempo. Se utiliza para definir covariables dependientes del tiempo empleando dos métodos generales:

- Para contrastar el supuesto de tasas de impacto proporcionales con respecto a una covariable particular o para estimar un modelo de Regresión de Cox extendido que permita impactos no proporcionales, se define la covariable dependiente del tiempo como una función de la variable de tiempo T_ y la covariable en cuestión.
- Algunas variables pueden tener valores distintos en períodos diferentes del tiempo, pero no están sistemáticamente relacionadas con el tiempo. En tales casos es necesario definir una covariable dependiente del tiempo segmentada, lo cual puede llevarse a cabo a partir de un conjunto de medidas usando las expresiones lógicas (1:verdadero, 0:falso).

Para las covariables dependientes del tiempo segmentadas, los casos en los que falte cualquiera de los valores quedarán eliminados del análisis. Por tanto, todos los casos deben tener valores para todos los puntos del tiempo medidos en la covariable, incluso para los puntos del tiempo posteriores a la eliminación del caso del conjunto bajo riesgo (ya sea por el evento o por la censura).

RESPUESTAS MÚLTIPLES

El PASW permite obtener frecuencias y porcentajes para variables que son múltiples. Una variable múltiple es aquella variable que puede tomar más de un valor para cada caso.



Definir conjuntos

Agrupar variables elementales en conjuntos de categorías múltiples y de dicotomías múltiples, para los que se pueden obtener tablas de frecuencias y tablas de contingencia. Se pueden definir hasta 20 conjuntos de respuestas múltiples. Cada conjunto debe tener un nombre exclusivo. Las variables elementales se pueden codificar como dicotomías o categorías:

- *Dicotomías.* En *Valor contado* debe introducirse un entero. Cada variable que tenga al menos una aparición del valor contado se convierte en una categoría del conjunto de dicotomías múltiples.
- *Categorías.* Crear un conjunto de categorías múltiples con el mismo rango de valores que las variables que lo componen. Deben introducirse dos enteros como valores máximo y mínimo del rango para las categorías del conjunto de categorías múltiples. El procedimiento suma cada valor entero distinto en el rango inclusivo para todas las variables que lo componen. Las categorías vacías no se tabulan. A cada conjunto de respuestas múltiples se le debe asignar un nombre exclusivo de hasta 7 caracteres. El procedimiento coloca delante del nombre asignado un signo dólar (\$). El nombre del conjunto de respuestas múltiples sólo se encuentra disponible para su uso en los procedimientos de respuestas múltiples.

Frecuencias

Produce tablas de frecuencias para conjuntos de respuestas múltiples. Es necesario definir previamente uno o más conjuntos de respuestas múltiples. Para los conjuntos de dicotomías múltiples, los nombres de categorías que se muestran en los resultados proceden de etiquetas de variable definidas para variables elementales del grupo.

Para los conjuntos de categorías múltiples, las etiquetas de categoría proceden de las etiquetas de valor de la primera variable del grupo. Si las categorías perdidas para la primera variable están presentes para otras variables del grupo, debe definirse una etiqueta de valor para las categorías perdidas.

Los casos con *Valores perdidos* se excluyen en base a tabla por tabla. Un caso se considera perdido para un conjunto de dicotomías (o categorías múltiples) sólo si ninguno de sus componentes contiene el valor (o tiene valores válidos en el rango). Se pueden excluir los casos con valores perdidos en cualquier variable de la tabulación del conjunto de dicotomías, o categorías, o ambas.

Tablas de contingencia

Presenta en forma de tabla de contingencia, conjuntos de respuestas múltiples, variables elementales o una combinación. También permite obtener porcentajes de casilla basados en casos o respuestas, modificar la gestión de los valores perdidos u obtener tablas de contingencia emparejadas.

Es necesario definir previamente uno o más conjuntos de respuestas múltiples. Los nombres en los resultados se muestran de la misma forma que en el procedimiento anterior. Las etiquetas de categoría por columnas se muestran en tres líneas, con un máximo de 8 caracteres por línea. Para evitar la división de palabras, se pueden invertir los elementos de las filas y las columnas o volver a definir las etiquetas.

Ejercicio N° 1

El ejercicio permitirá utilizar los principales procedimientos del menú **Clasificar**. Los datos utilizados son del siguiente ejemplo:

Supongamos que existen 5 terapias [Terapia 1 (1); Terapia 2 (2); Terapia 3 (3); Terapia 4 (4) y Terapia 5 (5)] para pacientes de cáncer que se le suministra uno por día. Cada paciente debe elegir 3 terapias como máximo. Además debe seleccionar los días de la semana [Lunes (1); Martes (2); Miércoles (3); Jueves (4) y Viernes (5)] en que recibirá la terapia. Los datos se encuentran en el archivo *RespuestasMúltiples.sav*.

Dentro del menú **Analizar** se encuentra el procedimiento **Respuestas múltiples**. El aspecto de dicho menú es el siguiente: **Analizar -> Respuestas múltiples (Múltiple Response)**.

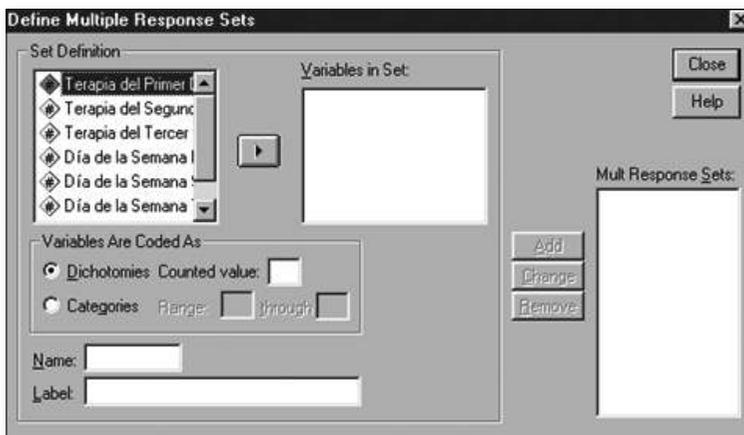


Para realizar el procedimiento respuestas múltiples se deben seguir los cuadros de diálogo que se presentan. A continuación se muestra una explicación detallada del cuadro de diálogo correspondiente a los principales procedimientos.

Definir Conjuntos

Para definir conjuntos se seleccionamos la opción **Analizar -> Respuestas múltiple->Definir conjuntos**, obteniéndose el cuadro de diálogo siguiente:

Lo primero que se debe hacer es definir cuáles son las variables que van a formar parte de la variable múltiple. El cuadro de la izquierda muestra la lista de variables del archivo:



Las variables seleccionadas se trasladan al cuadro del lado derecho, campo variables de conjuntos (Variables in Set:). En nuestro caso las variables son: Terapia del Primer Día, Terapia del Segundo Día y Terapia del Tercer Día. Como se muestra en el siguiente gráfico.

En el campo **Las variables están codificadas como (Variables Are Coded As)**, hay que indicar la forma en que se han introducido las variables que van a formar parte de la variable multirrespuesta, es decir, si se han codificado como dicotómicas o con tantas categorías como posibles respuesta tenga. En nuestro ejemplo son variables con 5 categorías cada una.

En el campo **Nombre (Name)** se especifica el nombre que se va dar a la variable múltiple. En nuestro ejemplo es Terapia.

En el campo Etiqueta (Label) se especifica una etiqueta para dicha variable. En nuestro ejemplo será Terapias aplicadas.

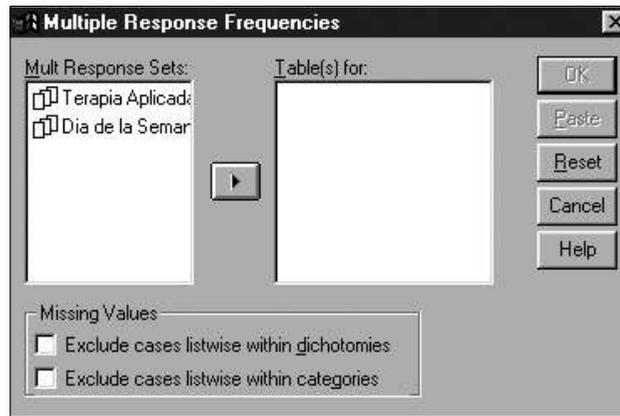
Una vez que se han rellenados estos campos, se presiona el botón Añadir (Add) para que la variable que se acaba de definir figure en el campo de **Conjunto de respuestas múltiples (Mult Response Sets:)**, como se puede observar en el siguiente gráfico:



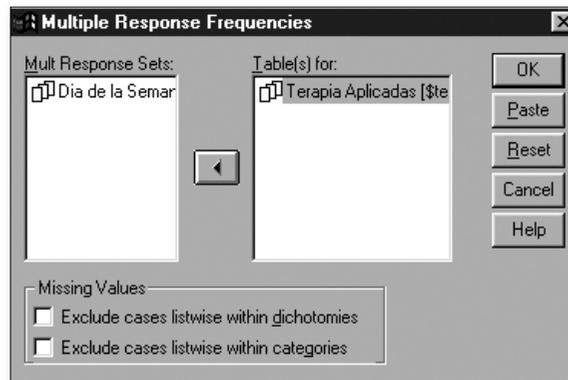
Una vez que se han definido las variables múltiples, se puede presionar Cerrar (Close). Entonces ya podemos obtener las frecuencias de dichas variables o cruzar variables de este tipo.

Frecuencias

Para obtener frecuencias y porcentajes de una o más variables múltiple, se selecciona la opción **Analizar -> Respuestas múltiple-> Frecuencias**, obteniéndose el cuadro de diálogo siguiente:



En el campo **Conjunto de respuestas múltiples (Mult Response Sets:)** figuran los nombres de las variables múltiples que previamente se han definido. De esta lista se selecciona las que se quieren analizar y se traslada a **Tablas para (Table[s] For:)**. En nuestro ejemplo hallaremos frecuencias Terapia aplicadas como se observa en el gráfico siguiente:



Para este tipo de análisis el PASW considera como missing aquellos caos que tienen valor missing en todas las variables que forman la variable múltiple. Si en el campo **Valores perdidos** se selecciona alguna de las 2 opciones, *Excluir casos según lista en las dicotómicas* o *Excluir casos según lista en las categorías*, el PASW excluye del análisis aquellos casos que tienen valor missing en 'alguna' de las variables que forman la variable múltiple.

Si se realiza el análisis de frecuencias de la variable múltiple Terapias aplicadas que está formada por tres variables (Terapia del Primer Día, Terapia del Segundo Día y Terapia del Tercer Día), el resultado será el siguiente:

| Group \$TERAPIA | Terapia Aplicadas | | | | |
|-----------------|-------------------|-------|------------------|--------------|--|
| Category label | Code | Count | Pct of Responses | Pct of Cases | |
| Terapia 1 | 1 | 20 | 22.2 | 66.7 | |
| Terapia 2 | 2 | 19 | 21.1 | 63.3 | |
| Terapia 3 | 3 | 17 | 18.9 | 56.7 | |
| Terapia 4 | 4 | 21 | 23.3 | 70.0 | |
| Terapia 5 | 5 | 13 | 14.4 | 43.3 | |
| | | ----- | ----- | ----- | |
| | Total responses | 90 | 100.0 | 300.0 | |

0 missing cases; 30 valid cases

Esta tabla contiene las frecuencias absolutas, los porcentajes calculados sobre el total de las respuestas (90).

Tablas de Contingencia

Para obtener tablas de contingencia de una o más variables múltiple, se selecciona la opción **Analizar** -> **Respuestas múltiple**-> **Tablas de Contingencia (Crosstabs)**, obteniéndose el cuadro de diálogo siguiente:

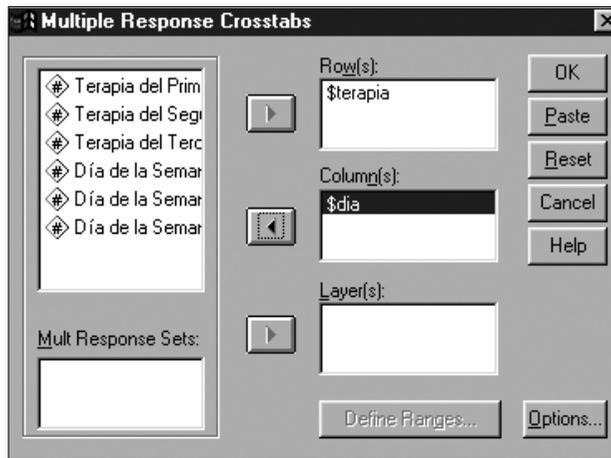


En el campo **Filas (Row[s]:)** se incluyen las variables que se van a representar en las filas de las tablas de frecuencia, y en el campo **Columnas (Column[s]:)** las variables que se representarán en las columnas de dichas tablas. En ambos campos se pueden incluir variables múltiples.

