

---

# ***Investigación Comercial Con SPSS***

---

***Diseño y procesamiento de encuestas, Plan de  
muestreo, Regresión y Correlación***

*Martha Lucía Sanclemente Daza*



**UNIVERSIDAD DEL CAUCA**

**ISBN: 978-958-44-1123-5**

# *Introducción*

---

La primera edición del libro se terminó en enero de 2007, y el objetivo era que el texto se convirtiera una guía para los estudiantes que cursan las asignaturas de Planes de Negocios, Formulación y Evaluación de Proyectos, Investigación de Mercados y para aquellos que están realizando su Trabajo de Grado. Hoy un año después el objetivo sigue siendo el mismo y para ello se ha ampliado el alcance de los temas introduciendo dos capítulos más (capítulo uno y el capítulo sexto), con temas no tratados en la versión anterior pero que considero son necesarios para lograr una comprensión integral sobre el tratamiento de la información. También se ampliaron los temas de los capítulos dos y el capítulo quinto. Se trabajará con la Versión 15 del programa SPSS, que incluye en esta nueva versión unos cambios, que lo convierten en uno de los mejores programas y el más versátil, para el procesamiento de información estadística, brindando mejores opciones en la presentación de las gráficas, tablas de frecuencias entre otros.

Hoy día se requiere contar con herramientas como el computador para el procesamiento y registro de la información recopilada en los trabajos de campo que se realizan en la investigación de mercados, e investigaciones de tipo económico y social. La amplia gama de Software estadístico que se consigue en el mercado como EViews, SAS, STATA y SPSS facilita el trabajo y disminuye el tiempo en actividades técnicas y mecánicas, dando una mayor holgura para la labor de análisis e interpretación de la información.

Este texto tiene doble propósito: uno es exponer temas de estadística como el Plan de Muestreo, Diseño de Cuestionarios o Encuestas y Análisis de Regresión, y el segundo es enseñar el manejo del programa SPSS aplicado a estos conceptos. De esta forma, el estudiante tendrá un conocimiento completo y mayor comprensión de los resultados que arroja el programa. Esta segunda edición del libro consta de ocho capítulos los cuales hacen referencia a aspectos teóricos y prácticos con una estructura sencilla que busca ante todo una forma didáctica de procesar la información mediante el método aprender haciendo.

En el **primer capítulo** se expone los conceptos básicos del diseño de investigación exploratoria, concluyente, descriptiva y los diseños de estudios transversales, también se exponen las principales fuentes de investigación y los métodos de recolección de información.

En el **capítulo dos** se presentan los conceptos generales para el diseño de formatos eficaces de encuestas para la recolección de la información, los tipos de preguntas y la estructuración de las mismas con un diseño apropiado para el manejo y aplicación en el programa SPSS. Se incluye los conceptos de escala nominal, ordinal, de intervalo y de relación, al igual que las técnicas de escala.

En el **capítulo tres** se exponen los principales conceptos del Plan de muestreo, los métodos de muestreo, el probabilístico y no probabilístico, el muestreo aleatorio simple, muestreo estratificado, además se presentan diversos aspectos para determinar el tamaño de la muestra. En este capítulo se asume que el estudiante tiene conocimiento de los conceptos básicos de estadística descriptiva.

En el **capítulo cuatro** se exponen la estructura y opciones generales del programa SPSS y las dos maneras de trabajar con él, desde el asistente y la programación con el visor de sintaxis. Se integra a este capítulo el trabajo con la opción de Muestras Complejas, diseñando un plan de muestreo, mediante una muestra aleatoria simple, haciendo uso de la base de datos de encuesta continua de hogares.

En el **capítulo quinto** se presentan la definición de variables en el programa, el procesamiento de los tipos de preguntas más frecuentes en la investigación de mercados, preguntas dicótomas, selección múltiple, preguntas abiertas, manejo de tablas de contingencias y gráficas interactivas.

En el **capítulo sexto** es un capítulo nuevo donde se considera la interpretación y análisis de la información, se explica ampliamente el análisis univariado y bivariado, al igual que los procedimientos para estimar los estadísticos necesarios para la interpretación de las escalas.

En el **capítulo séptimo** se exponen los conceptos de regresión y correlación simple, los cálculos de los parámetros, interpretación y la proyección de datos.

**Capítulo octavo** se explica la extensión del modelo de regresión lineal o formas funcionales de los modelos de regresión lineal, los modelos log-lineal y similogarítmicos (Log – Lin y Lin –log).

El texto termina con tres apéndices.

El **apéndice A** contiene la descripción general de la base de datos del archivo COMERCIO.SAV.

El **apéndice B** contiene la tabla de distribución normal

El **apéndice C** relaciona la tabla para determinar el tamaño de la muestra

## CONVENCIONES USADAS

 <b>Atención</b>	Corresponde a una precaución, avisa del problema y de posibles errores.
 <b>Nota</b>	Corresponde a una aclaración que se debe tener en cuenta para que los resultados sean los esperados.
 <b>Véase también</b>	Lo envía a un documento de referencia o a un tema ya visto.
 <b>En el CD-ROM</b>	Lo envía a un archivo que contiene el CD-ROM que se incluye en la contraportada de este libro.

# Capítulo Uno

## *Introducción a la investigación de mercados y diseños exploratorios y concluyentes*

---

### **1. ESTRUCTURA DE INVESTIGACIÓN**

Éste capítulo centra la atención en dar las pautas a modo de introducción, de los conceptos de la investigación de mercados, las características y utilidades de la misma. Se investiga con el propósito de identificar elementos que ayuden al proceso de toma de decisiones de una organización, evaluar la eficiencia de una determinada política empresarial.

- **¿CÓMO INICIAR EL PROYECTO?**

¿Cómo inicio el proyecto?, ¿Qué información necesito?, ¿Qué medio selecciono para la recopilación de la información?, son las preguntas más frecuentes que los estudiantes se hacen al iniciar la elaboración del proyecto. Este capítulo tiene como objetivo solucionar en parte o totalmente estas inquietudes. Un proyecto de inversión debe tener un orden lógico, que le permita al emprendedor ir avanzando en cada una de las etapas, haciendo uso de una planeación coherente de las actividades necesarias para lograr los objetivos planteados. Ahora centramos la atención en las técnicas de recopilación de la información necesaria para la investigación de mercados.

Thomas C. Kinneer y James R. Taylor señalan “La investigación de mercados, como la aplicación del método científico al marketing, y la característica del método científico es la recopilación, el análisis y la interpretación objetiva de los datos”<sup>1</sup>.

La American Marketing Association define la investigación de mercados como:

---

<sup>1</sup> KINNEAR. Thomas C. y TAYLOR. James R. Investigación de mercados, un enfoque aplicado. Santa Fé de Bogotá. McGrawHill. 1998. P. 6.

“la identificación, recopilación, análisis y difusión de la información de manera sistemática y objetiva, con el propósito de mejorar la toma de decisiones relacionadas con la identificación y solución de problemas y oportunidades de mercadotecnia”.

Es preciso analizar algunos aspectos de esta definición. En primer lugar la investigación de mercados incluye la identificación, recopilación, análisis de la información, actividades que requieren de un proceso que se resume en los siguientes pasos:

- Definición del problema.
- Formulación de un diseño de investigación
- Trabajo de campo o recopilación de datos
- Preparación y análisis de datos
- Preparación y presentación de los informes

### **1.1. DEFINICIÓN DEL PROBLEMA.**

El problema central de investigación de mercados para los proyectos de inversión radica en conocer la percepción que tiene el cliente sobre el producto o servicio que se pretende sacar al mercado, como también establecer las fortalezas y debilidades del proyecto frente a las empresas existentes o competencia.

No se encuentran elementos de tabla de ilustraciones.

- ¿Quién es el cliente?
- ¿Qué compra?
- ¿Cuánto compra?
- ¿Dónde compra?
- ¿Con que frecuencias?
- ¿Qué marcas son las preferidas?
- ¿Qué precio está dispuesto a pagar?

#### **1.1.1. PREGUNTAS FRENTE A LA COMPETENCIA**

- ¿Es pertinente sacar el nuevo producto al mercado?
- ¿Cuál es el precio de la competencia?
- ¿El mercado está en capacidad de recibir un nuevo producto?
- ¿Existe una demanda satisfecha?

## 1.2. FORMULACIÓN DEL DISEÑO DE INVESTIGACIÓN.

El segundo paso es el diseño de investigación. El diseño de investigación es la estructura o plano de ejecución, detalla los procedimientos necesarios para obtener la información requerida.

El diseño de investigación se clasifica en exploratoria y concluyente (figura 1.1). El objetivo principal de la investigación exploratoria es estructurar la idea del proyecto o negocio. La investigación concluyente se caracteriza por ser más estructurada, y se requiere realizar una planeación de los procedimientos a seguir en la investigación. En la tabla 1.1 se presenta las principales diferencias entre la investigación exploratoria y concluyente.

**Tabla 1.1. Diferencias entre investigación exploratoria y concluyente.**

	<b>EXPLORATORIA</b>	<b>CONCLUYENTE</b>
Objetivo	Proporciona ideas para la formulación de la hipótesis.	Prueba hipótesis específicas.
Características	La información se define en forma global y aproximada.	La información se define con claridad.
	Se trabaja con muestras no representativas y la mayor parte de los casos con muestras no probabilísticas.	La muestra es representativa y por lo general son muestras probabilísticas.
	El proceso de investigación es flexible.	El proceso de investigación es estructurado.
	La mayoría de los casos el análisis de los datos es cualitativo.	El análisis de los datos es cuantitativo.
Resultados	Son tentativos.	Son concluyentes, y son fuente de información para la toma de decisiones.

Figura 1.1. Formulación del diseño de investigación<sup>2</sup>.



**1.2.1. Investigación exploratoria.** Es la fase inicial de investigación, se caracteriza por la flexibilidad y versatilidad de sus métodos, porque no se emplean los protocolos y procedimientos de la investigación formal.

**1.2.2. Investigación concluyente.** La investigación concluyente está diseñada para suministrar información para la evaluación de cursos alternativos de acción, se clasifica en Investigación descriptiva y causal.

**3.2.2.1. Investigación descriptiva.** Esta investigación se caracteriza por delimitar las características del mercado o su funcionamiento y tiene los siguientes propósitos:

1. Describir las características más importantes de los consumidores, vendedores, organizaciones o áreas de mercado.
2. Estima el porcentaje de unidades que presentan cierto comportamiento en una población específica.
3. Determinar cómo se perciben las características del producto.

<sup>2</sup> Tomado de MALHOTRA. Naresh k. Investigación de mercados un enfoque práctico. Printice Hall. México. 1997. P. 87

4. Determinar el grado de asociación de las variables de mercado.
5. Da las bases para realizar las proyecciones de la demanda.

- **Diseño de estudios transversales.** Son los diseños descriptivos que se utilizan con mayor frecuencia en la investigación de mercados. Estos diseños de muestra representativa implican que la recopilación de información de los elementos de cualquier muestra se haga una sola vez.

- **Diseño de muestra representativa individual.** Se extrae una sola muestra de entrevistados de la población meta y la información se obtiene una sola vez de esta muestra.

- **Diseño de muestra representativa múltiple.** Hay dos o más muestras de entrevistados y la información se obtiene una sola vez de cada muestra. Con frecuencia, la información de cada una de las muestras se obtiene en diferentes periodos.

- **Diseño longitudinal.** En estos estudios una muestra fija de los elementos de una población se mide de manera repetida. Un diseño longitudinal difiere de un diseño transversal, en que la o las muestras permanecen iguales a través del tiempo. En otras palabras, el estudio se realiza constantemente con las mismas personas, éste tipo de diseño proporciona una información mas confiable de la situación y de los cambios que tienen lugar a través del tiempo. Un ejemplo de este tipo de investigación son las encuestas realizadas en las campañas políticas.

- **Investigación causal.** Esta investigación se utiliza para obtener evidencia de las relaciones de causa y efecto, en este caso los agentes de mercadotecnia continuamente toman decisiones con base en supuestas relaciones causales. La investigación causal es apropiada para los siguientes propósitos:

- Entender qué variables son la causa (variables independientes) y qué variables son el efecto (variables dependientes) de un fenómeno.
- Determinar la naturaleza de las relaciones entre las variables causales y el efecto que debe pronosticarse.

### **1.3. TRABAJO DE CAMPO O RECOPIACIÓN DE DATOS.**

La recopilación de datos incluye el trabajo realizado por el grupo o el investigador en el área de influencia del proyecto. Una selección y entrenamiento de los entrevistadores, como la continua supervisión y evaluación de estas actividades, reducen los errores en la recolección de datos.

#### 1.4. PREPARACIÓN Y ANÁLISIS DE DATOS.

La preparación de los datos incluye su edición, codificación, transcripción y verificación. La verificación, asegura que los datos de los cuestionarios originales se transcriben con exactitud, mientras que el análisis da mayor significado a la información recopilada. Hoy con los adelantos de los sistemas computacionales se encuentra software especializado para procesamiento de la información como SPSS, EVIEWS, STARGRAPH. El objetivo central de este texto es dar las bases necesarios para procesar la información en el programa SPSS.

#### 1.5. PREPARACIÓN DE LOS INFORMES.

El proyecto deberá documentarse en un informe escrito, que consigne de manera específica, las preguntas que se identificaron durante la investigación: el planteamiento, el diseño de la investigación, la recopilación de datos y los procedimientos de análisis de datos adoptados; así como la presentación de los resultados y los hallazgos más importantes.

#### 1.6. FUENTES DE INVESTIGACIÓN.

Existen dos fuentes básicas de datos (Figura 1.2.)

- Datos primarios o datos de los encuestados y
- Datos secundarios.

Figura 1.2. Fuentes de investigación.



**1.6.1. Datos de los encuestados o fuentes de información primarias.** Estos datos se recolectan específicamente para las necesidades inmediatas de la investigación. Uno de los medio de mayor difusión son los cuestionario, éstos se aplican en la mayoría de los casos en forma directa a cualquier persona o institución que se identifique como cliente

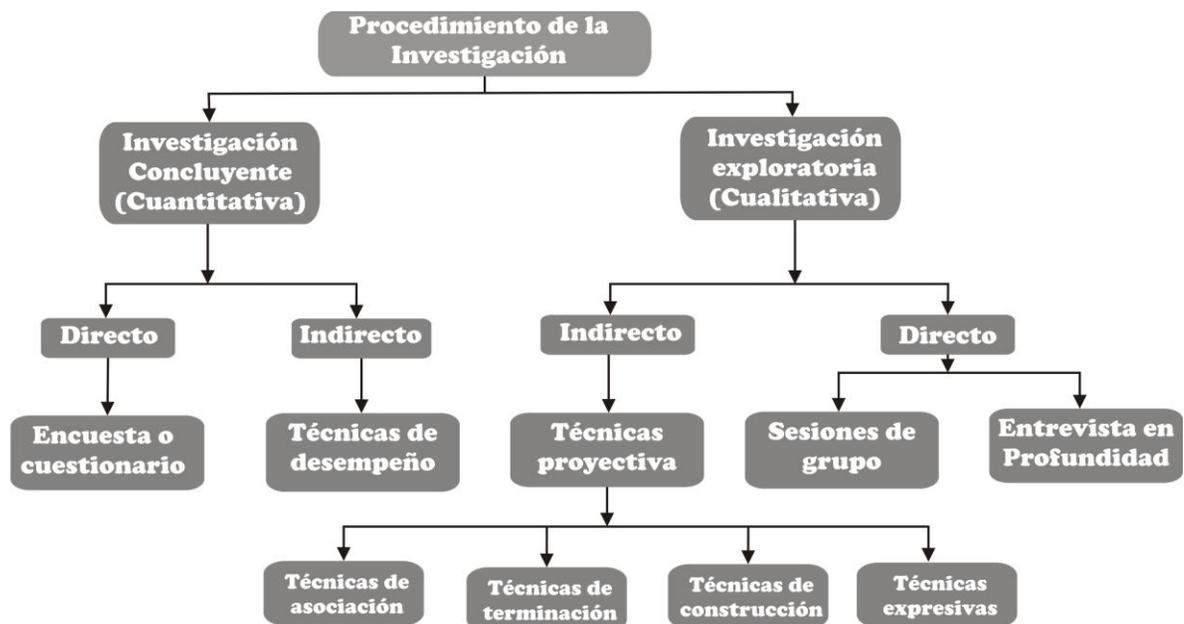
potencial. El cliente se constituyen en una fuente importante de información, es él quien nos comunica sus percepciones, motivaciones sobre el producto y/o servicio.

**1.6.2. Datos secundarios o fuentes secundarias.** Los datos secundarios son datos ya publicados y recolectados para propósitos diferentes de las necesidades inmediatas y específicas del proyecto. Estos datos se ubican en forma rápida en instituciones especializadas y centros de servicios como: Cámaras de Comercio, SENA, DANE, Secretarías de Agricultura, entre otros.

## 1.7. PROCEDIMIENTOS DE INVESTIGACIÓN EXPLORATORIA

Entre los procesos llevados a cabo en la investigación exploratoria (Figura 1.3.) se destacan las sesiones de grupo y las entrevistas en profundidad.

**Figura 1.3. Procedimientos de investigación.**



**1.7.1. Sesiones de grupo.** Es un procedimiento directo que se realiza con un grupo de personas que hacen parte de los clientes o consumidores, su propósito es obtener una visión general, al escuchar las opiniones y puntos de vistas. Con este procedimiento las personas que hacen parte de esta sesión se desinhiben y expresan sus percepciones sin restricciones, el valor de la técnica radica en los descubrimientos inesperados que con frecuencia, se obtienen en un debate libre.

**1.7.1.1. Características principales.** Para que una sesión de grupo tenga los resultados esperados debe cumplir con las siguientes características:

- El número de integrantes de las sesiones de grupo debe ser de 8 a 12 personas.
- El grupo debe ser homogénea, en términos de las características demográficas y socioeconómicas.
- El escenario físico para la sesión, debe ser informal, que relaje y aliente a los comentarios espontáneos.
- Debe servirse bebidas ligeras antes y durante la sesión.
- La duración de la sesión debe ser en promedio entre una y dos horas.
- Se recomienda grabar en video la sesión de la reunión.
- El moderador debe tener habilidad para mantener el interés de los integrantes del grupo y provocar que los participantes expresen sus ideas y sentimientos sobre el tema a investigar.

**1.7.2. Entrevistas en profundidad.** Es una forma no estructurada y directa de obtener información, pero a diferencia de las encuestas, las entrevistas en profundidad se realizan a una sola persona. Una entrevista en profundidad es una entrevista personal directa y no estructurada en la que el entrevistador de forma hábil hace preguntas para descubrir las motivaciones, creencias, actitudes y sentimientos sobre el tema.

**1.7.3. técnicas proyectivas.** Una técnica proyectiva es una forma no estructurada e indirecta de hacer preguntas que alientan a los entrevistados a que proyecten sus motivaciones, creencias, actitudes o sentimientos subyacentes respecto al tema de investigación. Las técnicas proyectivas son técnicas más especializadas y utilizadas en trabajos de psicología, donde se pide a los entrevistados que interpreten el comportamiento de otras personas, en lugar de describir el propio. Al interpretar el comportamiento de otros, los entrevistados proyectan de manera indirecta sus motivaciones, creencias, actitudes o sentimientos acerca de la situación analizada. Esta técnicas se clasifican como de asociación, terminación, construcción y expresión.

**2. Técnicas de asociación.** La asociación de palabras es la más conocida de estas técnicas. En ella, se le presenta al entrevistado una lista de palabras una por una, y se le pide que responda a cada una con la primera palabra que venga a su mente. La palabra de interés, que se llama palabra de prueba, se intercala en toda la lista, que también contiene algunas palabras neutrales, o de relleno, para ocultar el propósito.

**3. Técnicas de terminación.** Las más comunes en la investigación de mercados, son la terminación de enunciados y la terminación de historias. Se le pide al entrevistado que termine una situación de estímulo incompleta.

**4. Técnicas de construcción.** Son aquellas en la cual, se le piden a los entrevistados que construyan una respuesta en forma de historia, diálogo o descripción. Las dos

técnicas de construcción principales son las técnicas de respuestas a ilustraciones y las pruebas de caricaturas.

**5. Técnicas expresivas.** En las técnicas expresivas, se presenta a los entrevistados una situación verbal o visual y se les pide que relacionen los sentimientos y actitudes de otras personas con la situación. Los entrevistados no expresan sentimientos o actitudes propios, sino de otras personas. Las dos técnicas expresivas son la representación de papeles y la técnica de la tercera persona.

El estudiante que desee profundizar en éstos temas le recomiendo que consulte la bibliografía que se presenta al final de libro, de otra parte, es importante destacar que las técnicas proyectivas son objeto de investigaciones más especializadas, por tanto, los conceptos se salen del alcance de este texto.

## **1.8. PROCEDIMIENTOS DE INVESTIGACIÓN CONCLUYENTE**

Los diseños de investigación concluyente corresponden habitualmente con la última fase de la investigación. Después de agotadas las fuentes secundarias y realizado la investigación exploratoria, el objetivo principal de los diseños concluyentes es de probar o contrastar las hipótesis resultantes en las etapas anteriores.

### **Características del diseño concluyente**

- Constituyen procesos más estructurados y formales que los utilizados en la investigación exploratoria.
- Las muestras utilizadas son representativas y estadísticamente significativas.
- El análisis de datos es de tipo cuantitativo.
- Los resultados obtenidos se utilizan en la toma de decisiones.

## **1.9. DEFINICIÓN DE LA INFORMACIÓN NECESARIA.**

La investigación de mercados puede considerarse como una técnica de pronóstico, diseñada para facilitar el proceso de predicción del comportamiento del mercado y los datos que pueden obtenerse de los encuestados corresponden a:

- Comportamiento anterior.
- Actitudes.
- Características del encuestado<sup>3</sup>.

• **Comportamiento anterior.** Sin duda todas las evidencias del comportamiento anterior tienen influencia sobre el comportamiento futuro. En el estudio de investigación

---

<sup>3</sup> KINNEAR. Thomas C. Investigación de mercados – un enfoque aplicado. McGrawHill. Santa fé de Bogotá. 1998. P.323.

de mercados se recolecta evidencias sobre el comportamiento de un encuestado con respecto a la compra y al uso de un producto o una marca.

Entre las preguntas frecuentes con respecto al comportamiento anterior son:

- ¿Qué compró o utilizó?
- ¿Cuánto compró?
- ¿Cómo se compró?
- ¿Dónde se compró?
- ¿Cuándo se compró?

• **Actitudes.** Los datos sobre actitudes se utilizan para identificar segmentos de mercado, para desarrollar estrategias de “posicionamiento” y para evaluar programas publicitarios.

Una actitud tiene tres (3) componentes:

1. Componente cognoscitivo.
2. Componente afectivo.
3. Componente de comportamiento.

El **componente cognoscitivo** es el conocimiento que la persona tiene acerca del producto y/o servicio, el **componente afectivo** son los sentimientos de una persona acerca del objeto. Por ejemplo: los conceptos de lo bueno o malo del producto y **componente de comportamiento** es la disposición favorable de una persona para responder al objeto con su comportamiento.

• **Características del encuestado.** La descripción del encuestado implica analizar las características demográficas, socioeconómicas y psicológicas. Estas características están correlacionadas con el comportamiento de compra. Los datos de clasificación como edad, sexo, estado civil, tamaño del grupo familiar, ingresos, ocupación y nivel de escolaridad son útiles para clasificar al cliente o encuestado y al igual, para validar datos cualitativos de la información.

## **1.10. MÉTODOS DE RECOLECCIÓN DE INFORMACIÓN.**

Los métodos básicos para la recolección de información son:

- La observación y
- La comunicación.

**1.10.1. La observación.** Sin lugar a dudas es el mejor método de investigación, comprende el registro de los patrones de conducta de personas y usuarios en forma sistemática para obtener información sobre el problema a investigar. El observador no

pregunta ni se comunica con las personas que observa, y esta puede ser estructuradas, no estructuradas, directas e indirectas.

### **Métodos de observación**

- a) Observación personal.
- b) Observación mecánica.
- c) Auditorias.
- d) Análisis de contenido.
- e) Análisis de vestigios<sup>4</sup>.

**1.10.2. La comunicación.** El método de comunicación se basa en hacer preguntas a los encuestados, el instructivo de recolección de datos que se utiliza en este proceso es el cuestionario, el cuestionario se ha convertido en el instrumento predominante de recolección de datos en investigación de mercados y se identifican tres formas de aplicación de éste medio:

- Encuesta personal.
- Encuesta telefónica.
- Encuesta por correo o encuesta postal.

---

<sup>4</sup> Véase en MALHOTRA. Naresh K. Investigación de Mercados – un enfoque práctico. P. 214 a 222.



# Capítulo dos

## *Diseño de encuesta y técnicas de escalas*

---

### **INTRODUCCIÓN**

La encuesta es una de las principales técnicas cuantitativas empleadas habitualmente en la obtención de información primaria, en este capítulo se definen las etapas para la elaboración de una encuestas, los distintos tipos de encuestas (personal, postal y telefónica), también se proporcionan los conceptos para el diseño del cuestionario, haciendo énfasis en los distintos tipos de preguntas y escalas que se pueden emplear. Una vez diseñado el cuestionario debe someterse a una prueba piloto para realizar los cambios pertinentes antes de ser aplicado en forma definitiva.

### **2. ETAPAS PARA ELABORAR UNA ENCUESTA**

El procedimiento para elaborar una encuesta consta de una serie de etapas que se enuncian a continuación:

- **Determinación de los objetivos.** Esta fase es importante para el investigador puesto que le permite tener una orientación clara a las metas a logra con la investigación y la clase de información que requiere.
- **Determinación del tipo de encuesta.** Una vez definidos los objetivos el siguiente paso es elegir entre la encuesta personal, postal o telefónica, teniendo en cuenta las ventajas y desventajas de cada una de ellas.
- **Diseño del cuestionario.** Consiste en elaborar el formulario de la encuesta, seleccionando tipos de preguntas y tipos de escala. Este procedimiento se analizará en detalle en éste capítulo.
- **Codificación.** Se debe asignar códigos numéricos al cuestionario identificando las distintas variables de la encuesta. La codificación hace posible la posterior tabulación de los datos obtenidos.

- **Muestreo.** Consiste en la determinación del tamaño de la muestra y la selección de las unidades muestrales. El capítulo 3 está dedicado a este tema.

## 2.1. DETERMINACIÓN DE LOS OBJETIVOS.

Para definir los objetivos se deben tener en cuenta las siguientes consideraciones:

- Su formulación debe llevar a resultados concretos en el desarrollo de la investigación.
- El alcance de los mismos debe estar dentro de las posibilidades del investigador y no corresponder a acciones ajenas.
- Se recomienda formular un objetivo general y uno o varios objetivos específicos. El primero debe ofrecer resultados amplios o la meta que se persigue con la investigación; los objetivos específicos se refieren a situaciones particulares que inciden o forman parte de situaciones propias del objetivo general.
- El número de objetivos depende del alcance y los propósitos del estudio, y del criterio del investigador. Se recomienda no comprometerse con varios objetivos específicos, se debe ser concreto en las acciones planteadas.

## 2.2. DETERMINACIÓN DEL TIPO DE ENCUESTA.

**2.2.1. Encuesta personal.** En la encuesta personal el entrevistador y entrevistado se encuentran cara a cara para diligenciar el formulario. La encuesta puede realizarse en el hogar, en el trabajo, o en la calle. Una ventaja de la encuesta personal es que al existir interacción entre el entrevistador y el entrevistado es posible aplicar formularios complejos, dado que el entrevistador tiene más control sobre el manejo de la información.

Hay que señalar también que en la encuesta personal es posible tomar datos por observación. Se puede tomar datos sobre el comportamiento y actividades de los entrevistados, el ambiente de los lugares de compra (decoración, mobiliarios), en que medio se transporta o cualquier otro aspecto que llame la atención.

Una ventaja de la encuesta personal es la posibilidad de emplear material auxiliar como portafolio del servicio y/o del producto, fotografías, pruebas de degustación y utilización de tarjetas, estas últimas cuando se requiere que el entrevistado responda preguntas de escalas complejas. La utilización de tarjetas y pruebas de degustación se explicarán con mayor detalle con las preguntas de escala, adicionalmente se darán algunos consejos para aminorar los costos y el tiempo de aplicación de los cuestionarios.

Entre las desventajas de la encuesta personal hay que señalar los costos y el tiempo de aplicación de la misma. Hay que destacar que se requiere del desplazamiento,

entrenamiento y pago de los entrevistadores. Otro aspecto negativo puede ser menor sinceridad por parte de los entrevistados al encontrarse cara a cara con el entrevistador. Hay que destacar que con el avance las nuevas tecnologías computacionales, hoy día es posible presentar datos en tiempo real, con la aplicación por ejemplo del CAPI (“computer assisted personal interview”) **encuesta personal asistida por computador**, en el que el entrevistador va normalmente acompañado de un ordenador portátil, la encuesta se presenta en formato electrónica y las respuestas se introducen directa en el computador. De esta forma, es posible integrar el trabajo de campo con la tabulación de los datos.

**2.2.2. Encuesta telefónica.** En la encuesta telefónica entrevistador y entrevistado se comunican a través del teléfono, este es un método en alza en los últimos años especialmente gracias al empleo de nuevas tecnologías como CATI (“computer assisted telephone interview”) **encuesta asistida por teléfono y computador**, en la que el propio sistema se encarga de seleccionar aleatoriamente las llamadas telefónicas y en las que el entrevistador introduce las respuestas de los individuos directamente en el ordenador, integrando también el trabajo de campo y procesamiento de la información.

Una de las ventajas de la encuesta telefónica frente a la personal, es el bajo costo y la rapidez con que se realizan. La encuesta telefónica resulta económica para muestras dispersas y es un método rápido.

Un inconveniente que se presenta con este tipo de encuesta es que el cuestionario debe ser reducido y las preguntas y escalas utilizadas deben ser sencillas. No es posible emplear material auxiliar como las tarjetas. No es conveniente utilizar preguntas abiertas, se sugieren hacer preguntas cerradas.

**2.2.3. Encuesta postal.** En este caso la encuesta se envía por correo para que el entrevistado la diligencie y la remita posteriormente a su destino. La encuesta debe ir acompañada de una carta de presentación, en la que el investigador de a conocer su trabajo, comente las instrucciones de diligenciamiento *del cuestionario y motive al destinatario para que devuelva el cuestionario.*

Se aconseja incluir un sobre franqueado junto con los formularios para facilitar la devolución de la encuesta, adicionalmente hoy día existen alternativas para este tipo de encuestas como el empleo del correo electrónico, fax y fax- modem.

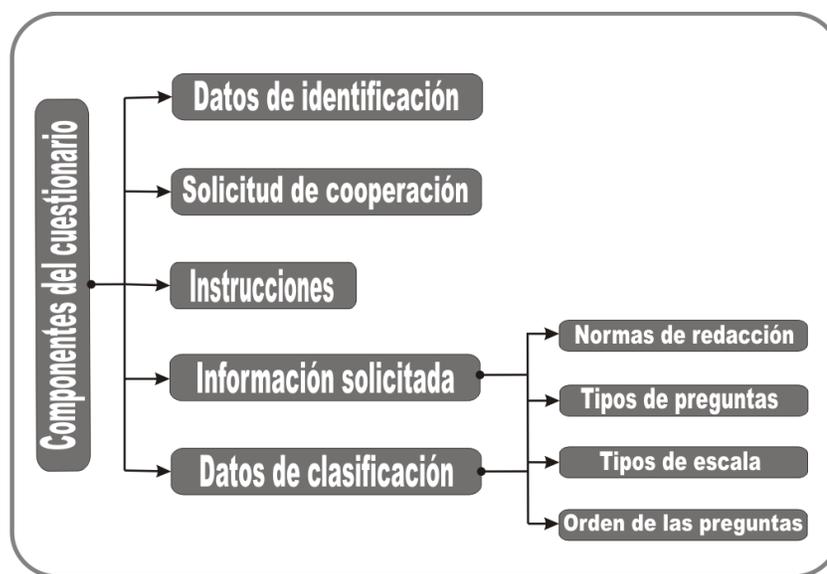
Un aspecto débil de la aplicación de la encuesta postal es el bajo índice de respuesta. En este tipo de encuesta el control de la muestra resulta ser difícil. No se tiene control de que la persona que diligencia el cuestionario sea realmente a quien va dirigido, otro aspecto a tener en cuenta es el nivel cultural del entrevistado. La encuesta postal no se puede realizar a cualquier persona independientemente de su nivel cultural. Es imprescindible asegurar como mínimo que sepan leer y escribir.

### 2.3. DISEÑO DE FORMULARIOS

Una encuesta es un plan formalizado para recolección de datos de encuestados. También se considera con un conjunto formal de preguntas para obtener información por parte de los entrevistados.

No existen pasos, principios o pautas que garanticen un cuestionario eficiente: el diseño de un cuestionario es una habilidad que el investigador aprende a través de la experiencia y no por medio de la lectura de una serie de pautas. A la hora de diseñar un cuestionario, el investigador debe tener en cuenta los objetivos que han motivado la aplicación de esta técnica. En el Figura 2.1 presenta los pasos a tener en cuenta en el diseño de un cuestionario y que a lo largo de éste capítulo se van a explicar.

Figura 2.1. Diseño de un cuestionario



- **Componentes del cuestionario**

El cuestionario tiene cinco secciones:

- Datos de identificación
- Solicitud de cooperación
- Instrucciones
- Información solicitada
- Datos de clasificación

- **Datos de identificación.** Es la primera sección del cuestionario se pide el nombre, la dirección y el número telefónico del encuestado. Usualmente, parte de toda esta

información se obtiene antes de la entrevista, de fuentes como directorios o contactos, aunque es importante considerar que estos datos por lo general el encuestado no los suministra, por tanto, se debe determinar si realmente la investigación requiere de ésta información. Pero datos adicionales como la hora y la fecha de la entrevista son importantes tomarlos, además del número o código del entrevistador.

- **Solicitud de cooperación.** Es un enunciado abierto para conseguir la colaboración del encuestado en relación con la entrevista. Este enunciado identifica primero al entrevistador y/o a la organización que hace la entrevista. En seguida se explica el propósito de estudio y se indica el tiempo requerido para completar la entrevista.

- **Las instrucciones.** Son comentarios que se le suministra al encuestado sobre cómo diligenciar las preguntas. Los comentarios aparecen directamente en el cuestionario, en una encuesta por correo o por computador vía Internet. En los casos de las entrevistas personales y por teléfono, las instrucciones se hacen en una hoja aparte con el título “Instrucciones para el entrevistado”, se explica el propósito del estudio, el plan de muestreo y otros aspectos del proceso de recolección de datos.

- **La información solicitada.** Constituye la parte principal del cuestionario. Se debe tener en cuenta los siguientes aspectos para el diseño del cuerpo del cuestionario:

- a. **La redacción de las preguntas.** La forma exacta como se redacte y se formule una pregunta tiene incidencia en los resultados de la investigación e inclusive puede invertir los resultados de la misma.

**Tenga en cuenta las siguientes sugerencias:**

- Emplee palabras sencillas
- Emplee palabras claras
- Evite preguntas que sugieran la respuesta
- Evite preguntas sesgadas
- Evite alternativas implícitas
- Evite supuestos implícitos
- Evite estimativos
- Evite preguntas de doble respuesta
- Considere el marco de referencia

- b. **La secuencia de las preguntas.** Una vez redactadas las preguntas, el siguiente paso es establecer la secuencia de las preguntas, es decir, el orden o flujo de preguntas en el cuestionario. La secuencia puede influir en la motivación del entrevistado. A continuación se dan algunas pautas para establecer un orden adecuado:

- c. **Utilice una pregunta introductoria simple e interesante:** La pregunta introductoria debe capturar inmediatamente el interés y la curiosidad del encuestado. La

pregunta introductoria por lo general no se relaciona con las necesidades de información del estudio; su único objetivo es lograr la cooperación del encuestado.

También se recomienda:

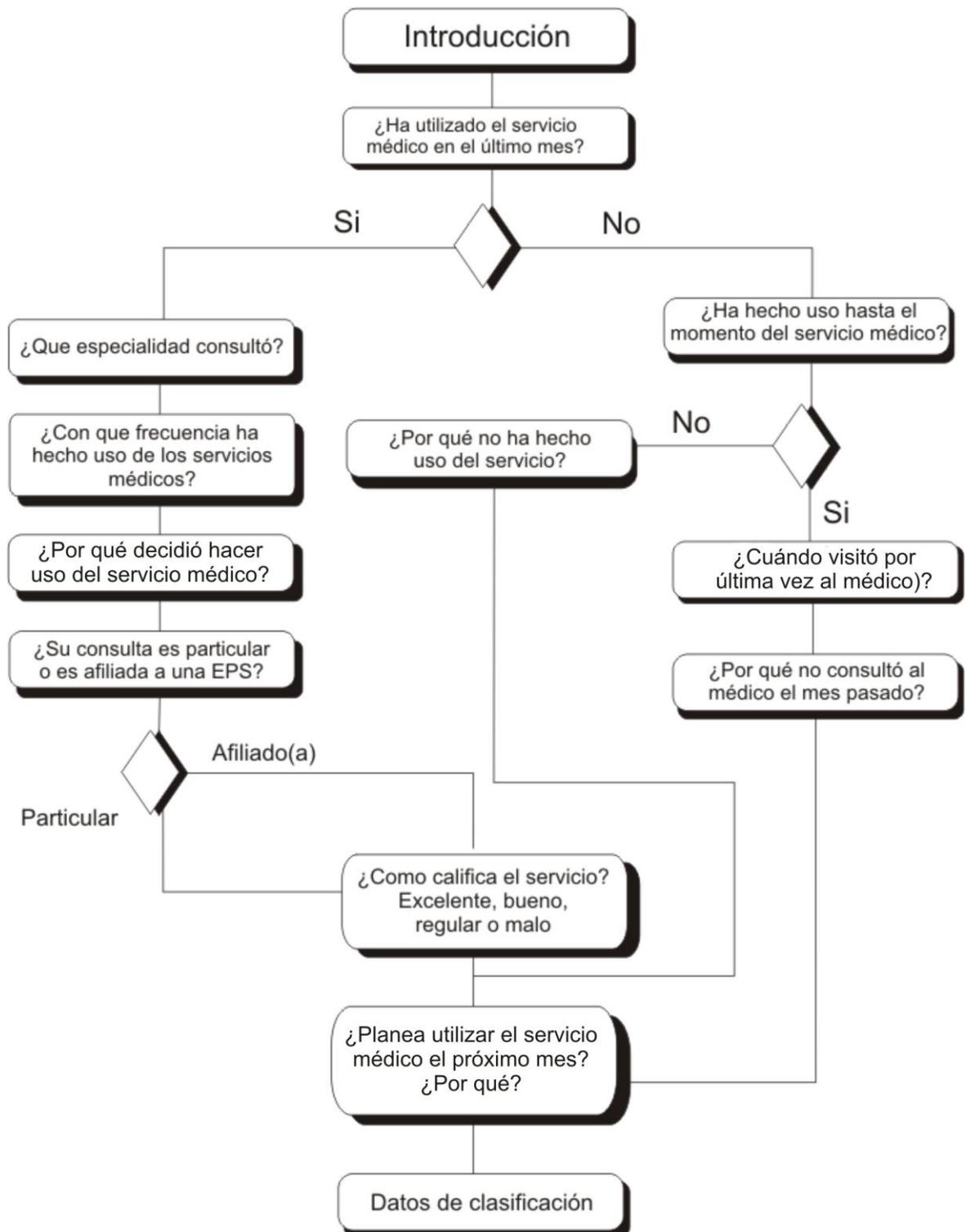
1. Formular primero las preguntas generales.
2. Colocar las preguntas no interesantes y difíciles al final de la secuencia.
3. Distribuir las preguntas en orden lógico.

Cuando no se tiene claro el flujo de información y cuando se formule preguntas a diferentes grupos dentro de la muestra, es aconsejable elaborar un diagrama de flujo de la secuencia de las preguntas. En la figura 2.2, se presenta un ejemplo. El diagrama de flujo puede ayudar a los investigadores a visualizar la estructura del cuestionario y a asegurarse que las preguntas tengan el orden adecuado.

**d. Las características físicas.** La apariencia física del cuestionario puede influir en el logro de la cooperación del encuestado, por ejemplo las encuestas por correo, la calidad del papel y la impresión determinan la primera reacción del encuestado hacia el cuestionario.

**e. Llevar a cabo una pre-prueba y revisión antes de aplicar el cuestionario definitivo.** Antes de que el cuestionario se aplique en el trabajo de campo, se recomienda ser sometido a una pre-prueba y revisión, aun cuando la encuesta se aplique por correo o por teléfono.

Figura 2.2. Ejemplo de flujo-grama para la elaboración de la encuesta.



## **2.4. TIPOS DE PREGUNTAS**

**2.4.1. Preguntas abiertas.** Es aquella que permite al entrevistado contestar libremente, una de las ventajas de este tipo de pregunta es que el encuestado responde con sus propias palabras, adicionalmente recoge información con un mínimo de indicaciones para el encuestado.

Ejemplo de pregunta abierta.

¿Qué le gustó más de la película?

Cuando el entrevistado no da una respuesta clara que satisfaga el entrevistador, entonces se debe realizar preguntas de seguimiento y profundización.

**2.4.2. Preguntas de profundización.** Se utilizan para obtener una respuesta amplia y completa a una pregunta abierta, deben ser utilizadas por los entrevistadores en forma rutinaria, como pregunta de seguimiento a preguntas abiertas, hasta que el encuestado no tenga más que añadir.

**Ejemplo de pregunta de profundización.**

¿Qué otro aspecto le gustó de la película?

¿Qué más me puede decir?

¿Qué otro punto puede añadir a su respuesta?

**2.4.2.1. Preguntas de clarificación.** Se utilizan para obtener una explicación más clara de una respuesta a una pregunta abierta.

Ejemplos de preguntas de clarificación.

¿Qué quiere decir con que a la película le faltó mas suspenso?

¿Explique lo que quiso decir con que la película no tenia una secuencia lógica?

**2.4.3. Preguntas cerradas.** La pregunta cerrada delimita las alternativas de respuesta en los cuestionarios y el entrevistador puede hacer que el encuestado seleccione una respuesta con las opciones suministradas.

**2.4.3.1. Preguntas de respuestas múltiples.** Una pregunta de selección múltiple requiere que el encuestado seleccione una respuesta de una lista suministrada en la misma pregunta. Al encuestado se le solicita que seleccione una o más de las alternativas presentadas.

- **Preguntas dicótomas.** Es aquella que ofrece dos alternativas de respuesta, mutuamente excluyentes.

Ejemplo:



SEXO

1. Hombre ( )      2. Mujer ( )

El diagrama muestra un recuadro gris con esquinas redondeadas. En el centro superior, el título "SEXO" está escrito en letras negras y mayúsculas. Debajo del título, hay dos íconos: a la izquierda, un ícono negro de un hombre; a la derecha, un ícono negro de una mujer. Debajo de cada ícono, hay una opción de respuesta numerada: "1. Hombre ( )" y "2. Mujer ( )".

Asegúrese que en realidad la pregunta solo tenga dos respuestas. A menudo se colocan las opciones “NO SABE NO CONSTESTA (NS/NC)” o “AMBOS” también pueden ser respuestas a la pregunta realizada.

- **Preguntas de respuesta múltiple.** Este tipo de pregunta ofrece distintas alternativas de respuesta, pero no son de carácter excluyente: el encuestado puede seleccionar varias opciones.

**Ejemplo.**

## Instrucciones

¿Del siguiente listado que le suministro conteste cual de las siguientes marcas de carnes frías son de su preferencia?. Marque en la opción uno (1) la marca de su mayor preferencia y maque en la opción dos (2) la marca de su preferencia en segundo lugar.



	Primer opción	Segunda opción
1. Zenú	<input type="checkbox"/>	<input type="checkbox"/>
2. Rica rondo	<input type="checkbox"/>	<input type="checkbox"/>
3. Americana	<input type="checkbox"/>	<input type="checkbox"/>
4. Berna	<input type="checkbox"/>	<input type="checkbox"/>
5. Pimpollo	<input type="checkbox"/>	<input type="checkbox"/>

En este tipo de preguntas se puede incluir una pregunta abierta como **¿Otras cual?**, esto con el fin de obtener más información, o cuando no se esta seguro que el listado de opciones incluya todas las categorías posibles.

**2.4.3.1.1. Preguntas de recuerdo respuesta espontánea.** Cuando la investigación va dirigida a medir el grado de reconocimiento de los consumidores frente a una serie de marcas o estímulos publicitarios.

Ejemplo.

**Pregunta**

Del listado de marcas de jugos naturales que le suministro, indique ¿Cuál o cuáles conoce, aun que no haya probado el producto?



	Si conoce	No conoce
1. Tutti - frutty	<input type="checkbox"/>	<input type="checkbox"/>
2. Postobón	<input type="checkbox"/>	<input type="checkbox"/>
3. Alpina	<input type="checkbox"/>	<input type="checkbox"/>
4. Colanta	<input type="checkbox"/>	<input type="checkbox"/>

**2.4.3.2. Preguntas de recuerdo sugerido.** Es una pregunta cerrada en la que se les nombra a los entrevistados una serie de marcas o estímulos publicitarios y él tiene que señalar si los conoce o no.