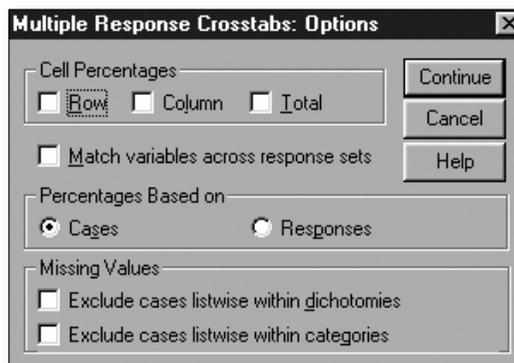
Entonces se obtendrá una tabla de frecuencias observadas cruzando la primera variable especificada en el campo **Filas** con la primera variable especificada en el campo **Columna**; se obtiene otra tabla cruzando la primera variable especificada en el campo **Filas** con la segunda variable especificada en el campo **Columna**, y así sucesivamente con todas las variables especificadas en el campo **Columnas**. Después se realizarán las tablas cruzando la segunda variable del campo **Filas** con cada una de las del campo **Columna**, y así sucesivamente.

Las tablas anteriores se pueden obtener en función de los valores que tomen otras variables que se especificarán en el campo **Capas (Layer[s])**.

En nuestro ejemplo hallaremos tablas de contingencia para la variable múltiple Terapia aplicada y Día de la semana, como se puede observar en la gráfica siguiente:



Si pulsamos el botón **Opciones (Options...)** se obtiene una ventana que contiene la lista de resultados que se pueden obtener además de las frecuencias observadas, es decir, el cuadro de diálogo siguiente:



BIBLIOGRAFÍA

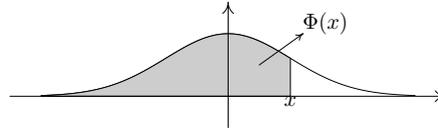
- Sabina, Carlos: El Proceso de Investigación, editorial PANAPO, Caracas 1992.
- Hernández, Sampieri, Roberto: Metodología de la Investigación, editorial Me Graw Hill. Bogotá, Colombia 1991.
- Tamayo y Tamayo, Mario: Metodología Formal de la Investigación Científica, editorial Limusa, S.A. México D.F. 1995.
- Bartolome, M. (2001): *Metodología cualitativa en Educación. Dossier de Doctorado*.
- Sandin M. P. (2003): *Investigación Cualitativa en Educación. Fundamentos y Tradiciones*. Madrid: McGraw Hill.
- Ackoff, R. (1953). The Desing of Social Research, University of Chicago, United States of America.
- Anderson, J., B. H. Durston y M. Poole (1993). Redacción de Tesis y Trabajos Escolares, 15ª. reimp., Ed. Diana, México.
- Baker, T. L. (1997). Doing Social Research, 2ª. ed., Ed. McGraw-Hill, United States of America.
- Berenson, M. L. y D. M. Levine (1994). Estadística para Administración y Economía, Ed. McGraw-Hill, México.
- Bunge, M. (1983). La Investigación Científica, 2ª. ed., Ed. Ariel, México.
- Campbell, D. T. y J. C. Stanley (1966). Experimental and Quasiexperimental Desings for Research, Rand McNally and Company, United States of America.
- Dankhe, G. L. (1989) 'Investigación y Comunicación', en C. Fernández-Collado y G. L. Dankhe (comps.), La Comunicación Humana: Ciencia Social, Ed. McGraw-Hill, México.
- Gallardo, J. (1998). Formulación y Evaluación de Proyectos, Ed. McGraw-Hill, México.
- Kazmier, L. J. (1998). Estadística aplicada a la Administración y a la Economía, 3ª. ed., Ed. McGraw-Hill, México.
- Webster, A. (1998). Estadística aplicada a la Empresa y a la Economía, 2ª. ed., Ed. McGraw Hill, México.
- Wierers, R. (1993). Investigación de Mercados, ed. Prentice-Hall Hispanoamericana, México.
- Walpole, R. E. y R. H. Myers (1996). Probabilidad y Estadística, 4ª. ed., Ed. McGraw Hill, México.
- Quezada Lucio, Nel (2009). Estadísticas con Pasw, Editorial Macro, Perú.



APÉNDICE

Tabla de la función de distribución Φ de una normal $N(0, 1)$ para $x \geq 0$

$$\Phi(x) = P[X \leq x] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt$$

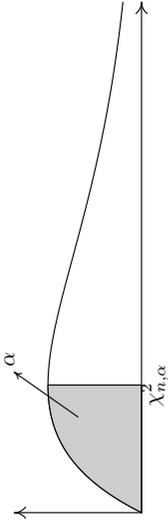


Si $x < 0 \implies \Phi(x) = 1 - \Phi(-x)$

| | 0.00 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 0.0 | 0.500000 | 0.503989 | 0.507978 | 0.511966 | 0.515953 | 0.519939 | 0.523922 | 0.527903 | 0.531881 | 0.535856 |
| 0.1 | 0.539828 | 0.543795 | 0.547758 | 0.551717 | 0.555670 | 0.559618 | 0.563559 | 0.567495 | 0.571424 | 0.575345 |
| 0.2 | 0.579260 | 0.583166 | 0.587064 | 0.590954 | 0.594835 | 0.598706 | 0.602568 | 0.606420 | 0.610261 | 0.614092 |
| 0.3 | 0.617911 | 0.621720 | 0.625516 | 0.629300 | 0.633072 | 0.636831 | 0.640576 | 0.644309 | 0.648027 | 0.651732 |
| 0.4 | 0.655422 | 0.659097 | 0.662757 | 0.666402 | 0.670031 | 0.673645 | 0.677242 | 0.680822 | 0.684386 | 0.687933 |
| 0.5 | 0.691462 | 0.694974 | 0.698468 | 0.701944 | 0.705401 | 0.708840 | 0.712260 | 0.715661 | 0.719043 | 0.722405 |
| 0.6 | 0.725747 | 0.729069 | 0.732371 | 0.735653 | 0.738914 | 0.742154 | 0.745373 | 0.748571 | 0.751748 | 0.754903 |
| 0.7 | 0.758036 | 0.761148 | 0.764238 | 0.767305 | 0.770350 | 0.773373 | 0.776373 | 0.779350 | 0.782305 | 0.785236 |
| 0.8 | 0.788145 | 0.791030 | 0.793892 | 0.796731 | 0.799546 | 0.802337 | 0.805105 | 0.807850 | 0.810570 | 0.813267 |
| 0.9 | 0.815940 | 0.818589 | 0.821214 | 0.823814 | 0.826391 | 0.828944 | 0.831472 | 0.833977 | 0.836457 | 0.838913 |
| 1.0 | 0.841345 | 0.843752 | 0.846136 | 0.848495 | 0.850830 | 0.853141 | 0.855428 | 0.857690 | 0.859929 | 0.862143 |
| 1.1 | 0.864334 | 0.866500 | 0.868643 | 0.870762 | 0.872857 | 0.874928 | 0.876976 | 0.879000 | 0.881000 | 0.882977 |
| 1.2 | 0.884930 | 0.886861 | 0.888768 | 0.890651 | 0.892512 | 0.894350 | 0.896165 | 0.897958 | 0.899727 | 0.901475 |
| 1.3 | 0.903200 | 0.904902 | 0.906582 | 0.908241 | 0.909877 | 0.911492 | 0.913085 | 0.914657 | 0.916207 | 0.917736 |
| 1.4 | 0.919243 | 0.920730 | 0.922196 | 0.923641 | 0.925066 | 0.926471 | 0.927855 | 0.929219 | 0.930563 | 0.931888 |
| 1.5 | 0.933193 | 0.934478 | 0.935745 | 0.936992 | 0.938220 | 0.939429 | 0.940620 | 0.941792 | 0.942947 | 0.944083 |
| 1.6 | 0.945201 | 0.946301 | 0.947384 | 0.948449 | 0.949497 | 0.950529 | 0.951543 | 0.952540 | 0.953521 | 0.954486 |
| 1.7 | 0.955435 | 0.956367 | 0.957284 | 0.958185 | 0.959070 | 0.959941 | 0.960796 | 0.961636 | 0.962462 | 0.963273 |
| 1.8 | 0.964070 | 0.964852 | 0.965620 | 0.966375 | 0.967116 | 0.967843 | 0.968557 | 0.969258 | 0.969946 | 0.970621 |
| 1.9 | 0.971283 | 0.971933 | 0.972571 | 0.973197 | 0.973810 | 0.974412 | 0.975002 | 0.975581 | 0.976148 | 0.976705 |
| 2.0 | 0.977250 | 0.977784 | 0.978308 | 0.978822 | 0.979325 | 0.979818 | 0.980301 | 0.980774 | 0.981237 | 0.981691 |
| 2.1 | 0.982136 | 0.982571 | 0.982997 | 0.983414 | 0.983823 | 0.984222 | 0.984614 | 0.984997 | 0.985371 | 0.985738 |
| 2.2 | 0.986097 | 0.986447 | 0.986791 | 0.987126 | 0.987455 | 0.987776 | 0.988089 | 0.988396 | 0.988696 | 0.988989 |
| 2.3 | 0.989276 | 0.989556 | 0.989830 | 0.990097 | 0.990358 | 0.990613 | 0.990863 | 0.991106 | 0.991344 | 0.991576 |
| 2.4 | 0.991802 | 0.992024 | 0.992240 | 0.992451 | 0.992656 | 0.992857 | 0.993053 | 0.993244 | 0.993431 | 0.993613 |
| 2.5 | 0.993790 | 0.993963 | 0.994132 | 0.994297 | 0.994457 | 0.994614 | 0.994766 | 0.994915 | 0.995060 | 0.995201 |
| 2.6 | 0.995339 | 0.995473 | 0.995604 | 0.995731 | 0.995855 | 0.995975 | 0.996093 | 0.996207 | 0.996319 | 0.996427 |
| 2.7 | 0.996533 | 0.996636 | 0.996736 | 0.996833 | 0.996928 | 0.997020 | 0.997110 | 0.997197 | 0.997282 | 0.997365 |
| 2.8 | 0.997445 | 0.997523 | 0.997599 | 0.997673 | 0.997744 | 0.997814 | 0.997882 | 0.997948 | 0.998012 | 0.998074 |
| 2.9 | 0.998134 | 0.998193 | 0.998250 | 0.998305 | 0.998359 | 0.998411 | 0.998462 | 0.998511 | 0.998559 | 0.998605 |
| 3.0 | 0.998650 | 0.998694 | 0.998736 | 0.998777 | 0.998817 | 0.998856 | 0.998893 | 0.998930 | 0.998965 | 0.998999 |
| 3.1 | 0.999032 | 0.999065 | 0.999096 | 0.999126 | 0.999155 | 0.999184 | 0.999211 | 0.999238 | 0.999264 | 0.999289 |
| 3.2 | 0.999313 | 0.999336 | 0.999359 | 0.999381 | 0.999402 | 0.999423 | 0.999443 | 0.999462 | 0.999481 | 0.999499 |
| 3.3 | 0.999517 | 0.999534 | 0.999550 | 0.999566 | 0.999581 | 0.999596 | 0.999610 | 0.999624 | 0.999638 | 0.999651 |
| 3.4 | 0.999663 | 0.999675 | 0.999687 | 0.999698 | 0.999709 | 0.999720 | 0.999730 | 0.999740 | 0.999749 | 0.999758 |
| 3.5 | 0.999767 | 0.999776 | 0.999784 | 0.999792 | 0.999800 | 0.999807 | 0.999815 | 0.999822 | 0.999828 | 0.999835 |
| 3.6 | 0.999841 | 0.999847 | 0.999853 | 0.999858 | 0.999864 | 0.999869 | 0.999874 | 0.999879 | 0.999883 | 0.999888 |
| 3.7 | 0.999892 | 0.999896 | 0.999900 | 0.999904 | 0.999908 | 0.999912 | 0.999915 | 0.999918 | 0.999922 | 0.999925 |
| 3.8 | 0.999928 | 0.999931 | 0.999933 | 0.999936 | 0.999938 | 0.999941 | 0.999943 | 0.999946 | 0.999948 | 0.999950 |
| 3.9 | 0.999952 | 0.999954 | 0.999956 | 0.999958 | 0.999959 | 0.999961 | 0.999963 | 0.999964 | 0.999966 | 0.999967 |
| 4.0 | 0.999968 | 0.999970 | 0.999971 | 0.999972 | 0.999973 | 0.999974 | 0.999975 | 0.999976 | 0.999977 | 0.999978 |

Inversa de la función de distribución χ^2 de Pearson:

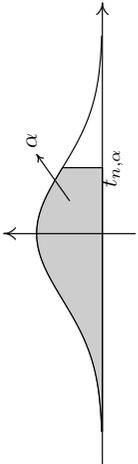
$$\mathcal{X}_n(x) = P[\chi_n^2 \leq x], \quad \mathcal{X}_n(\chi_{n,\alpha}^2) = \alpha, \quad \mathcal{X}_n^{-1}(\alpha) = \chi_{n,\alpha}^2$$



| $n \setminus \alpha$ | 0.005 | 0.01 | 0.025 | 0.05 | 0.1 | 0.2 | 0.3 | 0.5 | 0.7 | 0.8 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 |
|----------------------|----------|----------|----------|----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 1 | 0.00004 | 0.00010 | 0.00098 | 0.00393 | 0.01579 | 0.06418 | 0.14847 | 0.45494 | 1.07419 | 1.64237 | 2.70554 | 3.84146 | 5.02389 | 6.63490 | 7.87944 |
| 2 | 0.01002 | 0.02010 | 0.05064 | 0.10259 | 0.21072 | 0.44629 | 0.71335 | 1.38629 | 2.40795 | 3.21888 | 4.60517 | 5.99146 | 7.37776 | 9.21034 | 10.59663 |
| 3 | 0.07172 | 0.11483 | 0.21580 | 0.35185 | 0.58437 | 1.00517 | 1.42365 | 2.36597 | 3.66487 | 4.64163 | 6.25139 | 7.81473 | 9.34840 | 11.34487 | 12.83816 |
| 4 | 0.20699 | 0.29711 | 0.48442 | 0.71072 | 1.06362 | 1.64878 | 2.19470 | 3.35669 | 4.87843 | 5.98862 | 7.79494 | 9.48773 | 11.14329 | 13.27670 | 14.86026 |
| 5 | 0.41174 | 0.55430 | 0.83121 | 1.14548 | 1.61031 | 2.34253 | 2.99991 | 4.35146 | 6.06443 | 7.28928 | 9.23636 | 11.07050 | 12.83250 | 15.08627 | 16.74960 |
| 6 | 0.67573 | 0.87209 | 1.23734 | 1.63538 | 2.20413 | 3.07009 | 3.82755 | 5.34812 | 7.23114 | 8.55806 | 10.64464 | 12.59159 | 14.44938 | 16.81189 | 18.54758 |
| 7 | 0.98926 | 1.23904 | 1.68987 | 2.16735 | 2.83311 | 3.82232 | 4.67133 | 6.34581 | 8.38343 | 9.80325 | 12.01704 | 14.06714 | 16.01276 | 18.47531 | 20.27774 |
| 8 | 1.34441 | 1.64650 | 2.17973 | 2.73264 | 3.48954 | 4.59357 | 5.52742 | 7.34412 | 9.52446 | 11.03009 | 13.36157 | 15.50731 | 17.53455 | 20.09024 | 21.95495 |
| 9 | 1.73493 | 2.08790 | 2.70039 | 3.32511 | 4.16816 | 5.38005 | 6.39331 | 8.34283 | 10.65637 | 12.24215 | 14.68366 | 16.91898 | 19.02277 | 21.66599 | 23.58935 |
| 10 | 2.15586 | 2.55821 | 3.24697 | 3.94030 | 4.86518 | 6.17908 | 7.26722 | 9.34182 | 11.78072 | 13.44196 | 15.98718 | 18.30704 | 20.48318 | 23.20925 | 25.18818 |
| 11 | 2.60322 | 3.05348 | 3.81575 | 4.57481 | 5.57778 | 6.98867 | 8.14787 | 10.34100 | 12.89867 | 14.63142 | 17.27501 | 19.67514 | 21.92005 | 24.72497 | 26.75685 |
| 12 | 3.07382 | 3.57057 | 4.40379 | 5.22603 | 6.30380 | 7.80733 | 9.03428 | 11.34032 | 14.01110 | 15.81199 | 18.54935 | 21.02607 | 23.33666 | 26.21697 | 28.29952 |
| 13 | 3.56503 | 4.10692 | 5.00875 | 5.89186 | 7.04150 | 8.63386 | 9.92568 | 12.33976 | 15.11872 | 16.98480 | 19.81193 | 22.36203 | 24.73560 | 27.68825 | 29.81947 |
| 14 | 4.07467 | 4.66042 | 5.62873 | 6.57063 | 7.78953 | 9.46733 | 10.82148 | 13.33927 | 16.22210 | 18.15077 | 21.06414 | 23.68479 | 26.11895 | 29.14124 | 31.31935 |
| 15 | 4.60092 | 5.22935 | 6.26214 | 7.26094 | 8.54676 | 10.30696 | 11.72117 | 14.33886 | 17.32169 | 19.31066 | 22.30713 | 24.99579 | 27.48839 | 30.57791 | 32.80132 |
| 16 | 5.14221 | 5.81221 | 6.90766 | 7.96165 | 9.31224 | 11.15212 | 12.62435 | 16.33850 | 18.41789 | 20.46508 | 23.54183 | 26.29623 | 28.84535 | 31.99993 | 34.26719 |
| 17 | 5.69722 | 6.40776 | 7.56419 | 8.67176 | 10.08519 | 12.00227 | 13.53068 | 16.33818 | 19.51102 | 21.61456 | 24.76904 | 27.58711 | 30.19101 | 33.40866 | 35.71847 |
| 18 | 6.26480 | 7.01491 | 8.23075 | 9.39046 | 10.86494 | 12.85695 | 14.43986 | 17.33790 | 20.60135 | 22.75955 | 25.98942 | 28.86930 | 31.52638 | 34.80531 | 37.15045 |
| 19 | 6.84397 | 7.63273 | 8.90652 | 10.11701 | 11.65091 | 13.71579 | 15.35166 | 18.33765 | 21.68913 | 23.90042 | 27.20357 | 30.14353 | 32.85233 | 36.19087 | 38.58226 |
| 20 | 7.43384 | 8.26040 | 9.59078 | 10.85081 | 12.44261 | 14.57844 | 16.26586 | 19.33743 | 22.77454 | 25.03751 | 28.41198 | 31.41043 | 34.16961 | 37.56623 | 39.99685 |
| 21 | 8.03365 | 8.89720 | 10.28290 | 11.59131 | 13.23960 | 15.44461 | 17.18227 | 20.33723 | 23.85779 | 26.17110 | 29.61509 | 32.67057 | 35.47888 | 38.93217 | 41.40106 |
| 22 | 8.64272 | 9.54249 | 10.98232 | 12.33801 | 14.04149 | 16.31404 | 18.10072 | 21.33704 | 24.93902 | 27.30145 | 30.81328 | 33.92444 | 36.78071 | 40.28936 | 42.79566 |
| 23 | 9.26042 | 10.19572 | 11.68855 | 13.09051 | 14.84796 | 17.18651 | 19.02109 | 22.33688 | 26.01837 | 28.42879 | 32.00690 | 35.17246 | 38.07563 | 41.63840 | 44.18128 |
| 24 | 9.88623 | 10.85636 | 12.40115 | 13.84842 | 15.65868 | 18.06180 | 19.94323 | 23.33673 | 27.09596 | 29.55332 | 33.19624 | 36.41503 | 39.36408 | 42.97982 | 45.55851 |
| 25 | 10.51965 | 11.52398 | 13.11972 | 14.61141 | 16.47341 | 18.93975 | 20.86703 | 24.33659 | 28.17192 | 30.67520 | 34.38159 | 37.65248 | 40.64647 | 44.31410 | 46.92789 |
| 26 | 11.16024 | 12.19815 | 13.84390 | 15.37916 | 17.29188 | 19.82019 | 21.79240 | 25.33646 | 29.24633 | 31.79461 | 35.56317 | 38.88514 | 41.92317 | 45.64168 | 48.28988 |
| 27 | 11.80759 | 12.87850 | 14.57338 | 16.15140 | 18.11390 | 20.70298 | 22.71924 | 26.33634 | 30.31929 | 32.91169 | 36.74192 | 40.11327 | 43.19451 | 46.96294 | 50.99338 |
| 28 | 12.46134 | 13.56471 | 15.30786 | 16.92788 | 18.93924 | 21.58797 | 23.64746 | 27.33623 | 31.39088 | 34.02657 | 37.91592 | 41.33714 | 44.46079 | 48.27824 | 50.99338 |
| 29 | 13.12115 | 14.25645 | 16.04707 | 17.70837 | 19.76774 | 22.47505 | 24.57699 | 28.33613 | 32.46117 | 35.13936 | 39.08747 | 42.55697 | 45.72229 | 49.58788 | 52.33562 |
| 30 | 13.78672 | 14.95346 | 16.79077 | 18.49266 | 20.59923 | 23.36411 | 25.50776 | 29.33603 | 33.53023 | 36.25019 | 40.25602 | 43.77297 | 46.97924 | 50.89218 | 53.67196 |
| 40 | 20.70654 | 22.16426 | 24.43304 | 26.50930 | 29.05052 | 32.34495 | 34.87194 | 39.33534 | 44.16487 | 47.26854 | 51.80506 | 55.75848 | 59.34171 | 63.69074 | 66.76596 |
| 50 | 27.99075 | 29.70668 | 32.35736 | 34.76425 | 37.68865 | 41.44921 | 44.31331 | 49.33494 | 54.72279 | 58.16380 | 63.16712 | 67.50481 | 71.42020 | 76.15389 | 79.48998 |
| 60 | 35.53449 | 37.48485 | 40.48175 | 43.18796 | 46.45889 | 50.64062 | 53.80913 | 59.33467 | 65.22651 | 68.97207 | 74.39701 | 79.08194 | 83.29767 | 88.37942 | 91.95170 |
| 120 | 83.85157 | 86.92328 | 91.57264 | 95.70464 | 100.62363 | 106.80561 | 111.41857 | 119.33400 | 127.61590 | 132.80628 | 140.23257 | 146.56736 | 152.21140 | 158.95017 | 163.64818 |

Inversa de la función de distribución t de Student:

$$\mathcal{T}_n(x) = P[t_n \leq x], \quad \mathcal{T}_n(t_{n,\alpha}) = \alpha, \quad \mathcal{T}_n^{-1}(\alpha) = t_{n,\alpha}$$