**2.4.3.3. Preguntas filtro.** Permiten segmentar la muestra con la finalidad de realizar posteriormente preguntas diferentes a los segmentos.

**Tabla 2.1. Ejemplo pregunta filtro.**

## Pregunta

La División de Mercadeo de nuestra organización anualmente programa cuatro seminarios orientados a la atención al cliente, en diferentes fechas y horarios, esto con el fin de que el funcionario haga la capacitación en las fechas y horarios que mejor se acomoden a sus necesidades. En su caso, ¿Durante el último semestre de este año, ha recibido algún tipo de seminario de formación impartido por la división de Mercadeo?

1. Si

2. No

Si seleccionó la primera opción continúe con el cuestionario, de lo contrario pase a la pregunta 10.

Diligencie la siguiente tabla. Señale los temas vistos, fechas y horarios.

Temas	Fechas aa/mm/dd	Intensidad horaria

### 2.5. DATOS DE CLASIFICACIÓN DEL ENTREVISTADO.

**Se relaciona con las características del encuestado.** Estos datos los suministra directamente el encuestado en el caso de una encuesta por correo. En una encuesta de entrevista personal o telefónica, el entrevistador recolecta los datos del encuestado con anticipación. Generalmente estos datos se recolectan al final de la entrevista, ya que algunos encuestados pueden mostrarse renuentes a revelar datos personales como la edad, ingresos, estado civil, se sugiere que se realice al final hasta que se haya establecido una armonía con la entrevista. *Sin embargo, se sugiere tomar estos datos al inicio de la entrevista, para determinar si la persona califica como parte de la muestra del plan de muestreo.*

Ejemplo de datos de clasificación.

Sexo    1. Hombre        2. Mujer   

Edad en años cumplidos    \_\_\_\_\_    NS

Estado civil.

- |                       |                          |                 |                          |
|-----------------------|--------------------------|-----------------|--------------------------|
| 1. Soltero (a)        | <input type="checkbox"/> | 2. Casado (a)   | <input type="checkbox"/> |
| 3. Viviendo en pareja | <input type="checkbox"/> | 4. Separado (a) | <input type="checkbox"/> |
| 5. Divorciado (a)     | <input type="checkbox"/> | 6. Viudo (a)    | <input type="checkbox"/> |
| 7. NC                 | <input type="checkbox"/> |                 |                          |

¿Cuál es la relación con la cabeza familia?

- |                              |                          |
|------------------------------|--------------------------|
| 1. Cabeza de familia         | <input type="checkbox"/> |
| 2. Esposo (a), compañero (a) | <input type="checkbox"/> |
| 3. Hijo (a)                  | <input type="checkbox"/> |
| 4. Padre, madre.             | <input type="checkbox"/> |
| 5. Suegro (a)                | <input type="checkbox"/> |
| 6. Hermano (a).              | <input type="checkbox"/> |
| 7. Cuñado (a)                | <input type="checkbox"/> |
| 8. Otro                      | <input type="checkbox"/> |
| 9. NS/NC                     | <input type="checkbox"/> |

Número de miembros del hogar.

- |                |                          |           |                          |          |                          |           |                          |
|----------------|--------------------------|-----------|--------------------------|----------|--------------------------|-----------|--------------------------|
| 1. Uno         | <input type="checkbox"/> | 2. Dos    | <input type="checkbox"/> | 3. Tres  | <input type="checkbox"/> | 4. Cuatro | <input type="checkbox"/> |
| 5. Cinco       | <input type="checkbox"/> | 6. Seis   | <input type="checkbox"/> | 7. Siete | <input type="checkbox"/> | 8. Ocho   | <input type="checkbox"/> |
| 9. Nueve o más | <input type="checkbox"/> | 10. NS/NC | <input type="checkbox"/> |          |                          |           |                          |

Nivel de escolaridad

- |  |                          |
|--|--------------------------|
| 1. Menos de estudios primarios, no sabe leer         | <input type="checkbox"/> |
| 2. Menos de estudios primarios sabe leer             | <input type="checkbox"/> |
| 3. Estudios primarios completos, certificado escolar | <input type="checkbox"/> |
| 4. Bachiller   | <input type="checkbox"/> |
| 5. Universitarios o Técnicos de Grado Superior       | <input type="checkbox"/> |
| 6. NC  | <input type="checkbox"/> |

¿A qué religión pertenece usted?

- |             |                          |              |
|-------------|--------------------------|--------------|
| 1. Católica | <input type="checkbox"/> |              |
| 2. Otra     | <input type="checkbox"/> | ¿Cuál? _____ |
| 3. Ninguna  | <input type="checkbox"/> |              |
| 4. NS/NC    | <input type="checkbox"/> |              |

Según usted, ¿a qué clase social pertenece?

- |               |                          |               |                          |                  |                          |
|---------------|--------------------------|---------------|--------------------------|------------------|--------------------------|
| 1. Alta       | <input type="checkbox"/> | 2. Media alta | <input type="checkbox"/> | 3. Media - media | <input type="checkbox"/> |
| 4. Media baja | <input type="checkbox"/> | 5. Baja       | <input type="checkbox"/> | 6. NS            | <input type="checkbox"/> |
| 7. NC         | <input type="checkbox"/> |               |                          |                  |                          |

¿Marque el intervalo de sus ingresos mensuales que por todos los conceptos entren en su hogar?

- |                         |                          |
|-------------------------|--------------------------|
| 1. Hasta 45.000         | <input type="checkbox"/> |
| 2. De 45.000 a 75.000   | <input type="checkbox"/> |
| 3. De 75.001 a 100.000  | <input type="checkbox"/> |
| 4. De 100.001 a 150.000 | <input type="checkbox"/> |
| 5. De 150.001 a 200.000 | <input type="checkbox"/> |
| 6. De 200.001 a 275.000 | <input type="checkbox"/> |
| 7. De 275.001 a 350.000 | <input type="checkbox"/> |
| 8. De 350.001 a 450.000 | <input type="checkbox"/> |
| 9. Más de 450.000       | <input type="checkbox"/> |

## 2.6. TIPOS DE ESCALA.<sup>5</sup>

Existen cuatro escalas de medición primaria: nominal, ordinal, de intervalo y de relación. La figura 2.3 se resume las principales características de cada una de ellas, y en la tabla 2.3 se especifica las estadísticas permitidas.

Figura 2.3. Escalas primarias de medición.

Escala				
Nominal	Números asignados a los atletas	 058	 001	 859
Ordinal	Orden de clasificación de los atletas	 cuarto lugar	 Quinto lugar	 Primer lugar
De Intervalo	Calificación de desempeño con base en una escala de 0 a 100.	74.5	62.0	92.0
De Relación	Tiempo estimado hasta la meta, en segundos.	10.2	11.15	9.5

**2.6.1. Escala nominal.** Corresponde a una etiqueta, los números sirven sólo como etiqueta para identificar y clasificar los objetos. Por ejemplo el número que se le asigna a los deportistas en una competencia. Cada número se asigna sólo a un objeto y cada objeto tiene un solo número asignado. En la investigación de mercados, las escalas nominales se utilizan para identificar entrevistados, marcas, atributos y otros objetos.

Los números en una escala nominal no refleja la cantidad de la característica que poseen los objetos, por ejemplo un número alto asignado a un deportista de ninguna manera significa que este deportista tenga mayores cualidades que los demás. Las

<sup>5</sup> Este tema es una adaptación de los capítulos 8 y 9 de Malhotra. Narres K. "Investigación de Mercados un Enfoque Práctico". 2ª. Ed. Prentice Hall.1997. Algunos textos se tomaron textualmente, y los ejemplos se adaptaron conservando la estructura de las preguntas.

operaciones permitidas con los números en una escala nominal es el conteo. Estos incluyen porcentajes, estadísticos como: moda, chi – cuadrado y pruebas binomiales. *No tiene sentido calcular el promedio de sexo a los entrevistados.*

**2.6.2. Escalas ordinales.** Es una escala de clasificación en la que los números se asignan a los objetos para indicar el grado relativo en el que los objetos poseen cierta característica. Una escala ordinal indica la posición relativa, pero no la magnitud de la diferencias entre los objetos. En la investigación de mercados permite un orden en las preferencias. Las operaciones permitidas con esta escala son: percentiles, y estadísticos como la mediana, la correlación entre orden y clasificación

Ejemplo.

### Instrucciones

Quando compra ropa, ¿cuáles de las siguientes características considera más importantes?. Marque en primer lugar la que considere más importante, y en último lugar (séptimo) aquella menos importante.

No necesariamente tiene que seleccionar todas las opciones.

---

1<sup>ra</sup>
2<sup>da</sup>
3<sup>ra</sup>
4<sup>ta</sup>
5<sup>to</sup>
6<sup>ta</sup>
7<sup>ma</sup>

---

*Calidad	<input type="checkbox"/>						
*Diseños	<input type="checkbox"/>						
*Precio	<input type="checkbox"/>						
*Colores	<input type="checkbox"/>						
*Lugar de compra	<input type="checkbox"/>						
*Opción de cambio	<input type="checkbox"/>						
*Otros	<input type="checkbox"/>						

¿Cuál? \_\_\_\_\_

**2.6.3. Escalas de intervalo.** También denominadas ficticias, son escalas numéricas en las que el investigador establece a priori un rango de valoración (del 1 al 5, del 1 al 20) y el entrevistado tiene que puntuar, de forma individualizada, una serie de atributos o ítems utilizando dicho rango. Se permite otorgar la misma puntualización a varios de estos ítems.

Las técnicas estadísticas que pueden utilizarse en los datos de la escala de intervalos incluyen todas aquellas que pueden aplicarse a los datos nominales y ordinales, además de la media aritmética, la desviación estándar. Sin embargo ciertas estadísticas como la

media geométrica, la media armónica y el coeficiente de variación no son significativas en los datos de escala de intervalo.

EJEMPLO<sup>6</sup>

### Instrucciones

Utilizando una escala de 5 puntos en la que el 5 significa “mucho” y el 1 significa “nada”, ¿En que medida puede aplicarla a las condiciones de su trabajo actual?

Su trabajo es:

---

	Mucho				Nada	NS	NR
	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>8</b>	<b>9</b>
*Seguro y estable	<input type="checkbox"/>						
*Bien remunerado	<input type="checkbox"/>						
*flexible en los horarios de trabajo	<input type="checkbox"/>						
*Interesante y permite adquirir conocimientos	<input type="checkbox"/>						
*Útil para la sociedad	<input type="checkbox"/>						

**2.6.4. Escalas de relación.** Posee todas las propiedades de las escalas nominales, ordinales y de intervalos y, además, un punto cero absoluto. Con las escalas de relación se puede clasificar los objetos, ordenarlos y comparar los intervalos o diferencias, también se pueden calcular las relaciones de los valores de la escala. Algunos ejemplos de escalas de relación incluyen estatura, peso, edad y dinero. En mercadotecnia, las ventas, los costos, la participación en el mercado y el número de clientes son variables, que se miden con base en una escala de relación.

<sup>6</sup> VISAUTA VINACUA. Bienvenido. Análisis Estadístico con SPSS para Windows. Madrid.: Mc Graw Hill. 1997. Página 288.

Ejemplo.

### Instrucciones

Indique la cantidad en pesos que gasto el fin de semana en los almacenes y supermercados de la ciudad de Popayán.



Escala nominal		Escala de relación
No	Almacén	Pesos que se gasto el fin de semana
1.	Exito centro	\$ 200.000
2.	Exito panamericana.	\$ 0
3.	Carrefour	\$ 500.000
4.	Olimpica	\$ 0
5.	Carulla	\$ 50.000

El entrevistado gasto \$200.000 en el almacén Éxito centro y sólo \$ 50.000 en el supermercado Carulla, es decir, que en el almacén Éxito gastó cuatro veces más que en el supermercado Carulla, el punto cero es fijo, puesto que 0 significa que el entrevistado no gasto nada en este lugar.

Tabla 2.3. Escalas de medición.

Escala	Características básicas	Ejemplos	Estadísticas permitidas	
			Descriptivas	Deductivas
Nominal	Identifica la pertenencia a una categoría u otra	Clasificación por sexo, tipos de establecimiento, número de marcas	Frecuencia moda	Chi-cuadrado prueba binomial
Ordinal	Se establece un orden de preferencia.	Clasificación por preferencias, clase social, posición en el mercado.	Percentiles Mediana, Moda, frecuencias.	Correlación, orden clasificación, ANOVA.
De intervalo	Se valora una característica o atributo, asignando un valor dentro de un rango previamente establecido.	Actitudes, opiniones, números en un índice.	Frecuencia, Moda, mediana, media, desviación, varianza, rango, recorrido intercuartílico.	Correlaciones, pruebas t, ANOVA, regresión, análisis de factores.
De relación	Valor número real	Edad, ingresos, costos, ventas, participación en el mercado	Media, varianza, desviación típica, media geométrica, media armónica	Coefficiente de variación

Teniendo en cuenta estos cuatro tipos de escala, esta clasificación resulta imprescindible para determinar la forma adecuada del análisis de la información, recuerde que el

objetivo de este texto es hacer uso del software SPSS en la versión 15, para el procesamiento de la información. En tabla 2.4 se presenta la clasificación de las variables No Métricas y las Métricas según las escalas de medición.

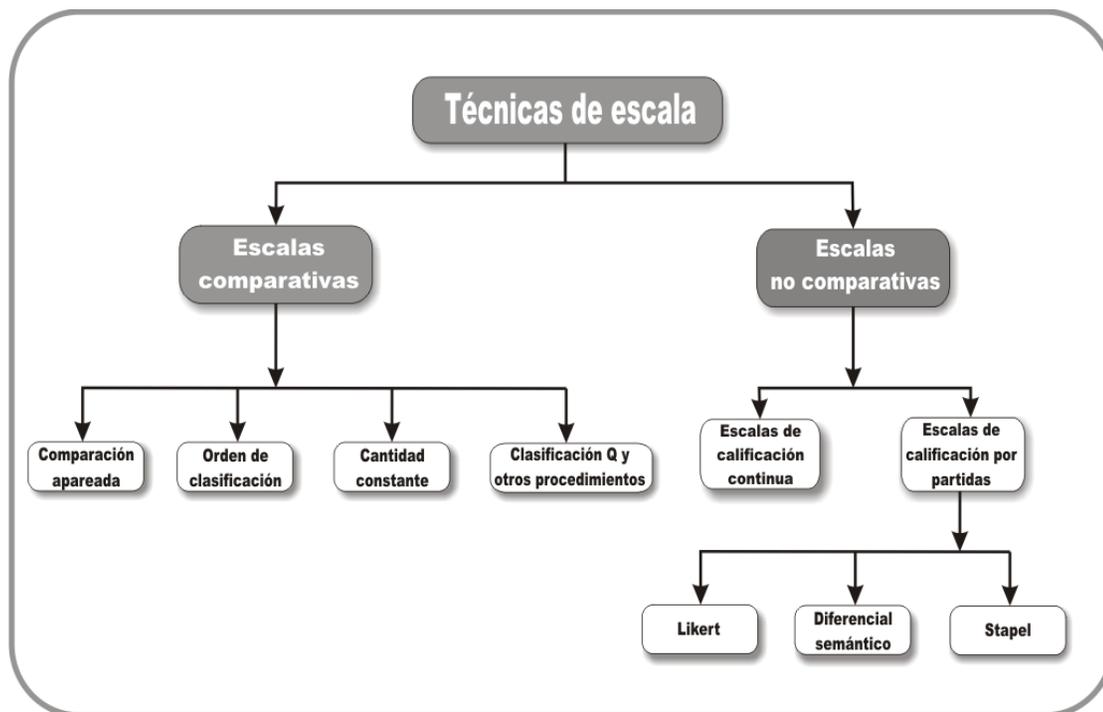
Tabla 2.4. Clasificación de escalas

NO MÉTRICAS	NOMINALES	Se identifica la pertenencia a una categoría u otra.
	ORDINALES	Se establece un orden de preferencia
MÉTRICAS	INTERVALO	Se valora una característica o atributo, asignando un valor dentro de un rango previamente establecido.
	DE RELACIÓN	Valor numérico real.

## 2.7. TÉCNICAS DE ESCALAS.

Las técnicas de escala que se utilizan con mayor frecuencia en la investigación de mercados se clasifican como escalas comparativas y no comparativas. Grafica 2.3.

Figura 2.3. Técnicas de escala.



**2.7.1. Técnicas de escala comparativas.** Comprenden la comparación directa de los objetos de estímulo. Por ejemplo se le puede preguntar al entrevistado si prefiere los carros o las motos. Los datos de esta escala deben interpretarse en términos relativos y

tienen solo propiedades ordinales o de orden por clasificación. Las escalas comparativas se conocen también como **escalas no métricas**.

**2.7.1.1. Escala de comparación apareada.** Se presenta al entrevistado dos objetos y se le pide que seleccione uno de acuerdo con cierto criterio. Los datos obtenidos son de naturaleza ordinal. Las escalas de comparación apareadas se utilizan con frecuencia cuando los objetos de estímulo son productos físicos.

**Figura 2.4. Ejemplo de pregunta de escala de comparación apareada.**



La figura 2.4, muestra los datos de una comparación apareada obtenida para evaluar las preferencias en las marcas de cerveza de los consumidores. El entrevistado realizó 13 comparaciones para evaluar cinco marcas. Para realizar todas las posibles comparaciones se puede utilizar la siguiente formula:

$$\text{Número de comparaciones} = \left[ \frac{n(n-1)}{2} \right]$$

**2.7.1.2. Escala por orden de clasificación.** Consiste en presentar a los entrevistados varios objetos en forma simultánea y se les pide los ordenen o clasifiquen de acuerdo a cierto criterio.

Figura 2.5. Ejemplo de escala por orden de calificación.

**Instrucciones**

Clasifique las marcas de champú en orden de preferencia. Comience por elegir aquella marca que más le agrada y asígnele el número 1. Después, busque la segunda marca que prefiere y asígnele el número 2. Continúe este procedimiento hasta que haya clasificado todas las marcas de champú en orden de preferencia. A la marca que menos prefiere debe asignarle el número 6.

NO DEBE ASIGNAR A DOS MARCAS EL MISMO NÚMERO DE CLASIFICACIÓN.

El criterio de preferencia depende únicamente de usted. No hay respuestas correctas ni incorrectas. Trate de ser consistente.



<u>Marca</u>	<u>Orden de clasificación</u>
1. Pantene	_____
2. Head & shoulders	_____
3. Johnson's	_____
4. Alert	_____
5. Dove	_____
6. Sedal	_____

**2.7.1.3. Escala de suma de constantes.** En esta técnica los entrevistados distribuyen una cantidad constante de unidades, como puntos o pesos, entre un conjunto de estímulos con respecto a cierto criterio.

Figura 2.6. Ejemplo de pregunta de escala de suma de constantes

**Instrucciones**

A continuación, se presentan seis características para la ubicación de nuestro centro comercial. Favor de distribuir 100 puntos entre las características, de modo que su distribución refleje la importancia relativa que da a cada una de ellas. Cuanto más puntos reciba una característica, es más importante. Si una característica no tiene ninguna importancia, asígnele cero puntos. Si una característica es dos veces más importante que alguna otra, debe recibir el doble de puntos.

RESPUESTA PROMEDIO DE TRES GRUPOS DE ENTREVISTADOS

Característica	Grupo I	Grupo II	Grupo III
1. Ubicación de los clientes	18	0	10
2. Servicios públicos	22	10	12
3. Vías de acceso	11	25	5
4. Servicio de transporte público	30	30	16
5. Seguridad de la zona	35	35	45
6. Actitud de los vecinos	0	0	12
<b>Suma</b>	<b>100</b>	<b>100</b>	<b>100</b>

**2.7.1.4. Escala de calificación Q y otros procedimientos.** Esta técnica se desarrolló para determinar con rapidez entre una cantidad relativa elevada de objetos.

**Figura 2.7. Ejemplo de clasificación Q<sup>7</sup>.**

Instrucciones						
Todos los días se escucha comentarios entre la gente acerca del porqué hay gente sin trabajo. ¿Hasta qué punto está usted de acuerdo o en desacuerdo con cada una de las siguientes razones con respecto al desempleo?						
	Muy de acuerdo	De acuerdo	Indiferente	En desacuerdo	Muy en desacuerdo	NS/NC
	5	4	3	2	1	9
* La crisis económica	( )	( )	( )	( )	( )	( )
* La política de empleo por parte del gobierno	( )	( )	( )	( )	( )	( )
* La mala gestión de los empresarios	( )	( )	( )	( )	( )	( )
* La comodidad de la gente, que sólo quiere buenos trabajos	( )	( )	( )	( )	( )	( )
* La falta de preparación del trabajador	( )	( )	( )	( )	( )	( )
* Las pocas ganas de trabajar de la gente	( )	( )	( )	( )	( )	( )
* El no saber buscar trabajo	( )	( )	( )	( )	( )	( )
* Que hay mucho pluriempleado	( )	( )	( )	( )	( )	( )
* El trabajo que hay no se reparte bien socialmente	( )	( )	( )	( )	( )	( )

Esta técnica emplea un procedimiento de orden de clasificación en el que los objetos se clasifican en columnas con base en la similitud respecto a cierto criterio. En la figura 2.7 se presenta un ejemplo donde se proporciona a los entrevistados algunas afirmaciones de actitud y se les pide que las coloquen en cinco (5) columnas, que van desde la “afirmación Muy de acuerdo” hasta a la afirmación “Muy en desacuerdo”. En esta técnica se recomienda que el número de objetos no debe ser menor de 60 ni mayor de 140, para efectos del ejemplo de colocó únicamente 9 objetos para analizar.

**2.7.1.5. Escala de estimación de la magnitud.** En esta técnica, se asigna números a los objetos de modo que las relaciones entre los números asignados reflejen las relaciones con base en el criterio que se especifica.

<sup>7</sup> Ibid. P. 287.

**Figura 2.8** Ejemplo de escala de estimación de la magnitud<sup>8</sup>.

Instrucciones	Por favor, indique su nivel de acuerdo con las frases que le suministro a continuación.					
	Muy de acuerdo				Muy en desacuerdo	NS/NC
	20	40	60	80	100	999
* La vida sólo tiene sentido cuando una persona se dedica plenamente a una causa ideal.	20	40	60	80	100	999
* De todas las religiones del mundo, probablemente sólo una es la verdadera.	20	40	60	80	100	999
* Un grupo en el que se toleran demasiadas diferencias de opinión entre sus miembros no puede durar mucho tiempo.	20	40	60	80	100	999
* En un mundo complicado como el actual lo mejor es atenerse a lo que nos digan las autoridades y expertos en quienes podamos confiar.	20	40	60	80	100	999
* Sólo mirando hacia el pasado encontramos solución a nuestros problemas actuales.	20	40	60	80	100	999
* El futuro es tan inseguro, que lo mejor que se puede hacer es vivir al día.	20	40	60	80	100	999
* En un sistema democrático como el nuestro los ciudadanos influyen realmente en las decisiones que toma el gobierno.	20	40	60	80	100	999
* Todo cambia tan rápidamente en estos tiempos que uno difícilmente puede ya distinguir entre lo que está bien y mal.	20	40	60	80	100	999

En la figura 2.8 se le pide a los entrevistados que indiquen si están de acuerdo con cada una de las afirmaciones que miden la actitud hacia los problemas económicos y sociales actuales, asignando a cada afirmación un número entre 20 y 100 a fin de indicar la intensidad de su acuerdo o desacuerdo.

**2.7.2. Escalas no comparativas.** Conocidas también como escalas métricas o monádicas, cada objeto se evalúa en forma independiente de los otros objetos del conjunto de estímulos. Los datos se encuentra en una escala de intervalos o relación.

**2.7.2.1. Escala de clasificación continua.** También conocida como escala de calificación gráfica, los entrevistados califican los objetos escribiendo una marca en la posición apropiada sobre una línea que va de un extremo de la variable de criterio al otro. De modo que los participantes no están limitados a seleccionar marcas que el investigador establece con anterioridad. La forma de una escala continua puede variar, la línea puede ser vertical u horizontal, los puntos de la escala pueden proporcionarse en

<sup>8</sup> Ibid. p.284.

forma de números o descripciones breves, y si se proporcionan, los puntos de la escala pueden ser pocos o muchos.

**Figura 2.8.1. Ejemplo de escala de clasificación continua.**

### Instrucciones

Por favor, marque en la siguiente escala el grado que considera están las relaciones interpersonales en su trabajo.

	Muy Buenas	Buenas	Ni buenas ni malas	Malas	Muy malas	NS/NC
	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>9</b>
*Entre directivos y trabajadores	<input type="radio"/>					
*Entre compañeros de trabajo	<input type="radio"/>					
*Entre trabajadores y público	<input type="radio"/>					

**2.7.2.2. Escala de clasificación de partidas.** Esta técnica consiste en darle a los entrevistados una escala que tiene una descripción breve o un número relacionado con cada categoría. Las categorías se ordenan en términos de posición en la escala; y se pide a los participantes seleccionen la categoría específica que describa mejor el objeto a calificar.

- **Escala de Likert.** Lleva el nombre de su creador, Rensis Likert, en esta técnica se pide a los entrevistados que indique un grado de acuerdo o desacuerdo con cada una de la serie de afirmaciones respecto a los objetos de estímulo. Regularmente cada partida de la escala tiene cinco categorías de respuesta, que van de “Por completo en desacuerdo” o “Por completo de acuerdo”. Para realizar el análisis, a cada afirmación se le asigna una calificación numérica, que puede ir desde -2 a +2 o de 1 a 5.

**Figura 2.9.** Ejemplo de escala de Likert.

### Instrucciones

A continuación, presentamos algunas opiniones sobre el almacén Exito panamericana. Favor de indicar si está de acuerdo o en desacuerdo con cada una de las características, marcando con X su elección en la escala siguiente:

---

	Por completo en desacuerdo	En desacuerdo	No estoy de acuerdo ni en desacuerdo	De acuerdo	Por completo de acuerdo	NS/NC
* Vende mercancías de alta calidad	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	9 <input type="checkbox"/>
* La atención al cliente es eficiente	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	9 <input type="checkbox"/>
* No ofrece variedad en marcas de los productos.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	9 <input type="checkbox"/>
* Tiene buenas políticas de crédito	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	9 <input type="checkbox"/>
* Tiene precios moderados en los electrodomésticos	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	9 <input type="checkbox"/>
* Vende una extensa variedad de mercancías	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	9 <input type="checkbox"/>
* El ambiente de la sección de restaurante es agradable	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	9 <input type="checkbox"/>
* Las políticas de crédito son terribles	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	9 <input type="checkbox"/>
* Los precios de los productos alimenticios son muy elevados	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>	9 <input type="checkbox"/>

Es importante emplear un procedimiento de calificación sólido, de modo que la calificación alta (baja) refleje una respuesta favorable (desfavorable) en forma consistente. Lo anterior requiere que las categorías que los entrevistados asignan a las afirmaciones negativas se califiquen revirtiendo la escala. En éste ejemplo la opción 4 “Tiene buenas políticas de crédito”, tiene una pregunta de seguimiento, que es la opción 8 “Las políticas de crédito son terribles”, es decir que el entrevistado debe ser consecuente con sus respuestas.

- **Escala de diferencial semántico.** Es una escala de calificación con siete puntos, cuyos puntos finales se relacionan con niveles bipolares que tienen un significado semántico. En una aplicación típica, los entrevistados califican los objetos con base en varias escalas de calificación repartidas en los siete puntos, limitadas en cada extremo con uno de dos adjetivos bipolares, como “Bueno” y “Malo”.

**Figura 2.10. Ejemplo de Escala de diferencial semántico.**

### Instrucciones

Favor de marcar con una (X) el espacio que indique mejor con cuánta precisión un adjetivo u otro describe el concepto de tiene sobre la selección Colombiana de fútbol.

Favor de asegurarse de marcar cada escala; no omita ninguna.

La selección Colombia es:



	1	2	3	4	5	6	7	
Organizada	<input type="checkbox"/>	Desorganizada						
Estrategica	<input type="checkbox"/>	Confiada						
Trabaja en equipo	<input type="checkbox"/>	Individualista						
Domina en el campo	<input type="checkbox"/>	Sumisos						
Acreditada	<input type="checkbox"/>	Desacreditada						

Las partidas individuales de una escala de diferencial semántico pueden calificarse ya sea de -3 a +3 o de 1 a 7. Los datos que resultan se caracterizan a través de un análisis de perfil, los valores de las medias o medianas, en cada escala de calificación, se calculan y comparan mediante un análisis de trazo o estadístico. Esto ayuda a determinar las diferencias y similitudes generales entre los objetos.

Esta escala se utiliza con frecuencia en la comparación de marcas, productos e imágenes de empresas. También se emplea para desarrollar estrategias de publicidad y promoción, así como los estudios para el desarrollo de nuevos productos.

- Escala de Stapel.** Esta escala lleva el nombre de su creador, Juan Stapel, es una escala de calificación unipolar con 10 categorías de -5 a +5, sin un punto neutral (cero). Esta escala por lo general se presenta en sentido vertical. Se le pide a los entrevistados que indiquen, mediante la selección de una categoría de respuesta numérica apropiada, con cuánta exactitud o inexactitud cada término describe el objeto. Cuanto más alto sean los números, el término describe el objeto con mayor exactitud.

**Figura 2.11.** Ejemplo de Escala de Stapel.

**Instrucciones**

Favor de evaluar con cuánta precisión cada palabra o frase describe el almacén CARREFOUR. Elija un número positivo para las frases que crea describen el almacén con exactitud. Cuanto más exacta considere sea la frase para describir el almacén, más alto debe ser el número positivo que seleccione. Debe elegir un número negativo para las frases que crea no la describen con precisión. Cuanto menos precisa considere que es la frase para describir el almacén, más bajo deberá ser el número negativo que elija. Puede seleccionar cualquier número, de +5 para las frases que considere muy exactas y -5 para las frases que considere muy inexactas.

+5 <input type="text"/>	+5 <input type="text"/>	+5 <input type="text"/>
+4 <input type="text"/>	+4 <input type="text"/>	+4 <input type="text"/>
+3 <input type="text"/>	+3 <input type="text"/>	+3 <input type="text"/>
+2 <input type="text"/>	+2 <input type="text"/>	+2 <input type="text"/>
+1 <input type="text"/>	+1 <input type="text"/>	+1 <input type="text"/>
Variedad en los productos ofrecidos	Precios elevados en los alimentos	Personal capacitado para atender al público
-1 <input type="text"/>	-1 <input type="text"/>	-1 <input type="text"/>
-2 <input type="text"/>	-2 <input type="text"/>	-2 <input type="text"/>
-3 <input type="text"/>	-3 <input type="text"/>	-3 <input type="text"/>
-4 <input type="text"/>	-4 <input type="text"/>	-4 <input type="text"/>
-5 <input type="text"/>	-5 <input type="text"/>	-5 <input type="text"/>

## 2.8. FORMA FÍSICA DE LA ESCALA

Las escalas pueden presentarse en diferentes formas. Pueden presentarse en sentido vertical u horizontal. Las categorías pueden expresarse por medio de círculos, cuadros, líneas discretas o unidades y pueden tener o no números asignados. Hasta el momento se han presentado gran variedad de este tipo de escalas.

Hay dos configuraciones de escala de clasificación únicas, que se emplean en la investigación de mercados, son la escala del termómetro y la escala de caras sonrientes.

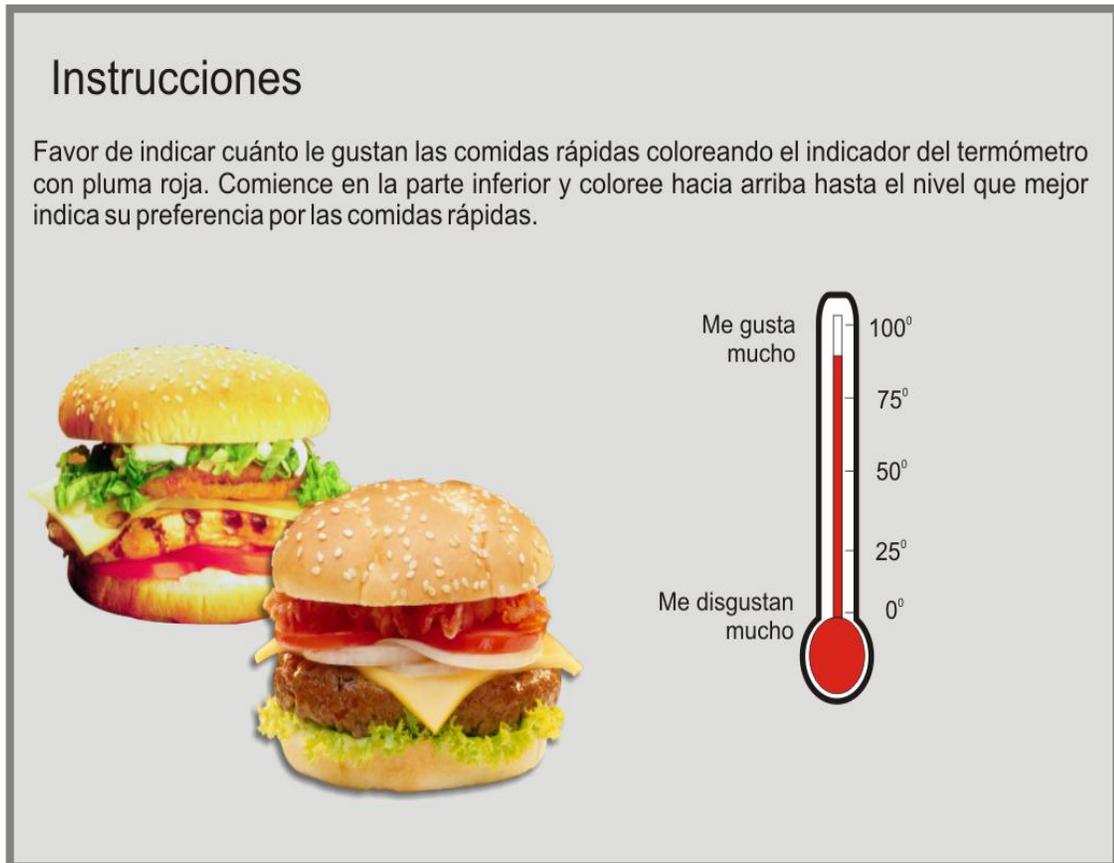
**2.8.1. Escala de caras sonrientes.** Son apropiadas para entrevistar a los niños, de modo que las caras más felices indican evaluaciones más favorables, en la Figura 2.12 se presenta un ejemplo dirigido a las niñas.

**Figura 2.12. Ejemplo de Escala de Caras Sonrientes, especial para niños.**



**2.8.2. Escala del termómetro.** Cuanta más alta sea la temperatura, más favorable es la evaluación.

**Figura 2.13.** Ejemplo de escala del termómetro



## 2.9. ESCALA DE INTENCIÓN DE COMPRA.

Como las ventas son usualmente la medida final del éxito de un producto, este tipo de pregunta es el que más se acerca a la evaluación del potencial de ventas en un contexto de encuestas. Los clientes necesitan que se les proporcione suficiente información sobre el servicio y/o producto, por ejemplo se debe mostrar el portafolio, se debe dar pruebas de degustación cuando es un producto es de consumo, con el fin de que el encuestado tenga una visión clara sobre el producto ofrecido y pueda responder con un mayor grado de certeza la pregunta de intención de compra.

Se recomienda descontar algún porcentaje de las respuestas: no todos los encuestados que dicen “definitivamente si lo compraría”, en realidad lo harán, de otra parte, también no todos los encuestados que contestan “Definitivamente no lo compraría” también lo harán.

**Figura 2.14. Escala de intención de compra.**

## Instrucciones

¿cuál de estas expresiones describe mejor qué tan interesado estaría usted en comprar este producto?



1. Definitivamente lo compraría
2. Probablemente lo compraría
3. Podría comprarlo o no comprarlo
4. Probablemente no lo compraría
5. Definitivamente no lo compraría

### **2.10. ESTUDIO DE CASO 1.**

Una estudiante de Ingeniería Agroindustrial de la Facultad de Ciencias Agropecuarias de la Universidad del Cauca elaboró un proyecto que consiste en implementar una empresa productora y comercializadora de productos cárnicos en la ciudad de Popayán. Los productos cárnicos son alimentos muy difundidos en el mercado, especialmente en los supermercados, que son los sitios de mayor afluencia de personas interesadas en consumir este tipo de productos, existen otros lugares como las salsamentarías, puestos de comidas rápidas y establecimientos de producción casera. La compra en los supermercados, es consecuencia de la mayor seguridad alimentaria que le representa al cliente comprar en estos sitios y además por factores como comodidad y seguridad personal.

**2.10.1. Descripción del problema.** En la actualidad, Popayán cuenta con una sola empresa dedicada exclusivamente al procesamiento de carne, lo cual no permite que los posibles consumidores tengan otras alternativas. También existen empresas dedicadas a comercializar productos cárnicos, su campo de acción está dirigido a clientela menos exigente, donde no se manejan altos requerimientos de calidad; un porcentaje elevado es distribuido a puestos callejeros, vendedores ambulantes y tenderos de barrio, quienes se encargan de venderlo al consumidor final.

**2.10.2. Formulación del problema.** ¿Qué incidencia tiene sobre el consumo de productos cárnicos y en la percepción del consumidor, el hecho de que se adopten políticas erradas en cuanto a la selección de materia prima, selección de proveedores, presentación del producto final, fijación de precios, atención al público, entre otros?

**2.10.3. Objetivo general.** Determinar la viabilidad para el montaje de una microempresa destinada al procesamiento de carnes en la ciudad de Popayán.

**2.10.4. Objetivos específicos.**

- Determinar e identificar los clientes potenciales, el mercado proveedor, el mercado competidor y su influencia en la viabilidad del proyecto.
- Diseñar e implementar la imagen corporativa de los productos en mención.
- Determinar la cantidad de productos cárnicos que la comunidad payanesa estaría en capacidad de adquirir.
- Definir las características generales de los productos cárnicos a ofrecer (Chorizo ahumando, jamón de cerdo y jamón de pescado), como: Tamaño, forma de empaque, marca, logotipo, slogan, tipo de envase, entre otros.
- Conocer características del mercado proveedor y el mercado competidor.
- Estimar el comportamiento futuro de la demanda y la oferta de los productos cárnicos.
- Determinar las estrategias de comercialización y promoción de los productos.
- Determinar la zona geográfica en la que se va a desarrollar el montaje de la microempresa de productos cárnicos.

**2.10.5. Diseño de la encuesta.**



**Alifres Ltda**  
**Colombia**

**Datos de clasificación**

## Encuesta a consumidores

Por favor tomese unos minutos para diligenciar este cuestionario, le solicitamos su colaboración para que nos de su opinión acerca de estos nuevos productos cárnicos que proyectamos sacar al mercado en el próximo año.

Su opinión es muy importante para nuestra empresa!!

Número de cuestionario: \_\_\_\_\_ Fecha: \_\_\_\_\_  
aa/mm/dd

---

Barrio: \_\_\_\_\_ Estrato: 1 ( ) 2 ( ) 3 ( ) 4 ( ) 5 ( ) 6 ( )

**Relación con la cabeza de familia**

Madre ( ) Padre ( ) Hijo (a) ( ) Otros ( ) ¿Cual?: \_\_\_\_\_

**Ocupación**

Empleado ( ) Profesional ( ) Ama de casa ( ) Otros ( ) ¿Cual?: \_\_\_\_\_

---

1. ¿Consumo productos cárnicos? Si  No

Si su respuesta es SI continúe con el cuestionario, de lo contrario pase a la pregunta No. 8.

2. ¿Cuál de las siguientes presentaciones de carne que se encuentran en el mercado consume con más frecuencia? ¿Cuál consume con frecuencia? y ¿Cuál con menor frecuencia?

	Consumo con más Frecuencia	Consumo con Frecuencia	Consumo con menor Frecuencia
A. Carne en canal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A. Embutidos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A. Enlatados	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. ¿Cuál es el producto que más consume? ¿Cuál consume con frecuencia? y ¿Cuál con menor frecuencia?

	Consumo con más Frecuencia	Consumo con Frecuencia	Consumo con menor Frecuencia
A. Chorizo corriente	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Chorizo de pollo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C. Jamón de cerdo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D. Jamón de pollo	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E. Costilla ahumada	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



## Encuesta a consumidores

**Alifres Ltda  
Colombia**

4. ¿Cuál es la cantidad de carne que consume expresada en libras y con que frecuencia lo hace?

	Diario	Semanal	Quincenal	Mensual
A. Chorizo corriente	_____	_____	_____	_____
B. Chorizo de pollo	_____	_____	_____	_____
C. Jamón de cerdo	_____	_____	_____	_____
D. Jamón de pollo	_____	_____	_____	_____
E. Costilla ahumada	_____	_____	_____	_____

5. ¿En que lugar compra preferiblemente estos productos? Seleccione dos opciones.

	Mayor Preferencia	De Preferencia
A. Salsamentarias	<input type="checkbox"/>	<input type="checkbox"/>
B. Supermercados	<input type="checkbox"/>	<input type="checkbox"/>
C. Puestos de comidas rápidas	<input type="checkbox"/>	<input type="checkbox"/>
D. Tiendas	<input type="checkbox"/>	<input type="checkbox"/>
E. Galerías	<input type="checkbox"/>	<input type="checkbox"/>

6. Del siguiente listado seleccione dos (2) razones por la cual usted prefiere estos sitios?

	Primera Razón	Segunda Razón
A. Comodidad	<input type="checkbox"/>	<input type="checkbox"/>
B. Economía	<input type="checkbox"/>	<input type="checkbox"/>
C. Costumbre	<input type="checkbox"/>	<input type="checkbox"/>
D. Seguridad	<input type="checkbox"/>	<input type="checkbox"/>
E. Otros	<input type="checkbox"/>	<input type="checkbox"/>

Si una de sus respuestas es la opción E especifique cuál o cuáles:

---



---



## Encuesta a consumidores

**Alifres Ltda  
Colombia**

7. Del siguiente listado de marcas seleccione su preferencia en primer lugar, segundo y tercero.

	Primera	Segunda	Tercera
A. Rica	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Zenu	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C. La blanca	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D. Elaboración casera	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E. Marcas de almacenes en cadena	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Pase a la pregunta No. 9.

8. ¿Por qué no consume productos cárnicos? Marque dos (2) opciones según sus motivos.

	Primero	Segundo
A. Por salud	<input type="checkbox"/>	<input type="checkbox"/>
B. Porque son muy costosos	<input type="checkbox"/>	<input type="checkbox"/>
C. Por costumbre	<input type="checkbox"/>	<input type="checkbox"/>
D. Otros	<input type="checkbox"/>	<input type="checkbox"/>

Si una de sus respuestas es la opción D especifique cuál o cuáles:

---



---

9. Por favor haga la prueba de degustación del **Jamón de Cerdo** y califique las siguientes características.

	Alta 	Normal 	Baja 
A. Cantidad de grasa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Cantidad de sal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Excelente 	Bueno (a) 	Regular 	Malo (a) 
A. Presentación	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. En general como le pareció el producto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



## Encuesta a consumidores

**Alifres Ltda**  
**Colombia**

10. Por favor haga la prueba de degustación del **Jamón de Pescado** y califique las siguientes características.

	Alta 	Normal 	Baja 
A. Cantidad de grasa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Cantidad de sal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Excelente 	Bueno (a) 	Regular 	Malo (a) 
A. Presentación	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. En general como le pareció el producto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

11. Por favor haga la prueba de degustación del **Chorizo de cerdo ahumado** y califique las siguientes características.

	Alta 	Normal 	Baja 
A. Cantidad de grasa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Cantidad de sal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Excelente 	Bueno (a) 	Regular 	Malo (a) 
A. Presentación	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. En general como le pareció el producto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



## Encuesta a consumidores

**Alifres Ltda  
Colombia**

12. Por favor haga la prueba de degustación del **Chorizo especial** y califique las siguientes características.

	Alta 	Normal 	Baja 
A. Cantidad de grasa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Cantidad de sal	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	Excelente 	Bueno (a) 	Regular 	Malo (a) 
A. Presentación	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. En general como le pareció el producto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

13. ¿Cuál de los anteriores productos le gustó más? Elija dos productos.

	Primera preferencia 	Segunda preferencia 
A. Jamón de cerdo	<input type="checkbox"/>	<input type="checkbox"/>
B. Jamón de pescado	<input type="checkbox"/>	<input type="checkbox"/>
C. Chorizo de cerdo ahumado	<input type="checkbox"/>	<input type="checkbox"/>
D. Chorizo especial	<input type="checkbox"/>	<input type="checkbox"/>

14. De las siguientes expresiones ¿Cuál describe que tan interesado (a) estaría usted en comprar y consumir nuestros productos? Marque una opción.