| $n \setminus \alpha$ | 0.55 | 0.6 | 0.65 | 0.7 | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 0.975 | 0.99 | 0.995 |
|----------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| 1 | 0.158384 | 0.324920 | 0.509525 | 0.726543 | 1.000000 | 1.376382 | 1.962610 | 3.077684 | 6.313752 | 12.706205 | 31.820516 | 63.656741 |
| 2 | 0.142134 | 0.288675 | 0.444750 | 0.617213 | 0.816497 | 1.060660 | 1.386218 | 2.149778 | 4.302653 | 8.324446 | 16.924843 | 31.820516 |
| 3 | 0.136598 | 0.276671 | 0.424202 | 0.584390 | 0.764892 | 0.978472 | 1.249778 | 1.637744 | 2.353363 | 3.182446 | 4.540703 | 5.840909 |
| 4 | 0.133830 | 0.270722 | 0.414163 | 0.568649 | 0.740697 | 0.940965 | 1.189567 | 1.533206 | 2.131847 | 2.776445 | 3.746947 | 4.604095 |
| 5 | 0.132175 | 0.267181 | 0.408229 | 0.559430 | 0.726687 | 0.919544 | 1.155767 | 1.475884 | 2.015048 | 2.570582 | 3.364930 | 4.032143 |
| 6 | 0.131076 | 0.264835 | 0.404313 | 0.553381 | 0.717558 | 0.905703 | 1.14157 | 1.439756 | 1.943180 | 2.446912 | 3.142668 | 3.707428 |
| 7 | 0.130293 | 0.263167 | 0.401538 | 0.549110 | 0.711142 | 0.896030 | 1.119159 | 1.414924 | 1.894579 | 2.364624 | 2.997952 | 3.499483 |
| 8 | 0.129707 | 0.261921 | 0.399469 | 0.545934 | 0.706387 | 0.888890 | 1.108145 | 1.396815 | 1.859548 | 2.306004 | 2.896459 | 3.355387 |
| 9 | 0.129253 | 0.260955 | 0.397868 | 0.543480 | 0.702722 | 0.883404 | 1.099716 | 1.383029 | 1.833113 | 2.262157 | 2.821438 | 3.249836 |
| 10 | 0.128890 | 0.260185 | 0.396591 | 0.541528 | 0.699812 | 0.879058 | 1.093058 | 1.372184 | 1.812461 | 2.228139 | 2.763769 | 3.169273 |
| 11 | 0.128594 | 0.259556 | 0.395551 | 0.539938 | 0.697445 | 0.875530 | 1.087666 | 1.363430 | 1.795885 | 2.200985 | 2.718079 | 3.105807 |
| 12 | 0.128347 | 0.259033 | 0.394686 | 0.538618 | 0.695483 | 0.872609 | 1.083211 | 1.356217 | 1.782288 | 2.178813 | 2.680998 | 3.054540 |
| 13 | 0.128139 | 0.258591 | 0.393955 | 0.537504 | 0.693829 | 0.870152 | 1.079469 | 1.350171 | 1.770933 | 2.160369 | 2.650309 | 3.012276 |
| 14 | 0.127961 | 0.258213 | 0.393331 | 0.536552 | 0.692417 | 0.868055 | 1.076280 | 1.345030 | 1.761310 | 2.144787 | 2.624494 | 2.976843 |
| 15 | 0.127806 | 0.257885 | 0.392790 | 0.535729 | 0.691197 | 0.866245 | 1.073531 | 1.340606 | 1.753050 | 2.131450 | 2.602480 | 2.946713 |
| 16 | 0.127671 | 0.257599 | 0.392318 | 0.535010 | 0.690132 | 0.864667 | 1.071137 | 1.336757 | 1.745884 | 2.119905 | 2.583487 | 2.920782 |
| 17 | 0.127552 | 0.257347 | 0.391902 | 0.534377 | 0.689195 | 0.863279 | 1.069033 | 1.333379 | 1.739607 | 2.109816 | 2.566934 | 2.898231 |
| 18 | 0.127447 | 0.257123 | 0.391533 | 0.533816 | 0.688364 | 0.862049 | 1.067170 | 1.330391 | 1.734064 | 2.100922 | 2.552380 | 2.878440 |
| 19 | 0.127352 | 0.256923 | 0.391202 | 0.533314 | 0.687621 | 0.860951 | 1.065507 | 1.327728 | 1.729133 | 2.093024 | 2.539483 | 2.860935 |
| 20 | 0.127267 | 0.256743 | 0.390906 | 0.532863 | 0.686954 | 0.859964 | 1.064016 | 1.325534 | 1.724718 | 2.085963 | 2.527977 | 2.845340 |
| 21 | 0.127190 | 0.256580 | 0.390637 | 0.532455 | 0.686352 | 0.859074 | 1.062670 | 1.323188 | 1.720743 | 2.079614 | 2.517648 | 2.831360 |
| 22 | 0.127120 | 0.256432 | 0.390394 | 0.532085 | 0.685805 | 0.858266 | 1.061449 | 1.321237 | 1.717144 | 2.073873 | 2.508325 | 2.818756 |
| 23 | 0.127056 | 0.256297 | 0.390171 | 0.531747 | 0.685306 | 0.857530 | 1.060337 | 1.319460 | 1.713872 | 2.068658 | 2.499867 | 2.807336 |
| 24 | 0.126998 | 0.256173 | 0.389967 | 0.531438 | 0.684850 | 0.856855 | 1.059319 | 1.317836 | 1.710882 | 2.063899 | 2.492159 | 2.796940 |
| 25 | 0.126944 | 0.256060 | 0.389780 | 0.531154 | 0.684430 | 0.856236 | 1.058384 | 1.316345 | 1.708141 | 2.059539 | 2.485107 | 2.787436 |
| 26 | 0.126895 | 0.255955 | 0.389607 | 0.530892 | 0.684043 | 0.855665 | 1.057523 | 1.314972 | 1.705618 | 2.055529 | 2.478630 | 2.778715 |
| 27 | 0.126849 | 0.255858 | 0.389448 | 0.530649 | 0.683685 | 0.855137 | 1.056727 | 1.313703 | 1.703288 | 2.051831 | 2.472660 | 2.770683 |
| 28 | 0.126806 | 0.255768 | 0.389299 | 0.530424 | 0.683353 | 0.854647 | 1.055989 | 1.312527 | 1.701131 | 2.048407 | 2.467140 | 2.763262 |
| 29 | 0.126767 | 0.255684 | 0.389161 | 0.530214 | 0.683044 | 0.854192 | 1.055302 | 1.311434 | 1.699127 | 2.045230 | 2.462021 | 2.756386 |
| 30 | 0.126730 | 0.255605 | 0.389032 | 0.530019 | 0.682756 | 0.853767 | 1.054662 | 1.310415 | 1.697261 | 2.042272 | 2.457262 | 2.749996 |
| 40 | 0.126462 | 0.255039 | 0.388100 | 0.528606 | 0.680673 | 0.850700 | 1.050046 | 1.303077 | 1.683851 | 2.021075 | 2.423257 | 2.704459 |
| 60 | 0.126194 | 0.254473 | 0.387170 | 0.527198 | 0.678601 | 0.847653 | 1.045469 | 1.295821 | 1.670649 | 2.000298 | 2.390119 | 2.660283 |
| 120 | 0.125928 | 0.253910 | 0.386244 | 0.525796 | 0.676540 | 0.844627 | 1.040932 | 1.288646 | 1.657051 | 1.979930 | 2.357825 | 2.617421 |
| ∞ | 0.125661 | 0.253347 | 0.385320 | 0.524401 | 0.674490 | 0.841621 | 1.036433 | 1.281552 | 1.644854 | 1.959964 | 2.326348 | 2.575829 |

Inversa de la función de distribución F de Snedecor:

$$\mathcal{F}_{m,n}(x) = P[F_{m,n} \leq x], \quad \mathcal{F}_{m,n}^{-1}(F_{m,n,\alpha}) = \alpha, \quad \mathcal{F}_{m,n,\alpha} = \mathcal{F}_{m,n}^{-1}(\alpha), \quad \alpha = 0,9$$

| $n \setminus m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 50 | 100 | ∞ |
|-----------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| 1 | 39.8635 | 49.5000 | 53.5932 | 55.8330 | 57.2401 | 58.2044 | 58.9060 | 59.4390 | 59.8576 | 60.1950 | 60.7052 | 61.2203 | 61.7403 | 62.0020 | 62.2650 | 62.6881 | 63.0073 | 63.3281 |
| 2 | 8.5263 | 9.0000 | 9.1618 | 9.2434 | 9.2926 | 9.3255 | 9.3491 | 9.3668 | 9.3805 | 9.3916 | 9.4081 | 9.4247 | 9.4413 | 9.4496 | 9.4579 | 9.4712 | 9.4812 | 9.4912 |
| 3 | 5.5383 | 5.4624 | 5.3908 | 5.3426 | 5.3092 | 5.2847 | 5.2662 | 5.2517 | 5.2400 | 5.2304 | 5.2156 | 5.2003 | 5.1845 | 5.1764 | 5.1681 | 5.1546 | 5.1443 | 5.1337 |
| 4 | 4.5448 | 4.3246 | 4.1909 | 4.1072 | 4.0506 | 4.0097 | 3.9790 | 3.9549 | 3.9357 | 3.9199 | 3.8955 | 3.8704 | 3.8443 | 3.8310 | 3.8174 | 3.7952 | 3.7782 | 3.7607 |
| 5 | 4.0604 | 3.7797 | 3.6195 | 3.5202 | 3.4530 | 3.4045 | 3.3679 | 3.3393 | 3.3163 | 3.2974 | 3.2682 | 3.2380 | 3.2066 | 3.1905 | 3.1741 | 3.1471 | 3.1263 | 3.1050 |
| 6 | 3.7759 | 3.4633 | 3.2888 | 3.1808 | 3.1075 | 3.0546 | 3.0145 | 2.9830 | 2.9577 | 2.9369 | 2.9047 | 2.8712 | 2.8363 | 2.8183 | 2.8000 | 2.7697 | 2.7463 | 2.7222 |
| 7 | 3.5894 | 3.2574 | 3.0741 | 2.9605 | 2.8833 | 2.8274 | 2.7849 | 2.7516 | 2.7247 | 2.7025 | 2.6681 | 2.6322 | 2.5947 | 2.5753 | 2.5555 | 2.5226 | 2.4971 | 2.4708 |
| 8 | 3.4579 | 3.1131 | 2.9238 | 2.8064 | 2.7264 | 2.6683 | 2.6241 | 2.5893 | 2.5612 | 2.5380 | 2.5020 | 2.4642 | 2.4246 | 2.4041 | 2.3830 | 2.3481 | 2.3208 | 2.2926 |
| 9 | 3.3603 | 3.0065 | 2.8129 | 2.6927 | 2.6106 | 2.5509 | 2.5053 | 2.4694 | 2.4403 | 2.4163 | 2.3789 | 2.3396 | 2.2983 | 2.2768 | 2.2547 | 2.2180 | 2.1892 | 2.1592 |
| 10 | 3.2850 | 2.9245 | 2.7277 | 2.6053 | 2.5216 | 2.4606 | 2.4140 | 2.3772 | 2.3473 | 2.3226 | 2.2841 | 2.2435 | 2.2007 | 2.1784 | 2.1554 | 2.1171 | 2.0869 | 2.0554 |
| 11 | 3.2252 | 2.8595 | 2.6602 | 2.5362 | 2.4512 | 2.3891 | 2.3416 | 2.3040 | 2.2735 | 2.2482 | 2.2087 | 2.1671 | 2.1230 | 2.1000 | 2.0762 | 2.0364 | 2.0050 | 1.9721 |
| 12 | 3.1765 | 2.8068 | 2.6055 | 2.4801 | 2.3940 | 2.3310 | 2.2828 | 2.2446 | 2.2135 | 2.1878 | 2.1474 | 2.1048 | 2.0597 | 2.0360 | 2.0115 | 1.9704 | 1.9379 | 1.9036 |
| 13 | 3.1362 | 2.7632 | 2.5603 | 2.4337 | 2.3467 | 2.2830 | 2.2341 | 2.1953 | 2.1638 | 2.1376 | 2.0966 | 2.0532 | 2.0070 | 1.9827 | 1.9576 | 1.9153 | 1.8817 | 1.8462 |
| 14 | 3.1022 | 2.7265 | 2.5222 | 2.3947 | 2.3069 | 2.2426 | 2.1931 | 2.1539 | 2.1220 | 2.0954 | 2.0537 | 2.0095 | 1.9625 | 1.9377 | 1.9119 | 1.8686 | 1.8340 | 1.7973 |
| 15 | 3.0732 | 2.6952 | 2.4898 | 2.3614 | 2.2730 | 2.2081 | 2.1582 | 2.1185 | 2.0862 | 2.0593 | 2.0171 | 1.9722 | 1.9243 | 1.8990 | 1.8728 | 1.8284 | 1.7929 | 1.7551 |
| 16 | 3.0481 | 2.6682 | 2.4618 | 2.3327 | 2.2438 | 2.1783 | 2.1280 | 2.0880 | 2.0553 | 2.0281 | 1.9854 | 1.9399 | 1.8913 | 1.8656 | 1.8388 | 1.7934 | 1.7570 | 1.7182 |
| 17 | 3.0262 | 2.6446 | 2.4374 | 2.3077 | 2.2183 | 2.1524 | 2.1017 | 2.0613 | 2.0284 | 2.0009 | 1.9577 | 1.9117 | 1.8624 | 1.8362 | 1.8090 | 1.7628 | 1.7255 | 1.6856 |
| 18 | 3.0070 | 2.6239 | 2.4160 | 2.2858 | 2.1958 | 2.1296 | 2.0785 | 2.0379 | 2.0047 | 1.9770 | 1.9333 | 1.8868 | 1.8368 | 1.8103 | 1.7827 | 1.7356 | 1.6976 | 1.6567 |
| 19 | 2.9899 | 2.6056 | 2.3970 | 2.2663 | 2.1760 | 2.1094 | 2.0580 | 2.0171 | 1.9836 | 1.9557 | 1.9117 | 1.8647 | 1.8142 | 1.7873 | 1.7592 | 1.7114 | 1.6726 | 1.6308 |
| 20 | 2.9747 | 2.5893 | 2.3801 | 2.2489 | 2.1582 | 2.0913 | 2.0397 | 1.9985 | 1.9649 | 1.9367 | 1.8924 | 1.8449 | 1.7938 | 1.7667 | 1.7382 | 1.6896 | 1.6501 | 1.6074 |
| 21 | 2.9610 | 2.5746 | 2.3649 | 2.2333 | 2.1423 | 2.0751 | 2.0233 | 1.9819 | 1.9480 | 1.9197 | 1.8750 | 1.8271 | 1.7756 | 1.7481 | 1.7193 | 1.6700 | 1.6298 | 1.5862 |
| 22 | 2.9486 | 2.5613 | 2.3512 | 2.2193 | 2.1279 | 2.0605 | 2.0084 | 1.9668 | 1.9327 | 1.9043 | 1.8593 | 1.8111 | 1.7590 | 1.7312 | 1.7021 | 1.6521 | 1.6113 | 1.5668 |
| 23 | 2.9374 | 2.5493 | 2.3387 | 2.2065 | 2.1149 | 2.0472 | 1.9949 | 1.9531 | 1.9189 | 1.8903 | 1.8450 | 1.7964 | 1.7439 | 1.7159 | 1.6864 | 1.6358 | 1.5944 | 1.5490 |
| 24 | 2.9271 | 2.5383 | 2.3274 | 2.1949 | 2.1030 | 2.0351 | 1.9826 | 1.9407 | 1.9063 | 1.8775 | 1.8319 | 1.7831 | 1.7302 | 1.7019 | 1.6721 | 1.6209 | 1.5788 | 1.5327 |
| 25 | 2.9177 | 2.5283 | 2.3170 | 2.1842 | 2.0922 | 2.0241 | 1.9714 | 1.9292 | 1.8947 | 1.8658 | 1.8200 | 1.7708 | 1.7175 | 1.6890 | 1.6589 | 1.6072 | 1.5645 | 1.5176 |
| 26 | 2.9091 | 2.5191 | 2.3075 | 2.1745 | 2.0822 | 2.0139 | 1.9610 | 1.9188 | 1.8841 | 1.8550 | 1.8090 | 1.7596 | 1.7059 | 1.6771 | 1.6468 | 1.5945 | 1.5513 | 1.5036 |
| 27 | 2.9012 | 2.5106 | 2.2987 | 2.1655 | 2.0730 | 2.0045 | 1.9515 | 1.9091 | 1.8743 | 1.8451 | 1.7989 | 1.7492 | 1.6951 | 1.6662 | 1.6356 | 1.5827 | 1.5390 | 1.4906 |
| 28 | 2.8938 | 2.5028 | 2.2906 | 2.1571 | 2.0645 | 1.9958 | 1.9427 | 1.9001 | 1.8652 | 1.8359 | 1.7895 | 1.7396 | 1.6852 | 1.6560 | 1.6252 | 1.5718 | 1.5276 | 1.4784 |
| 29 | 2.8870 | 2.4955 | 2.2831 | 2.1494 | 2.0566 | 1.9878 | 1.9345 | 1.8918 | 1.8568 | 1.8274 | 1.7808 | 1.7305 | 1.6759 | 1.6465 | 1.6155 | 1.5616 | 1.5169 | 1.4670 |
| 30 | 2.8807 | 2.4887 | 2.2761 | 2.1422 | 2.0492 | 1.9803 | 1.9269 | 1.8841 | 1.8490 | 1.8195 | 1.7727 | 1.7223 | 1.6673 | 1.6377 | 1.6065 | 1.5522 | 1.5069 | 1.4564 |
| 40 | 2.8354 | 2.4404 | 2.2261 | 2.0909 | 1.9968 | 1.9269 | 1.8725 | 1.8289 | 1.7929 | 1.7627 | 1.7146 | 1.6624 | 1.6052 | 1.5741 | 1.5411 | 1.4830 | 1.4336 | 1.3769 |
| 60 | 2.7911 | 2.3933 | 2.1774 | 2.0410 | 1.9457 | 1.8747 | 1.8194 | 1.7748 | 1.7380 | 1.7070 | 1.6574 | 1.6034 | 1.5435 | 1.5107 | 1.4755 | 1.4126 | 1.3576 | 1.2915 |
| 100 | 2.7564 | 2.3564 | 2.1394 | 2.0019 | 1.9057 | 1.8339 | 1.7778 | 1.7324 | 1.6949 | 1.6632 | 1.6124 | 1.5566 | 1.4943 | 1.4600 | 1.4227 | 1.3548 | 1.2934 | 1.2142 |
| 200 | 2.7308 | 2.3293 | 2.1114 | 1.9732 | 1.8763 | 1.8038 | 1.7470 | 1.7011 | 1.6630 | 1.6308 | 1.5789 | 1.5218 | 1.4575 | 1.4217 | 1.3826 | 1.3100 | 1.2418 | 1.1439 |
| ∞ | 2.7055 | 2.3026 | 2.0838 | 1.9449 | 1.8473 | 1.7741 | 1.7167 | 1.6702 | 1.6315 | 1.5987 | 1.5458 | 1.4871 | 1.4206 | 1.3832 | 1.3419 | 1.2633 | 1.1850 | 1.0000 |

Inversa de la función de distribución F de Snedecor:

$$\mathcal{F}_{m,n}(x) = P[F_{m,n} \leq x], \quad \mathcal{F}_{m,n}^{-1}(\alpha) = \alpha, \quad F_{m,n,\alpha} = \mathcal{F}_{m,n}^{-1}(\alpha), \quad \alpha = 0,95$$

| $n \setminus m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 50 | 100 | ∞ |
|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | 161.4476 | 199.5000 | 215.7073 | 224.5832 | 230.1619 | 233.9860 | 236.7684 | 238.8827 | 240.5433 | 241.8817 | 243.9060 | 245.9499 | 248.0131 | 249.0518 | 250.0951 | 251.7742 | 253.0411 | 254.3144 |
| 2 | 18.5128 | 19.0000 | 19.1643 | 19.2468 | 19.2964 | 19.3295 | 19.3532 | 19.3710 | 19.3848 | 19.3959 | 19.4125 | 19.4291 | 19.4458 | 19.4581 | 19.4624 | 19.4757 | 19.4857 | 19.4957 |
| 3 | 10.1280 | 9.5521 | 9.2766 | 9.1172 | 9.0135 | 8.9406 | 8.8867 | 8.8452 | 8.8123 | 8.7855 | 8.7446 | 8.7029 | 8.6602 | 8.6316 | 8.6166 | 8.5810 | 8.5539 | 8.5264 |
| 4 | 7.7086 | 6.9443 | 6.5914 | 6.3882 | 6.2561 | 6.1631 | 6.0942 | 6.0410 | 5.9988 | 5.9644 | 5.9117 | 5.8578 | 5.8025 | 5.7744 | 5.7459 | 5.6995 | 5.6641 | 5.6281 |
| 5 | 6.6079 | 5.7861 | 5.4095 | 5.1922 | 5.0503 | 4.9503 | 4.8759 | 4.8183 | 4.7725 | 4.7351 | 4.6777 | 4.6188 | 4.5581 | 4.5272 | 4.4957 | 4.4444 | 4.4051 | 4.3650 |
| 6 | 5.9874 | 5.1433 | 4.7571 | 4.5337 | 4.3874 | 4.2839 | 4.2067 | 4.1468 | 4.0990 | 4.0600 | 3.9999 | 3.9381 | 3.8742 | 3.8415 | 3.8082 | 3.7537 | 3.7117 | 3.6689 |
| 7 | 5.5914 | 4.7374 | 4.3468 | 4.1203 | 3.9715 | 3.8660 | 3.7870 | 3.7257 | 3.6767 | 3.6365 | 3.5747 | 3.5107 | 3.4445 | 3.4105 | 3.3758 | 3.3189 | 3.2749 | 3.2298 |
| 8 | 5.3177 | 4.4590 | 4.0662 | 3.8379 | 3.6875 | 3.5806 | 3.5005 | 3.4381 | 3.3881 | 3.3472 | 3.2839 | 3.2184 | 3.1503 | 3.1152 | 3.0794 | 3.0204 | 2.9747 | 2.9276 |
| 9 | 5.1174 | 4.2565 | 3.8625 | 3.6331 | 3.4817 | 3.3738 | 3.2927 | 3.2296 | 3.1789 | 3.1373 | 3.0729 | 3.0061 | 2.9365 | 2.9005 | 2.8637 | 2.8028 | 2.7556 | 2.7067 |
| 10 | 4.9646 | 4.1028 | 3.7083 | 3.4780 | 3.3258 | 3.2172 | 3.1355 | 3.0717 | 3.0204 | 2.9782 | 2.9130 | 2.8450 | 2.7740 | 2.7372 | 2.6996 | 2.6371 | 2.5884 | 2.5379 |
| 11 | 4.8443 | 3.9823 | 3.5874 | 3.3567 | 3.2039 | 3.0946 | 3.0123 | 2.9480 | 2.8962 | 2.8536 | 2.7876 | 2.7186 | 2.6464 | 2.6090 | 2.5705 | 2.5066 | 2.4566 | 2.4045 |
| 12 | 4.7472 | 3.8853 | 3.4903 | 3.2592 | 3.1059 | 2.9961 | 2.9134 | 2.8486 | 2.7964 | 2.7534 | 2.6866 | 2.6169 | 2.5436 | 2.5055 | 2.4663 | 2.4010 | 2.3498 | 2.2962 |
| 13 | 4.6672 | 3.8056 | 3.4105 | 3.1791 | 3.0254 | 2.9153 | 2.8321 | 2.7669 | 2.7144 | 2.6710 | 2.6037 | 2.5331 | 2.4589 | 2.4202 | 2.3803 | 2.3138 | 2.2614 | 2.2064 |
| 14 | 4.6001 | 3.7389 | 3.3439 | 3.1122 | 2.9582 | 2.8477 | 2.7642 | 2.6987 | 2.6458 | 2.6022 | 2.5342 | 2.4630 | 2.3879 | 2.3487 | 2.3082 | 2.2405 | 2.1870 | 2.1307 |
| 15 | 4.5431 | 3.6823 | 3.2874 | 3.0556 | 2.9013 | 2.7905 | 2.7066 | 2.6408 | 2.5876 | 2.5437 | 2.4753 | 2.4034 | 2.3275 | 2.2878 | 2.2468 | 2.1780 | 2.1234 | 2.0658 |
| 16 | 4.4940 | 3.6337 | 3.2389 | 3.0069 | 2.8524 | 2.7413 | 2.6572 | 2.5911 | 2.5377 | 2.4935 | 2.4247 | 2.3522 | 2.2756 | 2.2354 | 2.1938 | 2.1240 | 2.0685 | 2.0096 |
| 17 | 4.4513 | 3.5915 | 3.1968 | 2.9647 | 2.8100 | 2.6987 | 2.6143 | 2.5480 | 2.4943 | 2.4499 | 2.3807 | 2.3077 | 2.2304 | 2.1898 | 2.1477 | 2.0769 | 2.0204 | 1.9604 |
| 18 | 4.4139 | 3.5546 | 3.1599 | 2.9277 | 2.7729 | 2.6613 | 2.5767 | 2.5102 | 2.4563 | 2.4117 | 2.3421 | 2.2686 | 2.1906 | 2.1497 | 2.1071 | 2.0354 | 1.9780 | 1.9168 |
| 19 | 4.3807 | 3.5219 | 3.1274 | 2.8951 | 2.7401 | 2.6283 | 2.5435 | 2.4768 | 2.4227 | 2.3779 | 2.3080 | 2.2341 | 2.1555 | 2.1141 | 2.0712 | 1.9986 | 1.9403 | 1.8780 |
| 20 | 4.3512 | 3.4928 | 3.0984 | 2.8661 | 2.7109 | 2.5990 | 2.5140 | 2.4471 | 2.3928 | 2.3479 | 2.2776 | 2.2033 | 2.1242 | 2.0825 | 2.0391 | 1.9656 | 1.9066 | 1.8432 |
| 21 | 4.3248 | 3.4668 | 3.0725 | 2.8401 | 2.6848 | 2.5727 | 2.4876 | 2.4205 | 2.3660 | 2.3210 | 2.2504 | 2.1757 | 2.0960 | 2.0540 | 2.0102 | 1.9360 | 1.8761 | 1.8117 |
| 22 | 4.3009 | 3.4434 | 3.0491 | 2.8167 | 2.6613 | 2.5491 | 2.4638 | 2.3965 | 2.3419 | 2.2967 | 2.2258 | 2.1508 | 2.0707 | 2.0283 | 1.9842 | 1.9092 | 1.8486 | 1.7831 |
| 23 | 4.2793 | 3.4221 | 3.0280 | 2.7955 | 2.6400 | 2.5277 | 2.4422 | 2.3748 | 2.3201 | 2.2747 | 2.2036 | 2.1282 | 2.0476 | 2.0050 | 1.9605 | 1.8848 | 1.8234 | 1.7570 |
| 24 | 4.2597 | 3.4028 | 3.0088 | 2.7763 | 2.6207 | 2.5082 | 2.4226 | 2.3551 | 2.3002 | 2.2547 | 2.1834 | 2.1077 | 2.0267 | 1.9838 | 1.9390 | 1.8625 | 1.8005 | 1.7330 |
| 25 | 4.2417 | 3.3852 | 2.9912 | 2.7587 | 2.6030 | 2.4904 | 2.4047 | 2.3371 | 2.2821 | 2.2365 | 2.1649 | 2.0889 | 2.0075 | 1.9643 | 1.9192 | 1.8421 | 1.7794 | 1.7110 |
| 26 | 4.2252 | 3.3690 | 2.9752 | 2.7426 | 2.5868 | 2.4741 | 2.3883 | 2.3205 | 2.2655 | 2.2197 | 2.1479 | 2.0716 | 1.9898 | 1.9464 | 1.9010 | 1.8233 | 1.7599 | 1.6906 |
| 27 | 4.2100 | 3.3541 | 2.9604 | 2.7278 | 2.5719 | 2.4591 | 2.3732 | 2.3053 | 2.2501 | 2.2043 | 2.1323 | 2.0558 | 1.9736 | 1.9299 | 1.8842 | 1.8059 | 1.7419 | 1.6717 |
| 28 | 4.1960 | 3.3404 | 2.9467 | 2.7141 | 2.5581 | 2.4453 | 2.3593 | 2.2913 | 2.2360 | 2.1900 | 2.1179 | 2.0411 | 1.9586 | 1.9147 | 1.8687 | 1.7898 | 1.7251 | 1.6541 |
| 29 | 4.1830 | 3.3277 | 2.9340 | 2.7014 | 2.5454 | 2.4324 | 2.3463 | 2.2782 | 2.2229 | 2.1768 | 2.1045 | 2.0275 | 1.9446 | 1.9005 | 1.8543 | 1.7748 | 1.7096 | 1.6376 |
| 30 | 4.1709 | 3.3158 | 2.9223 | 2.6896 | 2.5336 | 2.4205 | 2.3343 | 2.2662 | 2.2107 | 2.1646 | 2.0921 | 2.0148 | 1.9317 | 1.8874 | 1.8409 | 1.7609 | 1.6950 | 1.6223 |
| 40 | 4.0847 | 3.2317 | 2.8387 | 2.6060 | 2.4495 | 2.3359 | 2.2490 | 2.1802 | 2.1240 | 2.0772 | 2.0035 | 1.9254 | 1.8389 | 1.7929 | 1.7444 | 1.6600 | 1.5892 | 1.5089 |
| 60 | 4.0012 | 3.1504 | 2.7581 | 2.5252 | 2.3683 | 2.2541 | 2.1665 | 2.0970 | 2.0401 | 1.9926 | 1.9174 | 1.8364 | 1.7480 | 1.7001 | 1.6491 | 1.5590 | 1.4814 | 1.3893 |
| 100 | 3.9361 | 3.0873 | 2.6955 | 2.4626 | 2.3053 | 2.1906 | 2.1025 | 2.0323 | 1.9748 | 1.9267 | 1.8503 | 1.7675 | 1.6764 | 1.6267 | 1.5733 | 1.4772 | 1.3917 | 1.2832 |
| 200 | 3.8884 | 3.0411 | 2.6498 | 2.4168 | 2.2592 | 2.1441 | 2.0556 | 1.9849 | 1.9269 | 1.8783 | 1.8008 | 1.7166 | 1.6233 | 1.5720 | 1.5164 | 1.4146 | 1.3206 | 1.1885 |
| ∞ | 3.8415 | 2.9957 | 2.6049 | 2.3719 | 2.2141 | 2.0986 | 2.0096 | 1.9384 | 1.8799 | 1.8307 | 1.7522 | 1.6664 | 1.5705 | 1.5173 | 1.4591 | 1.3501 | 1.2434 | 1.0000 |