- |                                    |                          |
|------------------------------------|--------------------------|
| A. Definitivamente si lo compraría | <input type="checkbox"/> |
| B. Probablemente lo compraría      | <input type="checkbox"/> |
| C. Podría comprarlo o no comprarlo | <input type="checkbox"/> |
| D. Probablemente no lo compraría   | <input type="checkbox"/> |
| E. Definitivamente no lo compraría | <input type="checkbox"/> |

**Recomendaciones**

---



---

## **2.11. ESTUDIO DE CASO 2.**

Una empresa dedicada a la confecciones de prendas de dotación está realizando una investigación de mercados; el medio para la recolección de la información es la encuesta. Los empresarios parten de tener claro el problema de investigación y establecen los siguientes objetivos:

**2.11.1. Objetivo General.** Determinar la viabilidad comercial del Proyecto de producción y comercialización de ropa de dotación en la ciudad de Popayán.

### **2.11.2. Objetivos Específicos:**

- Identificar las especificaciones y usos del producto
- Determinar el tamaño del mercado potencial para ropa de dotación en la ciudad de Popayán.
- Conocer el tipo de dotación de mayor utilización por parte de los clientes potenciales.
- Conocer las características y comportamiento de la oferta del mercado competidor.
- Establecer el perfil de las empresas que normalmente adquieren dotaciones de ropa.
- Determinar la política de precios que se seguirán.
- Determinar el canal de distribución a utilizar.
- Diseñar un sistema de promoción y publicidad para la óptima comercialización del producto.

### **2.11.3. Diseño de la encuesta**

	
No. de cuestionario: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> Fecha: <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> dd mm aa	<b>Encuesta a empresarios</b>
<p>Señor empresario le solicitamos que se tome unos minutos para diligenciar el siguiente cuestionario, le pedimos su colaboración para que nos exprese su opinión sobre las confecciones y diseños de ropa de labor. La presente encuesta tiene como objetivo mejorar nuestros productos y servicios.</p>	
<p><b>Su opinión es muy importante para nuestra organización!</b></p>	
<b>Datos de clasificación</b>	
Nombre de la empresa: _____	
Dirección: _____	
<b>Clasificación CHU</b>	
<b>EMPRESA AGROPECUARIA</b>	<b>EMPRESA INDUSTRIALES</b>
1. Agricultura, ganadería, caza y silvicultura <input type="checkbox"/>	1. Industria manufacturera <input type="checkbox"/>
2. Explotación de minas y canteras <input type="checkbox"/>	2. Construcción <input type="checkbox"/>
<b>EMPRESA SERVICIOS</b>	1. Otro <input type="checkbox"/>
1. Comercio <input type="checkbox"/>	¿Cual? _____
2. Servicio de hoteles <input type="checkbox"/>	Cargo del entrevistado: _____
3. Intermediación financiera <input type="checkbox"/>	_____
4. Transporte <input type="checkbox"/>	_____
5. Servicios sociales de salud <input type="checkbox"/>	_____
6. Educación <input type="checkbox"/>	
1. ¿Cuántos empleados hay en total en su empresa? <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	
2. ¿Usted otorga dotaciones a sus empleados?	
<input type="checkbox"/> 1. Si (Continúe con el cuestionario)	
<input type="checkbox"/> 2. No (Pase a la pregunta 14)	



Ormex

Confecciones de uniformes de labor

Diseños industriales exclusivo para comodidad del trabajador

## Encuesta a empresarios

3. ¿Normalmente a cuantos empleados se le da dotación?

4. ¿Anualmente con que frecuencia entrega usted dotación a sus empleados?

1. Una vez

2. Dos veces

3. Tres veces

4. Otro  Especifique: \_\_\_\_\_

5. ¿Existe una frecuencia adicional?

1. Si  ¿Cual? \_\_\_\_\_ ¿A quien se le da? \_\_\_\_\_

2. No  Continúe con el cuestionario

6. ¿En que consta la dotación que usted otorga a un empleado?

### Hombre

	Tipo de prenda		Cuantas prendas	Frecuencia adicional	Cuantas prendas
	1. Pantalón	<input style="width: 20px; height: 20px;" type="checkbox"/>	_____	_____	_____
	2. Camisa	<input style="width: 20px; height: 20px;" type="checkbox"/>	_____	_____	_____
	3. Overol	<input style="width: 20px; height: 20px;" type="checkbox"/>	_____	_____	_____
	4. Jeans	<input style="width: 20px; height: 20px;" type="checkbox"/>	_____	_____	_____
	5. Chaqueta	<input style="width: 20px; height: 20px;" type="checkbox"/>	_____	_____	_____
	6. Delantal	<input style="width: 20px; height: 20px;" type="checkbox"/>	_____	_____	_____
	7. Otro (s)	<input style="width: 20px; height: 20px;" type="checkbox"/>	_____	_____	_____

Especifique: \_\_\_\_\_

\_\_\_\_\_



## Encuesta a empresarios

### Mujer



Tipo de prenda	Cuantas prendas	Frecuencia adicional	Cuantas prendas
1. Blusa <input type="checkbox"/>	_____	_____	_____
2. Falda <input type="checkbox"/>	_____	_____	_____
3. Vestido <input type="checkbox"/>	_____	_____	_____
4. Chaqueta <input type="checkbox"/>	_____	_____	_____
5. Delantal <input type="checkbox"/>	_____	_____	_____
6. Jeans <input type="checkbox"/>	_____	_____	_____
7. Otro (s) <input type="checkbox"/>	_____	_____	_____

Especifique: \_\_\_\_\_

---

6. Aproximadamente ¿cual es el presupuesto anual asignado para la compra de ropa de dotación de esta empresa?

1. Menos de \$500.000 <input type="checkbox"/>
2. De \$ 500.000 - \$ 1.000.000 <input type="checkbox"/>
3. De \$ 1.000.001 - \$ 1.500.000 <input type="checkbox"/>
4. De \$ 1.500.001 - \$ 2.000.000 <input type="checkbox"/>
5. De \$ 2.000.001 - \$ 2.500.000 <input type="checkbox"/>
6. De \$ 2.500.001 - \$ 3.000.000 <input type="checkbox"/>
7. De \$ 3.000.001 - \$ 3.500.000 <input type="checkbox"/>
8. De \$ 3.500.001 - \$ 4.000.000 <input type="checkbox"/>
9. De \$ 4.000.001 - \$ 4.500.000 <input type="checkbox"/>
10. De \$ 4.500.001 - \$ 5.000.000 <input type="checkbox"/>
11. Mas de \$5.000.000 <input type="checkbox"/>



¿Cuanto? \_\_\_\_\_



## Encuesta a empresarios

8. ¿A quien le adquiere la ropa de dotación para sus empleados?. Especifique el nombre del proveedor.

1. Empresa especializada en confecciones	<input type="checkbox"/>	
2. Almacenes	<input type="checkbox"/>	
3. Modistas	<input type="checkbox"/>	
4. Otros	<input type="checkbox"/>	

9. Califique las características de su proveedor de ropa de dotación, en cuanto a:

	Excelente	Bueno	Regular	Mala (o)
1. Calidad del producto	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. Diseños	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. Variedad	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. Tiempo de entrega	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5. Precio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. Atención	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

¿Si respondió regular o malo, indique por que? \_\_\_\_\_

\_\_\_\_\_

10. ¿Como se entero del ultimo proveedor de ropa de dotación?

1. Revistas <input type="checkbox"/>		5. Publicidad especializada <input type="checkbox"/>
2. Periodicos <input type="checkbox"/>		6. Otro (s) <input type="checkbox"/> Especifique: _____
3. Visitas <input type="checkbox"/>		_____
4. Radio <input type="checkbox"/>		



## Encuesta a empresarios

11. ¿Cual de las siguientes características de sus proveedores considera mas importante?. Marque en primer lugar la que considera mas importante y en ultimo lugar (quinto) aquella menos importante.

No necesariamente tiene que marcar todas la opciones..

	1 <sup>ra.</sup>	2 <sup>da.</sup>	3 <sup>ra.</sup>	4 <sup>ta.</sup>	5 <sup>ta.</sup>
1. Cumplimiento en la entrega del pedido	<input type="checkbox"/>				
2. Facilidades de pago	<input type="checkbox"/>				
3. Trayectoria de la empresa	<input type="checkbox"/>				
4. Precios razonables	<input type="checkbox"/>				
5. Calidad de telas y acabados	<input type="checkbox"/>				

12. ¿Cual es el tramite formal de su firma para la compra de las dotaciones de sus empleados?.

Puede marcar mas de una opción.

1. Licitación     2. Contratos     3. Compra directa
4. Otro (s)     Especifique: \_\_\_\_\_

13. ¿Que requisitos exige su firma para realizar transacciones comerciales de este tipo?.

Puede marcar mas de una opción.

PERSONAS JURÍDICAS	1. Nit.	<input type="checkbox"/>	PERSONAS NATURALES	1. Rut	<input type="checkbox"/>
	2. Registro de camara de comercio	<input type="checkbox"/>		2. Antecedentes disciplinarios	<input type="checkbox"/>
	3. Facturación	<input type="checkbox"/>		3. Pasado judicial	<input type="checkbox"/>
	4. Pólizas de cumplimiento	<input type="checkbox"/>		4. Pólizas de cumplimiento	<input type="checkbox"/>
	5. Presentar diseños y muestras	<input type="checkbox"/>		5. Facturación	<input type="checkbox"/>
			5. Presentar diseños y muestras	<input type="checkbox"/>	

	<h2>Encuesta a empresarios</h2>
<p>Si su empresa exige otro(s) requisitos, por favor especifique cuales. _____</p>	
<p>14. ¿Por que no otorga dotaciones a sus empleados?</p>	
1. Porque la ley no me lo exige	<input type="checkbox"/>
2. Es un gasto muy alto para nuestra empresa	<input type="checkbox"/>
3. Otro(s)	<input type="checkbox"/> Especifique _____
<p>15. Por favor vea nuestro portafolio, donde puede observar el catalogo de productos y servicios que nuestra organización tiene a su disposición, así como las promociones para que su empresa se convierta en nuestro cliente.</p>	
<p>¿Cual de las siguiente expresiones describe que tan interesada estaría su empresa en comparar nuestros productos?.</p>	
1. Definitivamente si los compraría	<input type="checkbox"/>
2. Probablemente si los compraría	<input type="checkbox"/>
3. Podría comprarlos o no comprarlos	<input type="checkbox"/>
4. Probablemente no los compraría	<input type="checkbox"/>
5. Definitivamente no los compraría	<input type="checkbox"/>
	
<p><b>Sugerencias y recomendaciones.</b></p>	
<hr/> <hr/> <hr/>	
<p><b>Gracias por su colaboración</b></p>	

# *Capítulo tres*

## *Procedimientos de muestreo*

---

### **3. PLAN DE MUESTREO**

Todo estudio de mercado requiere de recolección de información. Ahora bien, dependiendo del objetivo u objetivos que se persiguen, puede realizarse una investigación exhaustiva o una investigación parcial.

En la investigación exhaustiva se deben observar todas las unidades que constituyen la población o universo. La enumeración total de una población en un tiempo dado recibe el nombre de CENSO. El censo requiere de una organización compleja y la ocupación de un gran número de personas en cada una de las diferentes etapas de investigación.

#### **Desventajas de realizar un censo**

- Se requiere demasiado tiempo para realizarlo.
- Los costos son elevados en las etapas de planeamiento, sistematización y análisis y publicación de la información.
- Tiene una alta probabilidad de cometerse el error sistemático.
- En algunos casos no se podrá obtener información especializada, por ejemplo, no se podrá dar pruebas de degustación a toda una población.
- Los estudiantes carecen de tiempo y recursos para realizar una investigación exhaustiva.

Las anteriores consideraciones obligan a buscar un método práctico para obtener información, en especial para poblaciones grandes. Este método consiste en realizar un estudio parcial o tomar una muestra de la población total.

#### **Ventajas de realizar una muestra**

- Reduce los costos y el tiempo
- Es aplicable en poblaciones infinitas y muy grandes

Al hablar de muestreo nos referimos al conjunto de técnicas estadísticas que estudian la forma de seleccionar una muestra lo suficientemente representativa de una población cuya información permita inferir las propiedades o características de toda la población cometiendo un error medible o contable. A partir de la muestra, seleccionada mediante un determinado método de muestreo, se estiman las características poblacionales (media, total, proporción, etc.), estas estimaciones se realizan a través de funciones matemáticas de la muestra denominadas estadísticos, que se convierten en variables aleatorias al considerar la variabilidad de las muestras. Los errores se cuantifican mediante varianzas, desviaciones típicas o errores cuadráticos medios de los estimadores, que miden la precisión de éstos. La metodología que permite inferir resultados, predicciones y generalizaciones sobre la población estadística, basándose en la información contenida en las muestras representativas previamente elegidas por métodos de muestreo formales, se denomina inferencia estadística.

A continuación se dan algunos conceptos generales de muestreo, se parte del supuesto de que los estudiantes tienen alguna formación en estadística básica para abordar el tema.

### **3.1. ELEMENTO.**

Es la unidad acerca de la cual se solicita información. Éste suministra la base del análisis que se llevará a cabo. Los elementos más comunes del muestro en investigación de mercados son los individuos. En otros casos, los elementos podrían ser productos, almacenes, empresas, familias, etc. Los elementos de cualquier muestra específica dependerán de los objetivos del estudio.

### **3.2. POBLACIÓN.**

Una población o universo, como también se llama, es el conjunto de todos los elementos definidos antes de selección de la muestra. Una población adecuadamente designada debe definirse en términos de: 1) elementos, 2) unidades de muestreo, 3) alcance y 4) tiempo: Por ejemplo, una encuesta de consumidores debe especificar la población pertinente de la siguiente manera:

- |                         |                                    |
|-------------------------|------------------------------------|
| 1. Elemento             | : Mujeres entre 18 – 50 años       |
| 2. Unidades de muestreo | : Mujeres entre 18 – 50 años       |
| 3. Alcance              | : Popayán                          |
| 4. Tiempo               | : 1 de mayo al 15 de junio de 2000 |

### 3.3. UNIDAD DE MUESTREO.

Es el elemento o los elementos disponibles para su selección en alguna etapa del proceso de muestreo. El ejemplo anterior es un tipo de muestreo simple, es decir, de una sola etapa, en este caso las unidades y los elementos de muestro son los mismos.

En procedimientos de muestreo más complejos pueden utilizarse diferentes niveles de unidades de muestreo, entonces las unidades de muestreo y los elementos difieren en todo, menos en la última etapa. Veamos el siguiente ejemplo:

La encuesta continúa de hogares realizada por el DANE, es Polietápica porque para la selección de los hogares objeto de investigación se seleccionan secuencialmente las unidades de muestreo.



**Departamento Administrativo  
Nacional de Estadística**

**Gran Encuesta Integrada de Hogares**

1. Elemento	: Hogares
2. Unidad de muestreo	
● Unidades primarias	: Municipios
● Unidades secundarias:	: Sectores urbanos y rurales
● Unidades terciarias:	: Manzanas
● Unidades cuartas:	: Hogares
3. Alcance	: Colombia
4. Tiempo	: Enero - marzo de 2007
5. Tipo de muestra	: Probabilística, estratificada, de conglomerados y polietápica.

### 3.4. MARCO MUESTRAL.

Es una lista de todas las unidades de muestreo disponibles para su selección en una etapa del proceso de muestreo. Uno de los procedimientos más creativos en un proyecto de investigación de mercados puede relacionarse con la especificación de un marco muestral.

Un marco puede ser una lista de alumnos, una lista de votantes inscritos, un directorio telefónico, una lista de empleados o incluso un mapa.

### 3.5. POBLACIÓN DE ESTUDIO.

La población de estudio es el conjunto de elementos del cual se saca la muestra. Existe varios tipos de muestreo, dependiendo si la población es finita o infinita, materia sobre la que existe amplia literatura estadística. En éste texto se considera solamente el muestreo en poblaciones finitas.

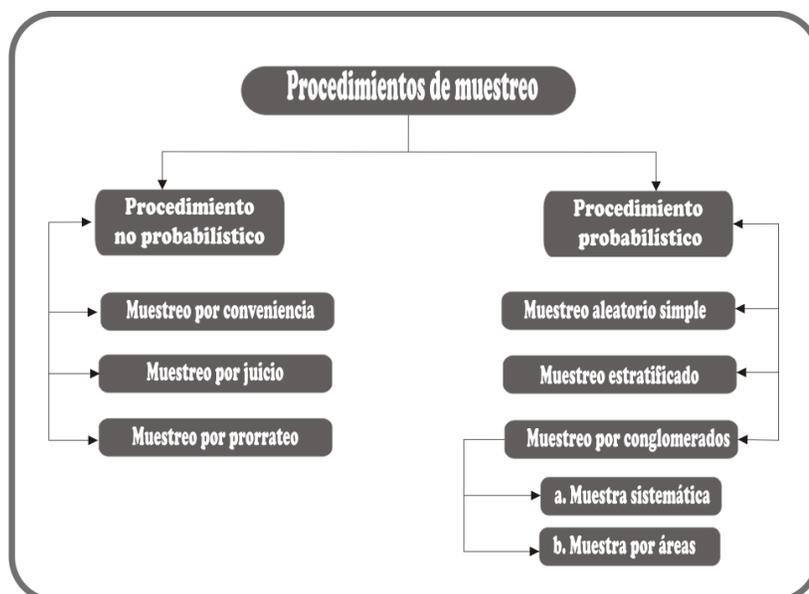
### 3.6. PROCEDIMIENTOS DE MUESTREO

Como se mencionó existen muchos procedimientos mediante los cuales los investigadores pueden seleccionar sus muestras, pero se debe establecer la diferencia entre una muestra probabilística y una muestra no probabilística.

**3.6.1. Procedimiento Probabilístico.** Cada elemento de la población tiene una probabilidad conocida de ser seleccionado para la muestra. Una muestra probabilística permite calcular el grado probable hasta el cual el valor de la muestra puede diferir del valor de la población de interés. Esta diferencia recibe el nombre de **error muestral**.

**3.6.2. Procedimiento no Probabilístico.** La selección de un elemento de la población que va a formar parte de la muestra se basa hasta cierto punto en el criterio del investigador o entrevistador de campo. No existe una posibilidad conocida para seleccionar cualquier elemento particular de la población.

**Figura 3.1.** Procedimientos de Muestreo



**3.6.3. Parámetro.** Un parámetro es una descripción de una medida de la población bajo estudio. Ejemplos: Edad promedio de los estudiantes, Ingreso promedio, etc.

**3.6.4. Estadístico.** Un estadístico es una descripción resumida de una medida en la muestra seleccionada. Así, la edad promedio de los estudiantes sería un estadístico si se mide a través de una muestra.

### 3.7. NOMENCLATURA UTILIZADA.

La nomenclatura que se utilizará será la siguiente:

**Tabla 3.1.** Nomenclatura

Parámetros / Estadísticos	Símbolo de Población	Símbolo Muestral
Media o promedio	$\mu$	$\bar{X}$
Varianza	$\sigma^2$	$S^2$
Probabilidad de ocurrencia del fenómeno de estudio.	$\pi$	p
Probabilidad de no ocurrencia del fenómeno de estudios	$(1-\pi)$	$(1-p)$ o q

### 3.8. DISTRIBUCIONES MUESTRALES.

Consideramos los sucesos elementales asociados a un fenómeno o experimento aleatorio dado  $S_1, S_2, \dots, S_m$ , entendiendo por suceso elemental los más simples posibles, es decir, aquellos que no pueden ser descompuestos en otros sucesos. El conjunto  $\{S_1, S_2, \dots, S_m\}$  se denomina espacio muestral asociado al fenómeno o experimento. Si consideramos como fenómeno o experimento la extracción aleatoria de muestras dentro de una población por un procedimiento o método de muestreo dado, podemos considerar como sucesos elementales las muestras obtenidas, constituyendo el conjunto de las mismas el espacio muestral.

Habitualmente en los métodos de muestreo se consideran iguales muestras con los mismos elementos, aunque estén colocados en orden diferente (el orden de colocación no interviene), es decir, cuando la selección se hace con reposición. Una muestra de tamaño  $n$  extraída de una población  $U=\{U_1, U_2, \dots, U_N\}$  de tamaño  $N$  mediante un método de muestreo dado, suele denotarse como  $s=\{u_1, u_2, \dots, u_n\}$ . De esta forma, El conjunto de las  $N^n$  muestras posibles de tamaño  $s$  que se pueden formar con los elementos de la población  $U$  es el espacio muestral  $S$ .

Cuando el método de muestreo se realiza sin reposición el conjunto de las  $\frac{N!}{(N-n)!n!}$  muestras posibles de tamaño n que se pueden formar con los elementos de la población U es el espacio muestral S.

Evidentemente, para establecer la probabilidad de todas las muestras posibles derivadas de un procedimiento de muestreo dado, será necesario conocer ese conjunto de muestras, es decir, será necesario delimitar tanto el método de muestreo como el espacio muestral derivado del mismo, por tanto el método aleatorio empleado para seleccionar la muestra define en el espacio muestral S una función de probabilidad P tal que:

$$\begin{aligned} P(S_i) &\geq 0 \forall i \\ \sum_s P(S_i) &= 1 \end{aligned} \quad (1)$$

Ahora consideremos una población de N elementos, con media  $\mu$  y desviación típica  $\sigma$ , si se obtiene M número de muestras posibles, de tamaño n, simbolizamos a cada media muestral por  $\bar{x}_1; \bar{x}_2; \bar{x}_3; \dots; \bar{x}_M$  y cada desviación típica muestral por:  $s_1; s_2; s_3 \dots \dots s_m$

**Teorema.** Dada una población, si extraemos todas las muestras posibles de un mismo tamaño, entonces las medias de la distribución de todas las medias muestrales posibles será igual a la media poblacional.

$$\mu_x = \frac{\sum \bar{X}_j}{M} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \dots + \bar{X}_M}{M} = \mu \quad (2)$$

La varianza de todas las medias muestrales se simboliza por:  $\sigma_{\bar{x}}^2$

El error estándar será simbolizado por:  $\sigma_{\bar{x}}$

$$\sigma_x = \sqrt{\frac{\sum (\bar{X}_j - \mu)^2}{M}} = \sqrt{\frac{(\bar{X}_1 - \mu)^2 + (\bar{X}_2 - \mu)^2 + \dots + (\bar{X}_M - \mu)^2}{M}} \quad (3)$$

Siendo  $\sigma_x = \frac{\sigma}{\sqrt{n}}$  para muestras grandes o sea  $n > 30$  y se denomina: error estándar de la media.

La media de todas las medias muestrales debe ser exactamente igual a la media poblacional ( $\mu$ ), debido a que la distribución de muestreo resulta de todas las muestras posibles que se pueden extraer de una población; por tal razón incluye a todos sus elementos.

Explicamos lo anterior mediante un ejemplo. Supongamos una población de 5 elementos ( $N=5$ ), los valores que toman las variables son:

$$X_1=2 \quad X_2=4 \quad X_3=6 \quad X_4=8 \quad X_5=10$$

Con los anteriores valores se calcula la media poblacional, la varianza y la desviación típica poblacional así:

$$\mu = \frac{\sum X_i}{N} = \frac{2+4+6+8+10}{5} = \frac{30}{5} = 6 \quad (4)$$

$$\sigma^2 = \frac{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2}{5} = \frac{40}{5} = 8 \quad (5)$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{8} = 2,83 \quad (6)$$

Ahora determinamos el número de muestras posibles ( $M$ ) de esta población, si el tamaño de la muestra que fijamos es de 2 y la selección se hace sin reposición se tiene:

$$M = c_n^N = \frac{N!}{(N-n)!n!} = \frac{5!}{(5-2)!2!} = 10 \quad (7)$$

Las combinaciones que se pueden obtener son las siguientes:

$$\begin{array}{llll} X_1X_2 & X_2X_3 & X_3X_4 & X_4X_5 \\ X_1X_3 & X_2X_4 & X_3X_5 & \\ X_1X_4 & X_2X_5 & & \\ X_1X_5 & & & \end{array}$$

Calculamos la media aritmética para cada una de las muestras:

$$\bar{x}_1 = \frac{x_1 + x_2}{n} = \frac{2+4}{2} = 3 \quad \bar{x}_6 = \frac{x_2 + x_4}{n} = \frac{4+8}{2} = 6$$

$$\bar{x}_2 = \frac{x_1 + x_3}{n} = \frac{2+6}{2} = 4 \quad \bar{x}_7 = \frac{x_2 + x_5}{n} = \frac{4+10}{2} = 7$$

$$\bar{x}_3 = \frac{x_1 + x_4}{n} = \frac{2+8}{2} = 5 \quad \bar{x}_8 = \frac{x_3 + x_4}{n} = \frac{6+8}{2} = 7$$

$$\bar{x}_4 = \frac{x_1 + x_5}{n} = \frac{2+10}{2} = 6 \qquad \bar{x}_9 = \frac{x_3 + x_5}{n} = \frac{6+10}{2} = 8$$

$$\bar{x}_5 = \frac{x_2 + x_3}{n} = \frac{4+6}{2} = 5 \qquad \bar{x}_{10} = \frac{x_4 + x_5}{n} = \frac{8+10}{2} = 9$$

La media de las medias muestrales será igual a:

$$\mu_{\bar{x}} = \frac{\sum \bar{x}_i}{M} = \frac{3+4+5+6+5+6+7+7+8+9}{10} = 6, \text{ es decir, } \mu_{\bar{x}} = \frac{\sum \bar{x}_i}{M} = \mu$$

La desviación típica de todas las medias muestrales, se calcula a continuación

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum (\bar{x}_i - \mu)^2}{M}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{(3-6)^2 + (4-6)^2 + (5-6)^2 + (6-6)^2 + (5-6)^2 + (6-6)^2 + (7-6)^2 + (7-6)^2 + (8-6)^2 + (9-6)^2}{10}} = \sqrt{\frac{30}{10}} = \sqrt{3} = 1.73$$

Lo anterior debe ser igual a  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ , esta expresión se denomina error estándar de la media, lo cual se cumple para muestras grandes, algunos autores consideran que una muestra es grande cuando  $n > 30$ , por esta razón los resultados difieren:

Con los datos poblacionales tenemos:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.8384}{1.4142} = 2 \neq 1.73$$

### 3.9. TEOREMA DEL LÍMITE CENTRAL.

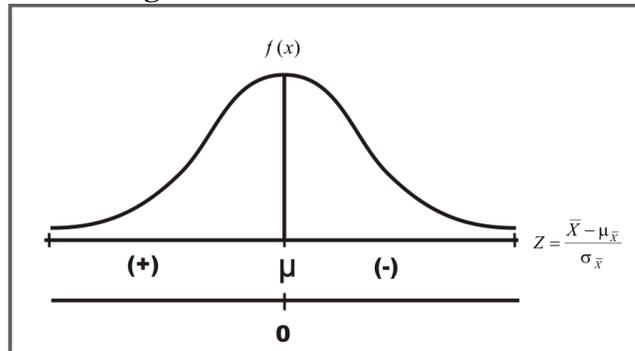
Se cumple cuando independientemente de la población origen la distribución de las medias aleatorias se aproxima a una distribución normal a medida que el tamaño de la muestra crece. Se podrá decir también que si las muestras provienen de una población que no es normal. Si el tamaño de la muestra es pequeño, la distribución obtenida de sus medias muestrales tendrá un comportamiento similar al de la población de donde se extrajeron. Por el contrario, si el tamaño muestral es grande, el comportamiento de estas medias muestrales será igual al de una distribución normal independientemente de la población de donde fueron extraídas.

En su forma más simple el teorema indica que si  $n$  variables aleatorias independientes tienen varianzas finitas, su suma, cuando se le expresa en medida estándar, tienden a estar normalmente distribuidas cuando  $n$  tiende a infinito.

De acuerdo al teorema, la varianza estadística para distribuciones de medias muestrales será:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Figura 3.2. Distribución normal

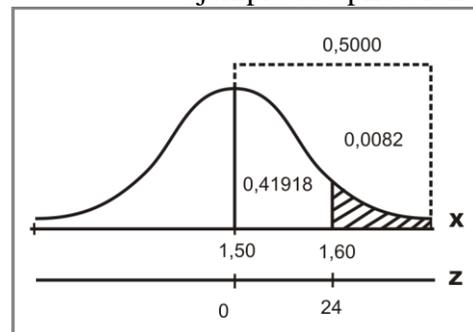


Es decir, se aproxima a una distribución normal

Ejemplo 1º.

La altura media de 400 alumnos de un plantel de secundaria es de 1,50 m. y su desviación típica es de 0,25 m. Determinar la probabilidad de que en una muestra de 36 alumnos, la media sea superior a 1,60 m.

Figura 3.3. Distribución normal ejemplo de aplicación



$$P(\bar{X} \geq 1,60) = ?$$

$$Z = \frac{1,60 - 1,50}{\frac{0,25}{\sqrt{36}}} = \frac{0,10}{\frac{0,25}{6}} = \frac{0,60}{0,25} = 2,40$$

$$Z = 2,40 \rightarrow A(0,4918)$$

$$P = 0,5000 - 0,4918 = 0,0082 = 0,82\%$$

<sup>9</sup> Martínez Bernardino. Ciro, Estadística y muestreo, 11ª ed., Ecoe Ediciones, Bogotá, D.C. 2002, p 321 y 322

### Ejercicio para realizar en clase

- Se tiene para la venta un lote de 1.000 pollos, con un peso promedio de 3,50 Kg. y una desviación estándar de 0,18 Kg., ¿Cuál es la probabilidad de que en una muestra aleatoria, 100 pollos de esta población, pesen entre 3,53 y 3,54 Kg.?
- Un fabricante de cierto champú para el cabello distribuye el tamaño profesional de su producto en 100 salones de belleza. Se ha determinado que el consumo promedio de su producto es de 2.800 cojines mensuales, con desviación estándar de 280 cojines. Si se toma una muestra probabilística de 36 salones, ¿cuál es la probabilidad de que el consumo promedio en un mes sea inferior a 2.700?

### 3.10. FACTOR DE CORRECCIÓN PARA POBLACIONES FINITAS.

En aquellos casos de poblaciones finitas, es decir, cuando se da información sobre el tamaño poblacional y cuando el tamaño de la muestra es mayor del 5% de la población, se puede aplicar el factor de corrección, representado de diferentes maneras; cualquiera de estas formas podrá ser aplicada:

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{N-n}{N}} = \sqrt{\frac{N}{N} - \frac{n}{N}} = \sqrt{1 - \frac{n}{N}} = \sqrt{1-f} \dots \text{donde} \dots f = \frac{n}{N} \quad (8)$$

En distribuciones de medias muestrales, la estandarización de Z, incluyendo el factor de corrección será:

$$Z = \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right) \sqrt{\frac{N-n}{N}}} \quad (9)$$

### 3.11. DISTRIBUCIÓN MUESTRAL DE UNA PROPORCIÓN.

En el análisis de una característica cualitativa o atributo se emplea la proporción de la ocurrencia (P) o éxito y no ocurrencias (Q) del fenómeno de estudio, siguiendo una distribución binomial.

Se define la proporción de éxitos o ocurrencia del fenómeno de estudio como:

$$P = \frac{\text{Numero de casos favorables o exitos}}{\text{total de casos posibles}} = \frac{\sum a_j}{n} \quad (10)$$

La nomenclatura que se utilizará es la siguiente:

$A = \sum A_j = NP$  Total de elementos que presentan la característica en la población

$P = \frac{A}{N} = \frac{\sum A_j}{N}$  Proporción de elementos que presentan la ocurrencia del fenómeno de estudio (éxito).

$Q = \frac{N - A}{N} = 1 - P$  Proporción de elementos que no presentan la ocurrencia del fenómeno estudio.

$$\text{Entonces: } P + Q = 1 \quad (11)$$

$$\text{Varianza de la proporción en la población } \sigma_p^2 = PQ \quad (12)$$

$$\text{Desviación estándar } \sigma_p = \sqrt{PQ} \quad (13)$$

$$\sigma_{\bar{p}} = \frac{\sigma_p}{\sqrt{n}} = \sqrt{\frac{PQ}{n}} = \text{Error estándar de la proporción} \quad (14)$$

Ejemplo 2<sup>10</sup>:

Se tiene que el 4% de las piezas producidas por cierta máquina son defectuosas, ¿cuál es la probabilidad de que en un grupo de 200 piezas, el 3% o más sean defectuosas?

Solución:

$$\mu_p = P = 0,04 \rightarrow \bar{P} = P = 0.04$$

$$\sigma_{\bar{p}} = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{(0,04)(0,96)}{200}} = 0,014$$

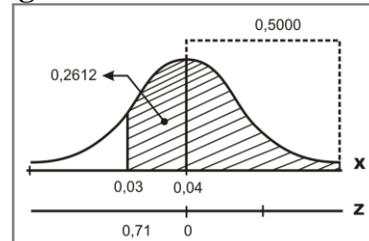
Se desea determinar la  $P(p \geq 0,03) = ?$

$$Z = \frac{p - \mu_p}{\sqrt{\frac{PQ}{n}}} = \frac{0,03 - 0,04}{\sqrt{\frac{(0,04)(0,96)}{200}}} = -0,71$$

$$Z = -0,71 \rightarrow A(0,2612)$$

$$P = 0,2612 + 0,5000 = 0,7612 = 76,12\%$$

**Figura 3.4.** Distribución normal



<sup>10</sup> *Ibíd.*, p. 332.

### Ejercicios para resolver en clase

- Se desea estudiar una muestra de 49 personas para saber la proporción de las mayores de 40 años; sabiendo que la proporción en la población es 0,4, ¿cuál es la probabilidad de que la proporción en la muestra sea menor de 0,5?
- Cuarenta y seis por ciento de los sindicatos del país están en contra de comerciar con la China Continental; ¿cuál es la probabilidad de que una encuesta a 100 sindicatos muestre que más del 52% tenga la misma posición?

### 3.12. TAMAÑO DE LA MUESTRA.

Para determinar **n** el tamaño de la muestra es necesario identificar los siguientes componentes o elementos técnicos:

**3.12.1. La Varianza** ( $\sigma^2_x$ ). En el caso de variables discretas ( $\sigma^2_x$ )=PQ.

**3.12.2. El Nivel de Confianza.** Tiene relación directa con el tamaño de la muestra, por lo tanto se dirá que a mayor nivel de confianza más grande debe ser el tamaño de la muestra. El nivel es fijado por el investigador, de acuerdo a su experiencia.

**3.12.3. Precisión de la Estimación.** Corresponde al margen de error que el investigador fija de acuerdo al conocimiento que tenga acerca del parámetro que piensa estimar. Se le conoce como error de muestreo (E), siendo:

$$E = Z \frac{\sigma}{\sqrt{n}} \quad E = Z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N}} \quad (15)$$

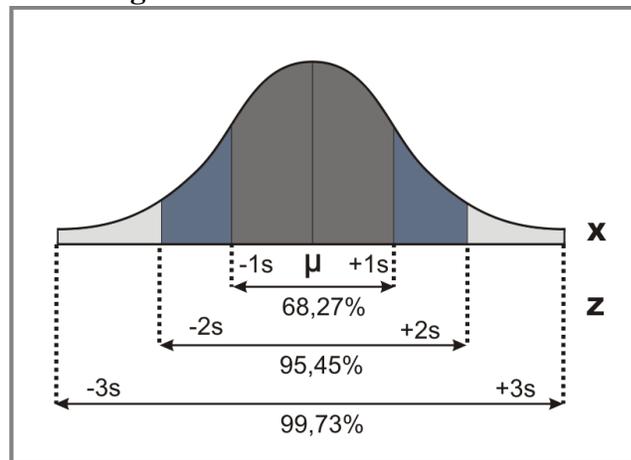
### 3.13. ESTIMACIÓN DE INTERVALOS DE CONFIANZA PARA PARÁMETROS DE POBLACIÓN.

Sean  $\mu_s$  y  $\sigma_s$  la media y la desviación (error típico) de la distribución de muestreo de un estadístico S. Entonces si la distribución de muestreo de S es aproximadamente normal (que como hemos visto es cierto para muchos estadísticos si el tamaño de la muestra  $n \geq 30$ ), se espera hallar un estadístico real S que esté en los intervalos.

$$[\mu_s - \sigma_s; \mu_s + \sigma_s] ; [\mu_s - 2\sigma_s; \mu_s + 2\sigma_s] ; [\mu_s - 3\sigma_s; \mu_s + 3\sigma_s]$$

Aproximadamente del 68.7%, 95,45% y 99,73% respectivamente.

**Figura 3.5.** Intervalos de confianza



### 3.14. CÁLCULO DE LOS COEFICIENTES DE CONFIANZA.

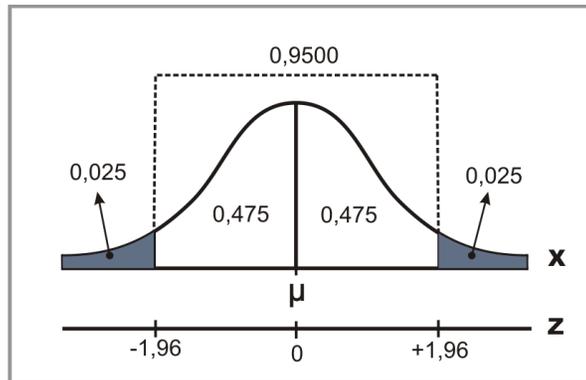
$S \pm 1.96\sigma_s$  son los límites de confianza 95% para S. El porcentaje de confianza suele llamarse nivel de confianza. Los números 1.96, 2.58, etc. en los límites de confianza se llaman **coeficientes de confianza o valores críticos** y se denota por Z. De los niveles de confianza podemos deducir los coeficientes de confianza y viceversa.

Para el calcular el Coeficiente de confianza para un nivel de confianza del 95%, haga el siguiente procedimiento:

- $1 - 0,95 = 0,0500$
- $0,0500 / 2 = 0,0250$

(Ésta operación se realiza por la tabla de distribución de frecuencia utilizada). Se busca en el cuerpo de la tabla apéndice B.

**Figura 3.6.** Distribución normal



En la Tabla 2 se dan algunos cálculos de los coeficientes de confianza más utilizados, si el estudiante requiere de algún coeficiente que no este especificado en la tabla, lo puede calcular haciendo uso de la tabla de distribución normal (Apéndice B) y de los procedimientos dados en la sección anterior.

**Tabla 3.2.** Niveles de confianza.

Nivel de confianza	99.73%	99%	98%	96%	95.45%	95%	90%	80%	68.27%	50%
Z	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.6745

### 3.13. TEORÍA ESTADÍSTICA DE LAS DECISIONES

A menudo no vemos obligados a tomar decisiones relativas a una población sobre la base de la información provenientes de muestras. Tales decisiones se llaman decisiones estadísticas. Por ejemplo, deseamos decidir basados en datos muestrales que producir el producto A es mejor que producir el producto B, o si la aceptación del producto C es lo suficientemente alta para invertir en determinado proyecto.

Al tomar una decisión, es útil hacer conjeturas sobre la hipótesis implicada. Las hipótesis estadísticas es un enunciado acerca de las distribuciones de probabilidad de las poblaciones.

**3.13.1. Hipótesis nula.** En muchos casos formulamos hipótesis estadística con el único propósito de rechazarla o invalidarla. Así, si queremos decidir si una producto es aceptado, formulamos la hipótesis de que este tiene aceptación (o sea  $p=0,5$ , donde  $p$  es la probabilidad de que definitivamente o probablemente lo compren). Este enunciado o conjetura se denomina hipótesis nula y se denotará por  $H_0$ .

**3.13.2. Hipótesis alternativa.** Toda hipótesis que difiera de una dada (hipótesis nula  $H_0$ ) se llamará hipótesis alternativa, si una hipótesis es  $p=0,5$ , la hipótesis alternativa podría ser  $p=0,7$ , es decir un  $p \neq 0,5$  o  $p > 0,5$ . Una hipótesis alternativa a la hipótesis nula se denotará por  $H_1$ .

**3.13.3. Contrastes de hipótesis y significación, o reglas de decisión.** Supongamos que vemos que los resultados hallados en una muestra aleatoria difieren notablemente de los esperados bajo la hipótesis (o sea, esperado sobre la base del azar, por teoría de muestreo), entonces diremos que las diferencias observadas son significativas y tendríamos que rechazar la hipótesis (o al menos no aceptarla ante la evidencia obtenida).

Los procedimientos que nos capacitan para determinar si las muestras observadas difieren significativamente de los resultados esperados, y por tanto nos ayudan a decidir si aceptamos o rechazamos las hipótesis planteadas, se llama contrastes (o tests) de hipótesis o de significación o reglas de decisión.

**3.13.4. Errores de tipo I y de tipo II.** Si rechazamos una hipótesis cuando debe ser aceptada, diremos que se ha cometido un error de Tipo I. De otra parte, si aceptamos una hipótesis que debe ser rechazada, diremos que se ha cometido un error de tipo II. Es decir, que ambos casos se ha producido un juicio erróneo.

**Tabla 3.3.** Tipos de error

	Aceptada	Rechazada
Hipótesis verdadera	Juicio acertado	Error de tipo I
Hipótesis falsa	Error de tipo II	Juicio acerado

Las reglas de decisión (o contraste de hipótesis) deben diseñarse de modo que minimicen los errores de la decisión. Esto no es una cuestión fácil, porque para cualquier tamaño de la muestra, un intento de disminuir un tipo de error suele ir acompañado de un crecimiento del otro tipo. La única forma de disminuir ambos es aumentar el tamaño de la muestra.

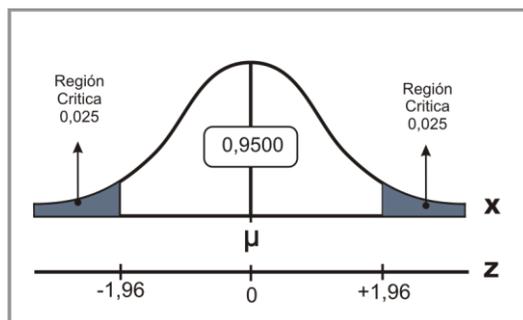
**3.13.5. Nivel de significación.** Al contrastar una hipótesis, la máxima probabilidad con la que estamos corriendo el riesgo de cometer un error de tipo I se llama nivel de significación del contraste. Esta probabilidad, denotada a menudo por  $\alpha$ , se suele especificar antes de tomar la muestra.

En la práctica, es frecuente un nivel de significación de 0,05 ó 0.01, si bien se usan otros valores. Si, por ejemplo, se escoge el nivel de significación 0,05 (o 5%) al diseñar una regla de decisión, entonces diremos que entre 100 muestras hay 5 oportunidades de rechazar la hipótesis, es decir, tenemos un 95% de confianza de que hemos tomado una decisión acertada. Dicho de otra forma, quiere decir, que la hipótesis tiene una probabilidad de 0.05 de ser falsa.

**3.13.6. Contraste mediante la distribución normal.** Para ilustrar los conceptos dados, supongamos que bajo cierta hipótesis la distribución de muestreo de un estadístico  $S$  es una distribución normal con media  $\mu$  y desviación típica  $\sigma$ , entonces la distribución de la variable tipificada  $z$ , estandarizada está dada por:

$$Z = \frac{S - \mu}{\sigma} \text{ donde } (0,1) \mu = 0 \text{ y } \sigma^2 = 1$$

**Figura 3.7.** Contraste de hipótesis



Como se ve en la Figura 3.7, se puede tener 95% de confianza de que si la hipótesis es verdadera, entonces el valor de  $z$  para un estadístico muestral  $S$  estará entre  $-1,96$  y  $1,96$  (porque el área bajo la curva normal entre esos valores es  $0,95$ ). Si al escoger una sola muestra al azar hallamos que el valor de  $z$  de sus estadístico está fuera de ese rango, debemos concluir que tal suceso puede ocurrir con una probabilidad de  $0,05$  (el área sombreada en la figura 3.7) si la hipótesis dada fuera cierta, entonces se puede afirmar que este  $z$  difiere de forma significativa de lo que se espera bajo la hipótesis y tendríamos que rechazar la hipótesis.

El área total sombreada  $0,05$  es el nivel de significación del contraste. Representa la probabilidad de equivocarnos al rechazar la hipótesis (o sea la probabilidad de cometer un error de tipo I).

El conjunto de  $z$  fuera del rango  $-1,96$  a  $1,96$  se llama región crítica de la hipótesis región de rechazo de la hipótesis, o región de significación. El conjunto de  $z$  en el rango  $-1,96$  a  $1,96$  se conoce como región de aceptación de la hipótesis o región de no significación.

Basado en lo anterior, se puede formular la siguiente regla de decisión o contrastes de hipótesis o significación:

Rechazar la hipótesis al nivel de significación 0,05 si el valor de z para el estadístico S está fuera del rango -1.96 a 1.96. Esto equivale a decir que el estadístico muestral observado es significativo al nivel 0,05.

**3.13.7. Curva de operación características.** Potencia de un contraste. Es posible evitar el riesgo de cometer un error de Tipo II, recurriendo a las curvas de operación características, o curvas OC, que son gráficos que muestran las probabilidades de error de Tipo II bajo diversas hipótesis, es decir, nos indica la potencia de un test a la hora de prevenir decisiones erróneas. Son útiles en el diseño de muestreo porque sugieren entre otras cosas el tamaño de muestra a calcular.

Mediante un ejemplo vamos a explicar la elaboración de estos gráficos.

### Ejemplo 3<sup>11</sup>

Para contrastar la hipótesis de que una moneda es buena, adoptamos la siguiente regla de decisión:

H<sub>0</sub>: Aceptarla si el número de caras en una sola muestra de 100 tiradas está entre 40 y 60 inclusive.

H<sub>1</sub>: Rechazar en caso contrario.

Como  $Np=100(1/2)$  y  $Nq(1/2)$ , la aproximación normal a la distribución binomial es correcta a la hora de evaluar la suma. La media y la desviación típica de número de caras en 100 tiradas son

$$\mu = Np = 100 \left[ \frac{1}{2} \right] = 50$$

$$\sigma = \sqrt{Npq} = \sqrt{(100)(1/2)(1/2)} = 5$$

En una escala continua, decir entre 40 y 60 inclusive es como decir entre 39,5 a 60,5 caras, luego:

$$39,5 \text{ en unidades estándar} = Z = \frac{S - \mu}{\sigma} = \frac{39,5 - 50}{5} = -2,10$$

$$60,5 \text{ en unidades estándar} = Z = \frac{S - \mu}{\sigma} = \frac{60,5 - 50}{5} = 2,10$$

La probabilidad es igual al área bajo la curva normal entre  $Z=-2,1$  y  $Z=2,1$

---

<sup>11</sup> Murray R. Spiegel. Estadística de Schaum. Segunda edición. Edit McGrawHill. pagina 235.

$2(\text{área entre } z=0 \text{ y } z=2)$  (Ver Apéndice B. Tabla de distribución normal)  
 $= 2(0,4821)=0.9642$

b) La probabilidad de no obtener entre 40 y 60 caras inclusive si la moneda es buena, es  $1-0.9642 = 0,0358$ . Luego la probabilidad de rechazar la hipótesis cuando sea correcta es 0.0358.

En la tabla 6.1 muestra los valores correspondientes a valores dados de p, es decir la probabilidad de aceptar la hipótesis y rechazarla, como se puede observar el punto ideal es cuando  $p=0,5$ .

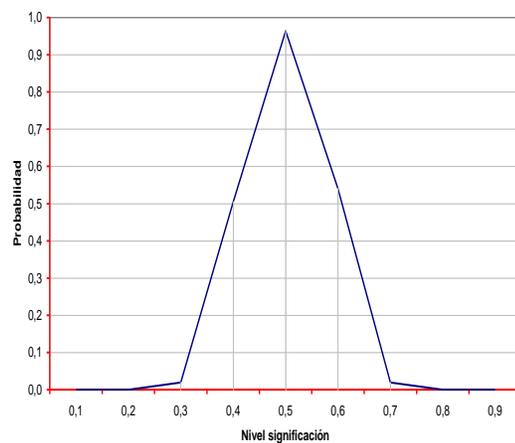
**Tabla 3.4.** Probabilidades de cometer un error de tipo I y II

p	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
Prob.	0,0000	0,0000	0,0192	0,5040	0,9642	0,5400	0,0192	0,0000	0,0000
1-Prob	1,0000	1,0000	0,9808	0,4960	0,0358	0,4600	0,9808	1,0000	1,0000

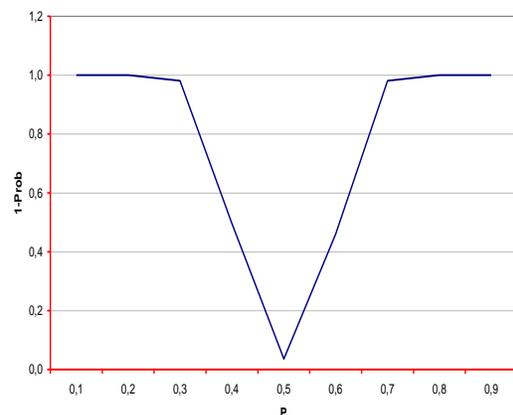
Graficando los valores de la tabla 3.4 se obtienen las figuras 3.8 (a) y 3.8 (b).

**Figura 3.8.**

(a) Curva de operación característica  
Curva OC



(b) Curva de Potencia de la regla de decisión



La grafica 3.8. (a) se llama la curva de operación característica, o curva OC, de la regla de decisión (o contraste de hipótesis), en general lo que se observa es que cuanto más agudo es el pico de la curva OC, mejor es la regla de decisión a la hora de rechazar hipótesis incorrectas.

La Figura 3.8 (b) se llama la curva de potencia de la regla de decisión se obtiene sin más que invertir la curva OC, luego ambas Figuras son equivalentes, esta curva indica la potencia de un test (o contraste) para rechazar hipótesis falsas.

### 3.14. TAMAÑO OPTIMO EN POBLACIONES FINITAS.

Partiendo de la ecuación (9)

$$Z = \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right) \sqrt{\frac{N-n}{N}}}$$

Se considera la diferencia entre  $\bar{X} - \mu$ , como el error máximo permitido en la muestra y se reemplaza por E, hecho este cambio se despeja la ecuación así:

$$Z = \frac{E}{\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N}}} ; E = \frac{Z\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N}} = E^2 = \left(\frac{Z^2\sigma^2}{n}\right) \left(\frac{N-n}{N}\right)$$

$$E^2 = \frac{Z^2\sigma^2N - Z^2\sigma^2n}{Nn}$$

$$E^2Nn = Z^2\sigma^2N - Z^2\sigma^2n$$

$$E^2Nn + Z^2\sigma^2n = Z^2\sigma^2N$$

$$n(NE^2 + Z^2\sigma^2) = Z^2\sigma^2N$$

$$n = \frac{Z^2N\sigma^2}{NE^2 + Z^2\sigma^2}$$

La formula que se utilizará para el cálculo de la muestra es:

$$n = \frac{Z^2N\sigma^2}{NE^2 + Z^2\sigma^2}$$

Realizando algunas manipulaciones algebraicas tenemos las siguientes formulas que son muy utilizadas también:

$$n = \frac{\sigma^2}{\left(\frac{E}{Z}\right)^2 + \frac{\sigma^2}{N}} \tag{16}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \text{ donde } n_0 = \frac{Z^2\sigma^2}{E^2} = \left(\frac{Z\sigma}{E}\right)^2$$

### 3.14.1. Formulas utilizadas con variables discretas (atributos).

$$n = \frac{Z^2 NPQ}{(N-1)E^2 + Z^2 PQ} \quad \text{ó} \quad n = \frac{PQ}{\left(\frac{E}{Z}\right)^2 + \frac{PQ}{N}}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad \text{Siendo} \quad n_0 = \frac{Z^2 PQ}{E^2} \quad (17)$$

### 3.15. MUESTREO ALEATORIO SIMPLE (M.A.S.).

El muestreo aleatorio simple sin reposición es un procedimiento de selección de muestras con probabilidades iguales, que consiste en obtener la muestra de forma aleatoria sin reposición a la población de las unidades previamente seleccionadas, teniendo presente que el orden de colocación de los elementos en las muestras no intervienen (es decir, que muestras con los mismos elementos colocados en orden distinto se consideran iguales). De esta forma, las muestras con elementos repetidos son imposibles. Como el procedimiento de selección es con probabilidades iguales, todas las muestras son equiprobables, y además se cumple que todas las unidades de la población tienen la misma probabilidad de pertenecer a la muestra  $\pi_i = n/N$ .

**3.15.1. Cálculo del tamaño de la muestra.** Para diseñar la muestra es indispensable contar con el marco muestral, es decir, la lista, mapa o otra especificación de las unidades que resulta de la información previamente disponible, respecto a la población sobre la cual se basan los esquemas particulares de muestreo. Para el ejemplo que se realizará el marco muestral lo constituye 50 tiendas (Tabla 3.5), cuya característica es que los establecimientos tienen activos mayores o iguales a \$1'000.000.

Cuando no se conoce la varianza poblacional de una o algunas de las características que tienen que ver con el objetivo de la investigación, se procede a tomar una muestra piloto, para estimar la muestra piloto no hay normas ni reglas específicas, esta es una decisión del investigador, en algunos casos depende del tamaño de la población, tiempo y costos. Para el ejercicio tomaremos 10 tiendas del marco muestral (Tabla 3.5), haciendo uso de una tabla de números aleatorios o de la generación de los números aleatorios de una calculadora. Sin tomar números superiores a 50 (total de la población) o repetidos se construyó la Tabla 3.6.

**Tabla 3.5.** Tiendas de la Ciudad de Popayán, discriminadas por el valor en activos, ingresos y propiedad del local.

# de ord.	Nombre o Razon Social	Direccion	Activos en millones de pesos	Ingresos en millones de pesos	Local propio
1	ANACONA AMELIA	-CARRERA 7E # 17BIS-17	1,000	13,888	no
2	ANACONA DE ANACONA LUCINDA	-CALLE 19 # 30-24	1,170	12,075	no
3	ANACONA PIAMBA LIOVIGILDO	-CALLE 63N NO. 7A-09	1,414	3,000	si
4	ANDRADE DASA WILLIAM JULIAN	-CALLE 68 # 10-92	1,150	0,450	no
5	ARIAS DE GAMBOA GRACIELA	-CARRERA 1AE # 9A-41	1,050	9,600	si
6	ARANGO MARULANDA JENNY LILIANA	-CARRERA 6 # 43N-50	7,350	11,550	si
7	BELTRAN PENA TITO GERARDINO	-CARRERA 11 # 12A-10	8,795	2,450	si
8	CABRERA SALAS HUGO HERNANDO	-CARRERA 6 # 12-51	10,500	521,794	no
9	CABRERA SALAS ONASIS ERNESTO	-CALLE 5 No 18-50 POPAYAN	7,200	117,197	si
10	AUSECHA ROJAS MARIA CONSUELO	-TRANSV. 19 # 10-121	1,000	7,518	si
11	BARBOSA LUZ DARY	-MANZANA 3 NO. 42A-11	1,200	1,200	no
12	BASTIDAS AVILA IRMA SIRLEY	-CALLE 29 BLOQUE H CASA No 5	1,500	4,900	si
13	BENAVIDES RODRIGUEZ BLANCA BELLANIRE	-CALLE 5B # 18-31	1,080	13,311	si
14	BENITEZ GUERRERO ENOELIO	-CARRERA 41 # 2-13	1,200	0,400	no
15	BERRIO BUITRAGO FABIO ELIAS	-CALLE 12 # 28A-04	1,120	14,300	no
16	BETANCOURT MOLANO GILMA	-MANZ. 25 #25-16 TOMAS CIPRIAN	1,270	11,270	no
17	BOLANOS BELALCAZAR LUIS EDUARDO	-CALLE 5 # 43-24	1,473	1,573	si
18	BOLANOS DE MARTINEZ ROSAURA	-GALERIA SUR PUESTO # 12	1,100	11,910	no
19	ACOSTA DE CERTUCHE PAULINA DUPERLY	-CALLE 7 # 19-114	1,900	11,900	no
20	ACOSTA DE PARAMO MATILDE	-CARRERA 2A No.7A-40	2,185	28,652	no
21	ACOSTA VILLEGAS ARNULFO	-CALLE 4 # 36-11	1,690	5,800	si
22	ACUNA REY OSCAR DARLEY	CRA 12 66N-72 BELLOHORIZONTE	2,300	44,308	si
23	AGREDO SOLIS RAUL	-CARRERA 41 # 4-11	1,547	2,160	si
24	ALVEAR DE SOLARTE MARIA MAGDALENA	-CALLE 12 13-03	2,200	6,938	no
25	AREVALO FIGUEROA JUAN BAUTISTA	-CARRERA 3 # 8-07	1,650	19,187	no
26	ARTUNDUAGA MOSQUERA EDGAR JESUS	-CALLE 7 # 21-62	2,140	0,945	no
27	ASTUDILLO JUSPIAN ADAN	-CALLE 9 # 17-25	2,800	15,500	si
28	AVILES SILVA OMAR	-CARRERA 9 # 7N-02	2,343	7,819	si
29	BARRERA DE CERON ISMENIA - SUCESTORES -	-CALLE 12 # 4-93	1,710	1,030	si
30	BOLANOS MORALES ROLANDO	-CARRERA 9 No 7-99	1,650	13,500	si
31	ACOSTA DIAZ NUMAR ARNEY	CALLE 17 # 6E-19 B/ LOS SAUCES.	4,600	10,500	si
32	AGREDO DE VASQUEZ MARTHA CLELIA	-CRA 3 # 9-84	3,300	16,720	si
33	ALEGRIA DE CALVACHE EMMA CLEMENCIA	-CARRERA 2 # 3-93	5,550	14,590	si
34	ASTORQUIZA ORDONEZ FABIO	-CALLE 5B # 18- 17	6,500	13,005	si
35	BEJARANO DINA ISABEL	-CALLE 20N # 8-47 POPAYAN	5,000	80,000	si
36	BENITEZ GUERRERO GELVI VIRGILIO	-CARRERA 41 # 2-13	3,600	24,166	si
37	CABRERA LARA JUAN CARLOS	PASAJE CENTRO COMERCIAL LOCAL 46	3,750	21,676	si
38	CALVACHE ROJAS LUIS ALBERTO	-CALLE 9 # 17-55	4,000	10,800	si
39	CEBALLOS CANO SANDRA LORENA	-CALLE 56N # 10-110	4,115	9,295	si
40	CERTUCHE BARRERA LILIA NIRA	-CALLE 70D # 7-15	6,920	31,650	si
41	ARTUNDUAGA ROJAS LUIS ALBERTO	-CALLE 5 # 27A-12	1,175	2,450	no
42	ASTAIZA DE PIEDRAHITA GEORGINA	-CALLE 8 # 5-28	1,433	14,900	si
43	ASTAIZA DE VIANA LUCILA	-CALLE 4 # 25-73	1,150	11,450	no
44	ASTUDILLO DE ANACONA IRMA	-CALLE 16A # 4-51	1,500	1,500	si
45	AGREDO CIFUENTES ANA LUCIA	-CARRERA 4 # 7-35	29,469	747,544	si
46	ALVAREZ LOPEZ LTDA	-CALLE 11N # 9-44	14,384	75,990	no
47	ANACONA ANACONA LUIS ANGEL	-CALLE 19 No 31-14	10,000	30,641	si
48	ANDRADE DIAZ MARIA ELENA	-CARRERA 9 # 63N-18	10,028	35,381	si
49	CASTANEDA CASTANEDA MARIA NIDIA	-CARRERA 7 # 12 -106	18,000	58,535	no
50	CASTANO CAMPO CLAUDIA MARCELA	-CARRERA 6A # 9N-92 B/BOLIVAR	23,178	111,957	no

Fuente: Cámara de Comercio del Cauca

**Tabla 3.6.** Muestra Piloto, muestreo aleatorio simple (M.A.S.)

# Aleatorio	048	033	024	003	018	035	039	041	044	049
Activos	10,028	5,550	2,200	1,414	1,100	5,000	4,115	1,175	1,500	18,000
Local	si	si	no	si	no	si	si	no	si	no

Se procede a calcular la varianza para la variable continua (Activos) de la siguiente forma:

**Tabla 3.7.** Calculo de la varianza, para variables continuas

No. De orden	Números aleatorios	Local propio	Activos en miles de pesos	$(X - \bar{X})^2$
1	48	si	10,028	25,198
2	33	si	5,550	0,294
3	24	no	2,200	7,886
4	3	si	1,414	12,918
5	18	no	1,100	15,274
6	35	si	5,000	0,000
7	39	si	4,115	0,798
8	41	no	1,175	14,693
9	44	si	1,500	12,307
10	49	no	18,000	168,787
Totales			50,082	258,156
			Media	Varianza
Media y Varianza			5,008	28,684

$$\bar{X} = \frac{\sum X_i}{n} = \frac{50.082}{10} = 5.008$$

Se estima un error del 10% del promedio de los activos así:

$$E = 0,10(\bar{X}) = 0,10 \times 5.008 = 0.5008$$

$$\text{Varianza muestral } S^2 = \frac{(X - \bar{X})^2}{n-1} = \frac{258.156}{9} = 28.684$$

Una vez calculados estos estadísticos se procede a calcular el tamaño de la muestra. Se trabajará con un nivel de confianza del 90%, es decir un valor de  $z=1,645$ .

$$n = \frac{Z^2 \times N \times S^2}{N \times E^2 + Z^2 \times S^2} = \frac{(1,645)^2 \times 50 \times 28.684}{50 \times (0,5008)^2 + (1,645)^2 \times 28.684} = 43,045 \approx 43$$

$$n = \frac{S^2}{\left(\frac{E}{Z}\right)^2 + \frac{S^2}{N}} = \frac{28.684}{\left(\frac{0.5008}{1.645}\right)^2 + \frac{28.684}{50}} = 43,045 \approx 43$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} \quad \text{donde} \quad n_0 = \frac{Z^2 S^2}{E^2} = \frac{\frac{Z^2 S^2}{E^2}}{1 + \frac{E^2}{N}} = \frac{\frac{(1,645)^2 * 28.684}{(0.5008)^2}}{1 + \frac{(0.5008)^2}{50}} = 43,045 \approx 43$$

Se observa más claramente en esta última fórmula que cuando  $N \rightarrow \infty$  (la fracción de muestreo  $n/N$  tiende a cero) el tamaño muestral  $n \rightarrow \frac{Z^2 S^2}{E^2} = n_0$  (n es inversamente proporcional al cuadrado del error de muestreo).

La expresión del tamaño muestral n puede ponerse en función de N y del valor  $n_0$  como se aprecia en la ecuación 18.

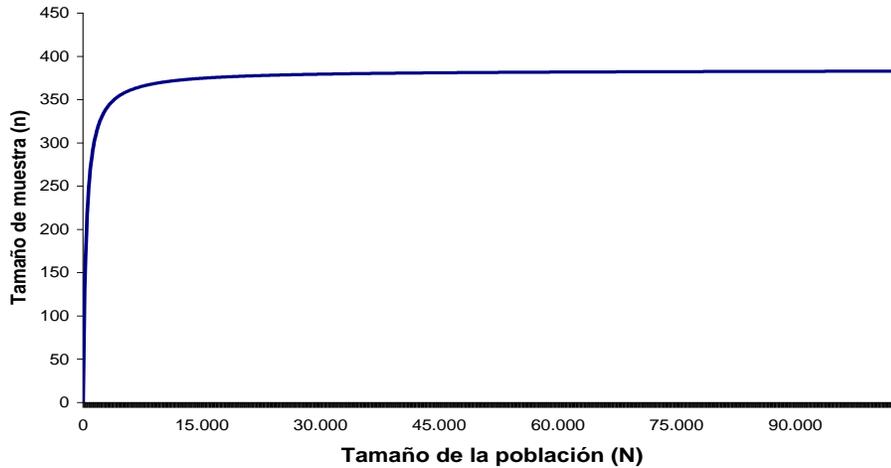
$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{n_0 N}{n_0 + N} = f(N) \tag{18}$$

Si representamos gráficamente la curva de la ecuación  $n = f(N)$  observamos que pasa por el origen de las coordenadas, ya que  $f(0) = 0$ , que tiene una asíntota paralela al eje de las X de la ecuación  $n = n_0$ , dado que el  $\lim_{N \rightarrow \infty} f(N) = n_0$ , que es siempre creciente puesto que la primera derivada:

$$f'(N) = \frac{n_0^2}{(n_0 + N)^2} \tag{19}$$

Es siempre positiva, que no tiene máximos ni mínimos dado que la ecuación definida por  $f'(N) = 0$  no tiene solución en N. Por tanto, la representación gráfica de  $n = f(N)$  es la siguiente:

**Figura 3.9.** Representación gráfica de la curva de la ecuación  $n = f(N)$



Como la curva  $n = f(N)$  es creciente, al aumentar el tamaño poblacional N también aumenta el tamaño muestral n necesario para un error de muestreo dado. Pero como n ha de ser un número entero y la curva  $n = n_0$  es una asíntota horizontal, desde un cierto N en adelante los aumentos de N no producen aumentos en n. (Apéndice C).

### FICHA TÉCNICA

Una vez calculado el tamaño de la muestra, se requiere hacer la ficha técnica donde se relaciona en forma resumida el plan de muestreo utilizado así:

## Ficha Técnica

1. Título	: Encuesta a tenderos
2. Responsables	: Estudiantes de administración de empresas de la Universidad del Cauca.
3. Alcance	: Ciudad de Popayán.
4. Unidad de análisis	: Tiendas con activos mayores a \$1.000.000.
5. Periodo	: Septiembre de 2003.
6. PLAN DE MUESTREO	
6.1. Población objetivo	: Tiendas ubicadas en la ciudad de Popayán
6.2. Marco	: Listado suministrado por la Cámara de Comercio del Cauca.
6.3. Unidades de muestreo	: Propietarios o administradores de las tiendas con patrimonio mayor o igual a \$ 1.000.000.
6.4. Nivel de confianza	: 90%
6.5. Coeficiente de conf.	: 1,645
6.6. Varianza muestral	: 28.684
6.7. Error máximo	: 5% sobre la media (0,5008)
6.8. Tipo de diseño	: Muestra probabilísticas.
6.9. Procedimiento	: Muestreo aleatorio simple (M.A.S.)
6.10. Variables utilizadas	: Promedio de ventas, tipo de productos vendidos.
6.11. Tipo de entrevista	: Entrevista directa con formulario en papel.
6.12. Tamaño poblacional	: 50

**Tamaño de la muestra = 43**

### 3.15.2. Tamaño de la muestra con la variable discreta (cualitativa) de proporción de tiendas con local propio

$$p = \frac{\sum a_i}{n} = \frac{6}{10} = 0,6$$

$$pq = 0,6 \times 0,4 = 0,24$$

$$n = \frac{Z^2 \times N \times p \times q}{N \times E^2 + Z^2 \times p \times q} = \frac{(1,645)^2 \times 50 \times 0,6 \times 0,40}{50 \times (0,10)^2 + (1,645)^2 \times 0,6 \times 0,40} = 28,25 \approx 28$$

## Ficha Técnica

1. Título	: Encuesta a tenderos
2. Responsables	: Estudiantes de administración de empresas de la Universidad del Cauca.
3. Alcance	: Ciudad de Popayán.
4. Unidad de análisis	: Tiendas con activos mayores a \$1.000.000.
5. Periodo	: Septiembre de 2003.
6. PLAN DE MUESTREO	
6.1. Población objetivo	: Tiendas ubicadas en la ciudad de Popayán
6.2. Marco	: Listado suministrado por la Cámara de Comercio del Cauca.
6.3. Unidades de muestreo	: Propietarios o administradores de las tiendas con patrimonio mayor o igual a \$ 1.000.000.
6.4. Nivel de confianza	: 90%
6.5. Coeficiente de conf.	: 1,645
6.6. Varianza muestral	: $P=0,6 - Q=0,40$
6.7. Error máximo	: 10%
6.8. Tipo de diseño	: Muestra probabilísticas.
6.9. Procedimiento	: Muestreo aleatorio simple (M.A.S.)
6.10. Variables utilizadas	: Promedio de ventas, tipo de productos vendidos.
6.11. Tipo de entrevista	: Entrevista directa con formulario en papel.
6.12. Tamaño poblacional	: 50

**Tamaño de la muestra = 28**

### 3.16. TAMAÑO OPTIMO DE MUESTRA CON VARIABLES DISCRETAS.

Teniendo en cuenta el punto 3.13 las curvas de operación características, se procede a calcular el tamaño de la muestra con diferentes valores de p así:

Estadísticos de la ficha técnica:

N = 50

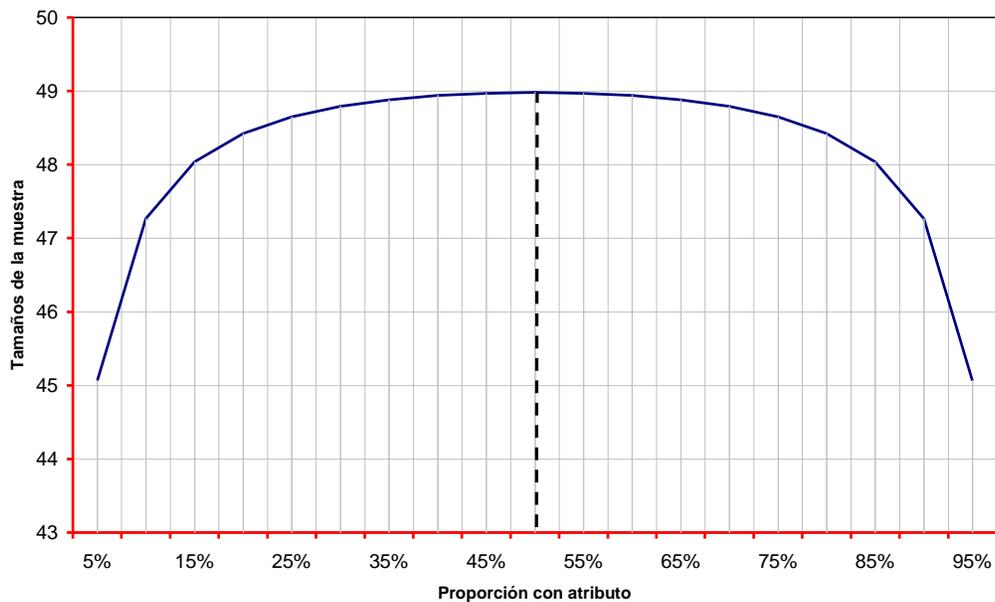
E= 2%

Z= 1,96

**Tabla 3.8.** Tamaño de la muestra, con diferentes valores de p.

p	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%
q	95%	90%	85%	80%	75%	70%	65%	60%	55%	50%	45%	40%	35%	30%	25%	20%	15%	10%	5%
n=	45	47	48	48	49	49	49	49	48,97	48,98	48,97	49	49	49	49	48	48	47	45

**Figura 3.10.** Tamaño de la muestra con diferentes proporciones de (p)



Fuente: Tabla No. 3.8. Tamaño de la muestra, con diferentes valores de p.

Como se aprecia en la figura 3.10 el tamaño de la muestra es óptimo cuando se toma las proporciones equivalentes a  $p=0,5$  y  $q=0,5$ <sup>12</sup>

### 3.17. HOMOGENEIDAD DE LA INFORMACIÓN.

Cuando hay cierto grado de homogeneidad en la característica investigada, el tamaño tiende a ser pequeño. El grado de homogeneidad se da cuando el coeficiente de variación es menor del 30%; en estos casos es recomendable la aplicación del muestreo aleatorio simple (m.a.s.), a continuación se calcula el coeficiente de variación.

$$S = \sqrt{S^2} = \sqrt{28.684} = 5,36 \quad (20)$$

$$\text{Coeficiente de Variación } CV = \frac{S}{\bar{X}} \times 100 = \frac{5,36}{5,008} \times 100 \cong 106,94\% \quad (21)$$

En este caso el coeficiente es mayor a 30% por tanto no es recomendable aplicar el muestreo aleatorio simple.

### 3.18. RELACIÓN DEL TAMAÑO DE LA MUESTRA CON EL NIVEL DE CONFIANZA Y EL ERROR.

En la tabla 3.9 se calculó el tamaño de la muestra dado los niveles de confianza del 96%, 95% y 80%, para errores del 4%, 5% y el 20% sobre el valor de la media ( $\bar{X} = 5$ ) y  $S^2 = 28,684$

**Tabla 3.9.** Cálculo del tamaño de la muestra

  
 Un mayor nivel de confianza menor tamaño de la muestra

N=50		NIVEL DE CONFIANZA		
		96%	95%	80%
		Z=2,05	Z=1,96	Z=1,28
ERROR	E=4%*5=0,2	49	49	48
	E=5%*5=0,25	48	48	47
	E=20%*5=1,0	35	33	23

↑ Un mayor error menor tamaño de la muestra

<sup>12</sup> Martínez Bernardina, expresa al respecto que "...se tiene la costumbre de tomar  $P=0,5$  con lo cual se obtiene el máximo valor posible de  $n$ ". Estadística y muestreo. 2002. p. 349.

Se concluye que un mayor nivel de confianza por ejemplo 96% y un error bajo en éste caso del 4% el tamaño de la muestra es de 49, que es casi el tamaño de la población; caso contrario ocurre con un nivel de confianza bajo como 80% y un error alto como del 20%, el tamaño de la muestra es de 23, es decir que la muestra se reduce considerablemente, lo cual puede traer como consecuencia un estudio o investigación poco confiable.

### **3.19. MUESTREO ESTRATIFICADO.**

Este es un método que permite una selección más eficiente que el obtenido mediante el Muestreo Aleatorio Simple (M.A.S), en especial cuando la característica que se investiga es de gran variabilidad, lo cual implica un tamaño muestral relativamente más grande.

Una muestra estratificada se selecciona de la siguiente manera:

Se divide la población en grupos o estratos mutuamente excluyentes y colectivamente exhaustivos. Los estratos son mutuamente excluyentes, sólo si los miembros de un estrato no pueden ser miembros de cualquier otro. Los estratos son colectivamente exhaustivos si se utilizan todas las categorías posibles de una variable para definirlos. Es decir, las categorías “Masculino” y “Femenino” definen el área completa de la variable “sexo”. Ninguna otra categoría es posible y, por tanto, los estratos son colectivamente exhaustivos.

Dependiendo de la manera como se distribuyen los elementos dentro de los estratos muestrales, hay tres métodos a saber:

1. Asignación o fijación igual
2. Asignación o fijación proporcional
3. Asignación o fijación óptima

Se trabajará con la asignación proporcional, por ser este método uno de los más utilizados en las investigaciones de mercado.

**3.19.1. Muestreo estratificado - asignación proporcional.** Este método permite determinar el tamaño óptimo de la muestra, así como los estimados puntuales y límites de confianza para el promedio, proporción, razón y proporciones en conglomerados. Con este método los tamaños muestrales en cada estrato se distribuyen en la misma proporción que las unidades en la población de cada uno de ellos; en otras palabras, el peso relativo dado por el número de unidades en cada estrato, en relación al total de elementos de la población, debe ser igual al obtenido en la muestra.

Antes de determinar el tamaño de la muestra, se elaboró la estratificación para la población de 50 tiendas. Se distribuyó la población de tiendas en dos estratos así:

**Tabla 3.10. Estrato I:** Tiendas con Activos menores e iguales a \$2.000.000

# de ord.	Nombre o Razon Social	Direccion	Activos en millones de pesos	Ingresos en millones de pesos	Local propio
1	ANACONA AMELIA	-CARRERA 7E # 17BIS-17	1,000	13,888	no
2	AUSECHA ROJAS MARIA CONSUELO	-TRANSV. 19 # 10-121	1,000	7,518	no
3	ARIAS DE GAMBOA GRACIELA	-CARRERA 1AE # 9A-41	1,050	9,600	si
4	BENAVIDES RODRIGUEZ BLANCA BELLANIRE	-CALLE 5B # 18-31	1,080	13,311	si
5	BOLANOS DE MARTINEZ ROSAURA	-GALERIA SUR PUESTO # 12	1,100	11,910	si
6	BERRIO BUITRAGO FABIO ELIAS	-CALLE 12 # 28A-04	1,120	14,300	si
7	ANDRADE DASA WILLIAM JULIAN	-CALLE 68 # 10-92	1,150	0,450	si
8	ASTAIZA DE VIANA LUCILA	-CALLE 4 # 25-73	1,150	11,450	si
9	ANACONA DE ANACONA LUCINDA	-CALLE 19 # 30-24	1,170	12,075	si
10	ARTUNDUAGA ROJAS LUIS ALBERTO	-CALLE 5 # 27A-12	1,175	2,450	no
11	BARBOSA LUZ DARY	-MANZANA 3 NO. 42A-11	1,200	1,200	no
12	BENITEZ GUERRERO ENOELIO	-CARRERA 41 # 2-13	1,200	0,400	no
13	BETANCOURT MOLANO GILMA	-MANZ. 25 #25-16 TOMAS CIPRIAN	1,270	11,270	si
14	ANACONA PIAMBA LIOVIGILDO	-CALLE 63N NO. 7A-09	1,414	3,000	no
15	ASTAIZA DE PIEDRAHITA GEORGINA	-CALLE 8 # 5-28	1,433	14,900	si
16	BOLANOS BELALCAZAR LUIS EDUARDO	-CALLE 5 # 43-24	1,473	1,573	no
17	ASTUDILLO DE ANACONA IRMA	-CALLE 16A # 4-51	1,500	1,500	si
18	BASTIDAS AVILA IRMA SIRLEY	-CALLE 29 BLOQUE H CASA No 5	1,500	4,900	si
19	AGREDO SOLIS RAUL	-CARRERA 41 # 4-11	1,547	2,160	no
20	AREVALO FIGUEROA JUAN BAUTISTA	-CARRERA 3 # 8-07	1,650	19,187	si
21	BOLANOS MORALES ROLANDO	-CARRERA 9 No 7-99	1,650	13,500	no
22	ACOSTA VILLEGAS ARNULFO	-CALLE 4 # 36-11	1,690	5,800	no
23	BARRERA DE CERON ISMENIA - SUCESTORES -	-CALLE 12 # 4-93	1,710	1,030	si
24	ACOSTA DE CERTUCHE PAULINA DUPERLY	-CALLE 7 # 19-114	1,900	11,900	no

Fuente: Cámara de Comercio del Cauca

**Tabla 3.11. Estrato II.** Tiendas con activos de más de \$2.000.0001

# de ord.	Nombre o Razon Social	Direccion	Activos en millones de pesos	Ingresos en millones de pesos	Local propio
1	ARTUNDUAGA MOSQUERA EDGAR JESUS	-CALLE 7 # 21-62	2,140	0,945	si
2	ACOSTA DE PARAMO MATILDE	-CARRERA 2A No.7A-40	2,185	28,652	si
3	ALVEAR DE SOLARTE MARIA MAGDALENA	-CALLE 12 13-03	2,200	6,938	si
4	ACUNA REY OSCAR DARLEY	CRA 12 66N-72 BELLOHORIZONTE	2,300	44,308	si
5	AVILES SILVA OMAR	-CARRERA 9 # 7N-02	2,343	7,819	si
6	ASTUDILLO JUSPIAN ADAN	-CALLE 9 # 17-25	2,800	15,500	no
7	AGREDO DE VASQUEZ MARTHA CLELIA	-CRA 3 # 9-84	3,300	16,720	si
8	BENITEZ GUERRERO GELVI VIRGILIO	-CARRERA 41 # 2-13	3,600	24,166	si
9	CABRERA LARA JUAN CARLOS	PASAJE CENTRO COMERCIAL LOCAL 46	3,750	21,676	si
10	CALVACHE ROJAS LUIS ALBERTO	-CALLE 9 # 17-55	4,000	10,800	si
11	CEBALLOS CANO SANDRA LORENA	-CALLE 56N # 10-110	4,115	9,295	si
12	ACOSTA DIAZ NUMAR ARNEY	CALLE 17 # 6E-19 B/ LOS SAUCES.	4,600	10,500	no
13	BEJARANO DINA ISABEL	-CALLE 20N # 8-47 POPAYAN	5,000	80,000	si
14	ALEGRIA DE CALVACHE EMMA CLEMENCIA	-CARRERA 2 # 3-93	5,550	14,590	no
15	ASTORQUIZA ORDONEZ FABIO	-CALLE 5B # 18- 17	6,500	13,005	no
16	CERTUCHE BARRERA LILIA NIRA	-CALLE 70D # 7-15	6,920	31,650	si
17	CABRERA SALAS ONASIS ERNESTO	-CALLE 5 No 18-50 POPAYAN	7,200	117,197	no
18	ARANGO MARULANDA JENNY LILIANA	-CARRERA 6 # 43N-50	7,350	11,550	si
19	BELTRAN PENA TITO GERARDINO	-CARRERA 11 # 12A-10	8,795	2,450	si
20	ANACONA ANACONA LUIS ANGEL	-CALLE 19 No 31-14	10,000	30,641	no
21	ANDRADE DIAZ MARIA ELENA	-CARRERA 9 # 63N-18	10,028	35,381	si
22	CABRERA SALAS HUGO HERNANDO	-CARRERA 6 # 12-51	10,500	521,794	si
23	ALVAREZ LOPEZ LTDA	-CALLE 11N # 9-44	14,384	75,990	no
24	CASTANEDA CASTANEDA MARIA NIDIA	-CARRERA 7 # 12 -106	18,000	58,535	no
25	CASTANO CAMPO CLAUDIA MARCELA	-CARRERA 6A # 9N-92 B/BOLIVAR	23,178	111,957	si
26	AGREDO CIFUENTES ANA LUCIA	-CARRERA 4 # 7-35	29,469	747,544	si

Fuente: Cámara de Comercio del Cauca

Nomenclatura para calcular las proporciones del muestreo estratificado

N	Total de unidades que constituyen la población objetivo.
$N_h$	Total de unidades que contiene cada estrato poblacional
$N_1, N_2, N_3$ etc.	Serán los tamaños poblacionales en los estratos 1, 2, 3 etc.
$\Sigma N_h = N = N_1 + N_2 + N_3 + \dots + N_M$	
$\bar{Y}_h = \frac{\Sigma Y_{hj}}{N_h}$	Media aritmética poblacional para cada estrato
$\bar{Y}_{st} = \frac{\Sigma Y_h N_h}{N} = \Sigma \bar{Y}_h W_h$	Media aritmética poblacional estratificada ponderada
$W_h = \frac{N_h}{N} \quad W_1 = \frac{N_1}{N}$ $W_2 = \frac{N_2}{N} \quad W_3 = \frac{N_3}{N}$	Proporción de elementos de cada estrato
$\Sigma W_h = W_1 + W_2 + W_3 + \dots + W_M = 1$	
$S^2_h = \frac{\Sigma Y^2_{hj} - N_h \bar{Y}_h^2}{N_h - 1}$	Varianza poblacional en cada estrato.
n =	Número de unidades que contiene la muestra total
nh =	Número de unidades que contiene la muestra en cada estrato muestral.
$\Sigma n_h = n = n_1 + n_2 + n_3 + \dots$	
$\bar{y}_h = \frac{\Sigma \bar{y}_{hj}}{n_h} \quad \bar{y}_1 = \frac{\Sigma \bar{y}_{1j}}{n_1}$	$\bar{y}_h =$ Media aritmética muestral para cada estrato.
$\bar{y}_2 = \frac{\Sigma \bar{y}_{2j}}{n_2} \quad \bar{y}_3 = \frac{\Sigma \bar{y}_{3j}}{n_3}$	
$\bar{y}_{st} = \frac{\Sigma N_h \bar{y}_h}{N} \dots \dots \dots \bar{y}_{st} = \Sigma \bar{y}_h W_h$	Media aritmética muestral estratificada.
$s^2_h = \frac{\Sigma y^2_{hj} - n_h \bar{y}_h^2}{n_h - 1}$	Varianza muestral en cada estrato

**3.19.2. Cálculo del tamaño de la muestra.** Se debe calcular la proporción de unidades en cada estrato y peso relativo:

ESTRATO	CRITERIO DE SELECCIÓN	No. Unid.	PROPORCIÓN
ESTRATO I	Activos menores o iguales a \$ 2.000.000	24	$W_1 = \frac{N_1}{N} = \frac{24}{50} = 0,48$
ESTRATO II	Activos mayores o iguales a \$ 2.000.0001.	26	$W_2 = \frac{N_2}{N} = \frac{26}{50} = 0,52$

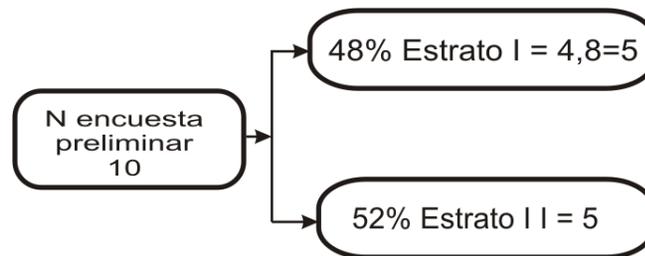
$$W_1 + W_2 = 0,48 + 0,52 = 1$$

$$\sum W_h = 1 \dots \text{ó} \dots 100\%$$

El tamaño de la muestra piloto es 10 encuestas.

En la asignación proporcional las 10 encuestas se distribuyen de la siguiente manera:

**Figura 3.11.** Distribución de número de encuestas por estrato (Variables continuas)



Determinados los tamaños muestrales para la encuesta preliminar, se procede a la selección de las unidades requeridas, que permitirán calcular las varianzas y el error. Haciendo uso de una tabla de números aleatorios para los dos estratos se seleccionó de la siguiente forma:

**Tabla 3.12.** Muestra piloto Estrato I.

No. De orden	Números aleatorios	Local propio	Activos en miles de pesos	$(X - \bar{X})^2$
1	24	no	1,900	0,072
2	20	si	1,650	0,000
3	14	no	1,414	0,047
4	22	no	1,690	0,004
5	17	si	1,500	0,017
Totales			8,154	46,879
			Media	Varianza
Media y Varianza			1,631	11,720

**Tabla 3.13.** Muestra piloto Estrato II.

No. De orden	Números aleatorios	Local propio	Activos en miles de pesos	$(X - \bar{X})^2$
1	24	no	18,000	98,804
2	6	no	2,800	27,668
3	13	si	5,000	9,364
4	10	si	4,000	16,484
5	22	si	10,500	5,954
Totales			40,300	186,956
			Media	Varianza
Media y Varianza			8,060	46,739

La media ponderada es igual a:

$$\bar{x}_{st} = \sum W_h \bar{x}_h = 0,48(1,631) + 0,52(8.060) = 4.974$$

El error de muestreo corresponde a 0,05  $E = 0,05(\bar{x}) = 0,05(4.974) = 0,2487$

Se calcula de la varianza ponderada:

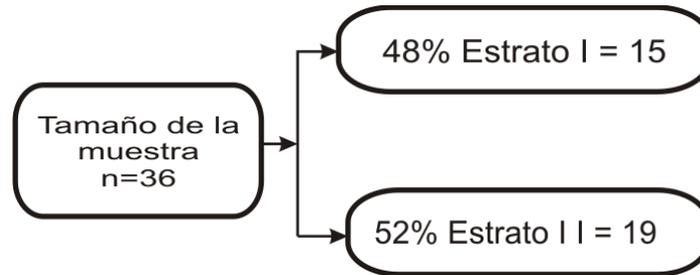
$$S^2_{st} = \sum W_h S^2_h = 0,48(11.72) + 0,52(0,46.739) = 29.93$$

El tamaño de la muestra es igual a:

$$n = \frac{Z^2 N S^2}{NE^2 + Z^2 S^2} = \frac{1,645^2 \times 50 \times 29.93}{(50 \times 0,806^2) + (1,96^2 \times 29.93)} = 35.687 \cong 36$$

Una vez calculado el tamaño de la muestra se procede a distribuir el tamaño de la muestra en forma proporcional a los estratos así:

**Figura 3.12.** Distribución de número de encuestas por estrato  
(Variables discretas)



Ahora veamos como sería el cálculo de  $n$  para variables *discretas o atributos*. Para ello consideramos como característica cualitativa la propiedad del local comercial, se establece un nivel confianza del 95% ( $Z=1,96$ ) y un error del 5%.

**Tabla 3.14.** Calculo de proporciones de la muestra piloto  
(Variables discretas).

ESTRATO I	ESTRATO II
$P = \frac{2}{5} = 0,4$	$P = \frac{3}{5} = 0,6$
$Q = \frac{3}{5} = 0,6$	$Q = \frac{2}{5} = 0,4$
$P + Q = 1$	$P + Q = 1$

Calculamos PQ ponderado así:

$$PQ_{st} = \sum W_h P_h Q_h = 0,48(0,40)(0,60) + 0,52(0,60)(0,40) = 0,24$$

Se procede a calcular el tamaño de la muestra así:

$$n = \frac{Z^2 NPQ}{NE^2 + Z^2 PQ} = \frac{1,96^2 \times 50 \times 0,24}{50 \times 0,05^2 + 1,96^2 \times 0,24} = 44$$



# *Capítulo cuatro*

## *Opciones generales y Plan de Muestreo con SPSS*

---

### **INTRODUCCIÓN**

SPSS para Windows proporciona un poderoso sistema de análisis estadístico y de gestión de datos en un entorno gráfico, utilizando menús descriptivos y cuadros de diálogo sencillos que realizan la mayor parte del trabajo. La mayoría de las tareas se pueden llevar a cabo simplemente situando el puntero del ratón en el lugar deseado y pulsando en el botón.

En este capítulo se suministran las bases generales del manejo del programa, haciendo uso del asistente, se inicia con una breve explicación de: instalación del programa, los formatos de los archivos que genera el programa, las opciones del menú principal y los tipos de ventanas del programa. También se explica la opción Muestras complejas de SPSS (opción presente en el programa a partir de la versión 12). Esta opción permite seleccionar una muestra de acuerdo con un diseño complejo e incorpora las especificaciones del diseño al análisis de los datos para asegurar la validez de los resultados.

### **4. ESTRUCTURA GENERALES DEL PROGRAMA SPSS**

SPSS para Windows es un programa de ordenador que se utiliza para realizar una gran variedad de análisis estadísticos, desde los más sencillos a los más avanzados. La versión del programa SPSS que se describe en este libro es la 15.

El propósito de este libro es dar los lineamientos básicos y no entrar profundizar en los conceptos estadísticos. Este documento está diseñado para enseñar el manejo esta herramienta en una forma rápida y eficiente para el procesamiento de encuestas, las técnicas de muestreo y el análisis de regresión.