Inversa de la función de distribución F de Snedecor:

$$F_{m,n}(x) = P[F_{m,n} \leq x], \quad F_{m,n}(F_{m,n,\alpha}) = \alpha, \quad F_{m,n,\alpha} = F_{m,n}^{-1}(\alpha), \quad \alpha = 0,975$$

| $n \setminus m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 50 | 100 | ∞ |
|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|-----------|
| 1 | 647.7890 | 799.5000 | 864.1630 | 899.8333 | 921.8479 | 937.1111 | 948.2169 | 956.6562 | 963.2846 | 968.6274 | 976.7079 | 984.8668 | 993.1028 | 997.2492 | 1001.4144 | 1008.1171 | 1013.1748 | 1018.2583 |
| 2 | 38.5063 | 39.0000 | 39.1655 | 39.2484 | 39.2982 | 39.3315 | 39.3552 | 39.3730 | 39.3869 | 39.3980 | 39.4146 | 39.4313 | 39.4479 | 39.4562 | 39.4646 | 39.4779 | 39.4879 | 39.4979 |
| 3 | 17.4434 | 16.0441 | 15.4392 | 15.1010 | 14.8848 | 14.7347 | 14.6244 | 14.5399 | 14.4731 | 14.4189 | 14.3766 | 14.2827 | 14.1674 | 14.1241 | 14.0805 | 14.0099 | 13.9563 | 13.9021 |
| 4 | 12.2179 | 10.6491 | 9.9792 | 9.6045 | 9.3645 | 9.1973 | 9.0744 | 8.9796 | 8.9047 | 8.8439 | 8.7512 | 8.6565 | 8.5599 | 8.5109 | 8.4613 | 8.3808 | 8.3195 | 8.2573 |
| 5 | 10.0070 | 8.4336 | 7.7636 | 7.3879 | 7.1464 | 6.9777 | 6.8531 | 6.7572 | 6.6811 | 6.6192 | 6.5245 | 6.4277 | 6.3286 | 6.2780 | 6.2269 | 6.1436 | 6.0800 | 6.0153 |
| 6 | 8.8131 | 7.2599 | 6.5988 | 6.2272 | 5.9876 | 5.8198 | 5.6955 | 5.5996 | 5.5234 | 5.4613 | 5.3662 | 5.2687 | 5.1684 | 5.1172 | 5.0652 | 4.9804 | 4.9154 | 4.8491 |
| 7 | 8.0727 | 6.5415 | 5.8898 | 5.5226 | 5.2852 | 5.1186 | 4.9949 | 4.8933 | 4.8232 | 4.7611 | 4.6658 | 4.5678 | 4.4667 | 4.4150 | 4.3624 | 4.2763 | 4.2101 | 4.1423 |
| 8 | 7.5709 | 6.0595 | 5.4160 | 5.0526 | 4.8173 | 4.6517 | 4.5286 | 4.4333 | 4.3572 | 4.2951 | 4.1997 | 4.1012 | 3.9995 | 3.9472 | 3.8940 | 3.8067 | 3.7393 | 3.6720 |
| 9 | 7.2093 | 5.7147 | 5.0781 | 4.7181 | 4.4844 | 4.3197 | 4.1970 | 4.1020 | 4.0260 | 3.9639 | 3.8682 | 3.7694 | 3.6669 | 3.6142 | 3.5604 | 3.4719 | 3.4034 | 3.3329 |
| 10 | 6.9367 | 5.4564 | 4.8256 | 4.4663 | 4.2361 | 4.0721 | 3.9498 | 3.8549 | 3.7790 | 3.7168 | 3.6209 | 3.5217 | 3.4185 | 3.3654 | 3.3110 | 3.2214 | 3.1517 | 3.0798 |
| 11 | 6.7241 | 5.2559 | 4.6300 | 4.2751 | 4.0440 | 3.8807 | 3.7586 | 3.6638 | 3.5879 | 3.5257 | 3.4296 | 3.3299 | 3.2261 | 3.1725 | 3.1176 | 3.0268 | 2.9561 | 2.8828 |
| 12 | 6.5538 | 5.0959 | 4.4742 | 4.1212 | 3.8911 | 3.7283 | 3.6065 | 3.5118 | 3.4358 | 3.3736 | 3.2773 | 3.1772 | 3.0728 | 3.0187 | 2.9633 | 2.8714 | 2.7996 | 2.7249 |
| 13 | 6.4143 | 4.9653 | 4.3472 | 3.9959 | 3.7667 | 3.6043 | 3.4827 | 3.3880 | 3.3120 | 3.2497 | 3.1532 | 3.0527 | 2.9477 | 2.8932 | 2.8372 | 2.7443 | 2.6715 | 2.5955 |
| 14 | 6.2979 | 4.8567 | 4.2417 | 3.8919 | 3.6634 | 3.5014 | 3.3799 | 3.2853 | 3.2093 | 3.1469 | 3.0502 | 2.9493 | 2.8437 | 2.7888 | 2.7324 | 2.6384 | 2.5646 | 2.4872 |
| 15 | 6.1995 | 4.7650 | 4.1528 | 3.8043 | 3.5764 | 3.4147 | 3.2934 | 3.1987 | 3.1227 | 3.0602 | 2.9633 | 2.8621 | 2.7559 | 2.7006 | 2.6437 | 2.5488 | 2.4739 | 2.3953 |
| 16 | 6.1151 | 4.6867 | 4.0768 | 3.7294 | 3.5021 | 3.3406 | 3.2194 | 3.1248 | 3.0488 | 2.9862 | 2.8890 | 2.7875 | 2.6808 | 2.6252 | 2.5678 | 2.4719 | 2.3961 | 2.3163 |
| 17 | 6.0420 | 4.6189 | 4.0112 | 3.6648 | 3.4379 | 3.2767 | 3.1556 | 3.0610 | 2.9849 | 2.9222 | 2.8249 | 2.7230 | 2.6158 | 2.5598 | 2.5020 | 2.4053 | 2.3285 | 2.2474 |
| 18 | 5.9781 | 4.5597 | 3.9539 | 3.6083 | 3.3820 | 3.2209 | 3.0999 | 3.0053 | 2.9291 | 2.8664 | 2.7689 | 2.6667 | 2.5590 | 2.5027 | 2.4445 | 2.3468 | 2.2692 | 2.1869 |
| 19 | 5.9216 | 4.5075 | 3.9034 | 3.5587 | 3.3327 | 3.1718 | 3.0509 | 2.9563 | 2.8801 | 2.8172 | 2.7196 | 2.6171 | 2.5089 | 2.4523 | 2.3937 | 2.2952 | 2.2167 | 2.1333 |
| 20 | 5.8715 | 4.4613 | 3.8587 | 3.5147 | 3.2891 | 3.1283 | 3.0074 | 2.9128 | 2.8365 | 2.7737 | 2.6758 | 2.5731 | 2.4645 | 2.4076 | 2.3486 | 2.2493 | 2.1699 | 2.0853 |
| 21 | 5.8266 | 4.4199 | 3.8188 | 3.4754 | 3.2501 | 3.0895 | 2.9686 | 2.8740 | 2.7977 | 2.7348 | 2.6368 | 2.5338 | 2.4247 | 2.3675 | 2.3082 | 2.2081 | 2.1280 | 2.0422 |
| 22 | 5.7863 | 4.3828 | 3.7829 | 3.4401 | 3.2151 | 3.0546 | 2.9338 | 2.8392 | 2.7628 | 2.6998 | 2.6017 | 2.4984 | 2.3890 | 2.3315 | 2.2718 | 2.1710 | 2.0901 | 2.0032 |
| 23 | 5.7498 | 4.3492 | 3.7505 | 3.4083 | 3.1835 | 3.0232 | 2.9023 | 2.8077 | 2.7313 | 2.6682 | 2.5699 | 2.4664 | 2.3567 | 2.2989 | 2.2389 | 2.1374 | 2.0557 | 1.9677 |
| 24 | 5.7166 | 4.3187 | 3.7211 | 3.3794 | 3.1548 | 2.9946 | 2.8738 | 2.7791 | 2.7027 | 2.6396 | 2.5411 | 2.4374 | 2.3273 | 2.2693 | 2.2090 | 2.1067 | 2.0243 | 1.9353 |
| 25 | 5.6864 | 4.2909 | 3.6943 | 3.3530 | 3.1287 | 2.9685 | 2.8478 | 2.7531 | 2.6766 | 2.6135 | 2.5149 | 2.4110 | 2.3005 | 2.2422 | 2.1816 | 2.0787 | 1.9955 | 1.9055 |
| 26 | 5.6586 | 4.2655 | 3.6697 | 3.3289 | 3.1048 | 2.9447 | 2.8240 | 2.7293 | 2.6528 | 2.5896 | 2.4908 | 2.3867 | 2.2759 | 2.2174 | 2.1565 | 2.0530 | 1.9691 | 1.8781 |
| 27 | 5.6331 | 4.2421 | 3.6472 | 3.3067 | 3.0828 | 2.9228 | 2.8021 | 2.7074 | 2.6309 | 2.5676 | 2.4688 | 2.3644 | 2.2533 | 2.1946 | 2.1334 | 2.0293 | 1.9447 | 1.8527 |
| 28 | 5.6096 | 4.2205 | 3.6264 | 3.2863 | 3.0626 | 2.9027 | 2.7820 | 2.6872 | 2.6106 | 2.5473 | 2.4484 | 2.3438 | 2.2324 | 2.1735 | 2.1121 | 2.0073 | 1.9221 | 1.8291 |
| 29 | 5.5878 | 4.2006 | 3.6072 | 3.2674 | 3.0438 | 2.8840 | 2.7633 | 2.6686 | 2.5919 | 2.5286 | 2.4295 | 2.3248 | 2.2131 | 2.1540 | 2.0923 | 1.9870 | 1.9011 | 1.8072 |
| 30 | 5.5675 | 4.1821 | 3.5894 | 3.2499 | 3.0265 | 2.8667 | 2.7460 | 2.6513 | 2.5746 | 2.5112 | 2.4120 | 2.3072 | 2.1952 | 2.1359 | 2.0739 | 1.9681 | 1.8816 | 1.7867 |
| 40 | 5.4239 | 4.0510 | 3.4633 | 3.1261 | 2.9037 | 2.7444 | 2.6238 | 2.5289 | 2.4519 | 2.3882 | 2.2882 | 2.1819 | 2.0677 | 2.0069 | 1.9429 | 1.8374 | 1.7405 | 1.6371 |
| 60 | 5.2856 | 3.9253 | 3.3425 | 3.0077 | 2.7863 | 2.6274 | 2.5068 | 2.4117 | 2.3344 | 2.2702 | 2.1692 | 2.0613 | 1.9445 | 1.8817 | 1.8152 | 1.6985 | 1.5990 | 1.4821 |
| 100 | 5.1786 | 3.8284 | 3.2496 | 2.9166 | 2.6961 | 2.5374 | 2.4168 | 2.3215 | 2.2439 | 2.1793 | 2.0773 | 1.9679 | 1.8486 | 1.7839 | 1.7148 | 1.5917 | 1.4832 | 1.3473 |
| 200 | 5.1004 | 3.7578 | 3.1820 | 2.8503 | 2.6304 | 2.4720 | 2.3513 | 2.2558 | 2.1780 | 2.1130 | 2.0103 | 1.8996 | 1.7780 | 1.7117 | 1.6403 | 1.5108 | 1.3927 | 1.2290 |
| ∞ | 5.0239 | 3.6889 | 3.1161 | 2.7858 | 2.5665 | 2.4082 | 2.2875 | 2.1918 | 2.1136 | 2.0483 | 1.9447 | 1.8326 | 1.7085 | 1.6402 | 1.5660 | 1.4284 | 1.2956 | 1.0000 |