Los destinatarios de este texto son los alumnos que están cursando la asignatura de formulación y evaluación de proyectos, planes de negocios, investigación de mercados y

aquellos que estén elaborando su trabajo de grado y que requieran realizar análisis de datos.

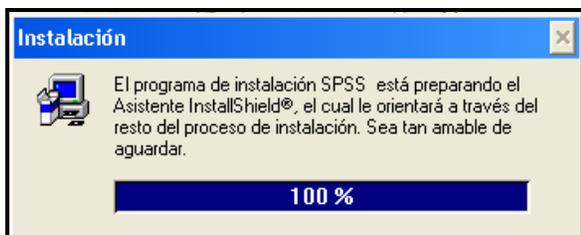
Para instalar el programa, inserte el CD en su computadora, mediante el explorador de Windows, ubique el archivo **Setup** y de doble clic sobre el nombre del archivo. Figura 4.1

**Figura 4.1.** Grupo de programas de SPSS



Inmediatamente se inicia el proceso de instalación. (Figura 4.2).

**Figura 4.2.** Inicio de instalación del programa



Una vez instalado, se crea un grupo de programas (figura 4.3), y automáticamente se ubica en el menú de inicio de Windows (Figura 4.4).

Figura 4.3. Grupo de programas de SPSS

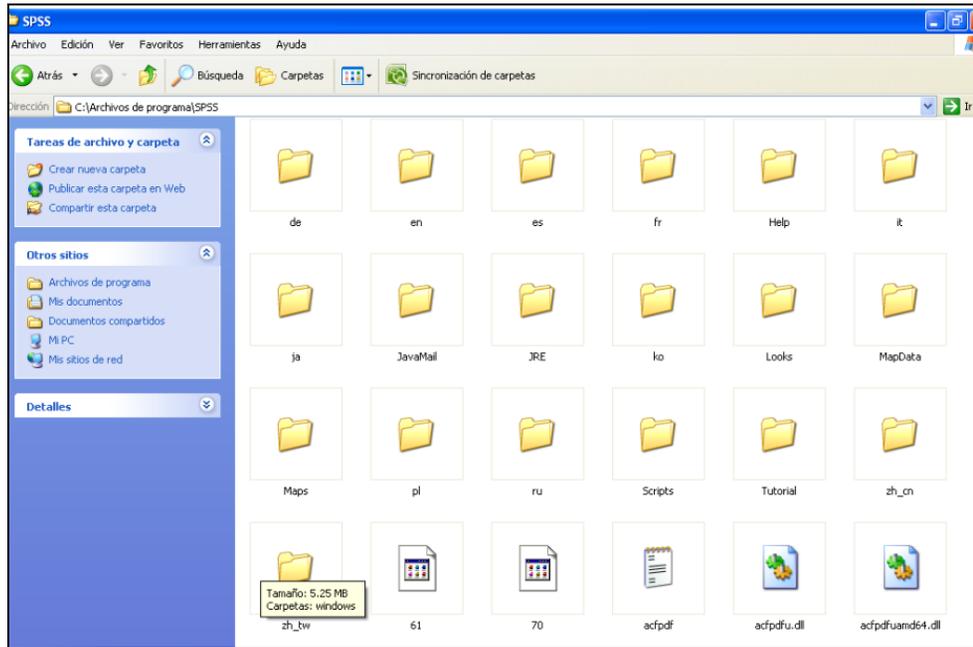


Figura 4.4. Iniciar el programa SPSS





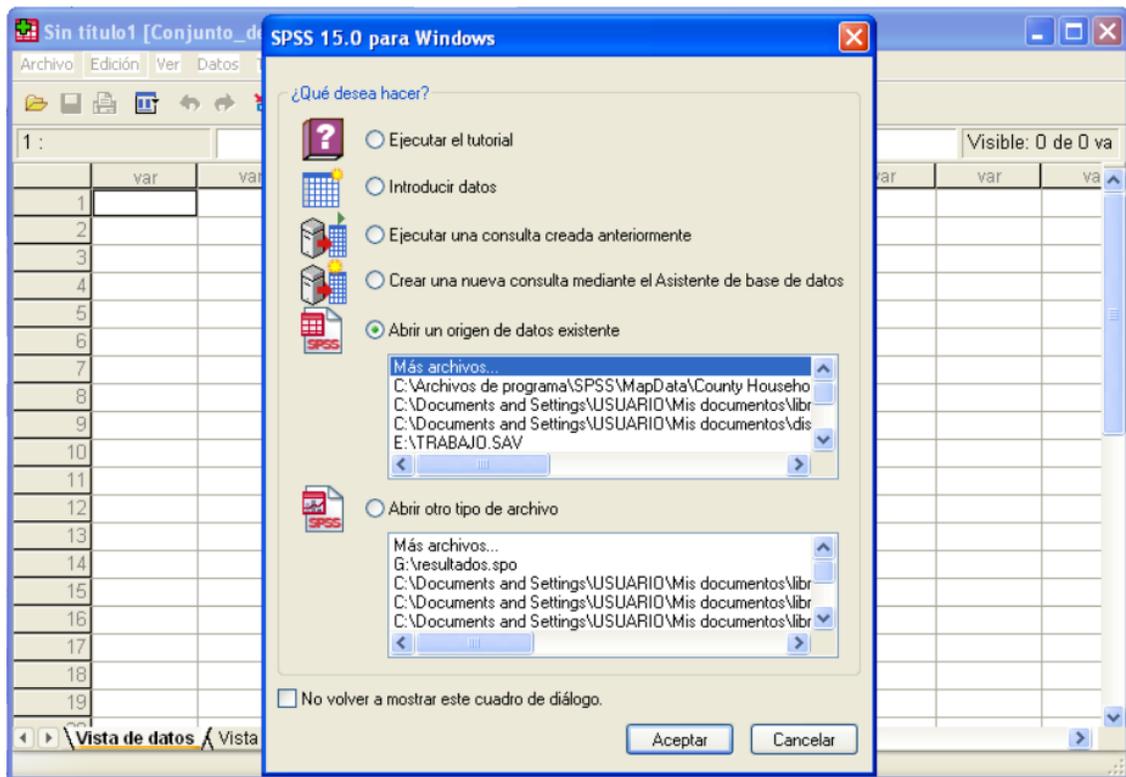
Para entrar en SPSS haga clic en **Inicio**, ubíquese en la sección de **Programas** y busque el Icono **SPSS para Windows**, y de doble clic en: **SPSS 15.0 para Windows**

Otra forma de acceder al programa es dar doble clic en el icono que aparece en el grupo de archivo que crea el SPSS una vez instalado.



Cuando se accede al programa SPSS, aparece la ventana de Editor de datos y la ventana del Visor SPSS para Windows (Figura 4.5).

**Figura 4.5.** Ventana de aplicación del SPSS.



Cuando se abre el programa aparece un cuadro de diálogo con la sección **¿Qué desea hacer?**, que le permite abrir el último archivo que trabajó, simplemente haciendo doble clic sobre el nombre del archivo. También le da la oportunidad de activar la opción **“No volver a mostrar este cuadro de diálogo”** (Figura 4.5).

#### 4.1. ARCHIVOS QUE GENERA EL SPSS.

Los archivos que genera el SPSS son de tres tipos:

**De datos:** tienen extensión *.sav* y está en formato SPSS.

**De gráficos:** tiene extensión *.cht* (diagrama de barras, histogramas, etc.).

**De texto:** tiene extensión *.sps* (archivo de sintaxis) o *.lst* (archivo de resultados)

#### 4.2. VENTANAS DEL SPSS.

Existen diversos tipos de ventanas de SPSS:

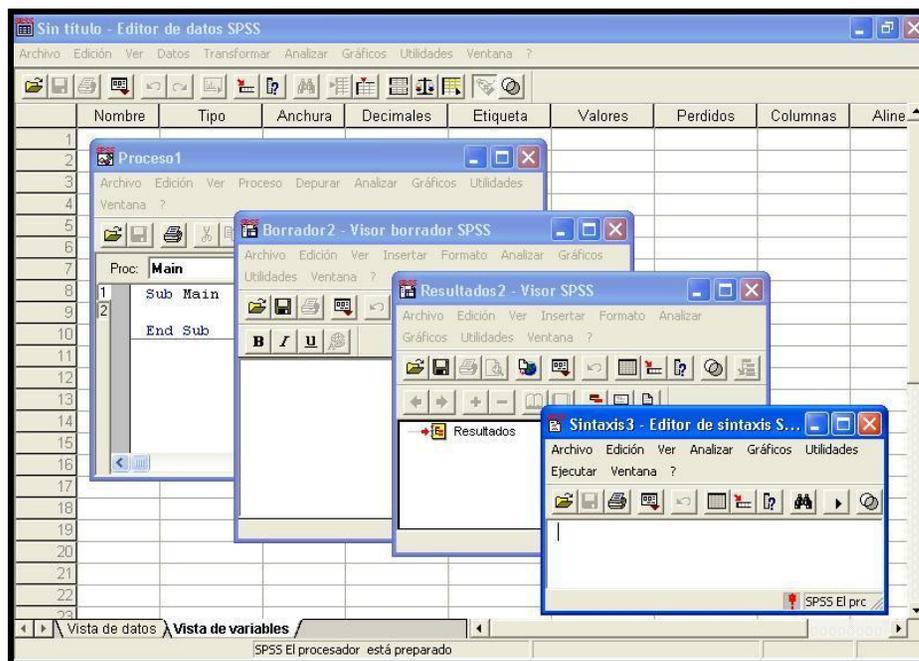
- **Editor de datos.** Esta ventana se abre automáticamente cuando se inicia una sesión de SPSS. El programa no permite tener más de un archivo de datos abierto al mismo tiempo. Esta ventana muestra el contenido del archivo de datos. Con el editor de datos puede crear nuevos archivos de datos o modificar los existentes.
- **Visor.** La ventana del visor se abre automáticamente la primera vez que se ejecuta un procedimiento que genere resultados. Todas las tablas, los gráficos y los resultados estadísticos se muestran en el visor.
- **Visor de borrador.** Los resultados pueden mostrarse como texto normal (en lugar de tablas pivote interactivas) en el visor de borrador.
- **Editor de tablas pivote.** Con el editor de tablas pivote es posible modificar los resultados mostrados en este tipo de tablas, se puede editar el texto, intercambiar los datos de las filas y las columnas, añadir colores, crear tablas multidimensionales y ocultar y mostrar resultados de manera selectiva.
- **Editor de gráficos.** Con el editor de gráficos se puede modificar los gráficos y diagramas de alta resolución en las ventanas de los gráficos. Permite cambiar los colores, seleccionar diferentes tipos de fuentes y tamaños, intercambiar los ejes horizontal y vertical, rotar diagramas de dispersión, graficar en 3-D e incluso cambiar el tipo de gráfico.
- **Editor de resultados de texto.** Los resultados de texto que no aparecen en las tablas pivote pueden modificarse con el Editor de resultados de texto. Puede editar los resultados y cambiar las características de las fuentes (tipo, estilo, color y tamaño).

- **Editor de sintaxis.** Puede pegar las selecciones del cuadro de diálogo en una ventana de sintaxis, donde aparecerán en formas de sintaxis de comandos. En el editor de sintaxis puede ver las funciones especiales de SPSS que no se encuentran disponibles en los cuadros de diálogo. Esto datos se pueden guardar los comandos en un archivo para utilizarlos en sesiones de SPSS posteriores.
- **Editor de procesos.** El proceso y la automatización OLE permite personalizar y automatizar muchas tareas en SPSS. Se puede utilizar el Editor de procesos para crear y modificar los procesos básicos.

Con SPSS se puede trabajar de dos maneras, con cuadros de diálogos o con comandos. Cada una de las ventanas tiene una barra de menús desplegable y una barra de iconos. Tanto la barra de menús como la de iconos son distintas en cada ventana, aunque algunos elementos de los menús y algunos iconos están disponibles en varias de ellas.

- **Editor de datos.** Contiene la barra de menús con las distintas opciones del programa.

**Figura 4.6.** Ventanas del SPSS.



### **4.3. MENÚ PRINCIPAL.**

En la barra de menús están agrupadas las diversas posibilidades que ofrece esta ventana. Cada elemento de la barra tiene un menú desplegable. Para acceder al menú basta seleccionarlo con el ratón o pulsar la tecla **Alt** + la letra subrayada.

Desde el menú se accede a la mayoría de las funciones de SPSS. El menú principal tiene nueve opciones:

- **Archivo:** Esta opción le permite crear un archivo nuevo; abrir uno existente, grabar, leer datos creados con otras aplicaciones, etc.
- **Edición:** Contiene las habituales opciones de Windows para copiar datos, buscar, etc.
- **Ver:** Permite cambiar la apariencia de los datos de su pantalla.
- **Datos:** Contiene opciones para hacer cambios que afectan a todo el archivo de datos (unir archivos, transponer variables y casos, crear subconjuntos de casos, etc.). Estos cambios son temporales mientras no se guarde explícitamente el archivo.
- **Transformar:** Cambios sobre variables seleccionadas, creación de nuevas variables. Estos cambios son temporales mientras no se guarde explícitamente el archivo.
- **Analizar:** Desde esta opción se ejecuta todos los procedimientos estadísticos.
- **Gráficos:** Esta opción le permite crear gráfico de barras, de torta, histogramas, etc. A partir de la versión 8 se encuentran las graficas interactivas.
- **Utilidades:** Cambiar fuentes, obtener información completa del archivo de datos, acceder a un índice de comandos SPSS, etc.
- **Ventana:** Ordenar, seleccionar, controlar atributos de las ventanas abiertas.
- **Ayuda?:** Abre un archivo estándar de ayuda Windows.

### **4.4. BARRA DE HERRAMIENTAS.**

Cada ventana de SPSS tiene su propia barra de herramientas que proporciona acceso rápido y fácil a las tareas más habituales. Las pistas proporcionan una breve descripción de cada herramienta cuando se sitúa el puntero del ratón sobre ellas.

**Figura 4.7.** Barra de herramientas de editor de Datos con pista de ayuda.



Otras barras de herramientas encontradas en el programa SPSS:

**Figura 4.8.** Barra estándar del Visor



**Figura 4.9.** Barra Titulares de visor



**Figura 4.10.** Barra de Formato de gráficos.



**Figura 4.11.** Barra Estándar de gráficos



A continuación se describen cada uno de los iconos de la barra de herramientas del editor de datos, ya que es la más utilizada.

#### 4.5. HERRAMIENTAS DE EDITOR DE DATOS



Abre el cuadro de diálogo para el tipo de documento activo en pantalla: datos, salida, sintaxis, gráficos.



Graba una base de datos (sav), un análisis (lst), texto (sps) o un gráfico (cht) que está activo en pantalla. Si previamente no tenía nombre, va al cuadro de diálogo para nombrarlo.



Imprime el documento activo en pantalla. En todos, excepto en los gráficos, permite seleccionar parte del documento.



Recupera cuadros de diálogo.



Deshace la última opción que se trabajó en el programa.



Activa la ventana de edición de datos para mostrar los datos del sujeto seleccionado.



Información sobre variables. Desde este icono se puede pegar nombres de variables en la ventana de sintaxis o activar la ventana de edición para ver la variable seleccionada.



Para buscar un dato de una cierta variable, seleccione la variable y elija este icono – Buscar, o digite ctrl. + F.



Inserta un nuevo caso (fila) en la base de datos encima de la fila activa.



Permite segmentar el archivo



Pondera los casos



Selecciona casos.



Conmuta entre la opción visualizar los valores de la variable seleccionada o visualizar sus etiquetas.



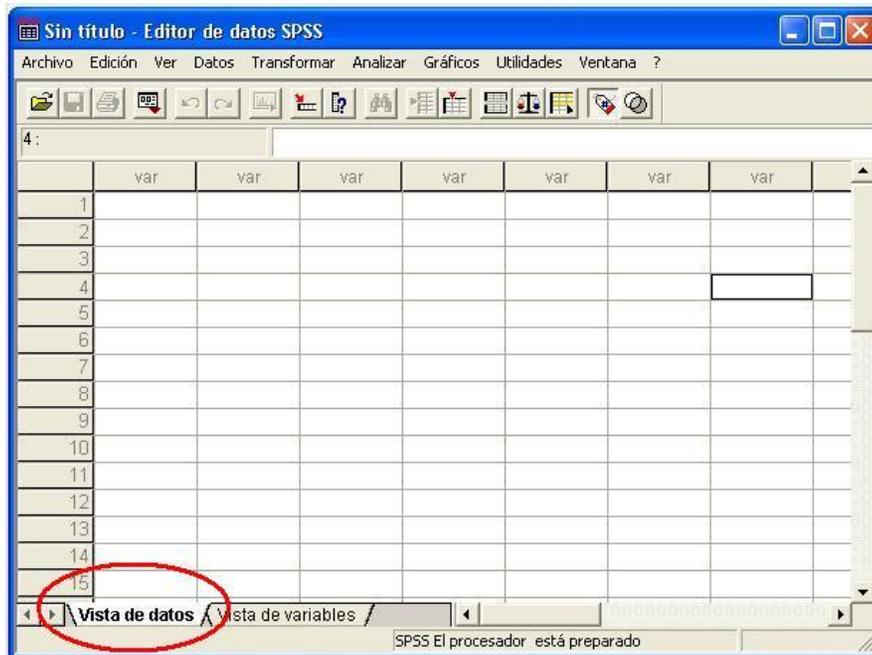
Selecciona un conjunto de variables predefinido para utilizarlo en el análisis.

#### 4.6. VENTANA DE EDICIÓN DE DATOS.

Cuando se inicia el trabajo con el SPSS se establece en forma predeterminada el editor de datos, en ésta ventana en la parte inferior hay dos solapas: **vista de datos** (para introducir datos) y **vista de variables** (para definir las variables). Si se encuentra en el editor de datos, se puede cambiar a **vista de variables** o viceversa simplemente tecleando **Ctrl** + T.

Dentro de las dimensiones de la base de datos (Filas x Columnas) no existen casillas en blanco. Si alguna casilla está vacía, se considerará como un dato perdido en el caso de variables numéricas o como una cadena de caracteres vacía en el caso de variables categóricas.

**Figura 4.12.** Editor de Datos



#### 4.7. SPSS Y EL MUESTREO<sup>13</sup>

La opción Muestras complejas de SPSS (opción presente en el programa a partir de la versión 12) permite seleccionar una muestra de acuerdo con un diseño complejo e incorporar las especificaciones del diseño al análisis de los datos para asegurar la validez de los resultados.

En SPSS, una muestra aleatoria simple, las unidades de muestreo individuales se seleccionan aleatoriamente con la misma probabilidad y sin reposición (SR) directamente a partir de la totalidad de la población. Pero si se quiere realizar un procedimiento de una muestra compleja hay que tener en cuenta los siguientes aspectos:

- **Estratificación:** El muestreo estratificado requiere seleccionar muestras independientes dentro de los subgrupos excluyentes de la población o estratos. Con la estratificación, puede asegurar que los tamaños muestrales de los subgrupos de interés

<sup>13</sup> El material de esta sección es una recopilación y adaptación del capítulo 10 del texto "Muestreo estadístico", Pérez López. César. 2005, p. 351.

son adecuados, mejorar la precisión de las estimaciones globales y utilizar distintos métodos de muestreo entre los estratos identificados.

- **Conglomerados:** El muestreo por conglomerados implica la selección de grupos de unidades muestrales o conglomerados. Por ejemplo un conglomerado puede ser universidades, hospitales, etc., y las unidades muestrales pueden ser alumnos y pacientes.
- **Múltiples etapas:** En el muestreo polietápico, se selecciona una muestra de primera etapa basada en conglomerados, a continuación, se crea una muestra de segunda etapa extrayendo submuestras a partir de los conglomerados seleccionados.
- **Muestro no aleatorio:** El muestreo no aleatorio se aplica cuando es difícil obtener la muestra aleatoriamente, las unidades se pueden muestrear sistemáticamente (con un intervalo fijo) o secuencialmente.
- **Probabilidad de selección desigual:** Cuando se muestran conglomerados que contienen números de unidades desiguales, puede utilizar el muestreo probabilístico proporcional al tamaño (PPS) para que la probabilidad de elección del conglomerado sea igual a la proporción de unidades que contiene. El muestreo PPS también puede utilizar esquemas de ponderación más generales para seleccionar unidades.
- **Muestreo no restringido:** El muestreo no restringido selecciona las unidades con reposición (CR), por lo que se puede seleccionar más de una vez una unidad individual para la muestra.
- **Ponderaciones muestrales:** Las ponderaciones muestrales se calculan automáticamente al extraer una muestra compleja y de forma ideal se corresponden con la frecuencia que cada unidad muestral representa en la población objetivo. Por lo tanto, la suma de las ponderaciones muestrales debe estimar el tamaño de la población

#### 4.7.1. Creación del plan de muestreo y almacenamiento de los datos de la muestra.

Haciendo uso del asistente de muestreo que le guiará paso a paso para crear, modificar y ejecutar un archivo de plan de muestreo, antes de iniciar, se debe tener definida la población objetivo, unidad de muestreo, el elemento, una vez, definidos los aspectos anteriores, se procede a realizar un diseño de muestreo adecuado. Se explicará como crear un plan de muestreo estratificado al 10% por estratos.



En el CD-ROM

En el CD-ROM que acompaña a éste libro encontrará el archivo *Etapa84.sav*, correspondiente a una etapa de la encuesta continua de hogares realizada por el DANE (Departamento Administrativo Nacional de Estadística) aplicada en el segundo trimestre de 1994.

El archivo con que se trabajará este tema corresponde a la etapa 84 de la Encuesta Continua de Hogares, realizada por el DANE en el segundo trimestre de 1994. El objetivo principal de la encuesta continua de hogares es proporcionar información básica sobre el tamaño y estructura de la fuerza de trabajo (empleo, desempleo e inactividad) de la población del país. Además permite obtener variables de población como: parentesco, sexo, estado civil y educación.

La cobertura geográfica de la encuesta corresponde a las 5 regiones del país es decir:

- Región Atlántica
- Región Oriental
- Región Central
- Región Pacífica y
- Región Bogotá

Se encuestan los hogares de las ciudades con sus áreas metropolitanas.

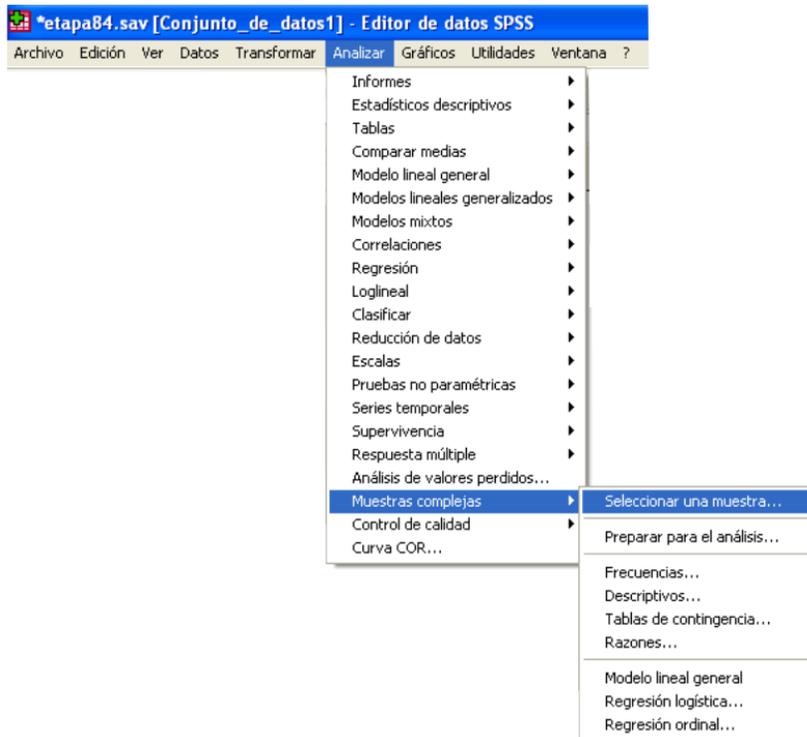
En el archivo se agrupo todas las muestras aplicadas en el país que corresponde a una muestra desagregada por sexo así:

Hombres	Mujeres	Total
39.241	44.013	83.254

Ahora bien, para crear un plan de muestreo, elija en el menú principal

Analizar → Muestra complejas → Seleccionar una muestra (Figura 4.13), inmediatamente ingresamos al asistente de muestreo (Figura 4.14).

**Figura 4.13.** Muestras complejas



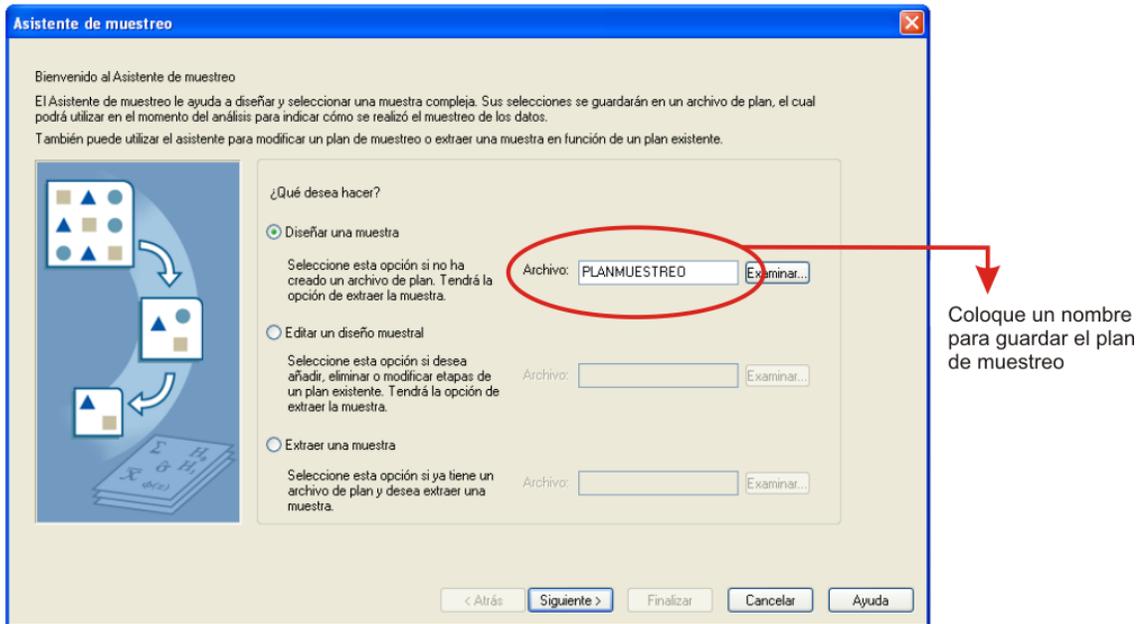
En el asistente de muestreo seleccione

- Diseñar una muestra → y elija un nombre de archivo, en este caso se colocó PLANMUESTREO.SPS

 <p><b>Nota</b></p>	<p>Se recomienda colocar el nombre de archivo con la extensión, con el fin de facilitar la lectura del archivo, una vez se termine todo el proceso.</p> <p>En este caso el nombre es PLANMUESTREO.SPS, se colocó la extensión SPS puesto que se trata de un archivo de texto (Archivo de sintaxis).</p>
--	---

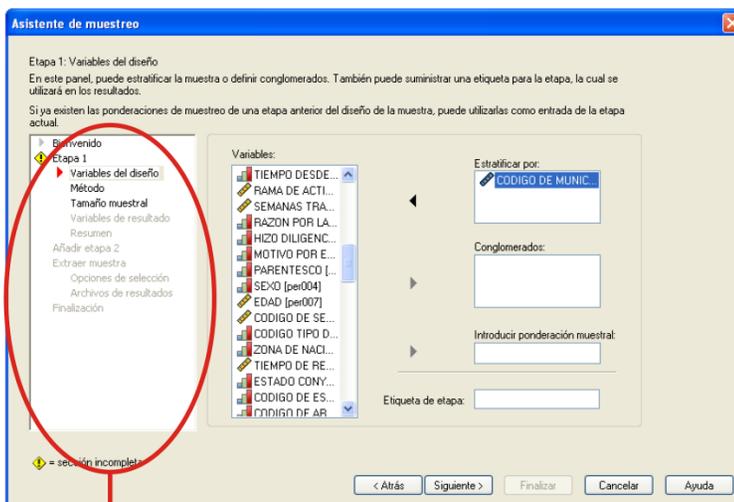
- Pulse **siguiente** para determinar las variables de diseño.

Figura 4.14. Asistente de muestreo



Este paso permite seleccionar las variables de estratificación y conglomeración en el campo *Variables* arrastrándolas a los campos *Estratificar por* y *Conglomerados* respectivamente, y definir ponderaciones.

Figura 4.15. Etapa 1. Variables de diseño.



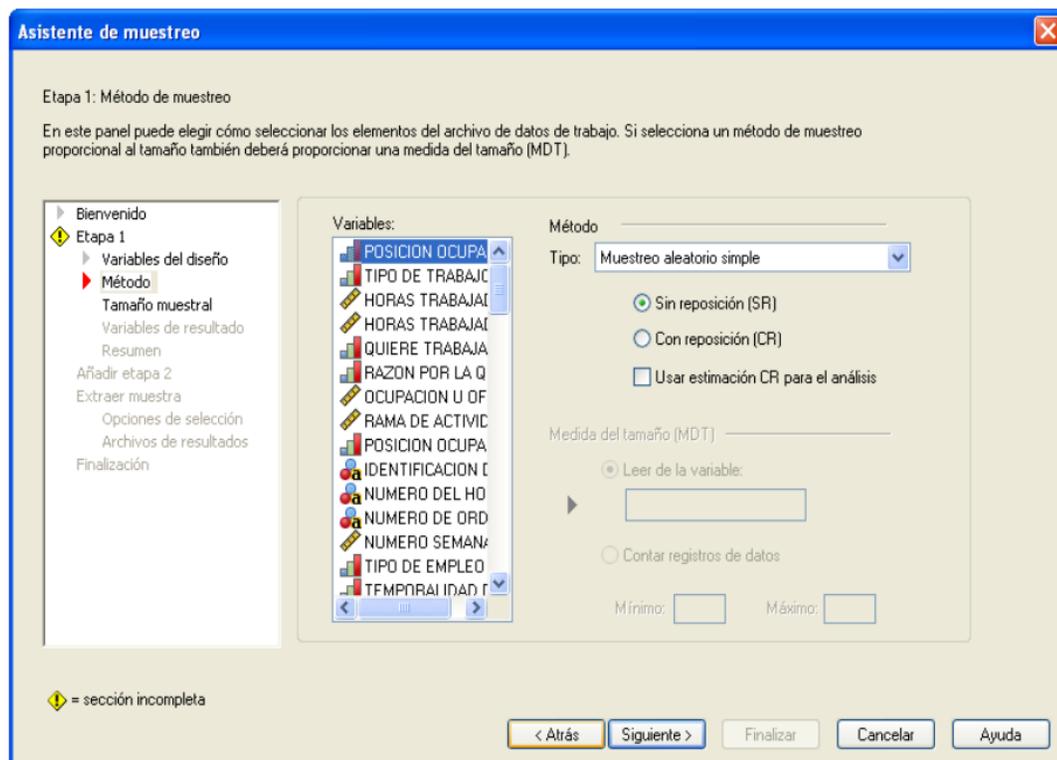
En esta parte del sistema se muestra un esquema de los títulos de todos los pasos. Se puede pasar al siguiente paso al pulsar un nombre de los títulos activandos en el esquema.

También puede especificar una etiqueta para la etapa en el campo *Etiqueta de etapa* (se utiliza en los resultados para facilitar la identificación de la información por etapas).

- En este caso la variable que se toma para estratificación es **CODIGO DE MUNICIPIO**

A continuación pulse *siguiente* para ir al paso *Método* (Figura 4.16), o puede pulsar Método en la parte izquierda de la pantalla del Asistente.

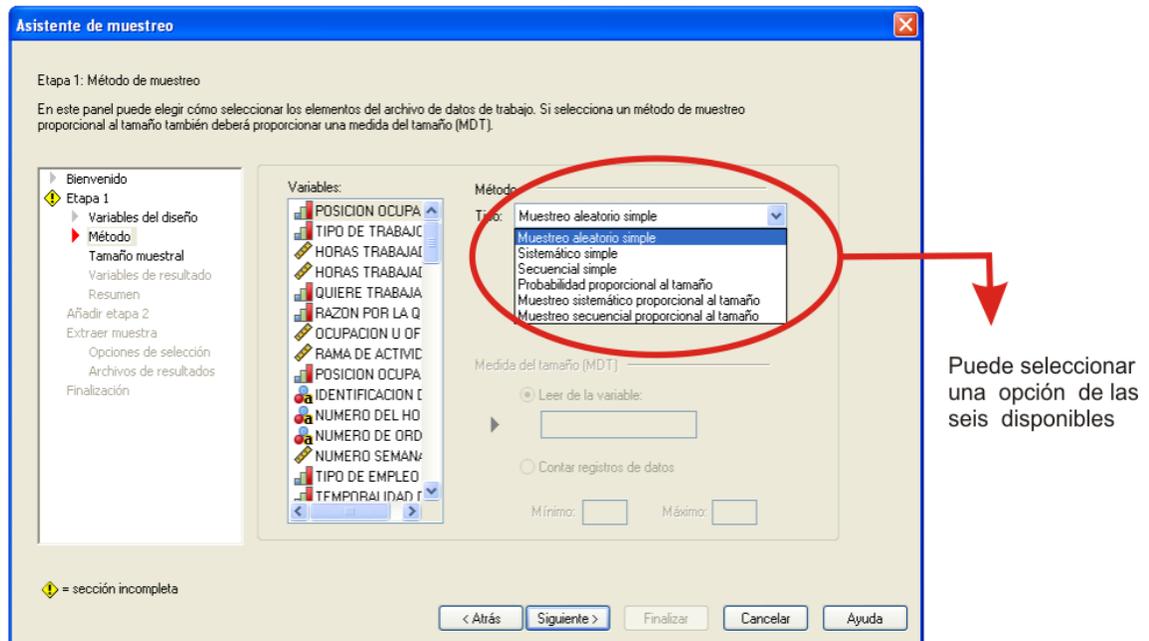
**Figura 4.16.** Método de muestreo



En el paso Método (Figura 4.17), se elige el tipo de muestreo, hay seis (6) opciones:

- Muestreo aleatorio simple
- Sistemático simple
- Secuencial simple
- Probabilidad proporcional al tamaño
- Muestreo sistemático proporcional al tamaño
- Muestreo secuencial proporcional al tamaño

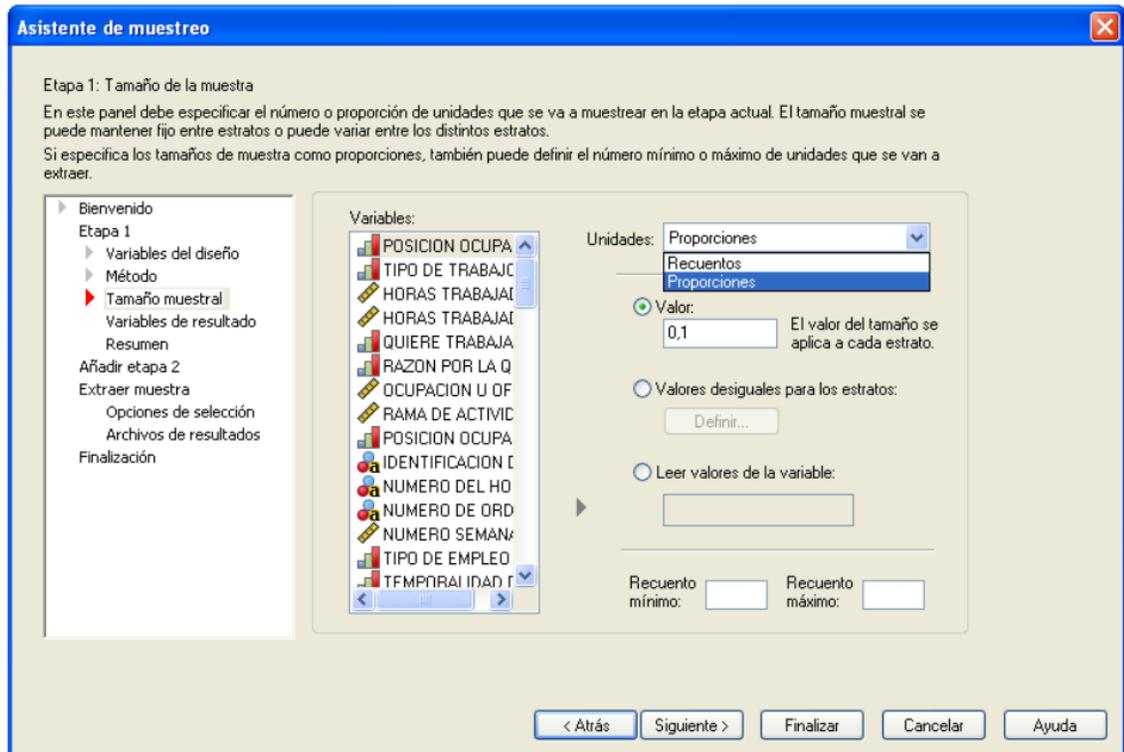
Figura 4.17. Selección del método de muestreo.



En el **muestreo aleatorio Simple** las unidades se seleccionan con probabilidad igual. Se puede seleccionar con o sin reposición. En el **muestreo Sistemático Simple** las unidades se seleccionan con un intervalo fijo en todo el marco muestral (o en los estratos, si se han especificado) y se extraer sin reposición. Se selecciona una unidad aleatoriamente dentro del primer intervalo como el punto inicial. En el **Muestreo secuencial simple** las unidades se seleccionan de forma secuencial con probabilidad igual y sin reposición. El **muestreo con probabilidad proporcional al tamaño** es un método de primera etapa que selecciona unidades de forma aleatoria con probabilidad proporcional al tamaño. Se puede seleccionar cualquier unidad con reposición; sólo se puede realizar muestreo sin reposición de los conglomerados. El **Muestreo sistemático proporcional al tamaño** es un método de primera etapa que selecciona unidades de forma sistemática con probabilidad proporcional al tamaño. Se selecciona sin reposición. El **Muestreo secuencial proporcional al tamaño** es un método de primera etapa que selecciona unidades de forma secuencial con probabilidad proporcional al tamaño del conglomerado y sin reposición.

Para el ejemplo que se está realizando se seleccionó **Muestreo secuencial simple**.

- Pulse siguiente, o pulse **Tamaño muestral** en las opciones de la parte izquierda del ayudante (Figura 4.18).

**Figura 4.18.** Tamaño de la muestra

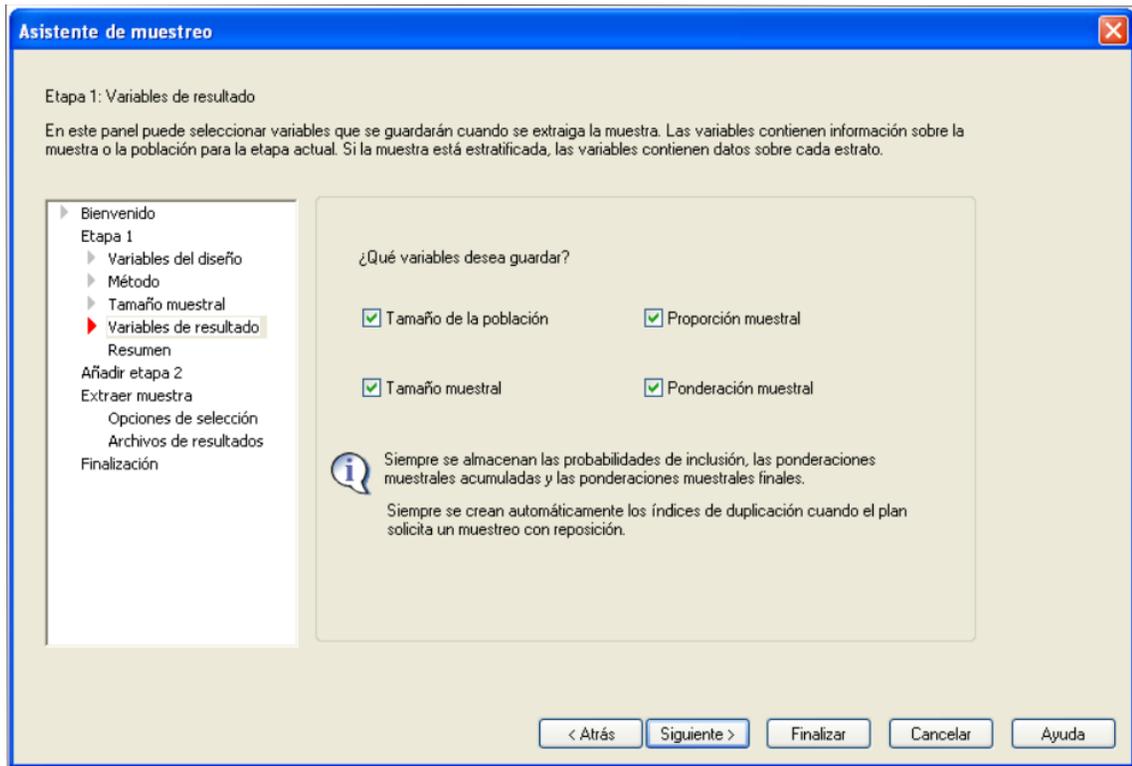
Este paso permite especificar el número o la proporción de unidades que se van a muestrear dentro de la etapa actual. El tamaño de muestra puede ser fijo o variar entre estratos. En el campo **Unidades** puede especificar un tamaño muestral exacto o una proporción de unidades a muestrear. El campo **Valor** se aplica un valor particular a todos los estratos. Si se selecciona **Recuentos** como unidad métrica, se debe introducir un entero positivo. Si selecciona **Proporciones**, se deberá introducir un valor no negativo entre 0,1 y 1. El campo **Leer valores de la variable** permite seleccionar variable numérica que contenga los valores del tamaño para los estratos. Si se selecciona **Proporciones**, se tiene que diligenciar las opciones **Recuento mínimo** y **Recuento máximo** para el número de unidades muestreadas.

- Para el ejemplo registre:

**Unidades** = Proporciones  
**Valor** = 0,1

- Pulse **siguiete** o se puede pulsar la opción **Variables de Resultado** de la parte izquierda del asistente (Figura 4.19)

Figura 4.19. Variables de resultado

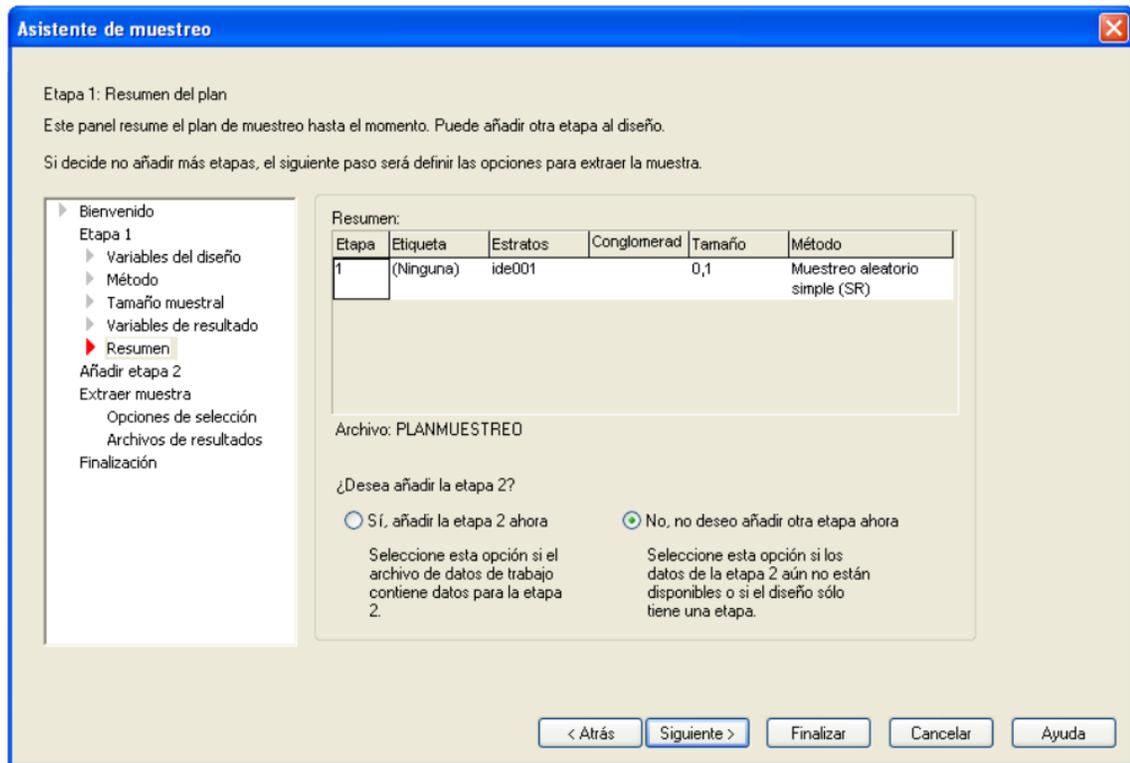


Este paso permite elegir las variables que desean guardar cuando se extraiga la muestra.

**Tamaño de la poblacional** recoge el número estimado de unidades en la población de una etapa dada. La **proporción muestral** recoge la tasa de la muestra en una etapa dada. **Tamaño muestral** recoge el número de unidades extraídas en una etapa dada. **Proporción muestral** recoge la inversa de las probabilidades de inclusión.

- Seleccione las opciones que se muestran en la figura 4.19
- Pulse **continuar** o pulse Resumen en la parte izquierda de la pantalla del asistente (Figura 4.20).

**Figura 4.20.** Resumen del plan



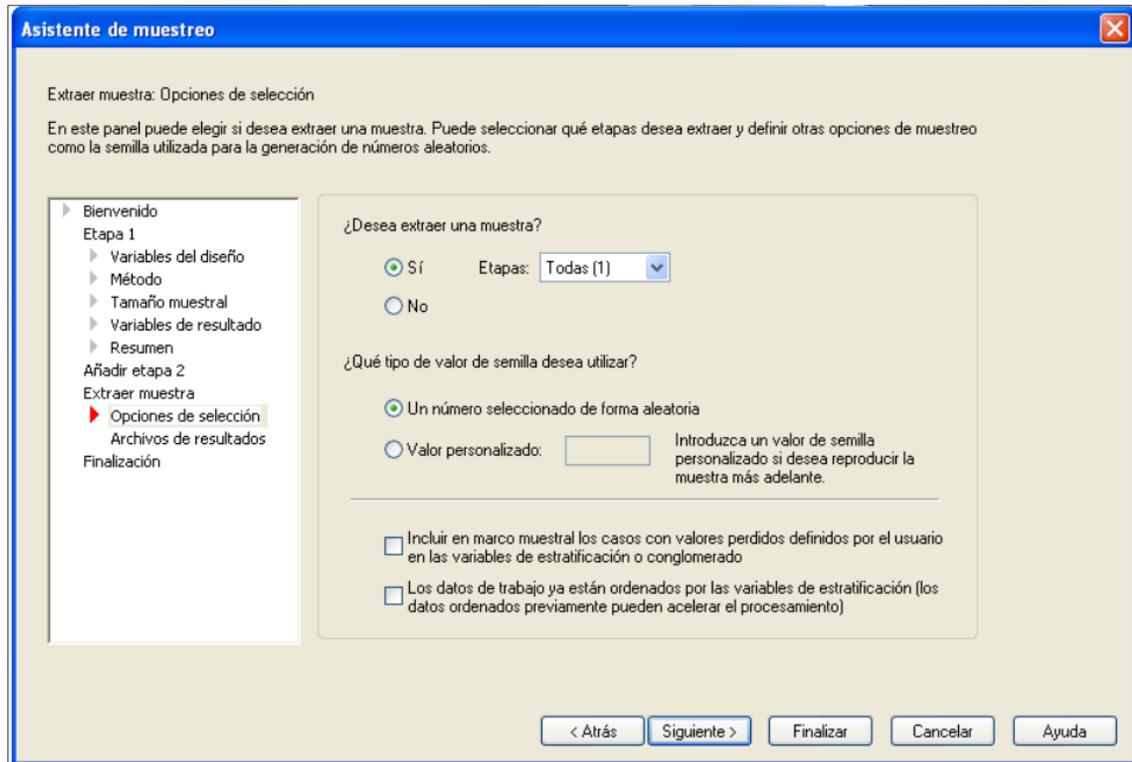
Este es el último paso de cada etapa que proporciona un resumen de las especificaciones del diseño muestral hasta la etapa actual. A partir de aquí, puede pasar a la siguiente etapa (creándole si es necesario en Añadir etapa 2) o definir las opciones para extraer la muestra.

- Seleccione No, no deseo añadir otra etapa ahora.

En este punto ya estamos en condiciones de *extraer la muestra* según el diseño definido en los pasos anteriores.

- Presione *siguiente* o pulse *Extraer muestra* → *Opciones de selección* en la parte izquierda de la pantalla del Asistente de muestreo. También puede controlar otras opciones del muestreo, como la *semilla aleatoria* y el *tratamiento de los valores perdidos*.

**Figura 4.21.** Extraer muestra: Opciones de selección



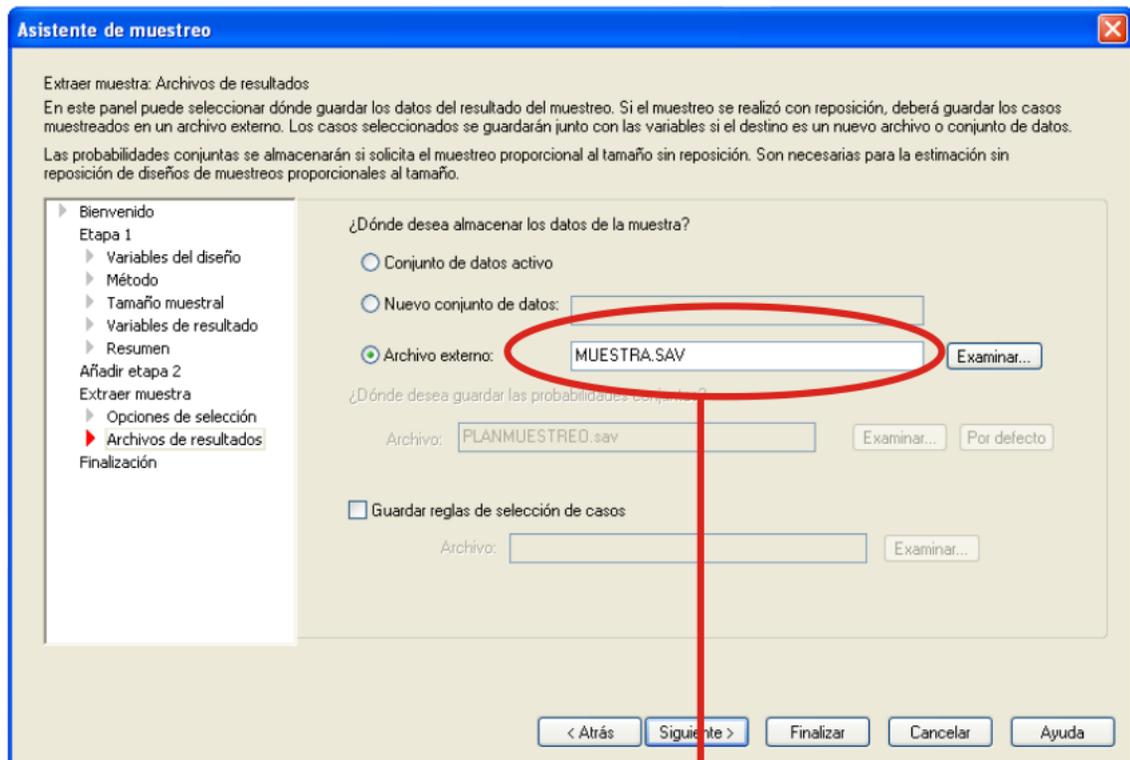
Con los datos ya ordenados permite acelerar el proceso de selección si el marco está clasificado previamente por los valores de las variables de estratificación.

- Active *Si* en la opción *¿Desea extraer una muestra?*
- Active *Un número seleccionado de forma aleatoria* en la sesión *¿Qué tipo de valor de semilla desea utilizar?*

Después de este procedimiento se debe guardar los resultados adecuadamente. Para esto se selecciona:

- Presione *siguiete* o *Archivos de resultados* en la parte izquierda de la pantalla del Asistente de muestreo, este paso permite elegir dónde dirigir los casos muestreados, las variables de ponderación, las probabilidades conjuntas y las reglas de selección de casos.

Figura 4.22. Extraer la muestra



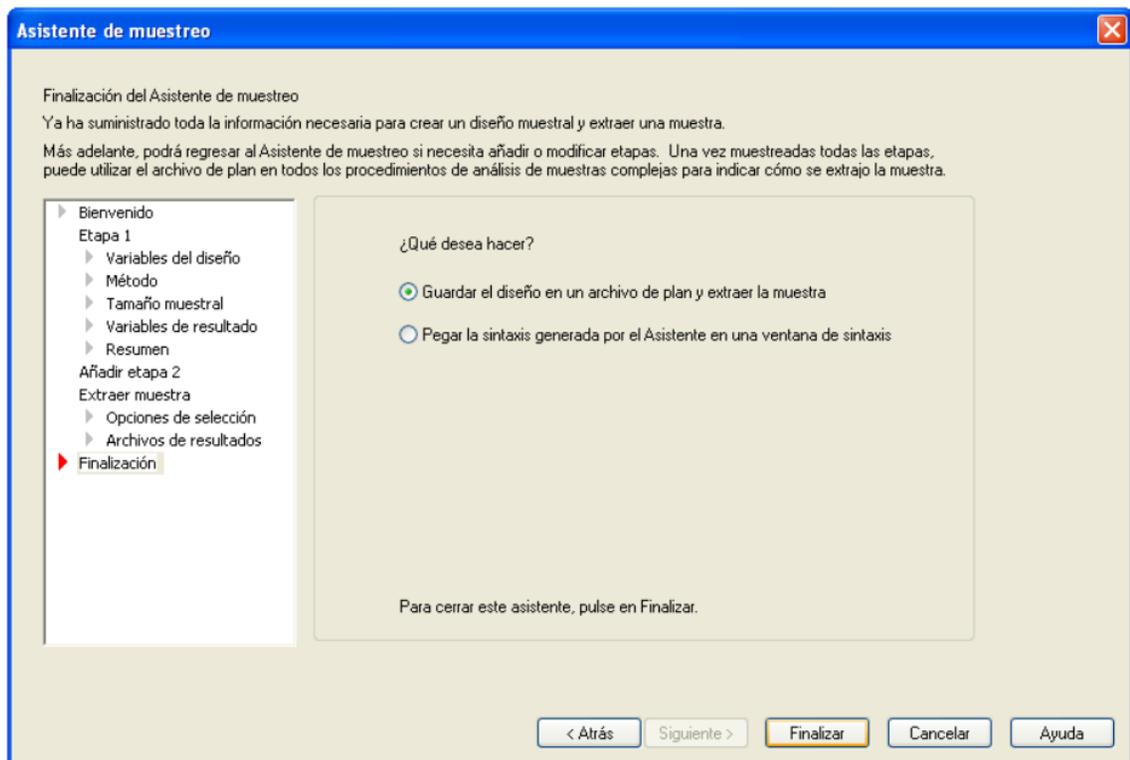
Esta opción permite guardar los datos de la muestra en un archivo, se debe especificar el nombre junto con la extensión SAV, con el fin de leer fácilmente el nuevo archivo.

La opción *¿dónde desea almacenar los datos de la muestra?*, permite determinar dónde se escribe el resultado de la muestra. Se puede añadir al archivo de datos de trabajo o guardar en un archivo externo. Si se especifica un archivo externo se debe colocar la extensión *SAV*. La opción de *¿dónde desea guardar las probabilidades conjuntas?* permite determinar dónde se escriben las probabilidades conjuntas. Las probabilidades conjuntas se producen si se seleccionan la probabilidad proporcional al tamaño sin reposición, el muestreo de Brewer proporcional al tamaño. El campo Guardar reglas de selección de casos, sí se construye la muestra por etapas, es posible que quiera guardar las reglas de selección de casos en un archivo texto.

- Seleccione Archivo externo
- Coloque el nombre MUSTRA.SAV

- Presione **Siguiente** o **finalización** en la parte izquierda de la pantalla del Asistente de muestreo (Figura 4.23)

**Figura 4.23.** Finalización del Asistente de muestreo

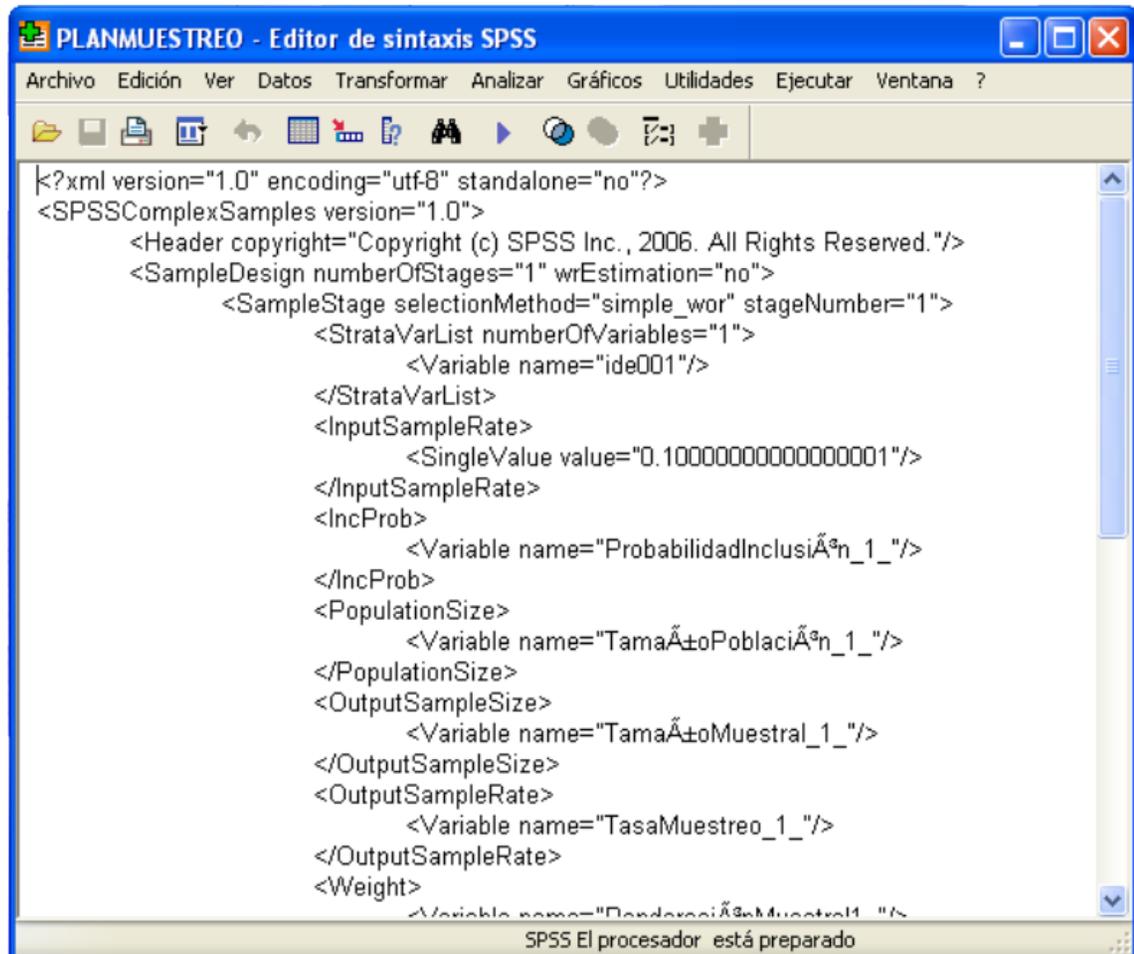


Puede guardar el archivo de plan y extraer la muestra ahora o pegar las selecciones en una ventana de sintaxis. Al pulsar en Finalización (figura 4.23)

- Seleccione **Guardar el diseño en un archivo de plan y extraer la muestra** de la sesión **¿Qué desea hacer?**
- Y por último presione finalizar

Concluido todo el procedimiento se obtiene un resumen para las etapas (Figura 4.24).

Figura 4.24. Ventana de Sintaxis del procedimiento.



The image shows a screenshot of the SPSS Syntax Editor window titled "PLANMUESTREO - Editor de sintaxis SPSS". The window contains XML code for defining a complex sample design. The code is as follows:

```
<?xml version="1.0" encoding="utf-8" standalone="no"?>
<SPSSComplexSamples version="1.0">
  <Header copyright="Copyright (c) SPSS Inc., 2006. All Rights Reserved."/>
  <SampleDesign numberOfStages="1" wrEstimation="no">
    <SampleStage selectionMethod="simple_wor" stageNumber="1">
      <StrataVarList numberOfVariables="1">
        <Variable name="ide001"/>
      </StrataVarList>
      <InputSampleRate>
        <SingleValue value="0.100000000000000001"/>
      </InputSampleRate>
      <IncProb>
        <Variable name="ProbabilidadInclusiÃ³n_1_"/>
      </IncProb>
      <PopulationSize>
        <Variable name="TamaÃ±oPoblaciÃ³n_1_"/>
      </PopulationSize>
      <OutputSampleSize>
        <Variable name="TamaÃ±oMuestral_1_"/>
      </OutputSampleSize>
      <OutputSampleRate>
        <Variable name="TasaMuestreo_1_"/>
      </OutputSampleRate>
      <Weight>
        <Variable name="PonderaciÃ³nMuestral1_"/>
      </Weight>
    </SampleStage>
  </SampleDesign>
</SPSSComplexSamples>
```

At the bottom of the window, the status bar indicates "SPSS El procesador esta preparado".

**Figura 4.25.** Resumen de las etapas

**Resumen para la etapa 1**

CODIGO DE MUNICIPIO	Número de unidades muestreadas		Proporción de unidades muestreadas	
	Solicitados	Reales	Solicitados	Reales
1,00	950	950	10,0%	10,0%
2,00	437	437	10,0%	10,0%
3,00	1677	1677	10,0%	10,0%
4,00	335	335	10,0%	10,0%
5,00	795	795	10,0%	10,0%
6,00	928	928	10,0%	10,0%
7,00	424	424	10,0%	10,0%
8,00	443	443	10,0%	10,0%
9,00	386	386	10,0%	10,0%
10,00	150	150	10,0%	10,0%
11,00	462	462	10,0%	10,0%
12,00	37	37	10,0%	9,9%
13,00	8	8	10,0%	9,8%
14,00	42	42	10,0%	10,1%
20,00	47	47	10,0%	10,1%
21,00	29	29	10,0%	10,0%
22,00	20	20	10,0%	10,0%
23,00	165	165	10,0%	10,0%
24,00	100	100	10,0%	10,0%
25,00	237	237	10,0%	10,0%
26,00	13	13	10,0%	10,3%
27,00	33	33	10,0%	10,1%
28,00	14	14	10,0%	9,8%
29,00	230	230	10,0%	10,0%
62,00	28	28	10,0%	10,1%
63,00	173	173	10,0%	10,0%
64,00	68	68	10,0%	10,0%
65,00	36	36	10,0%	9,9%
83,00	60	60	10,0%	10,0%

Archivo del plan: C:\Archivos de programa\SPSS\PLANMUESTREO

También se puede obtener un resumen sobre las distintas etapas de selección de la muestra, que se presenta a continuación.

**Figura 4.26.** Resumen de las etapas de selección.

Resumen			Etapa 1
Variables del diseño	Estratificación	1	CODIGO DE MUNICIPIO
Información de la muestra	Método de selección		Muestreo aleatorio simple sin reposición
	Medida del tamaño		
	Proporción de unidades muestreadas		,1
	Variables creadas o modificadas	Probabilidad de inclusión (selección) según etapa	Probabilidad de inclusión (selección) para la etapa 1
		Ponderación de muestreo acumulada según etapa	Ponderación de muestreo acumulada para la etapa 1
		Tamaño de la población según etapa	TamañoPoblación_1_
		Tamaño de la muestra según etapa	TamañoMuestral_1_
		Tasa de muestreo según etapa	TasaMuestreo_1_
Ponderación de muestreo según etapa	PonderaciónMuestral1_		
Información sobre el análisis	Supuestos del estimador		Muestreo de probabilidad igual sin reposición
	Probabilidad de inclusión		A partir de la variable Probabilidad de inclusión (selección) para la etapa 1

Archivo del plan: C:\Archivos de programa\SPSS\PLANMUESTREO  
Variable de ponderación: Ponderación de muestreo final

Usted encontrará los dos archivos creados a través del procedimiento en el subdirectorio direccionado para guardar los archivos.

**Figura 4.27. Archivos creados**



- De doble clic sobre el icono del archivo PLANMUESTREO, de inmediato se abre el archivo de sintaxis mostrado en la figura 4.24.
- De doble clic sobre el icono del archivo MUESTRA para ver el nuevo archivo de datos. (Figura 4.28)

Figura 4.28. Archivo Muestra

	ocu009	ocu001	ocu002	ocu003	ocu004	ocu005	ocu006	ocu007	ocu008	
1	.	.	.	.	.	.	.	.	.	01001'
2	3,00	1,00	84,00	84,00	2,00	,00	58,00	83,00	2,00	01001'
3	2,00	1,00	48,00	48,00	2,00	,00	95,00	33,00	2,00	01001'
4	.	.	.	.	.	.	.	.	.	01001'
5	2,00	1,00	54,00	54,00	2,00	,00	84,00	95,00	2,00	01002'
6	5,00	1,00	40,00	40,00	2,00	,00	98,00	71,00	5,00	01002'
7	.	.	.	.	.	.	.	.	.	01002'
8	3,00	1,00	45,00	45,00	2,00	,00	39,00	83,00	2,00	01002'
9	2,00	1,00	56,00	56,00	2,00	,00	93,00	61,00	2,00	01003'
10	.	.	.	.	.	.	.	.	.	01003'
11	5,00	1,00	50,00	60,00	2,00	,00	45,00	62,00	5,00	01003'
12	.	.	.	.	.	.	.	.	.	01003'
13	.	.	.	.	.	.	.	.	.	01003'
14	.	.	.	.	.	.	.	.	.	01003'
15	5,00	1,00	43,00	35,00	1,00	5,00	45,00	62,00	5,00	01003'
16	5,00	1,00	70,00	64,00	2,00	,00	98,00	71,00	5,00	01003'
17	.	.	.	.	.	.	.	.	.	01003'
18	5,00	1,00	80,00	80,00	2,00	,00	45,00	62,00	5,00	01003'
19	.	.	.	.	.	.	.	.	.	01003'

A continuación se presenta las nuevas especificaciones del archivo MUESTRA.

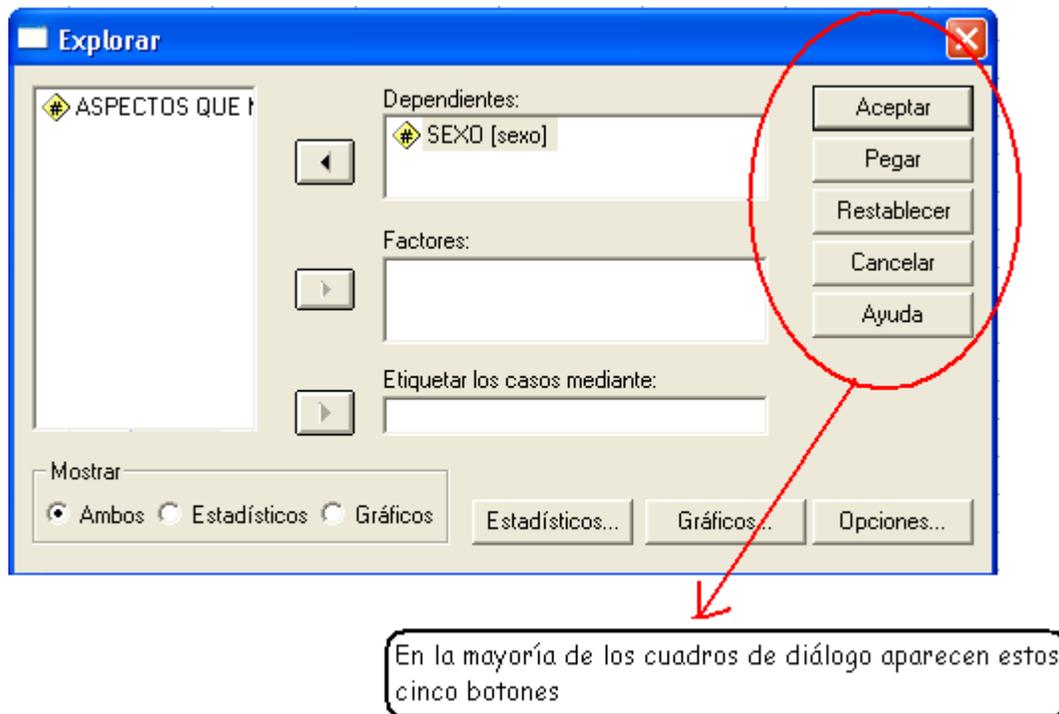
Hombres	Mujeres	Total
3.914	4.413	8.327

Como se puede observar corresponde al 10% del archivo original.



**5.1.1. Cuadros de Diálogo.** El SPSS trabaja con cuadros de diálogos, en la mayoría de éstos hay cinco botones, que tienen aplicación a unos comandos directos de ejecución. (Figura 5.2).

**Figura 5.2.** Cuadro de diálogo



**ACEPTAR:** Realiza las operaciones de acuerdo a los procedimientos establecidos.

**PEGAR:** Escribe la sintaxis correspondiente a los procedimientos realizados con el asistente.

**RESTABLECER:** Cuando se usa por segunda vez un cuadro de diálogo éste conserva las definiciones realizadas, si se pulsa éste botón se eliminan dichas definiciones.

**CANCELAR:** Cancela las definiciones en curso y sale del cuadro de diálogo.

**AYUDA:** Provee ayuda sobre el cuadro de diálogo.

## 5.2. PREGUNTA CERRADA – DICÓTOMA.

**5.2.1. Nombre de la Variable.** Para iniciar la explicación de la definición de variables, se hace con una pregunta dicótoma.

SEXO 1. Hombre ( ) 2. Mujer ( ). Esta pregunta tiene dos opciones.



Para definir las variables, haga clic en la solapa de la parte inferior “**Vista Variables**” y en la columna de **Nombre** escribe el nombre de la variable. En este caso se le asignará el nombre “sexo” (Figura 5.3). Si la escribe en mayúsculas SPSS lo pasará a minúsculas

### 5.2.1.1. Características de los nombres de las variables.

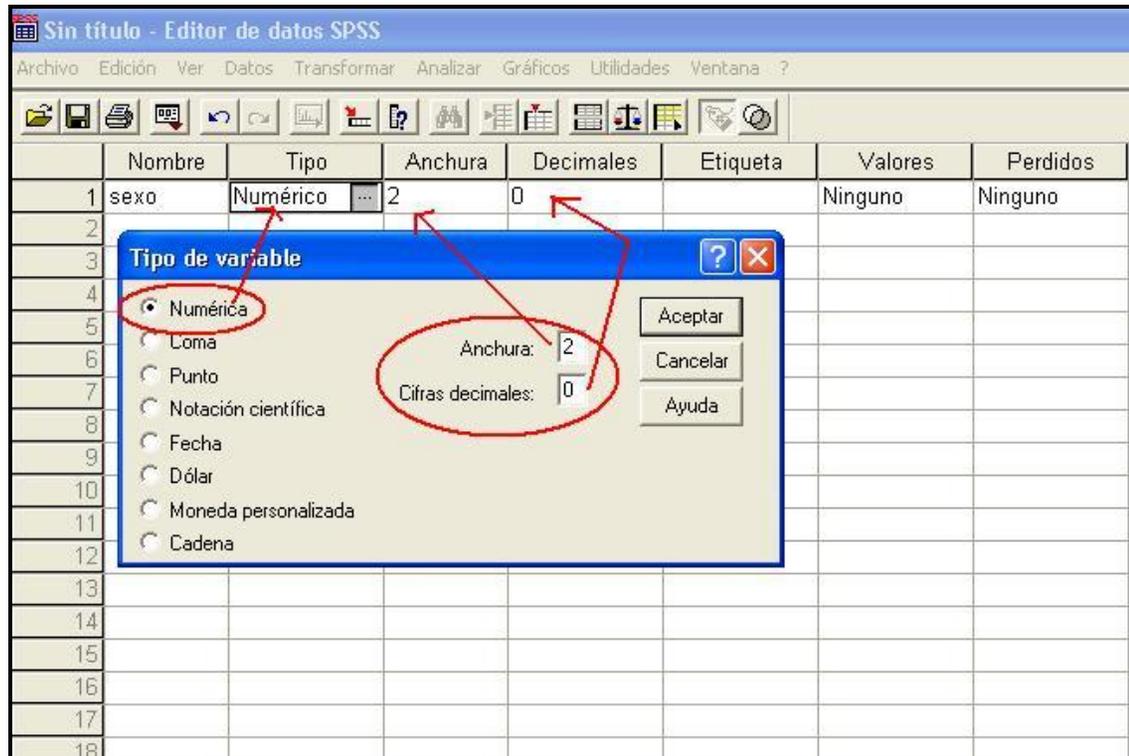
Los nombres de las variables deben cumplir los siguientes requisitos:

- Máximo ocho caracteres.
- Deben comenzar con una letra y no pueden acabar con un punto.
- No puede contener espacios en blanco ni caracteres especiales (¡,?,\*,...).
- No puede haber dos nombres de variables repetidos.
- El programa no distingue entre mayúsculas y minúsculas.
- No se puede utilizar las palabras: ALL, LT, AND, NE, BY, NOT, EQ, OR, GE, TO, GT, WIDT, LE.

**5.2.2. Tipo de Variable.** Cuando se desplace a la columna “**Tipo**” observará dos cosas: que SPSS automáticamente la define como una variable numérica con una anchura de 8 dígitos, de los cuales 2 son decimales, cambie los aspectos, de “tipo”, “Anchura” y “Decimales”, como aparece en la Figura 5.3. Luego haga clic en **Aceptar** para continuar con el procedimiento.

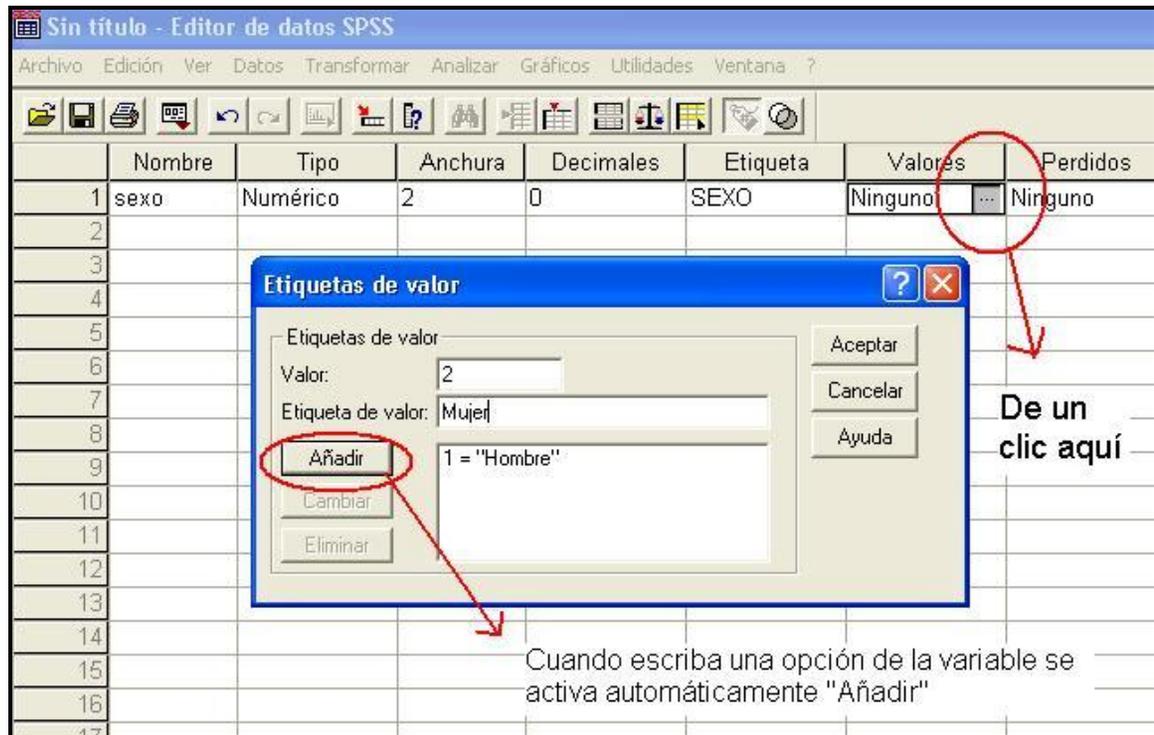
	<b>Atención</b>	Cuando la anchura es 2, diligencie primero la opción de decimales colocando cero (0) y luego digite la opción anchura dos caracteres. Recuerde que el SPSS coloca en forma predeterminada la anchura de 8 y dos caracteres los utiliza para los decimales.
---	-----------------	--

**Figura 5.3.** Definición de Variables



**5.2.3. Etiquetas.** A continuación, en la columna “**Etiqueta**” se escribe la descripción de la variable. En esta columna sí se distingue entre mayúsculas y minúsculas. La etiqueta de variable añade una descripción a la variable, en esta opción se escribe “SEXO” en mayúsculas. La etiqueta automáticamente corresponde al título de las tablas de frecuencia y las gráficas que se generan cuando se procesa la pregunta.

**Figura 5.4.** Opción de Etiquetas de valor



Para definir los valores de la variable SEXO sitúese en la columna “**Valores**” y pulse en el botón . Aparece el cuadro de diálogo Etiquetas de valor (figura 5.4).

**5.2.4. Etiquetas de Valores.** En la casilla Valor, digite el número 1, a continuación pulse la tecla . (o pase con el ratón hasta la casilla de Etiqueta de valor) y escribir “Hombre”, el lector puede notar que una vez digitada la etiqueta de valor, se activa la opción AÑADIR. Pulse el botón Añadir, para completar el procedimiento; repita el procedimiento para la opción “Mujer” (Figura 5.4).

**5.2.5. Perdidos.** En SPSS, existen dos tipos de valores perdidos:

- Los del sistema. Cualquier casilla en blanco dentro de la matriz de datos.
- Los del usuario. En este caso debemos definir los valores perdidos y especificar el valor que asignamos como perdidos.

 <p><b>Véase también</b></p>	<p>Para más detalles consulte el punto 5.3.1 en la sección <b>valores perdidos</b>.</p>
---	---

**5.2.6. Columns – Alineación – Medida.** Las opciones “Columns”, “Alineación” y “Medida”, corresponden a la apariencia que tendrá la pantalla de vista de datos, y estas son fácilmente manipulables tanto en vista de datos como en vista de variables. Digite los siguientes datos:

“Columns”= 8

“Alineación”=Derecha

“Medida”= Nominal

### 5.3. PREGUNTA DE SELECCIÓN MÚLTIPLE.

Para explicar cómo se especifica este tipo de pregunta en el programa SPSS, se realizará con una pregunta de tipo social, referente a los aspectos que más le preocupan y los aspectos que menos le preocupan de la cotidianidad de su vida. Es importante recordar que de acuerdo a la forma cómo se redacte la pregunta, es decir, la claridad con que se realice la pregunta, el encuestado tendrá más elementos para dar una respuesta acertada, de esta forma, los resultados tendrán una mayor confiabilidad.

#### Ejemplo<sup>14</sup>

Todo el mundo tiene siempre alguna preocupación o está algo más preocupado por unas cosas que por otras. En estos momentos, y de estos cuatro aspectos de su vida, ¿cuál es el que más le preocupa, aunque sea sólo un poco más? ¿y el que menos le preocupa de estos cuatro aspectos, aunque sólo sea un poco menos que los demás?

	Aspecto que más le preocupa	Aspecto que menos le preocupa
1. Su vida afectiva	( )	( )
2. Su trabajo	( )	( )
3. La armonía relacionada con su familia.	( )	( )
4. Su salud.	( )	( )
9. NS/NC	( )	( )

En este caso se le da dos alternativas para que conteste el encuestado; una para los aspectos que más le preocupan y la otra para la respuesta de los aspectos que menos le preocupan, entonces se debe definir dos variables. Realice el siguiente procedimiento:

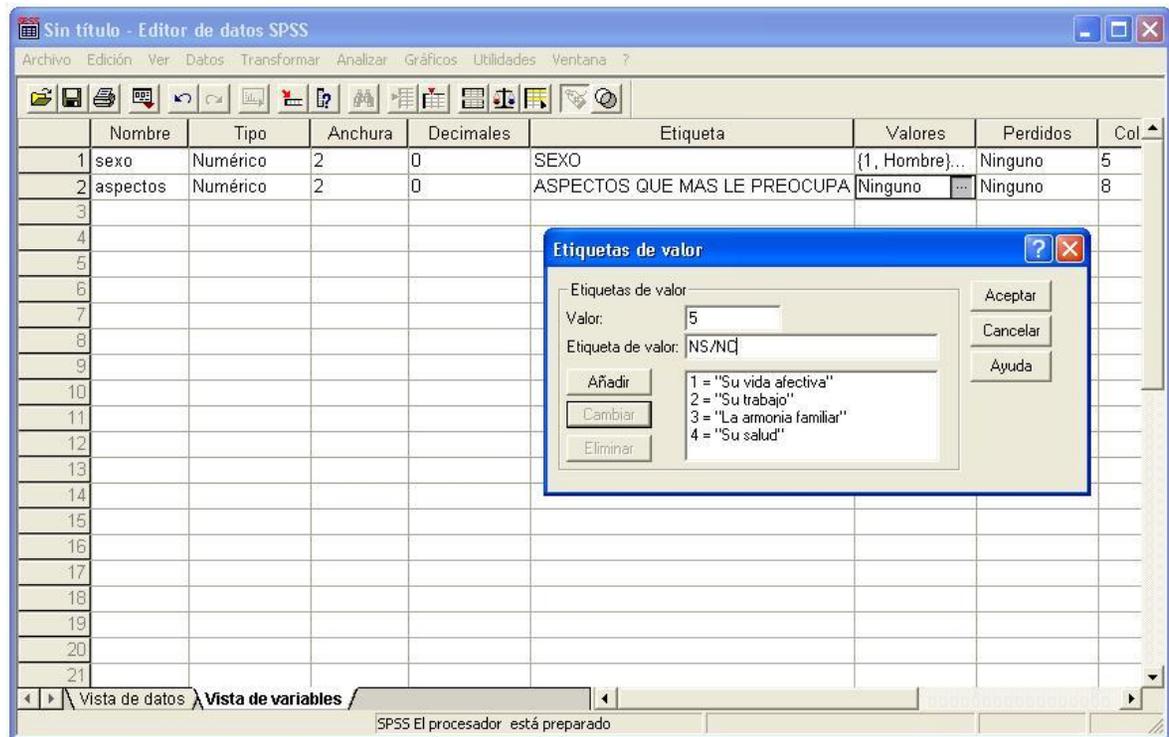
**Variable** “Aspectos que más le preocupan”

De acuerdo con lo visto en la pregunta dicotómica, procese la pregunta como aparece en la Figura 5.5.

---

<sup>14</sup> Visauta Vinacua. Bienvenido, Análisis Estadístico con SPSS para Windows, Estadística Básica. 1997, p. 279.

**Figura 5.5.** Caracterización de la Pregunta de Selección Múltiple.

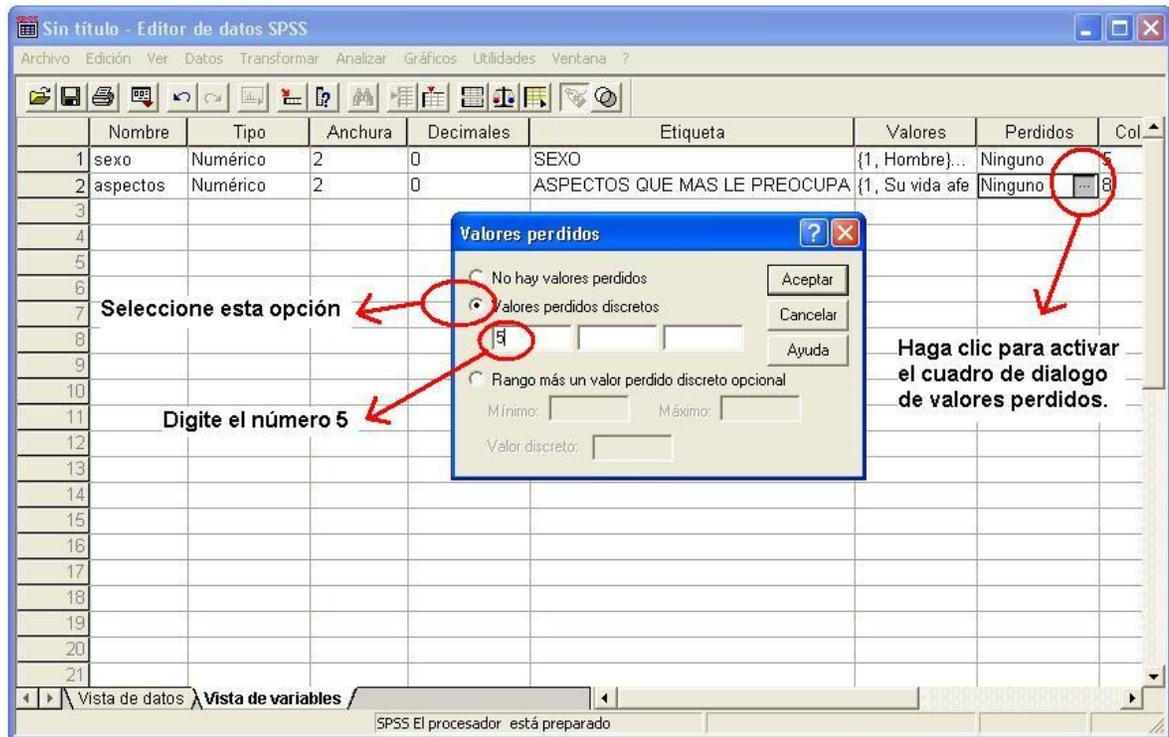


- Nombre = “Aspectos”
- Tipo = Numérico
- Anchura = 2
- Decimales = 0
- Etiqueta = “Aspectos que más le preocupan”

**Valores**

- 1 = “Su vida afectiva”
- 2 = “Su trabajo”
- 3 = “La armonía familiar”
- 4 = “Su salud”
- 5 = “NS/NC”

**Figura 5.6.** Cuadro de diálogo de valores perdidos



**5.3.1. Valores Perdidos.** En la pregunta se formuló la opción de NS/NC, considerada como variable o valor perdido. Para procesar esta opción se debe pulsar en el botón ... de la columna “Perdidos”, y a continuación se desplegará un cuadro de diálogo (figura 5.6.) active el botón “Valores perdidos discretos”. Digite 5 en el primer espacio.

**Complete el procesamiento de la pregunta así:**

- Alineación = Derecha
- Medida = Nominal

**5.3.2. Menú Contextual (Botón secundario del ratón).**



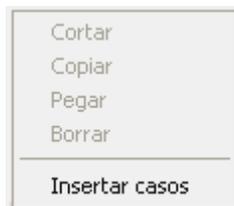
Como se debe crear otra variable para procesar **los aspectos que menos le preocupan**, pero si analiza, se dará cuenta que la estructura de la variable es la misma, únicamente cambiaría la etiqueta con “**ASPECTOS QUE MENOS LE PREOCUPAN**” y desde luego se le asignará otro nombre, muchas acciones de este tipo y de otros se pueden facilitar en gran medida empleando el menú contextual, que se activa al pulsar el botón derecho (secundario) del ratón. El menú contextual depende del lugar en el que se active. Se elige una acción de este menú situándose sobre ella y pulsando el botón derecho del ratón.