Inversa de la función de distribución F de Snedecor:

$$\mathcal{F}_{m,n}(x) = P[F_{m,n} \leq x], \quad \mathcal{F}_{m,n}^{-1}(\alpha) = \alpha, \quad F_{m,n,\alpha} = \mathcal{F}_{m,n}^{-1}(\alpha), \quad \alpha = 0,99$$

| $n \setminus m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 50 | 100 | ∞ |
|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | 4052.181 | 4999.500 | 5403.352 | 5624.583 | 5763.650 | 5858.986 | 5928.356 | 5981.070 | 6022.473 | 6055.847 | 6106.321 | 6157.285 | 6208.730 | 6234.631 | 6260.649 | 6302.517 | 6334.110 | 6365.864 |
| 2 | 98.503 | 99.000 | 99.166 | 99.249 | 99.299 | 99.333 | 99.356 | 99.374 | 99.388 | 99.399 | 99.416 | 99.432 | 99.449 | 99.458 | 99.466 | 99.479 | 99.489 | 99.499 |
| 3 | 34.116 | 30.817 | 29.457 | 28.710 | 28.237 | 27.911 | 27.672 | 27.489 | 27.345 | 27.229 | 27.052 | 26.872 | 26.690 | 26.598 | 26.505 | 26.354 | 26.240 | 26.125 |
| 4 | 21.198 | 18.000 | 16.694 | 15.977 | 15.522 | 15.207 | 14.976 | 14.799 | 14.659 | 14.546 | 14.374 | 14.198 | 14.020 | 13.838 | 13.650 | 13.490 | 13.377 | 13.463 |
| 5 | 16.258 | 13.274 | 12.060 | 11.392 | 10.967 | 10.672 | 10.456 | 10.289 | 10.158 | 10.051 | 9.888 | 9.722 | 9.553 | 9.466 | 9.379 | 9.238 | 9.130 | 9.020 |
| 6 | 13.745 | 10.925 | 9.780 | 9.148 | 8.746 | 8.466 | 8.260 | 8.102 | 7.976 | 7.874 | 7.718 | 7.559 | 7.396 | 7.313 | 7.229 | 7.091 | 6.987 | 6.880 |
| 7 | 12.246 | 9.547 | 8.451 | 7.847 | 7.460 | 7.191 | 6.993 | 6.840 | 6.719 | 6.620 | 6.469 | 6.314 | 6.155 | 6.074 | 5.992 | 5.858 | 5.755 | 5.650 |
| 8 | 11.259 | 8.649 | 7.591 | 7.006 | 6.632 | 6.371 | 6.178 | 6.029 | 5.911 | 5.814 | 5.667 | 5.515 | 5.359 | 5.279 | 5.198 | 5.065 | 4.963 | 4.859 |
| 9 | 10.561 | 8.022 | 6.992 | 6.422 | 6.057 | 5.802 | 5.613 | 5.467 | 5.351 | 5.257 | 5.111 | 4.962 | 4.808 | 4.729 | 4.649 | 4.517 | 4.415 | 4.311 |
| 10 | 10.044 | 7.559 | 6.552 | 5.994 | 5.636 | 5.386 | 5.200 | 5.057 | 4.942 | 4.849 | 4.706 | 4.558 | 4.405 | 4.327 | 4.247 | 4.115 | 4.014 | 3.909 |
| 11 | 9.646 | 7.206 | 6.217 | 5.668 | 5.316 | 5.069 | 4.886 | 4.744 | 4.632 | 4.539 | 4.397 | 4.251 | 4.099 | 4.021 | 3.941 | 3.810 | 3.708 | 3.602 |
| 12 | 9.330 | 6.927 | 5.953 | 5.412 | 5.064 | 4.821 | 4.640 | 4.499 | 4.388 | 4.296 | 4.155 | 4.010 | 3.858 | 3.780 | 3.701 | 3.569 | 3.467 | 3.361 |
| 13 | 9.074 | 6.701 | 5.739 | 5.205 | 4.862 | 4.620 | 4.441 | 4.302 | 4.191 | 4.100 | 3.960 | 3.815 | 3.665 | 3.587 | 3.507 | 3.375 | 3.272 | 3.165 |
| 14 | 8.862 | 6.515 | 5.564 | 5.035 | 4.695 | 4.456 | 4.278 | 4.140 | 4.030 | 3.939 | 3.800 | 3.656 | 3.505 | 3.427 | 3.348 | 3.215 | 3.112 | 3.004 |
| 15 | 8.683 | 6.359 | 5.417 | 4.893 | 4.556 | 4.318 | 4.142 | 4.004 | 3.895 | 3.805 | 3.666 | 3.522 | 3.372 | 3.294 | 3.214 | 3.081 | 2.977 | 2.868 |
| 16 | 8.531 | 6.226 | 5.292 | 4.773 | 4.437 | 4.202 | 4.026 | 3.890 | 3.780 | 3.691 | 3.553 | 3.409 | 3.259 | 3.181 | 3.101 | 2.967 | 2.863 | 2.753 |
| 17 | 8.400 | 6.112 | 5.185 | 4.669 | 4.336 | 4.102 | 3.927 | 3.791 | 3.682 | 3.593 | 3.455 | 3.312 | 3.162 | 3.084 | 3.003 | 2.869 | 2.764 | 2.653 |
| 18 | 8.285 | 6.013 | 5.092 | 4.579 | 4.248 | 4.015 | 3.841 | 3.705 | 3.597 | 3.508 | 3.371 | 3.227 | 3.077 | 2.999 | 2.919 | 2.784 | 2.678 | 2.566 |
| 19 | 8.185 | 5.926 | 5.010 | 4.500 | 4.171 | 3.939 | 3.765 | 3.631 | 3.522 | 3.434 | 3.297 | 3.153 | 3.003 | 2.925 | 2.844 | 2.709 | 2.602 | 2.489 |
| 20 | 8.096 | 5.849 | 4.938 | 4.431 | 4.103 | 3.871 | 3.699 | 3.564 | 3.457 | 3.368 | 3.231 | 3.088 | 2.938 | 2.859 | 2.778 | 2.643 | 2.535 | 2.421 |
| 21 | 8.017 | 5.780 | 4.874 | 4.369 | 4.042 | 3.812 | 3.640 | 3.506 | 3.398 | 3.310 | 3.173 | 3.030 | 2.880 | 2.801 | 2.720 | 2.584 | 2.475 | 2.360 |
| 22 | 7.945 | 5.719 | 4.817 | 4.313 | 3.988 | 3.758 | 3.587 | 3.453 | 3.346 | 3.258 | 3.121 | 2.978 | 2.827 | 2.749 | 2.667 | 2.531 | 2.422 | 2.305 |
| 23 | 7.881 | 5.664 | 4.765 | 4.264 | 3.939 | 3.710 | 3.539 | 3.406 | 3.299 | 3.211 | 3.074 | 2.931 | 2.780 | 2.702 | 2.620 | 2.483 | 2.373 | 2.256 |
| 24 | 7.823 | 5.618 | 4.718 | 4.218 | 3.895 | 3.667 | 3.496 | 3.363 | 3.256 | 3.168 | 3.032 | 2.889 | 2.738 | 2.659 | 2.577 | 2.440 | 2.329 | 2.211 |
| 25 | 7.770 | 5.568 | 4.675 | 4.177 | 3.855 | 3.627 | 3.457 | 3.324 | 3.217 | 3.129 | 2.993 | 2.850 | 2.699 | 2.620 | 2.538 | 2.400 | 2.289 | 2.169 |
| 26 | 7.721 | 5.526 | 4.637 | 4.140 | 3.818 | 3.591 | 3.421 | 3.288 | 3.182 | 3.094 | 2.958 | 2.815 | 2.664 | 2.585 | 2.503 | 2.364 | 2.252 | 2.131 |
| 27 | 7.677 | 5.488 | 4.601 | 4.106 | 3.785 | 3.558 | 3.388 | 3.256 | 3.149 | 3.062 | 2.926 | 2.783 | 2.632 | 2.552 | 2.470 | 2.330 | 2.218 | 2.097 |
| 28 | 7.636 | 5.453 | 4.568 | 4.074 | 3.754 | 3.528 | 3.358 | 3.226 | 3.120 | 3.032 | 2.896 | 2.753 | 2.602 | 2.522 | 2.440 | 2.300 | 2.187 | 2.064 |
| 29 | 7.598 | 5.420 | 4.538 | 4.045 | 3.725 | 3.499 | 3.330 | 3.198 | 3.092 | 3.005 | 2.868 | 2.726 | 2.574 | 2.495 | 2.412 | 2.271 | 2.158 | 2.034 |
| 30 | 7.562 | 5.390 | 4.510 | 4.018 | 3.699 | 3.473 | 3.304 | 3.173 | 3.067 | 2.979 | 2.843 | 2.700 | 2.549 | 2.469 | 2.386 | 2.245 | 2.131 | 2.006 |
| 40 | 7.314 | 5.178 | 4.313 | 3.828 | 3.514 | 3.291 | 3.124 | 2.993 | 2.888 | 2.801 | 2.665 | 2.522 | 2.369 | 2.288 | 2.203 | 2.058 | 1.938 | 1.805 |
| 60 | 7.077 | 4.977 | 4.126 | 3.649 | 3.339 | 3.119 | 2.953 | 2.823 | 2.718 | 2.632 | 2.496 | 2.352 | 2.198 | 2.115 | 2.028 | 1.877 | 1.749 | 1.601 |
| 100 | 6.895 | 4.824 | 3.984 | 3.513 | 3.206 | 2.988 | 2.823 | 2.694 | 2.590 | 2.503 | 2.368 | 2.223 | 2.067 | 1.983 | 1.893 | 1.735 | 1.598 | 1.424 |
| 200 | 6.763 | 4.713 | 3.881 | 3.414 | 3.110 | 2.893 | 2.730 | 2.601 | 2.497 | 2.411 | 2.275 | 2.129 | 1.971 | 1.886 | 1.794 | 1.629 | 1.481 | 1.279 |
| ∞ | 6.635 | 4.605 | 3.782 | 3.319 | 3.017 | 2.802 | 2.639 | 2.511 | 2.407 | 2.321 | 2.185 | 2.039 | 1.878 | 1.791 | 1.696 | 1.523 | 1.358 | 1.000 |