Veamos las diferentes opciones del menú contextual existentes en el SPSS, Cuando se pulsa el botón derecho del ratón en:

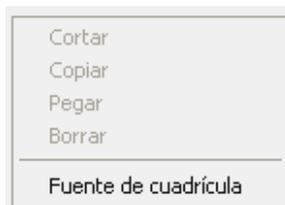
*El nombre de una variable*



*En el número de caso*

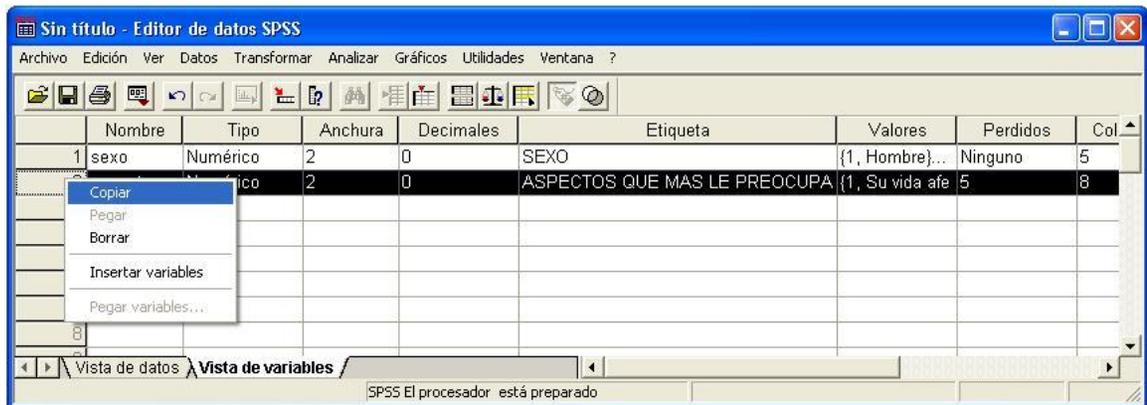


*Sobre una celdilla*

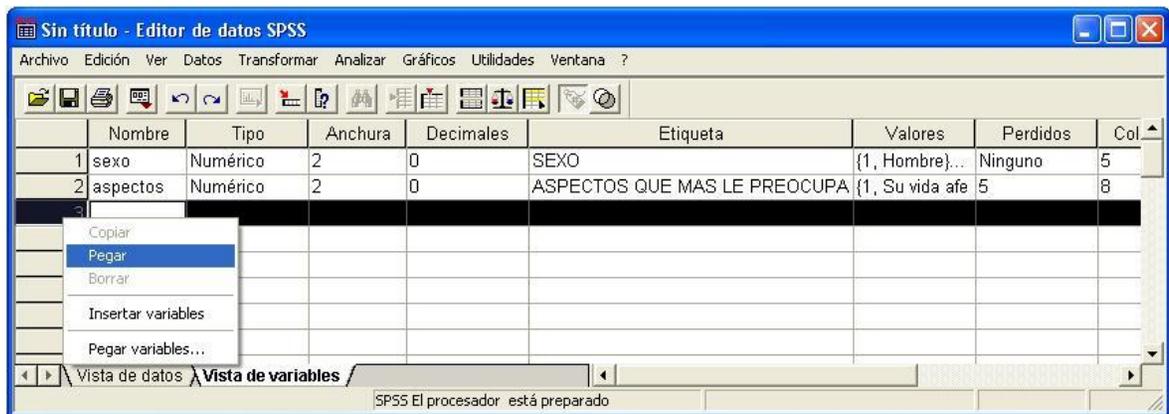


**Copiar una Variable.** Para copiar una variable se debe ubicar en el número del caso de la variable (Figura 5.7), luego hacer clic en el botón secundario del ratón, entonces se despliega el menú contextual y se selecciona la opción copiar. Luego se ubica en el número del caso donde se quiere copiar la variable (Figura 5.8), hacer clic en el botón secundario (botón derecho) y se selecciona la opción pegar.

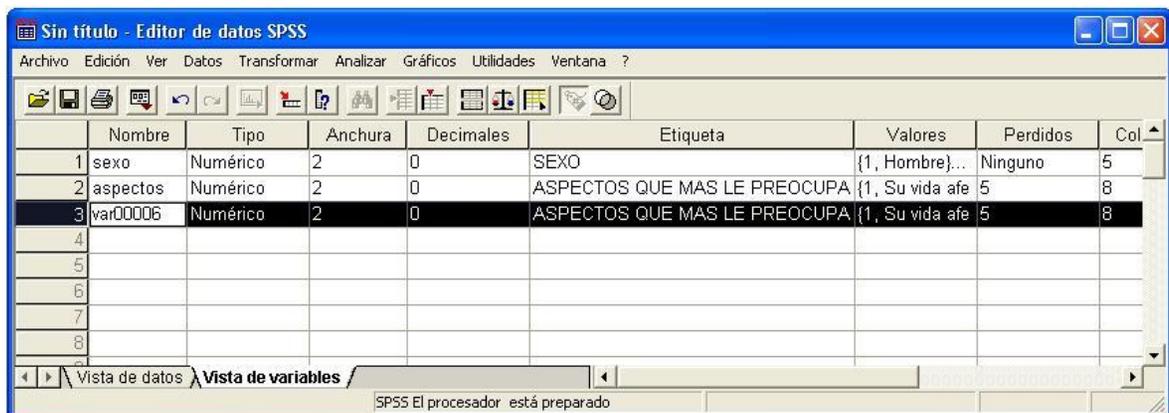
**Figura 5.7.** Menú contextual – pulsado en una variable



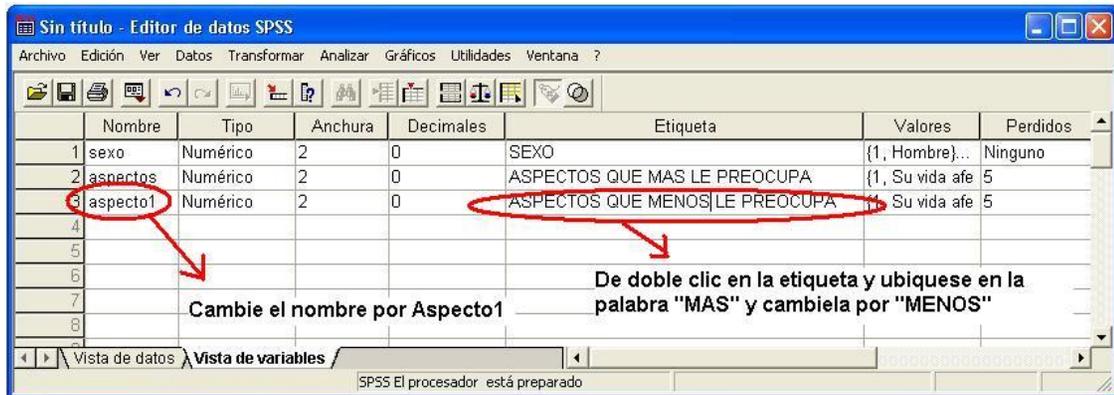
**Figura 5.8.** Selección de la opción pegar del menú contextual



**Figura 5.9.** Aspecto de la variable una vez pegada



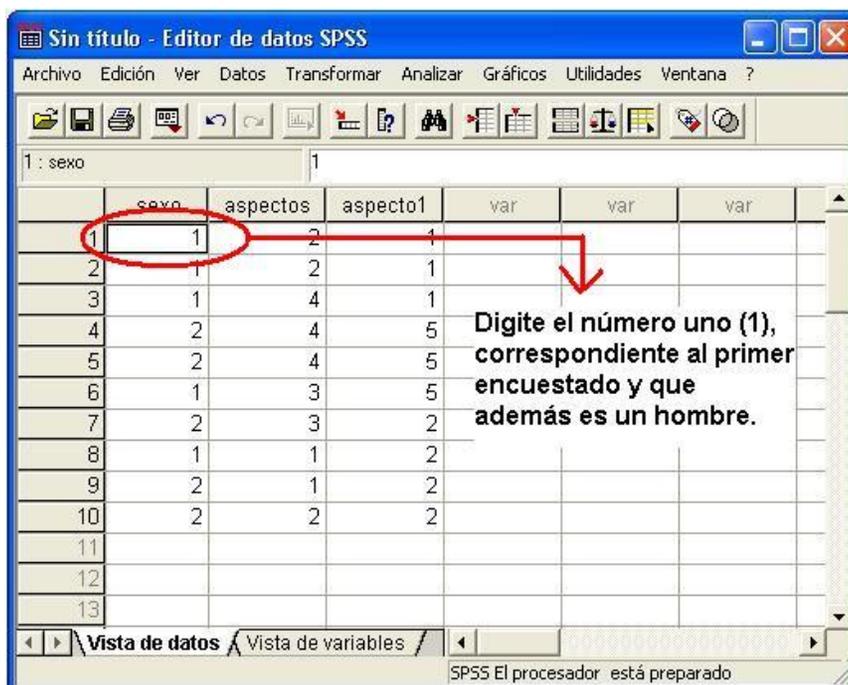
**Figura 5.10.** Realice los cambios del nombre de la variable y la etiqueta



Cuando se pega la variable, ésta tendrá un aspecto similar al que aparece en la Figura 5.9, es decir, al nombre de la variable se le asignó **var00007** en este caso, hay que cambiar este nombre por **ASPECTO1** y la etiqueta por “**ASPECTOS QUE MENOS LE PREOCUPAN**”(Figura 5.10).

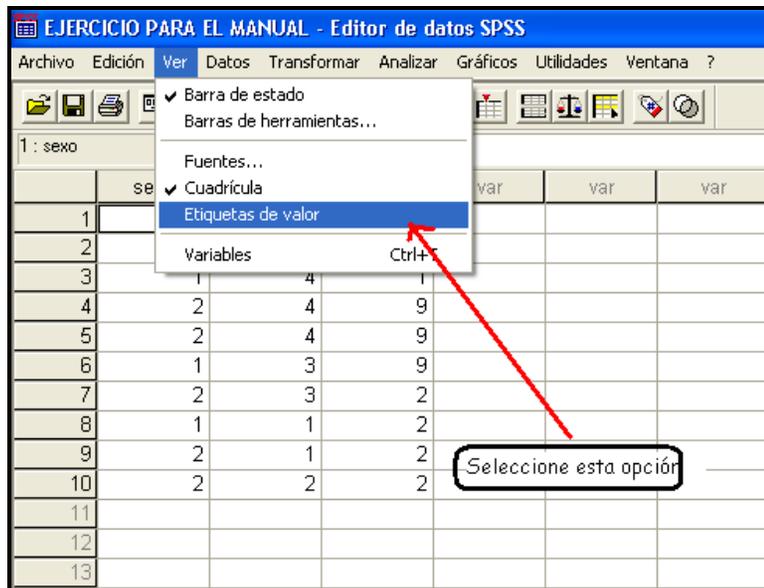
**5.3.3. Digite los datos.** Una vez definidas las variables se procede a digitar los datos. En la ventana de edición de datos, la celda activa aparece marcada con un borde más grueso e identificado en la parte superior izquierda de la pantalla (figura 5.11). Los valores digitados son: uno (1) que corresponde a la opción “Hombre” y dos (2) para la opción “Mujer”, esto para la variable **sexo**.

**Figura 5.11.** Ingreso de datos.

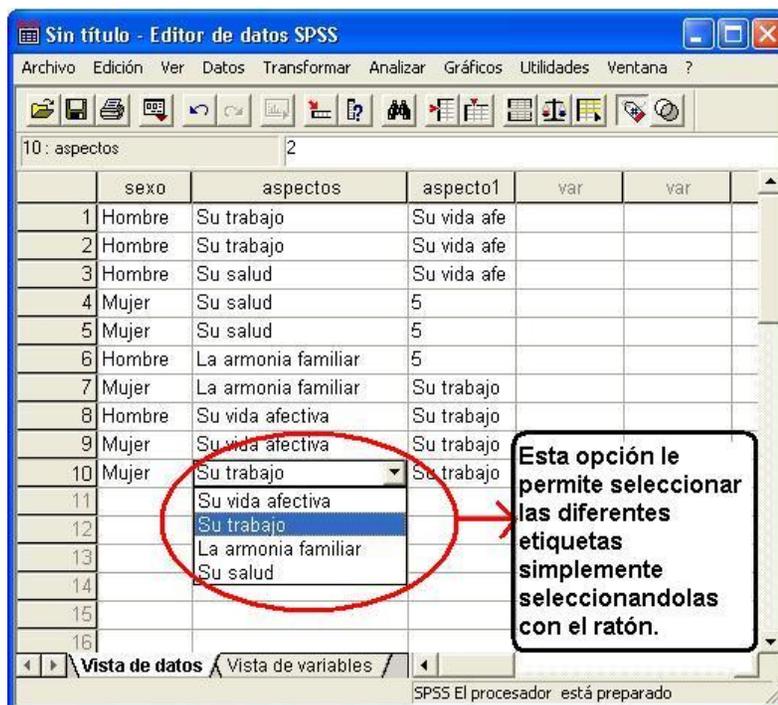


Hay otra forma de ver la información que se está digitando mediante la opción “VER” se activa la opción “Etiquetas de Valor” (Figuras 5.12 y 5.13).

**Figura 5.12.** La opción Etiquetas de valor del menú ver.

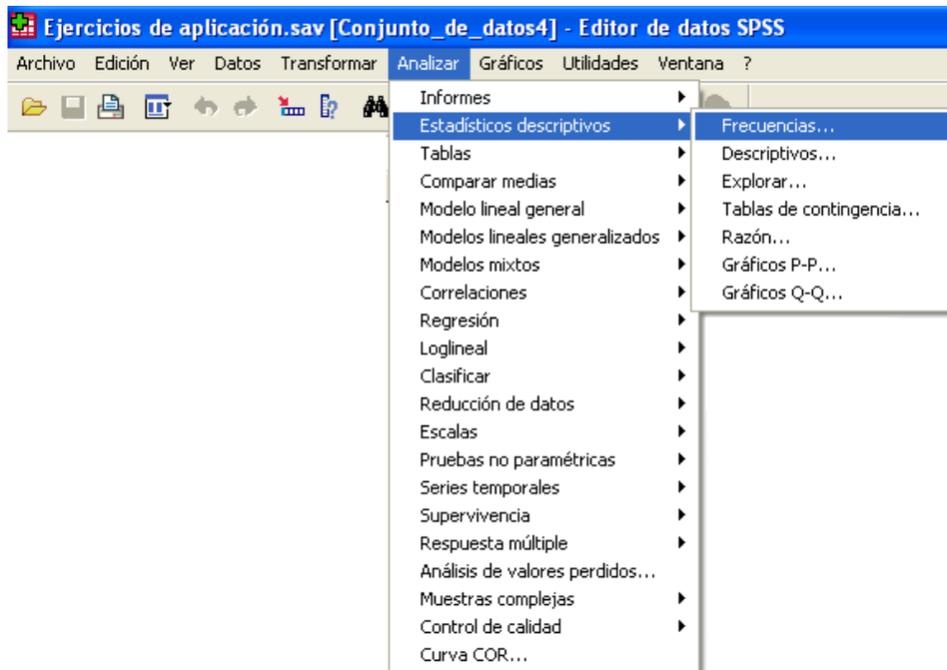


**Figura 5.13.** Aspectos de los datos con la opción etiquetas de valor activada.



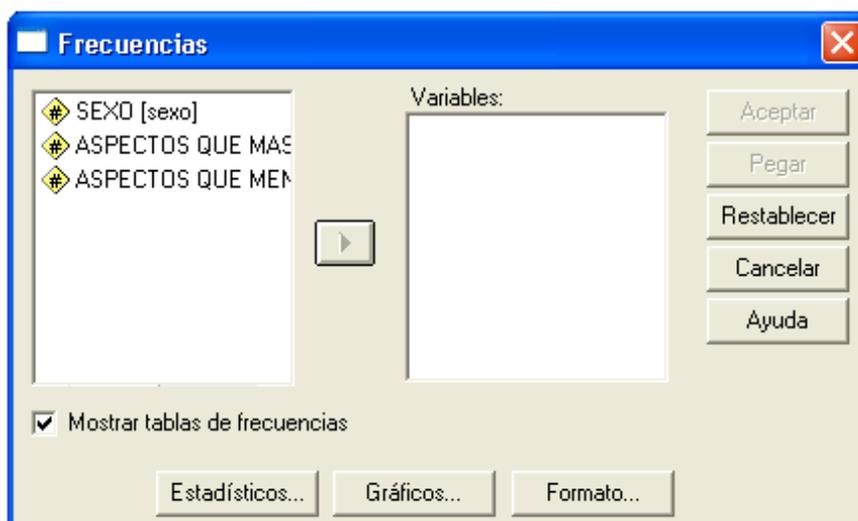
**5.3.4. Análisis de la información.** Para realizar el análisis o procesamiento de la información se selecciona la opción **Analizar**, de menú principal, luego seleccione la opción **Estadísticos descriptivos**, luego **Frecuencias**.

**Figura 5.14.** Proceso de análisis de la información.



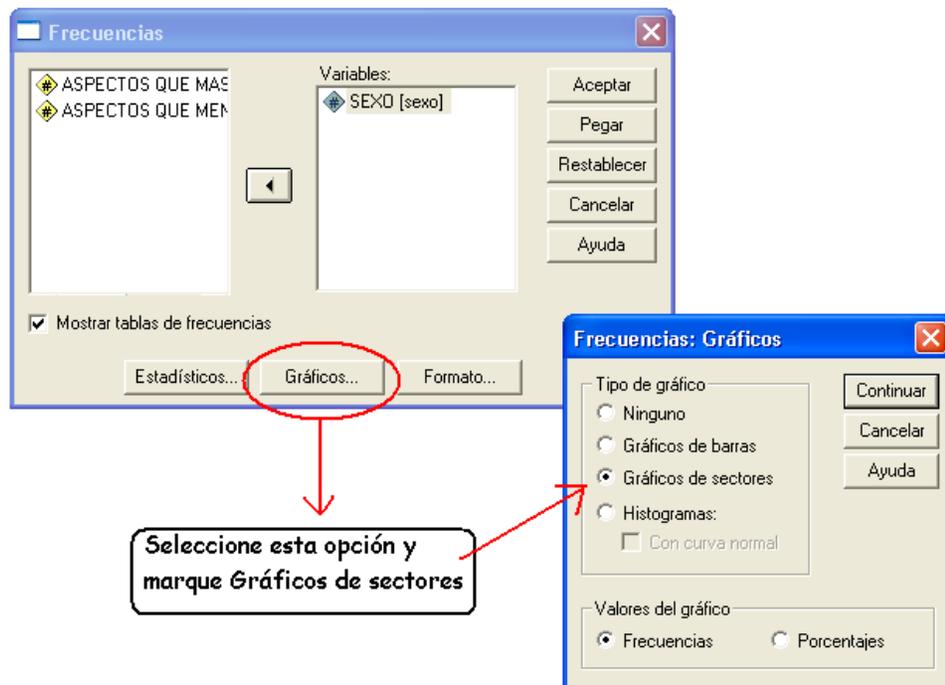
Cuando selecciona **Frecuencias**, aparece un cuadro de diálogo (Figura 5.15), realice las siguientes instrucciones:

**Figura 5.15.** Cuadro de dialogo frecuencias.



1. Seleccione la variable que va a procesar, de esta forma se activará el icono de flecha de dirección 
2. Haga clic en  para pasar la variable a la sección de **variables**, complete el procedimiento como se presenta en la figura 5.16.

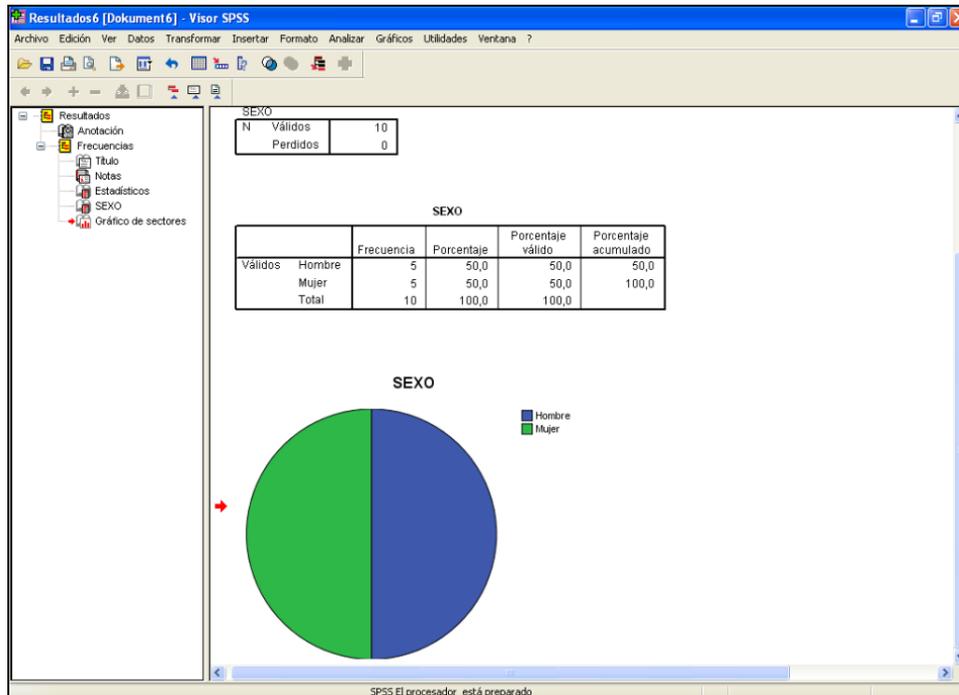
**Figura 5.16.** Cuadro de Diálogo Frecuencias y Gráficos



3. Una vez completo el procedimiento, presione **Continuar**
4. Luego presione **Aceptar**

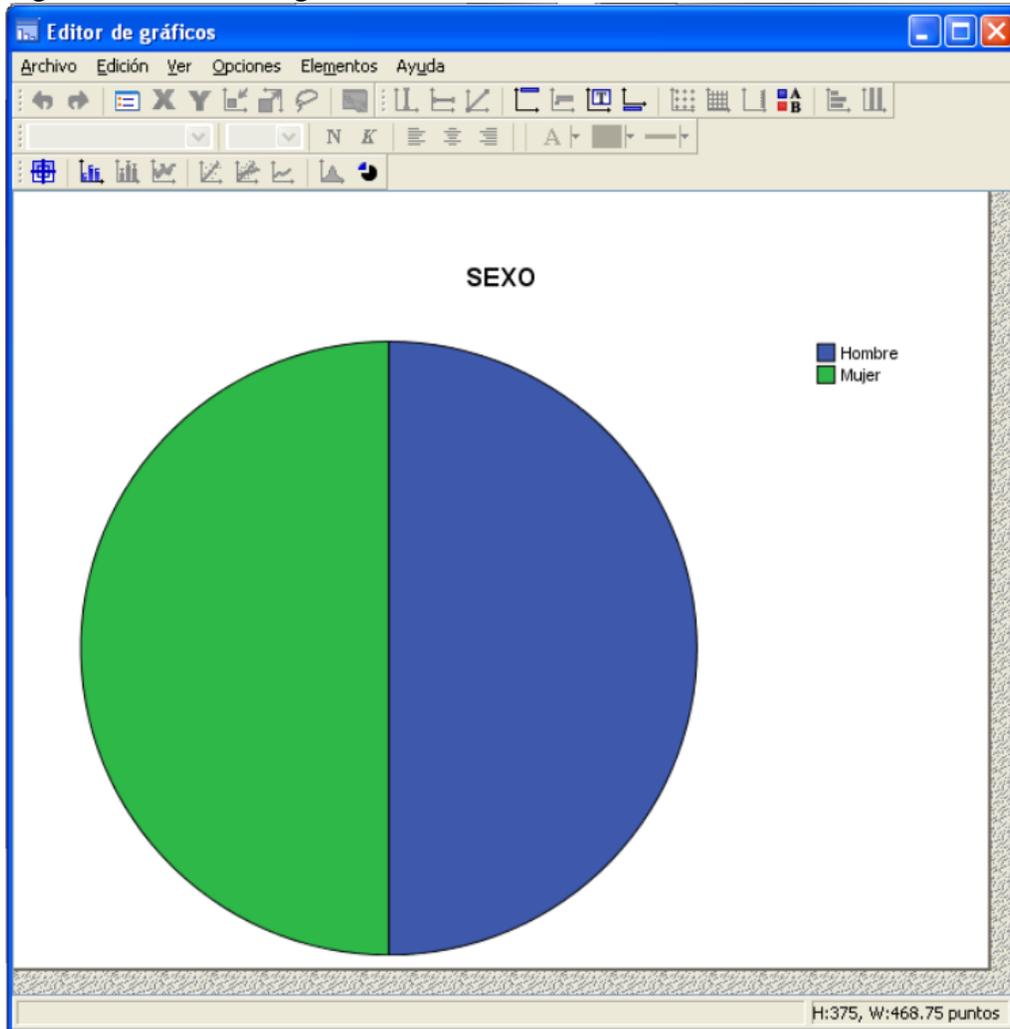
La pantalla de su computadora debe tener una apariencia similar a la de la figura 5.17. Recuerde que debe grabar los resultados con otro nombre, es decir, que cada vez que trabaje con el programa SPSS, usted tendrá dos archivos, uno de datos y otro de resultados.

Figura 5.17. Visor de Resultados.



**5.3.5. Edición de gráficos.** Para editar el gráfico, pulse dos veces en el gráfico para abrirlo en el editor de gráficos (Figura 5.18).

Figura 5.18. Editor de gráfico



Puede editar los gráficos de diversas formas:

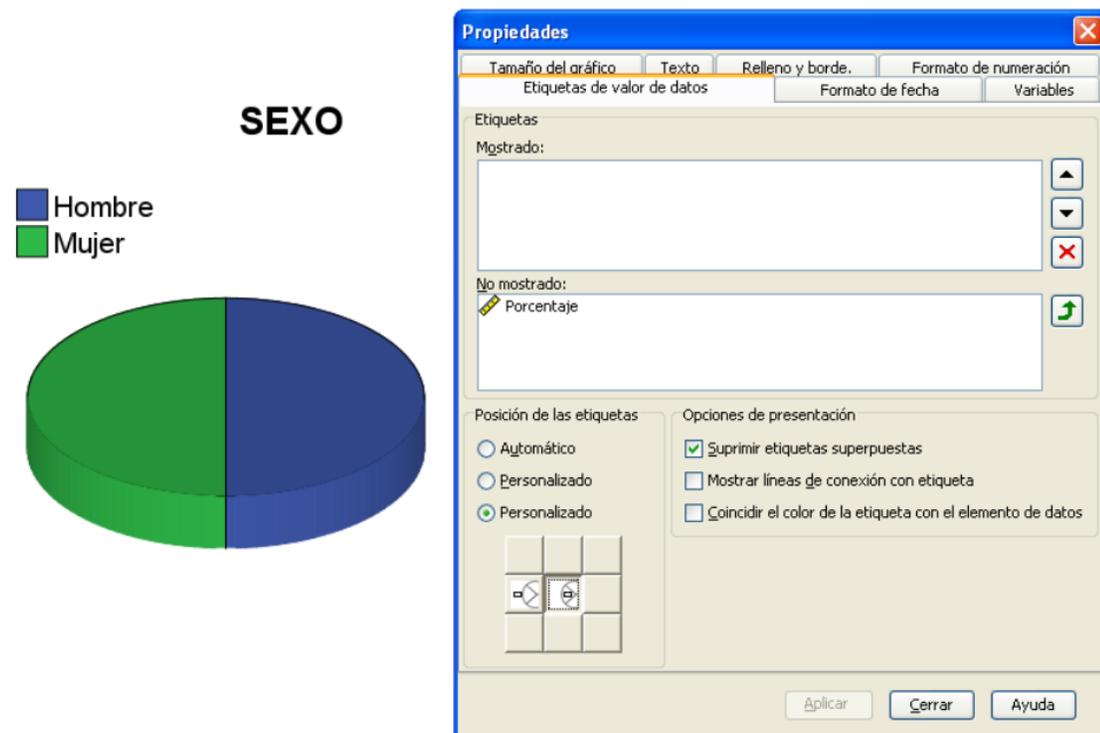
- Se puede cambiar los colores
- Cambiar el formato de los números de las etiquetas
- Editar los textos
- Mostrar las etiquetas de valores de datos
- Utilizar plantillas gráficas

A continuación se explica como mostrar los porcentajes y valores de las variables de la gráfica.

**5.3.5.1. Mostrar los porcentajes y valores de las variables graficadas.** Para mostrar valores y porcentajes elija en el menú del editor de gráficos las siguientes opciones:

- Elementos
- Mostrar etiquetas de datos.

**Figura 5.19.** Opción mostrar etiqueta de valor



Para mostrar las etiquetas de valor seleccione Porcentajes de la sesión *No mostrado* y trásela a la sesión de *Mostrado*, luego de este procedimiento su grafico debe tener una apariencia como el de la Figura 5.19.

**5.3.5.2. Editor de propiedades del grafico de sectores.** Para trabajar con el editor de propiedades del grafico de sectores, pulse dos veces sobre el grafico, (más exactamente los dos sectores del pastel en este caso, deben estar seleccionados en el momento en que usted pulsa las dos veces) realizada esta actividad aparece la ventana de editor de grafico (Figura 4.21).

También puede trabajar con el editor de propiedades del grafico de sectores, utilizando el menú del editor de gráficos. Para trabajar con el menú debe tener previamente el grafico seleccionado (es decir, los sectores en este caso del pastel deben estar seleccionados), realizada esta labor, seleccione del menú de gráficos las siguientes opciones.

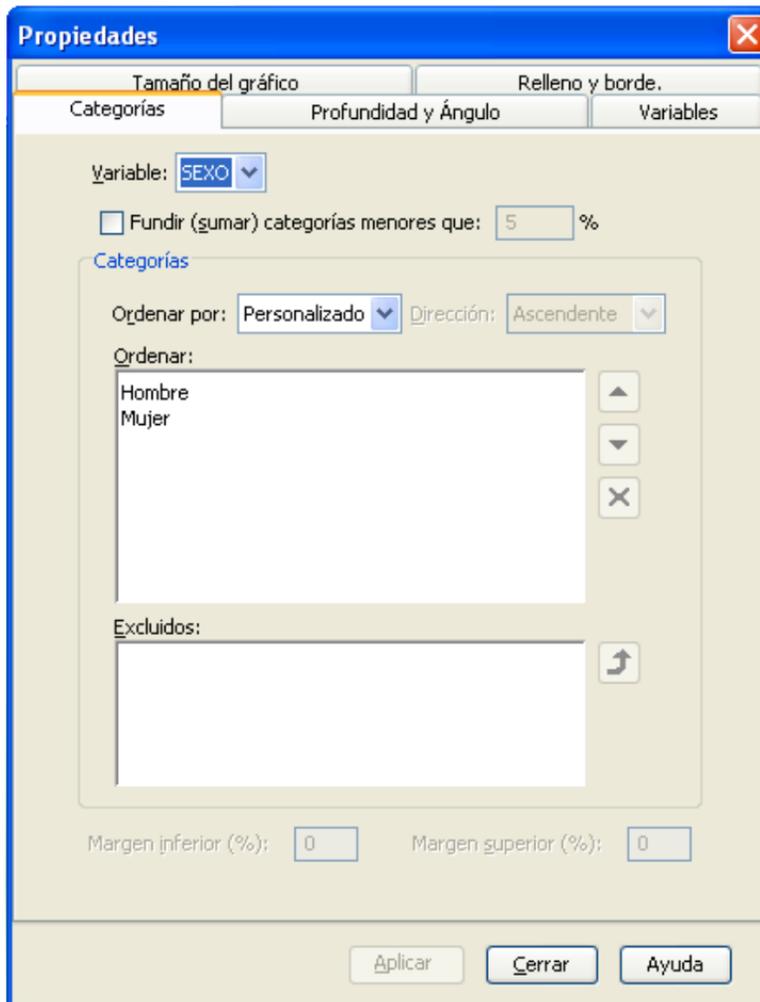
- Edición
- Propiedades (Figura 5.20)

Figura 5.20. Menú edición



Una vez hecha la selección aparece la ventana de propiedades que presenta las siguientes pestañas (Figura 4.21):

Figura 5.21. Ventana de propiedades del grafico de sectores.

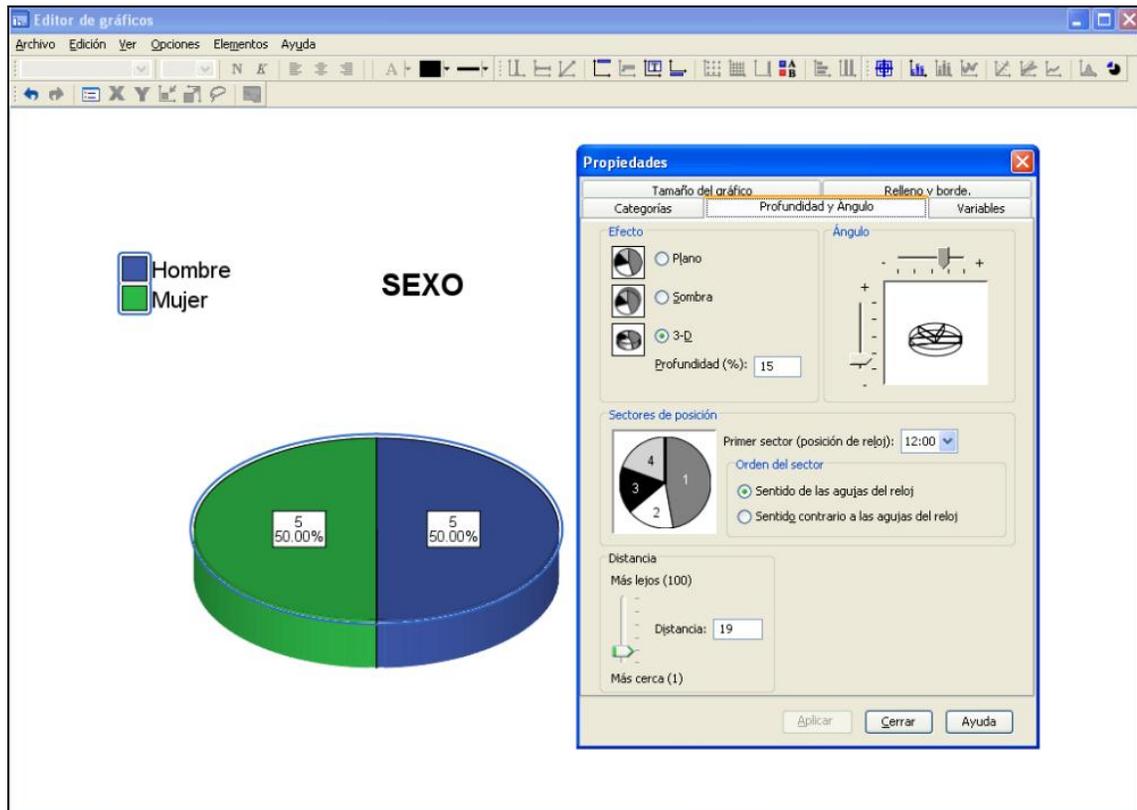


- Categorías
- Profundidad y ángulo
- Variables
- Tamaño del Gráfico y
- Relleno y borde

Con estas opciones se podrá cambiar el grafico a tres dimensiones, también modificar la presentación de las variables, modificar el tamaño del gráfico, entre otros.

**5.3.5.3. Transformar el grafico en tres dimensiones.** Para transformar el grafico en tres dimensiones seleccione la pestaña de propiedades. Profundidad y ángulo.

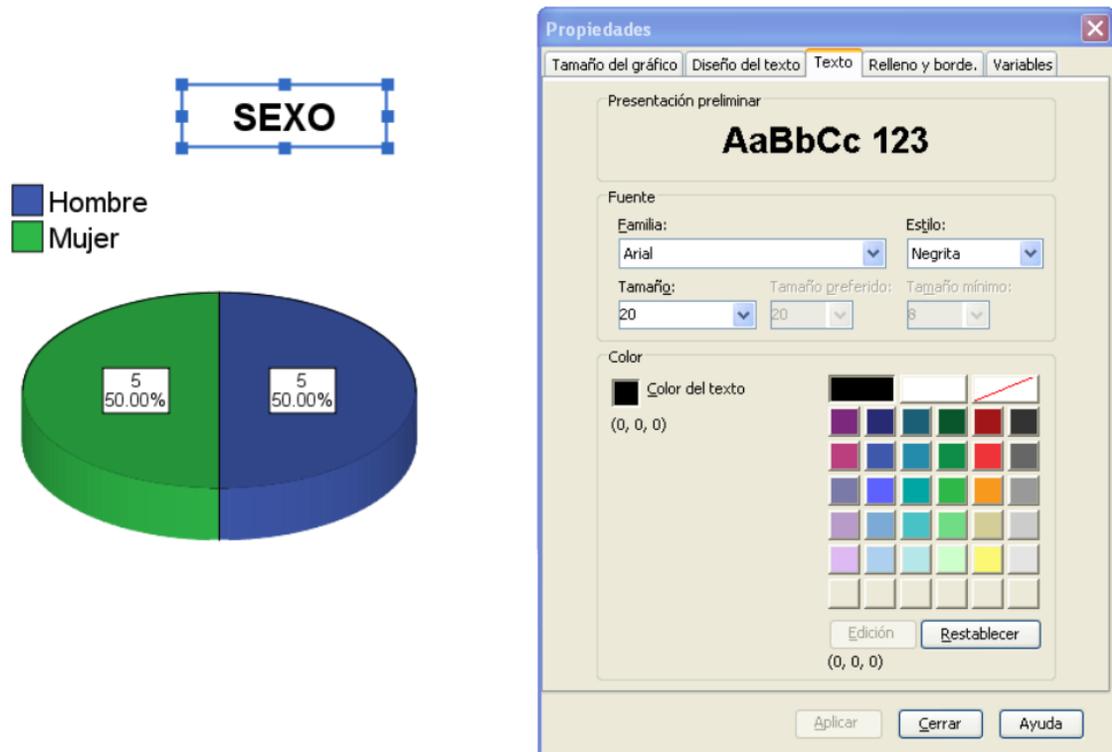
Figura 5.22. Opciones de la pestaña Profundidad y Ángulo.



- Seleccione 3-D en la sesión de EFECTO
- Digite 15 en la sesión de PROFUNDIDAD
- Digite 19 en la sesión de DISTANCIA
- Deje las demás opciones como aparece en la Figura 5.22.

Estas pestañas cambian según el elemento del gráfico que seleccione en el editor de gráficos. En la Figura 5.23 se presenta el editor del título del gráfico.

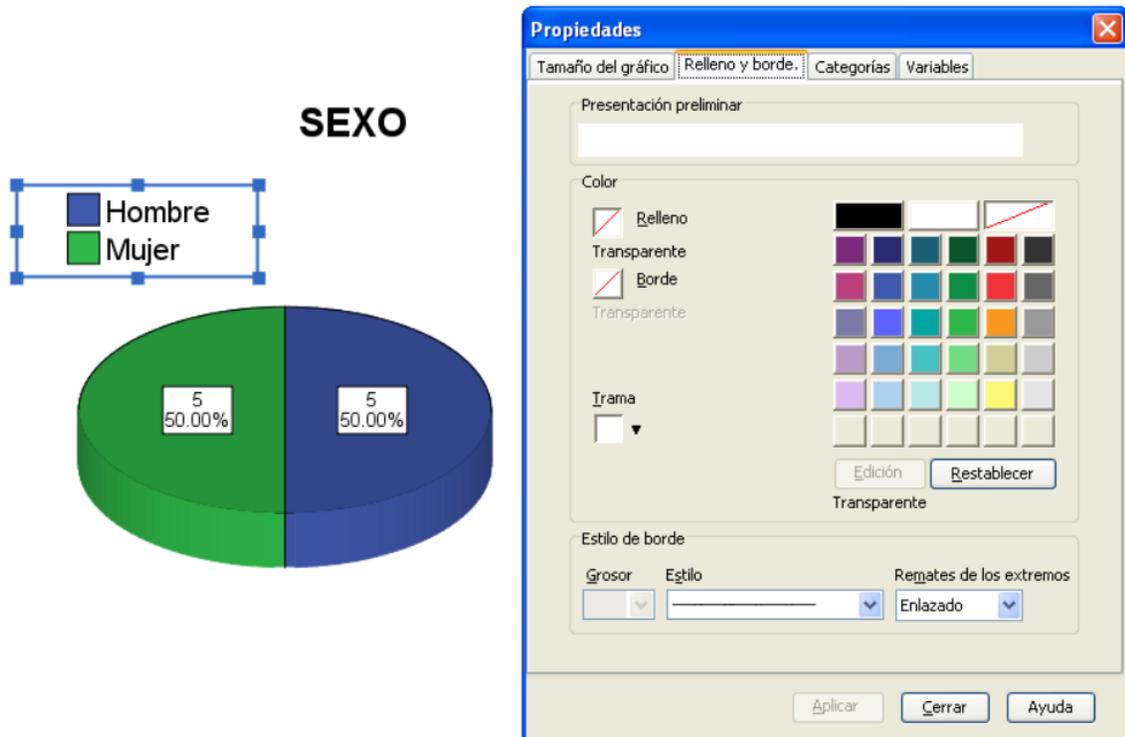
Figura 5.23. Editor del título del gráfico.



Seleccione previamente el título de gráfico luego haga el siguiente procedimiento:

- Seleccione del menú del editor de gráfico la opción *EDICIÓN*
- Seleccione el submenú *PROPIEDADES*, con el fin de tener acceso a la ventana de propiedades de títulos. Figura 5.23.

Figura 5.24. Editor de propiedades de leyendas de gráficos.



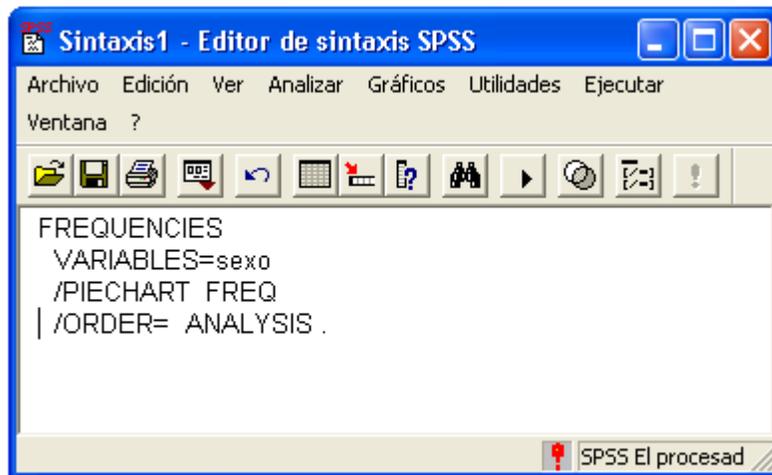
Seleccione previamente la leyenda del gráfico luego haga el siguiente procedimiento:

- Seleccione del menú del editor de gráfico la opción *EDICIÓN*
- Seleccione el submenú *PROPIEDADES*, con el fin de tener acceso a la ventana de propiedades de la leyenda. Figura 5.24.

**5.3.6. Lenguaje de comandos (Sintaxis).** Hay dos maneras de trabajar con SPSS para Windows: una es emplear los cuadros de diálogos (tal como se ha hecho hasta ahora), y la otra consiste en usar el lenguaje de comandos (sintaxis). También se puede utilizar una estrategia mixta realizando algunas acciones con cuadros de diálogos y otras con comandos.

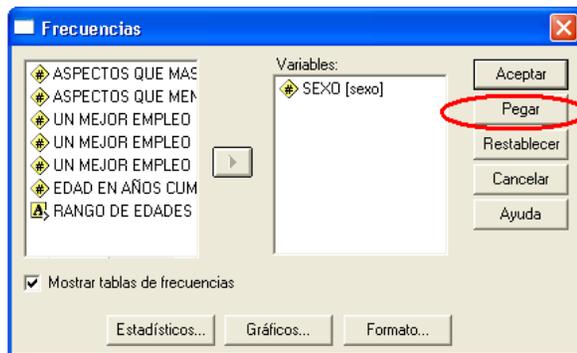
La principal ventaja de los comandos es que permite la automatización del trabajo. Los comandos aparecen en la ventana de sintaxis. Para abrir la ventana de sintaxis desde el editor de datos elija **Archivo/Nuevo/Sintaxis**.

**Figura 5.25.** Editor de Sintaxis



En esta figura se presenta la sintaxis del procesamiento de la variable sexo, realizado en la sección anterior mediante los cuadros de diálogo. La sintaxis se obtuvo activando la opción **pegar** del cuadro de diálogo **Frecuencias** (Figura 5.26).

**Figura 5.26.** Opción pegar para obtener la sintaxis de un procedimiento



**Figura 5.27.** Forma de ejecutar los comandos



Para ejecutar los comandos de sintaxis active la opción **Ejecutar** y luego de un clic en **Todo** (Figura 5.27).

Esta opción de cuadros de sintaxis se deja a libre discreción de las personas que gustan de la programación.

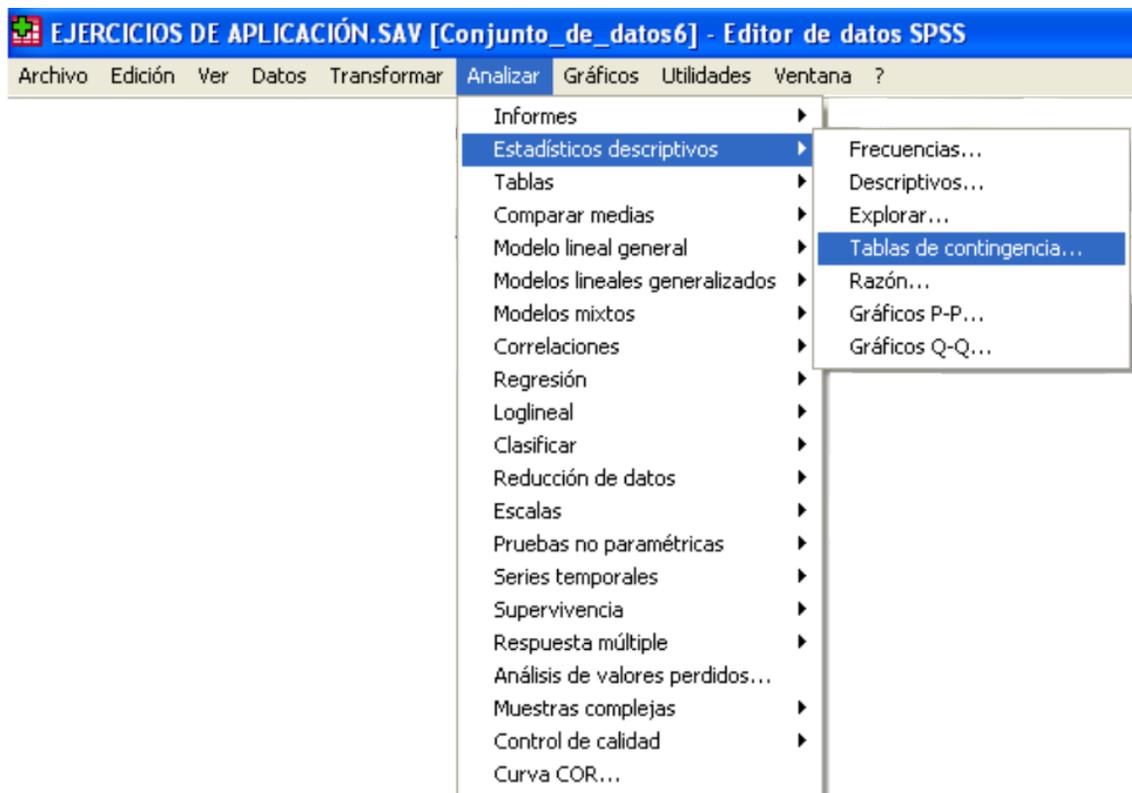
**5.3.7. Relación entre variables -Tablas de contingencia.** El SPSS puede cruzar hasta tres (3) variables; además las tablas de contingencia se emplean para hallar simultáneamente distribuciones de frecuencia de dos variables y para encontrar índices estadísticos que midan la fuerza de la asociación entre las variables.