Inversa de la función de distribución F de Snedecor:

$$F_{m,n}(x) = P[F_{m,n} \leq x], \quad F_{m,n}(F_{m,n,\alpha}) = \alpha, \quad F_{m,n,\alpha} = F_{m,n}^{-1}(\alpha), \quad \alpha = 0,995$$

| $n \setminus m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 50 | 100 | ∞ |
|-----------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 1 | 16210.72 | 19999.50 | 21614.74 | 22499.58 | 23055.80 | 23437.11 | 23714.57 | 23925.41 | 24091.00 | 24224.49 | 24426.37 | 24630.21 | 24835.97 | 24939.57 | 25043.63 | 25211.09 | 25337.45 | 25464.46 |
| 2 | 198.50 | 199.00 | 199.17 | 199.25 | 199.30 | 199.33 | 199.36 | 199.37 | 199.39 | 199.40 | 199.42 | 199.43 | 199.45 | 199.46 | 199.47 | 199.48 | 199.49 | 199.50 |
| 3 | 55.55 | 49.80 | 47.47 | 46.19 | 45.39 | 44.84 | 44.43 | 44.13 | 43.88 | 43.69 | 43.39 | 43.08 | 42.78 | 42.62 | 42.47 | 42.21 | 42.02 | 41.83 |
| 4 | 31.33 | 26.28 | 24.26 | 23.15 | 22.46 | 21.97 | 21.62 | 21.35 | 21.14 | 20.97 | 20.70 | 20.44 | 20.17 | 20.03 | 19.89 | 19.67 | 19.50 | 19.32 |
| 5 | 22.78 | 18.31 | 16.53 | 15.56 | 14.94 | 14.51 | 14.20 | 13.96 | 13.77 | 13.62 | 13.38 | 13.15 | 12.90 | 12.78 | 12.66 | 12.45 | 12.30 | 12.14 |
| 6 | 16.24 | 12.40 | 10.88 | 10.05 | 9.52 | 9.16 | 8.89 | 8.68 | 8.51 | 8.38 | 8.18 | 7.97 | 7.75 | 7.64 | 7.53 | 7.35 | 7.22 | 7.08 |
| 8 | 14.69 | 11.04 | 9.60 | 8.81 | 8.30 | 7.95 | 7.69 | 7.50 | 7.34 | 7.21 | 7.01 | 6.81 | 6.61 | 6.50 | 6.40 | 6.22 | 6.09 | 5.95 |
| 9 | 13.61 | 10.11 | 8.72 | 7.96 | 7.47 | 7.13 | 6.88 | 6.69 | 6.54 | 6.42 | 6.23 | 6.03 | 5.83 | 5.73 | 5.62 | 5.45 | 5.32 | 5.19 |
| 10 | 12.83 | 9.43 | 8.08 | 7.34 | 6.87 | 6.54 | 6.30 | 6.12 | 5.97 | 5.85 | 5.66 | 5.47 | 5.27 | 5.17 | 5.07 | 4.90 | 4.77 | 4.64 |
| 11 | 12.23 | 8.91 | 7.60 | 6.88 | 6.42 | 6.10 | 5.86 | 5.68 | 5.54 | 5.42 | 5.24 | 5.05 | 4.86 | 4.76 | 4.65 | 4.49 | 4.36 | 4.23 |
| 12 | 11.75 | 8.51 | 7.23 | 6.52 | 6.07 | 5.76 | 5.52 | 5.34 | 5.20 | 5.09 | 4.91 | 4.72 | 4.53 | 4.43 | 4.33 | 4.17 | 4.04 | 3.90 |
| 13 | 11.37 | 8.19 | 6.93 | 6.23 | 5.79 | 5.48 | 5.25 | 5.08 | 4.94 | 4.82 | 4.64 | 4.46 | 4.27 | 4.17 | 4.07 | 3.91 | 3.78 | 3.65 |
| 14 | 11.06 | 7.92 | 6.68 | 6.00 | 5.56 | 5.26 | 5.03 | 4.86 | 4.72 | 4.60 | 4.43 | 4.25 | 4.06 | 3.96 | 3.86 | 3.70 | 3.57 | 3.44 |
| 15 | 10.80 | 7.70 | 6.48 | 5.80 | 5.37 | 5.07 | 4.85 | 4.67 | 4.54 | 4.42 | 4.25 | 4.07 | 3.88 | 3.79 | 3.69 | 3.52 | 3.39 | 3.26 |
| 16 | 10.58 | 7.51 | 6.30 | 5.64 | 5.21 | 4.91 | 4.69 | 4.52 | 4.38 | 4.27 | 4.10 | 3.92 | 3.73 | 3.64 | 3.54 | 3.37 | 3.25 | 3.11 |
| 17 | 10.38 | 7.35 | 6.16 | 5.50 | 5.07 | 4.78 | 4.56 | 4.39 | 4.25 | 4.14 | 3.97 | 3.79 | 3.61 | 3.51 | 3.41 | 3.25 | 3.12 | 2.98 |
| 18 | 10.22 | 7.21 | 6.03 | 5.37 | 4.96 | 4.66 | 4.44 | 4.28 | 4.14 | 4.03 | 3.86 | 3.68 | 3.50 | 3.40 | 3.30 | 3.14 | 3.01 | 2.87 |
| 19 | 10.07 | 7.09 | 5.92 | 5.27 | 4.85 | 4.56 | 4.34 | 4.18 | 4.04 | 3.93 | 3.76 | 3.59 | 3.40 | 3.31 | 3.21 | 3.04 | 2.91 | 2.78 |
| 20 | 9.94 | 6.99 | 5.82 | 5.17 | 4.76 | 4.47 | 4.26 | 4.09 | 3.96 | 3.85 | 3.68 | 3.50 | 3.32 | 3.22 | 3.12 | 2.96 | 2.83 | 2.69 |
| 21 | 9.83 | 6.89 | 5.73 | 5.09 | 4.68 | 4.39 | 4.18 | 4.01 | 3.88 | 3.77 | 3.60 | 3.43 | 3.24 | 3.15 | 3.05 | 2.88 | 2.75 | 2.61 |
| 22 | 9.73 | 6.81 | 5.65 | 5.02 | 4.61 | 4.32 | 4.11 | 3.94 | 3.81 | 3.70 | 3.54 | 3.36 | 3.18 | 3.08 | 2.98 | 2.82 | 2.69 | 2.55 |
| 23 | 9.63 | 6.73 | 5.58 | 4.95 | 4.54 | 4.26 | 4.05 | 3.88 | 3.75 | 3.64 | 3.47 | 3.30 | 3.12 | 3.02 | 2.92 | 2.76 | 2.62 | 2.48 |
| 24 | 9.55 | 6.66 | 5.52 | 4.89 | 4.49 | 4.20 | 3.99 | 3.83 | 3.69 | 3.59 | 3.42 | 3.25 | 3.06 | 2.97 | 2.87 | 2.70 | 2.57 | 2.43 |
| 25 | 9.48 | 6.60 | 5.46 | 4.84 | 4.43 | 4.15 | 3.94 | 3.78 | 3.64 | 3.54 | 3.37 | 3.20 | 3.01 | 2.92 | 2.82 | 2.65 | 2.52 | 2.38 |
| 26 | 9.41 | 6.54 | 5.41 | 4.79 | 4.38 | 4.10 | 3.89 | 3.73 | 3.60 | 3.49 | 3.33 | 3.15 | 2.97 | 2.87 | 2.77 | 2.61 | 2.47 | 2.33 |
| 27 | 9.34 | 6.49 | 5.36 | 4.74 | 4.34 | 4.06 | 3.85 | 3.69 | 3.56 | 3.45 | 3.28 | 3.11 | 2.93 | 2.83 | 2.73 | 2.57 | 2.43 | 2.29 |
| 28 | 9.28 | 6.44 | 5.32 | 4.70 | 4.30 | 4.02 | 3.81 | 3.65 | 3.52 | 3.41 | 3.25 | 3.07 | 2.89 | 2.79 | 2.69 | 2.53 | 2.39 | 2.25 |
| 29 | 9.23 | 6.40 | 5.28 | 4.66 | 4.26 | 3.98 | 3.77 | 3.61 | 3.48 | 3.38 | 3.21 | 3.04 | 2.86 | 2.76 | 2.66 | 2.49 | 2.36 | 2.21 |
| 30 | 9.18 | 6.35 | 5.24 | 4.62 | 4.23 | 3.95 | 3.74 | 3.58 | 3.45 | 3.34 | 3.18 | 3.01 | 2.82 | 2.73 | 2.63 | 2.46 | 2.32 | 2.18 |
| 40 | 8.83 | 6.07 | 4.98 | 4.37 | 3.99 | 3.71 | 3.51 | 3.35 | 3.22 | 3.12 | 2.95 | 2.78 | 2.60 | 2.50 | 2.40 | 2.23 | 2.09 | 1.93 |
| 60 | 8.49 | 5.79 | 4.73 | 4.14 | 3.76 | 3.49 | 3.29 | 3.13 | 3.01 | 2.90 | 2.74 | 2.57 | 2.39 | 2.29 | 2.19 | 2.01 | 1.86 | 1.69 |
| 100 | 8.24 | 5.59 | 4.54 | 3.96 | 3.59 | 3.33 | 3.13 | 2.97 | 2.85 | 2.74 | 2.58 | 2.41 | 2.23 | 2.13 | 2.02 | 1.84 | 1.68 | 1.49 |
| 200 | 8.06 | 5.44 | 4.41 | 3.84 | 3.47 | 3.21 | 3.01 | 2.86 | 2.73 | 2.63 | 2.47 | 2.30 | 2.11 | 2.01 | 1.90 | 1.71 | 1.54 | 1.31 |
| ∞ | 7.88 | 5.30 | 4.28 | 3.72 | 3.35 | 3.09 | 2.90 | 2.74 | 2.62 | 2.52 | 2.36 | 2.19 | 2.00 | 1.90 | 1.79 | 1.59 | 1.40 | 1.00 |

NÚMEROS ALEATORIOS

Es una tabla de números entre 0 y 9 cuyo orden no obedece ninguna regla de formación; ellas se pueden leer individualmente o en grupos y en cualquier orden, en columna vertical hacia abajo o hacia arriba, en fila, diagonalmente a la derecha o la izquierda. Si se desea formar números aleatorios en un determinado rango, basta con calcular la proporción; otra forma de usarlo es sumando dos números tomados de alguna posición o multiplicarlos.

Una tabla de números aleatorios es útil para seleccionar al azar los individuos de una población conocida que deben formar parte de una muestra.

Por ejemplo: Para ser presentadas estas cifras se agrupan en números de 4 dígitos, formando bloques de 5 filas y 10 columnas, facilitando de esta forma su lectura que puede iniciarse desde cualquier parte de la tabla.

4251 5149 4751 4847 4249 4648 5047 4847 5156 8789
4849 5051 5046 4756 4738 5350 4746 4847 4846 2346
5692 9870 3583 8997 1533 6466 8830 7271 3809 4256
2080 3828 7880 0586 8482 7811 6807 3309 2729 2235
1039 3382 7600 1077 4455 8806 1822 1669 7501 8330

6477 5289 4092 4223 6454 7632 7577 2816 9002 2365
4554 6146 4846 4647 5034 4646 5139 5355 5249 2224
0772 2160 7236 0812 4195 5589 0830 8261 9232 0902
0092 1629 0377 3590 2209 4839 6332 1490 3092 2390
7315 3365 7203 1231 0546 6612 1038 1425 2709 3092

5775 7517 8974 3961 2183 5295 3096 8536 9442 2392
5500 2276 6307 2346 1285 7000 5306 0414 3383 2303
3251 8902 8843 2112 8567 8131 8116 5270 5994 9092
4675 1435 2192 0874 2897 0262 5092 5541 4014 2113
3543 6130 4247 4859 2660 7852 9096 0578 0097 1324

3521 8772 6612 0721 3899 2999 1263 7017 8057 3443
5573 9396 3464 1702 9204 3389 5678 2589 0288 6343
7478 7569 7551 3380 2152 5411 2647 7242 2800 3432
3339 2854 9691 9562 3252 9848 6030 8472 2266 3255
5505 8474 3167 8552 5409 1556 4247 4652 2953 9854

6381 2086 5457 7703 2758 2963 8167 6712 9820 5324
0935 5565 2315 8030 7651 5189 0075 9353 1921 0222
2605 3973 8204 4143 2677 0034 8601 3340 8383 3243
7277 9889 0390 5579 4620 5650 0210 2082 4664 5643

5484 3900 3485 0741 9069 5920 4326 7704 6525 1249

7227 0104 4141 1521 9104 5563 1392 8238 4882 2324
8506 6348 4612 8252 1062 1757 0964 2983 2244 7654
5086 0303 7423 3298 3979 2831 2257 1508 7642 1245
3690 2492 7171 7720 6509 7549 2330 5733 4730 4534
0813 6790 6858 1489 2669 3743 1901 4971 8280 0835

6905 7127 5933 1137 7583 6450 5658 7678 3444 3754
8387 5323 3753 1859 6043 0294 5110 6340 9137 6323
4094 4957 0163 9717 4118 4276 9465 8820 4127 0202
4951 3781 5101 1815 7068 6379 7252 1086 8919 2093
9047 0199 5068 7447 1664 9278 1708 3625 2864 0204

7274 9512 0074 6677 8676 0222 3335 1976 1645 3203
9192 4011 0255 5458 6942 8043 6201 1587 0972 0243
0554 1690 6333 1931 9433 2661 8690 2313 6999 3094
9231 5627 1815 7171 8036 1832 2031 6298 6073 9044
3995 9677 7765 3194 3222 4191 2734 4469 8617 3233

2402 6250 9362 7373 4757 1716 1942 0417 5921 5345
5295 7385 5474 2123 7035 9983 5192 1840 6176 5756
5177 1191 2106 3351 5057 0967 4538 1246 3374 0304
4344 4044 4549 4443 4249 4948 4151 5152 4240 4737
7343 4706 4440 4646 4548 4742 4746 5253 4749 4689



Impreso en los Talleres Gráficos de



MACRO[®]
EMPRESA EDITORA

Surquillo

☎ 243-2003 – 827*2650

Abril 2010