Las variables Sexo, Aspectos que más le preocupan y Aspectos que menos le preocupan, corresponden a variables nominales. Se va a estudiar la relación entre la variable sexo y los Aspectos que más le preocupan, con el fin de hallar una distribución de frecuencia cruzada (tabla de contingencia).

Para construir una tabla de contingencia realice los siguientes pasos:

1. Seleccione **Analizar**.
2. **Estadísticos descriptivos** (Figura 5.28)
3. **Tablas de contingencia**, aparecerá el cuadro de diálogo (Figura 5.29).

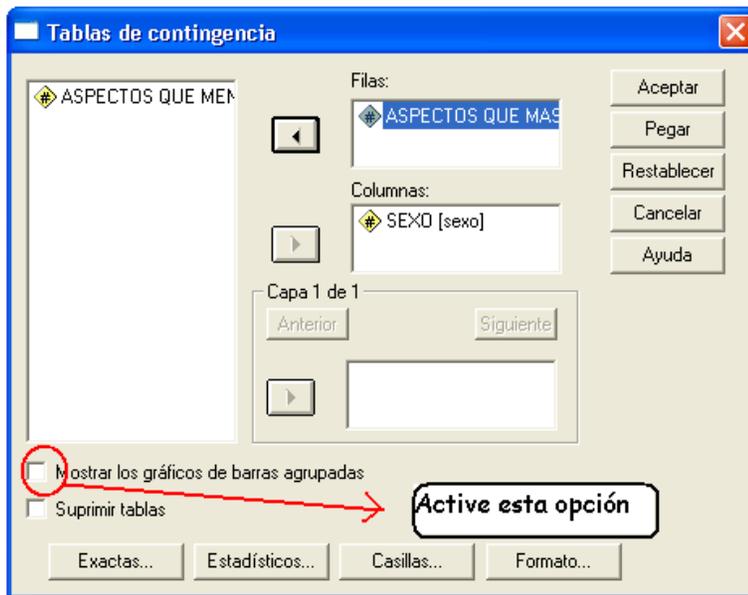
**Figura 5.28.** Tablas de contingencia.



4. En el cuadro de diálogo seleccione la variable Sexo, trasládela a la sección "Columnas".

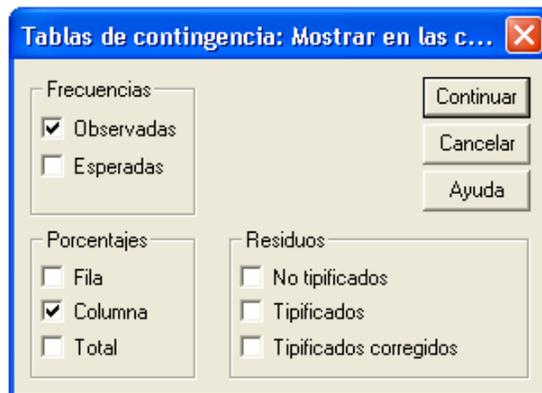
5. Seleccione la variable “Aspectos que más le preocupan” y luego trasládelala a la sección “Filas”.
6. Active la opción de **Mostrar los gráficos de barras agrupadas**, figura 5.29.

**Figura 5.29.** Tablas de contingencias.



7. Haga clic en la opción **Casillas** y active **Columnas** en la sección de **Porcentajes**. Esta opción permite ver la tabla de frecuencias con porcentajes por la variable que se ubicó en las columnas. Figura 5.30.

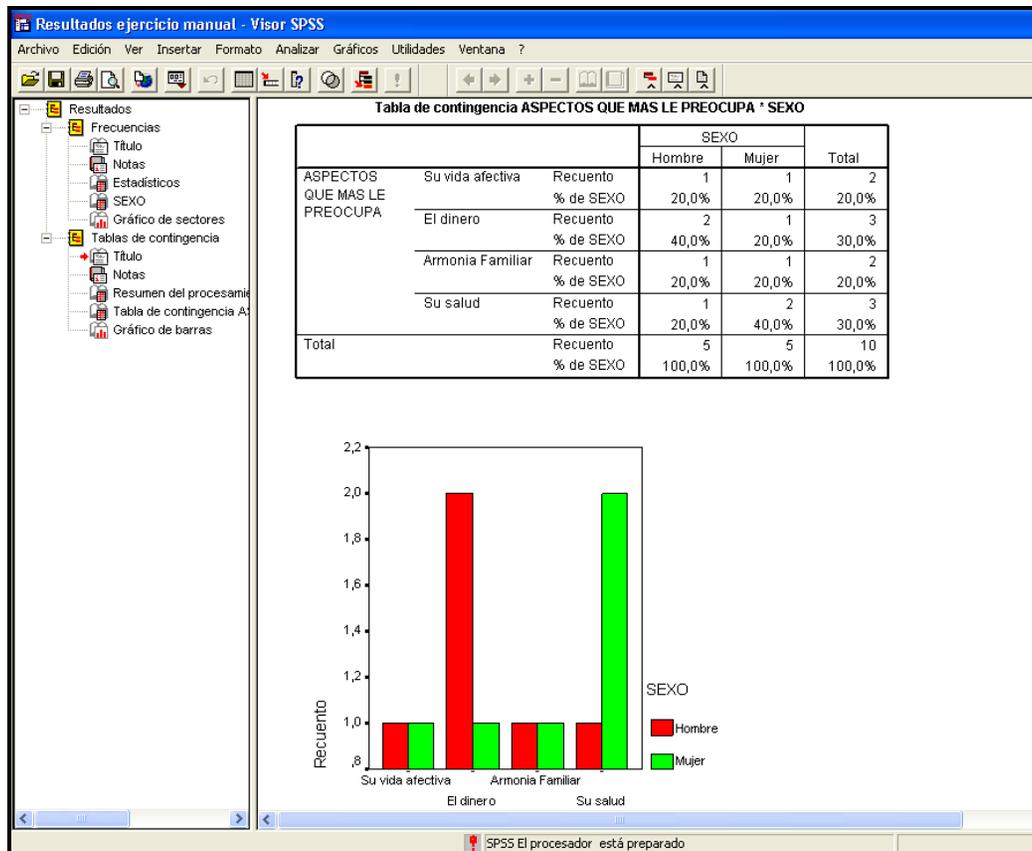
**Figura 5.30.** Cuadro de diálogo de Casillas Sección porcentajes.



8. Haga clic en **Continuar**.
9. Haga clic en **Aceptar**.

La información procesada debe ser similar a la que aparece en el visor de resultados, de la Figura 5.31.

**Figura 5.31.** Visor de resultados.



Como se observa, los porcentajes están referenciados sobre un 100% de hombres y un 100% de mujeres.

**5.3.8. Lenguaje de Comandos (Sintaxis).** El procedimiento anterior se presenta en el cuadro de sintaxis, para los estudiantes que gustan de la programación.

**Figura 5.32.** Sintaxis de la opción de tablas de contingencia.



**5.3.9. Tablas de Contingencia Respuestas Múltiples.** En algunas ocasiones una pregunta puede tener más de una respuesta. Al conjunto de las posibles respuestas se llamará variable de respuesta múltiple. Para ello se puede hallar distribuciones de frecuencias y tablas cruzadas de frecuencias. Este tipo de preguntas permiten que el encuestado seleccione un número determinado de respuestas ante una variedad de opciones.

**Pregunta<sup>15</sup>.** Al evaluar si un empleo es mejor o peor que otro ¿qué es lo que para usted es realmente más importante?, ¿Y lo más importante en segundo lugar? ¿Y en tercer lugar?.

	OPCIONES		
	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>
1 Que sea seguro (estable en el tiempo)	( )	( )	( )
2 Una buena paga (retribución, sueldo)	( )	( )	( )
3 Buenas oportunidades para ascender	( )	( )	( )
4 Que deje mucho tiempo libre	( )	( )	( )
5 Que sea interesante	( )	( )	( )
6 Que permita trabajar con independencia	( )	( )	( )
7 Que sirva para ayudar a la gente	( )	( )	( )
8 Que sea útil para la sociedad	( )	( )	( )
9 Que tenga horario flexible	( )	( )	( )
10 Que me permita adquirir más conocimientos	( )	( )	( )
11 Buenas relaciones con los compañeros	( )	( )	( )
12 Buen ambiente externo	( )	( )	( )
13 Que sea un trabajo variado	( )	( )	( )
14 Otra	( )	( )	( )
15 NS/NC	( )	( )	( )

<sup>15</sup> *Ibíd.* p. 286.

El encuestado tiene la oportunidad de seleccionar tres opciones, entonces se debe crear tres variables.

Realice las siguientes instrucciones:

1. Crear la primera variable con el nombre = Trabajo1
2. Tipo = Numérico
3. Anchura = 2
4. Decimales = 0
5. Etiqueta = UN MEJOR EMPLEO OPCIÓN 1
6. Valores =

- 1 = Que sea seguro (estable en el tiempo)
- 2 = Una buena paga (retribución, sueldo)
- 3 = Buenas oportunidades para ascender
- 4 = Que deje mucho tiempo libre
- 5 = Que sea interesante
- 6 = Que permita trabajar con independencia
- 7 = Que sirva para ayudar a la gente
- 8 = Que sea útil para la sociedad
- 9 = Que tenga horario flexible
- 10 = Que me permita adquirir más conocimientos
- 11 = Buenas relaciones con los compañeros
- 12 = Buen ambiente externo
- 13 = Que sea un trabajo variado
- 14 = Otra
- 15 = NS/NC

7. Perdidos = 15
8. Columnas = 8
9. Alineación = Derecha
10. Medida = Escala

El visor de “Vista de variables”, debe quedar igual como en la figura 55.

**Figura 5.33.** Especificación de pregunta de selección múltiple.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alinea
1	sexo	Numérico	2	0	SEXO	{1, Hombre}...	Ninguno	7	Izquierc
2	aspectos	Numérico	2	0	ASPECTOS QUE MAS LE PREOCUPA	{1, Su vida afe	5	16	Izquierc
3	aspecto1	Numérico	2	0	ASPECTOS QUE MENOS LE PREOCUPA	{1, Su vida afe	5	8	Izquierc
4	trabajo1	Numérico	2	0	UN MEJOR EMPLEO OPCION 1	{1, Que sea se	15	8	Izquierc
5									
6									
7									
8									

11. Copiar haciendo uso del menú interactivo, las dos variables, pero realice los siguientes cambios:

1. Nombre: Trabajo2. Para la segunda variable.
2. Etiqueta: UN MEJOR EMPLEO OPCIÓN 2
3. Nombre: Trabajo3. Para la tercera variable.
4. Etiqueta: UN MEJOR EMPLEO OPCIÓN 3. El visor de variables debe tener una apariencia similar al presentado en la figura 5.34.

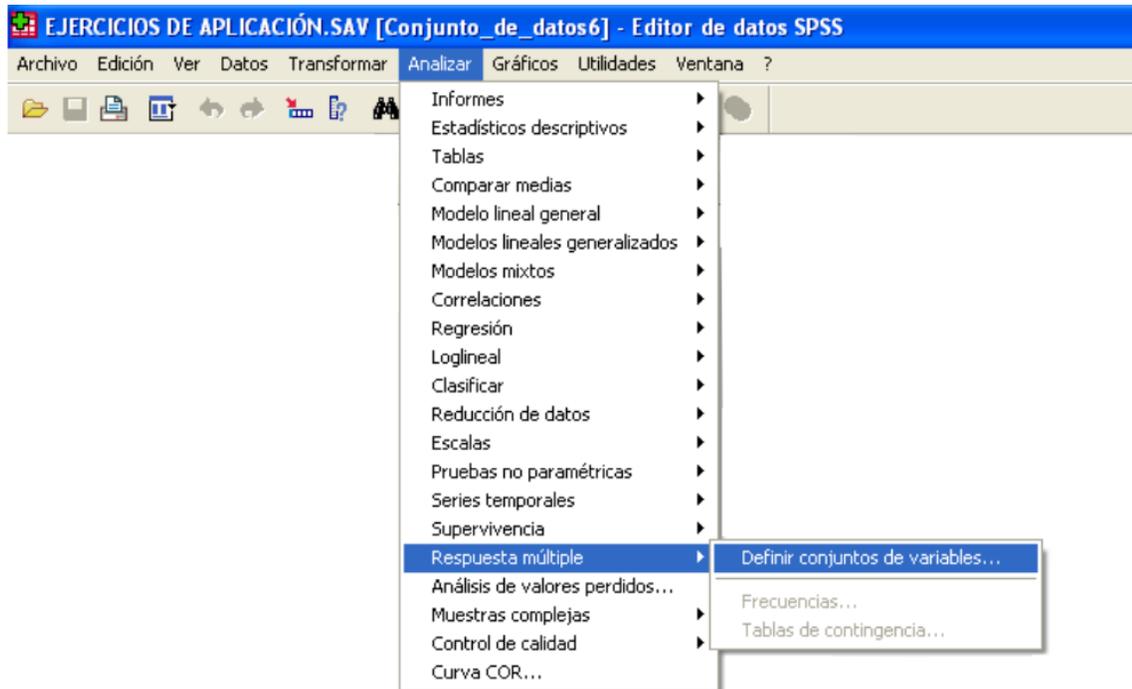
**Figura 5.34.** Caracterización de la Variable de Selección Múltiple.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alinea
1	sexo	Numérico	2	0	SEXO	{1, Hombre}...	Ninguno	7	Izquierc
2	aspectos	Numérico	2	0	ASPECTOS QUE MAS LE PREOCUPA	{1, Su vida afe	5	16	Izquierc
3	aspecto1	Numérico	2	0	ASPECTOS QUE MENOS LE PREOCUPA	{1, Su vida afe	5	8	Izquierc
4	trabajo1	Numérico	2	0	UN MEJOR EMPLEO OPCION 1	{1, Que sea se	15	8	Izquierc
5	trabajo2	Numérico	2	0	UN MEJOR EMPLEO OPCION 2	{1, Que sea se	15	8	Izquierc
6	trabajo3	Numérico	2	0	UN MEJOR EMPLEO OPCION 3	{1, Que sea se	15	8	Izquierc
7									
8									

El procedimiento para la creación de tablas de contingencias de múltiples respuestas es el siguiente:

1. Seleccione **Analizar** del menú principal.
2. **Respuestas múltiples.**
3. **Definir conjuntos de variables.** Cuando usted selecciona esta opción aparece el cuadro de diálogo de la Figura 5.36.

**Figura 5.35.** Pregunta de selección múltiple.



En el cuadro de diálogo (Figura 5.36) seleccione las tres variables, trasládelas a la sección **Variables de Conjunto**, active la opción de **Categorías**, coloque en **Rango = 1 Hasta 14**, Tenga en cuenta que la opción 15 no se tiene en cuenta, por corresponder a una variable tratada como perdida.

**Complete el procedimiento con:**

**Nombre:** TRABAJO

**Etiqueta:** OPCIONES DE MEJOR EMPLEO

**Figura 5.36.** Definición de conjuntos de respuesta múltiple.



Haga clic en **Añadir**, y el nombre quedará definido como **\$trabajo**, luego haga clic en **cerrar**.

**5.3.10. Procesamiento de pregunta de selección múltiple.** Para procesar este tipo de variable realice las siguientes instrucciones:

1. Seleccione **Analizar**.
2. **Múltiples Respuestas**.
3. **Frecuencias**, cuando usted selecciona esta opción aparece un cuadro de dialogo, (figura 5.37).

**Figura 5.37.** Frecuencias de respuesta múltiple.



4. En el cuadro de diálogo seleccione la variable y trasládela a la sección **Tablas para:**
5. Haga clic en **Aceptar**.
6. Los resultados tendrán una apariencia como la mostrada en la figura 5.38.

**Figura 5.38 -A** Resultados de la Pregunta de Respuesta Múltiple

**Resumen de los casos**

	Casos					
	Válidos		Perdidos		Total	
	Nº	Porcentaje	Nº	Porcentaje	Nº	Porcentaje
\$Trabajo <sup>a</sup>	10	100,0%	0	,0%	10	100,0%

a. Agrupación

Figura 5.38 -B

## Frecuencias \$TRABAJO

		Respuestas		Porcentaje de casos
		Nº	Porcentaje	
OPCIONES DE MEJOR EMPLEO	Que sea seguro	8	26,7%	80,0%
	Una buena paga	6	20,0%	60,0%
	Buenas oportunidades para ascender	2	6,7%	20,0%
	Que deje mucho tiempo libre	1	3,3%	10,0%
	Que sea interesante	2	6,7%	20,0%
	Que me permita trabajar con independencia	1	3,3%	10,0%
	Que sirva para ayudar a la gente	3	10,0%	30,0%
	Que sea útil para la sociedad	3	10,0%	30,0%
	Buen ambiente externo	2	6,7%	20,0%
	Que sea un trabajo variado	2	6,7%	20,0%
	Total	30	100,0%	300,0%

a. Agrupación

**Interpretación.** Hay 8 personas que seleccionaron la opción “QUE SEA SEGURO” y hay 30 respuestas en total (Figura 5.38-B).

**La columna Respuestas – porcentaje.** Se calcula sobre el total de respuestas es decir:

$$\text{Calculo del porcentaje} = \frac{8}{30} \times 100 = 26,7\%$$

Es decir, que el 27% aproximadamente de los encuestados consideran que un buen empleo es el que ofrece seguridad y permanencia.

La columna Porcentaje de casos. Se calcula sobre el total de casos válidos (Figura 5.38-A) así: Son 10 casos válidos (Ver cuadro resumen de casos=, entonces el calculo se hace de la siguiente manera:

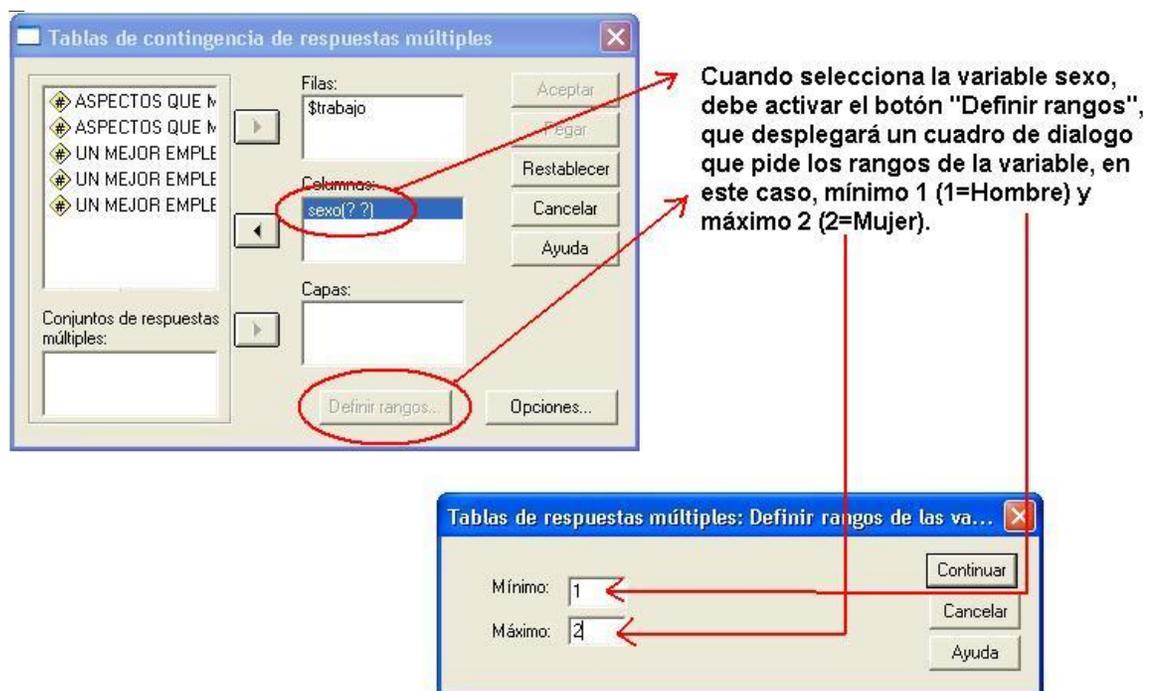
$$\text{Calculo del porcentaje de casos} = \frac{8}{10} \times 100 = 80\%$$

Es decir, que el 80% de las respuestas válidas consideran que un buen empleo es el que ofrece seguridad y permanencia.

**5.3.11. Opción de Tablas de Contingencia con Respuesta Múltiple.** Además de la distribución de frecuencia, se puede hallar tablas de distribución de frecuencia cruzadas (tablas de contingencia). Para ello debe realizar el siguiente procedimiento.

1. Seleccione **Analizar**.
2. **Respuestas Múltiples**.
3. **Tablas de contingencia**, aparece un cuadro de diálogo (Figura 5.39).

**Figura No. 5.39.** Tabla de contingencia de Respuesta múltiple.



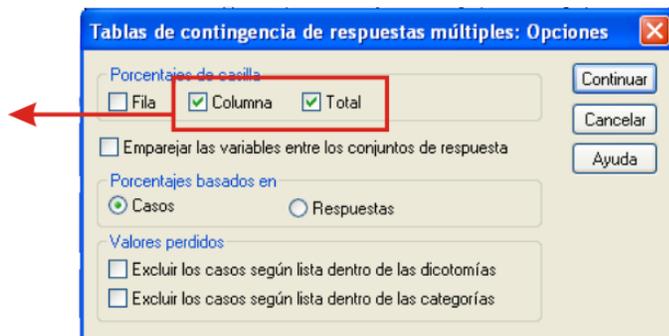
En el cuadro de diálogo (Figura 5.39), realice el siguiente procedimiento:

1. Seleccione la variable creada (\$trabajo) que se encuentra en la sección de Conjuntos de respuestas múltiples y trásládela a la sección de filas.
2. Seleccione la variable sexo y trásládela a la sección de columnas.
3. La variable tendrá el siguiente aspecto (sexo(.??)), es decir, que el SPSS solicita el rango de esta variable.
4. Haga clic en el botón **Definir rangos**. Luego aparecerá un cuadro de diálogo.
5. En el nuevo cuadro de dialogo digite en Mínimo=1 y Máximo=2 (Figura 5.39).

Figura 5.39-B. Ventana de opciones



Active las opciones columna y total, de la sección porcentajes de casilla



6. Presione opciones de la ventana de Tablas de contingencia de respuesta múltiples.
7. En la ventana de Opciones de tablas de contingencias, active las opciones COLUMNA y TOTAL de la sección de porcentajes de casilla.
8. Active CASOS de la sección de “porcentajes basados en”.
9. Haga clic en continuar para pasar a la ventana anterior.
10. Presione aceptar, para ver los resultados, que deben tener un aspecto similar a los mostrados en la Figura 5.40.

**Figura 5.40.** Resultados de tabla de contingencia

**Tabla de contingencia \$TRABAJO\*sexo**

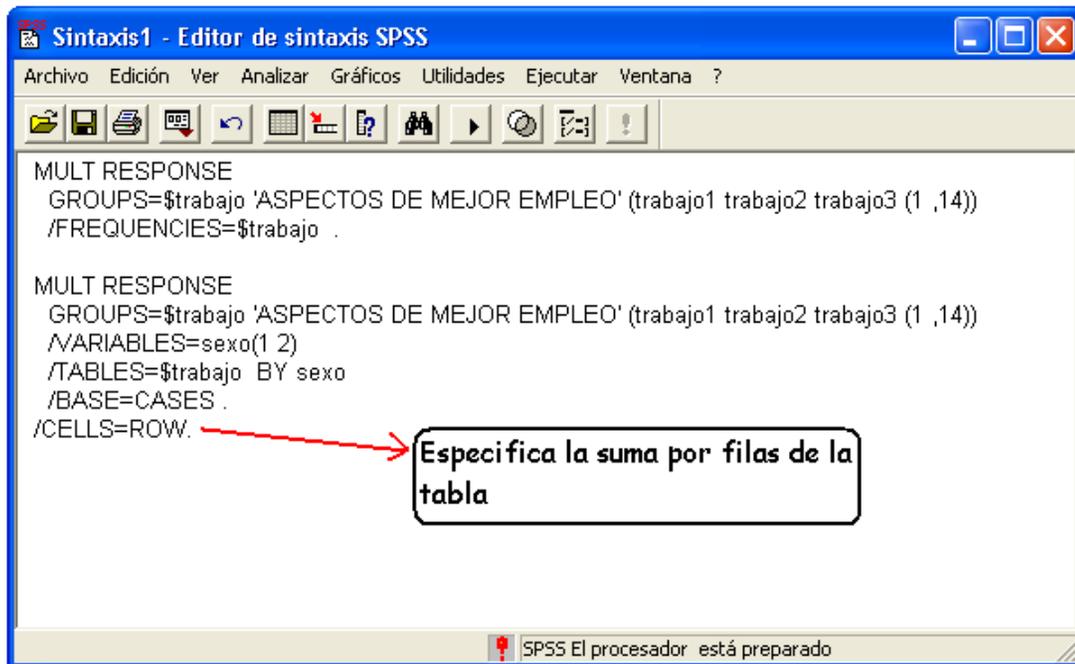
			SEXO		Total
			Hombre	Mujer	
OPCIONES DE MEJOR EMPLEO	Que sea seguro	Recuento	5	3	8
		% dentro de sexo	100,0%	60,0%	
		% del total	50,0%	30,0%	80,0%
	Una buena paga	Recuento	3	3	6
		% dentro de sexo	60,0%	60,0%	
		% del total	30,0%	30,0%	60,0%
	Buenas oportunidades para ascender	Recuento	2	0	2
		% dentro de sexo	40,0%	,0%	
		% del total	20,0%	,0%	20,0%
	Que deje mucho tiempo libre	Recuento	0	1	1
		% dentro de sexo	,0%	20,0%	
		% del total	,0%	10,0%	10,0%
Que sea interesante	Recuento	0	2	2	
	% dentro de sexo	,0%	40,0%		
	% del total	,0%	20,0%	20,0%	
Que me permita trabajar con independenciam	Recuento	0	1	1	
	% dentro de sexo	,0%	20,0%		
	% del total	,0%	10,0%	10,0%	
Que sirva para ayudar a la gente	Recuento	2	1	3	
	% dentro de sexo	40,0%	20,0%		
	% del total	20,0%	10,0%	30,0%	
Que sea útil para la sociedad	Recuento	1	2	3	
	% dentro de sexo	20,0%	40,0%		
	% del total	10,0%	20,0%	30,0%	
Buen ambiente externo	Recuento	1	1	2	
	% dentro de sexo	20,0%	20,0%		
	% del total	10,0%	10,0%	20,0%	
Que sea un trabajo variado	Recuento	1	1	2	
	% dentro de sexo	20,0%	20,0%		
	% del total	10,0%	10,0%	20,0%	
Total	Recuento	5	5	10	
	% del total	50,0%	50,0%	100,0%	

Los porcentajes y los totales se basan en los encuestados.

a. Agrupación

**5.3.12. Lenguaje de Comandos (Sintaxis), de Pregunta de Respuesta Múltiple.** El procedimiento anterior se presenta en el cuadro de sintaxis para los estudiantes que gustan de la programación.

**Figura 5.41.** Sintaxis para tabla de frecuencia y tablas de contingencia para variables de múltiples respuestas.



#### 5.4. PREGUNTA ABIERTA.

Las preguntas abiertas son necesarias en algunos casos, por ejemplo, en la variable edad se aconseja dejarla abierta con el fin calcular los Estadísticos Descriptivos como la media, la desviación estándar, mediana y moda entre otros.

Esta variable se trabajará inicialmente como abierta, luego se crearán unos rangos que permitan agrupar los datos en forma menos extensa, para su posterior presentación.

**Pregunta:** Edad en años cumplidos \_\_\_\_\_

Crear la variable con las siguientes especificaciones:

1. Nombre = edad
2. Tipo = numérico
3. Anchura = 2
4. Decimales = 0
5. Etiqueta = EDAD EN AÑOS CUMPLIDOS
6. Valor = Ninguno
7. Perdidos = Ninguno
8. Columna = 8

- 9. Alineación = Derecha
- 10. Medida = Escala

Una vez realizado los procedimientos, el visor de variable debe tener un aspecto similar como aparece en la Figura 5.42.

**Figura 5.42.** Creación de una variable abierta.

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	sexo	Numérico	2	0	SEXO	{1, Hombre}...	Ninguno	4	Izquierda	Ordinal
2	aspectos	Numérico	2	0	ASPECTOS QUE MAS LE PREOCUPA	{1, Su vida afectiva}...	5	7	Derecha	Ordinal
3	aspecto1	Numérico	2	0	ASPECTOS QUE MENOS LE PREOCUPA	{1, Su vida afectiva}...	5	6	Derecha	Ordinal
4	trabajo1	Numérico	2	0	UN MEJOR EMPLEO OPCION 1	{1, Que sea seguro}...	15	7	Derecha	Ordinal
5	trabajo2	Numérico	2	0	UN MEJOR EMPLEO OPCION 2	{1, Que sea seguro}...	15	8	Derecha	Ordinal
6	trabajo3	Numérico	2	0	UN MEJOR EMPLEO OPCION 3	{1, Que sea seguro}...	15	8	Derecha	Ordinal
7	edad	Numérico	2	0	EDAD EN AÑOS CUMPLIDOS	Ninguno	Ninguno	8	Derecha	Escala

Luego se procede a digitar la información.

**Figura 5.43.** Digitación de la Información de la Variable Edad.

	sexo	aspectos	aspecto1	trabajo1	trabajo2	trabajo3	edad	var
1	Hombre	El dinero	Su vida afe	Que sea se	Una buena	Que sirva p	18	
2	Hombre	El dinero	Su vida afe	Que sea se	Buenas op	Que sirva p	20	
3	Hombre	Su salud	Su vida afe	Que sea se	Que sea úti	Buen ambi	21	
4	Mujer	Su salud	NS/NC	Que sea int	Que sirva p	Que sea un	23	
5	Mujer	Su salud	NS/NC	Una buena	Que deje m	Que sea úti	25	
6	Hombre	Armonia Fa	NS/NC	Que sea se	Una buena	Buenas op	27	
7	Mujer	Armonia Fa	El dinero	Que me pe	Que sea se	Una buena	30	
8	Hombre	Su vida afe	El dinero	Que sea un	Que sea se	Una buena	32	
9	Mujer	Su vida afe	El dinero	Que sea se	Que sea úti	Buen ambi	34	
10	Mujer	El dinero	El dinero	Que sea se	Una buena	Que sea int	36	

**5.4.1. Procesamiento de la Información - Tabla de Frecuencias.** Para procesar la información realice los siguientes pasos:

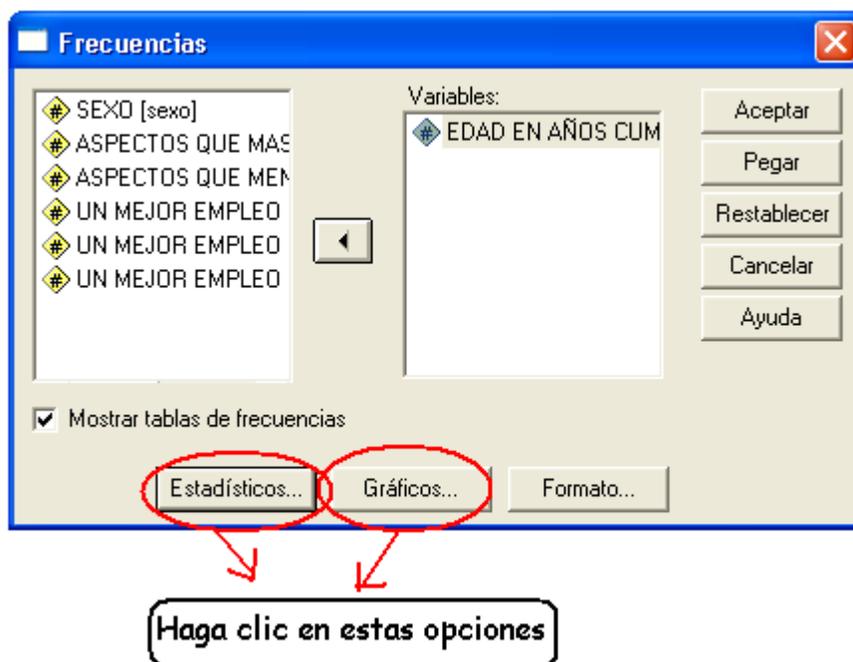
- 1. Seleccione **Analizar**.
- 2. **Estadísticos Descriptivos**.

**3. Frecuencias**, cuando se selecciona esta opción aparece el cuadro de diálogo (Figura 5.44).

En el cuadro de diálogo seleccione:

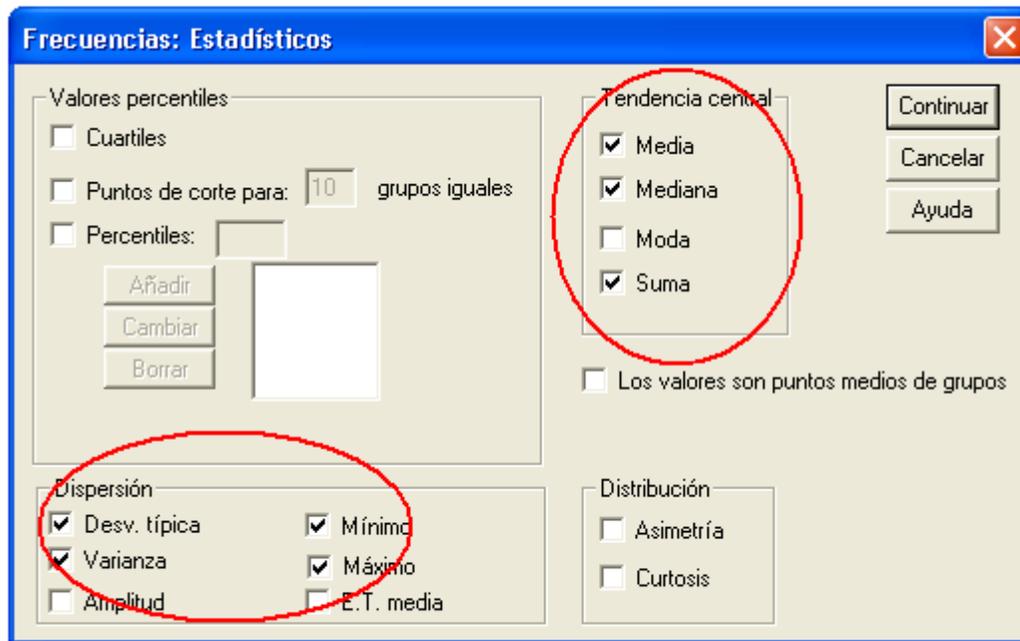
1. La variable EDAD EN AÑOS CUMPLIDOS y trasládelo a la sección de variables.
2. La opción de **estadísticos**. Cuando selecciona esta opción aparece el cuadro de diálogo (Figura 5.45).

**Figura 5.44.** Cuadro de diálogo de Frecuencias



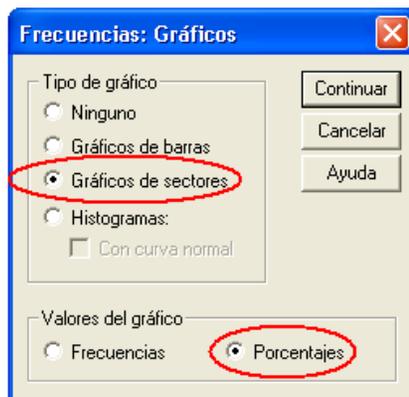
En el cuadro de diálogo de estadístico active Media, Mediana, Suma, Desviación Típica, Varianza, Mínimo y Máximo.

Figura 5.45. Cuadro de dialogo Frecuencias – Estadísticos



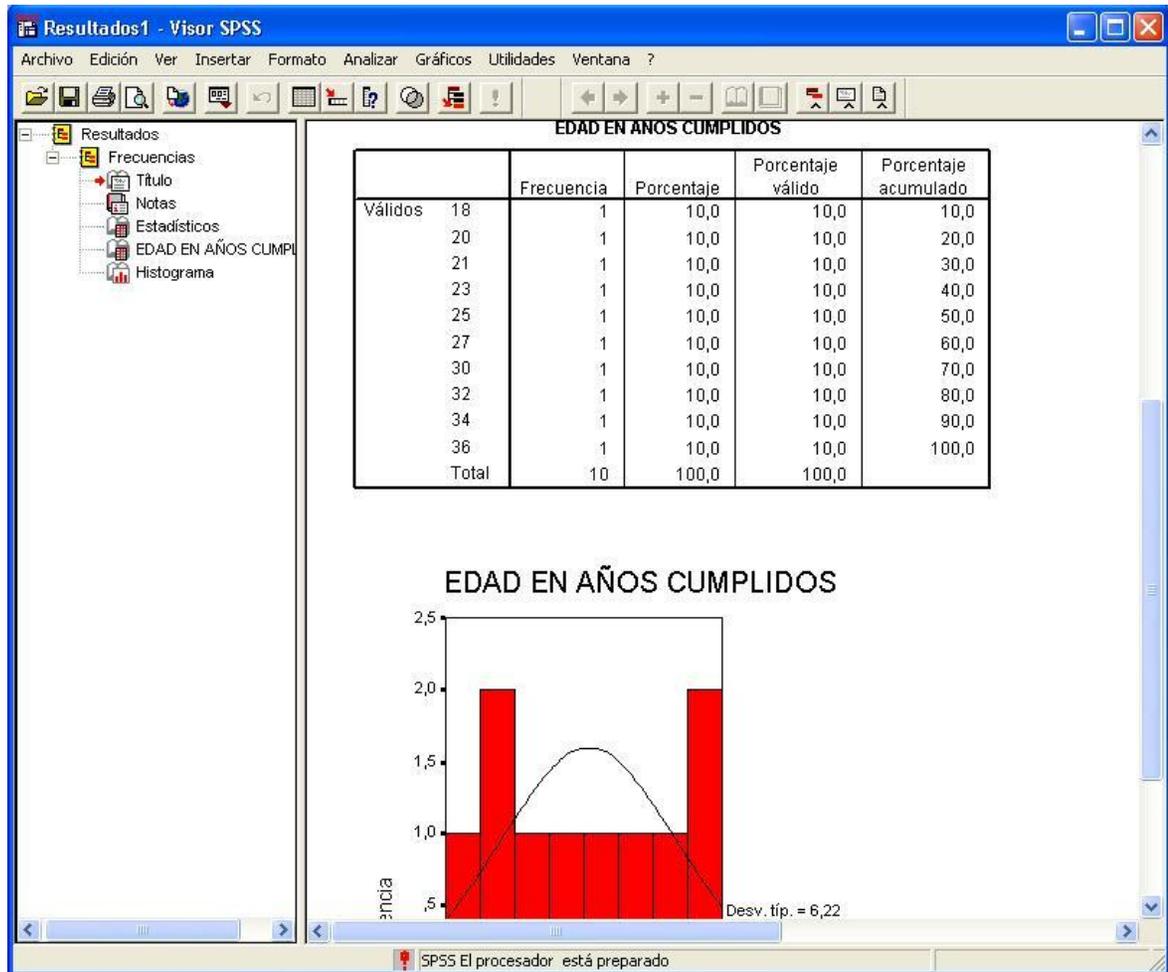
Luego seleccione la opción de gráficos, active Gráfico de sectores y Porcentajes.

Figura 5.46. Cuadro de diálogo de Frecuencias – Gráficos



Los resultados se presentan en la Figura 5.47.

**Figura 5.47.** Visor de Resultados de la Variable Edad.



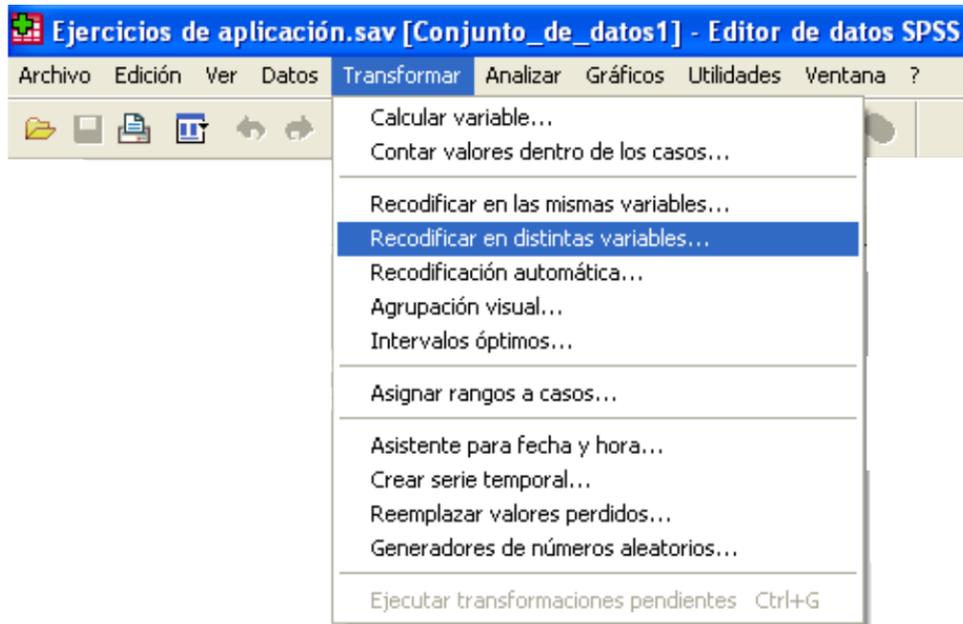
## 5.5. TRANSFORMACIÓN DE LA VARIABLE - EDAD EN RANGOS.

Para crear rangos de variables numéricas abiertas, como el caso de la variable EDAD, este procedimiento se realiza mediante algunas opciones del menú **Transformar**.

Para crear los rangos realice el siguiente procedimiento:

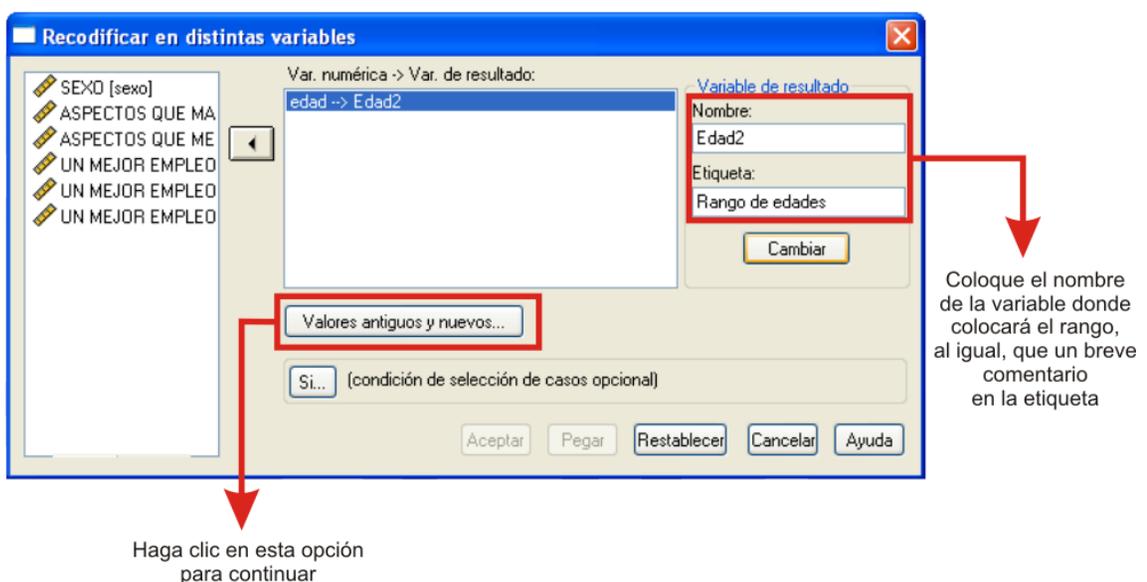
1. Elegir **Transformar**, del menú principal.
2. **Re codificar en distinta variable** Figura 5.49.

Figura 5.48. Creación y transformación en una nueva variable



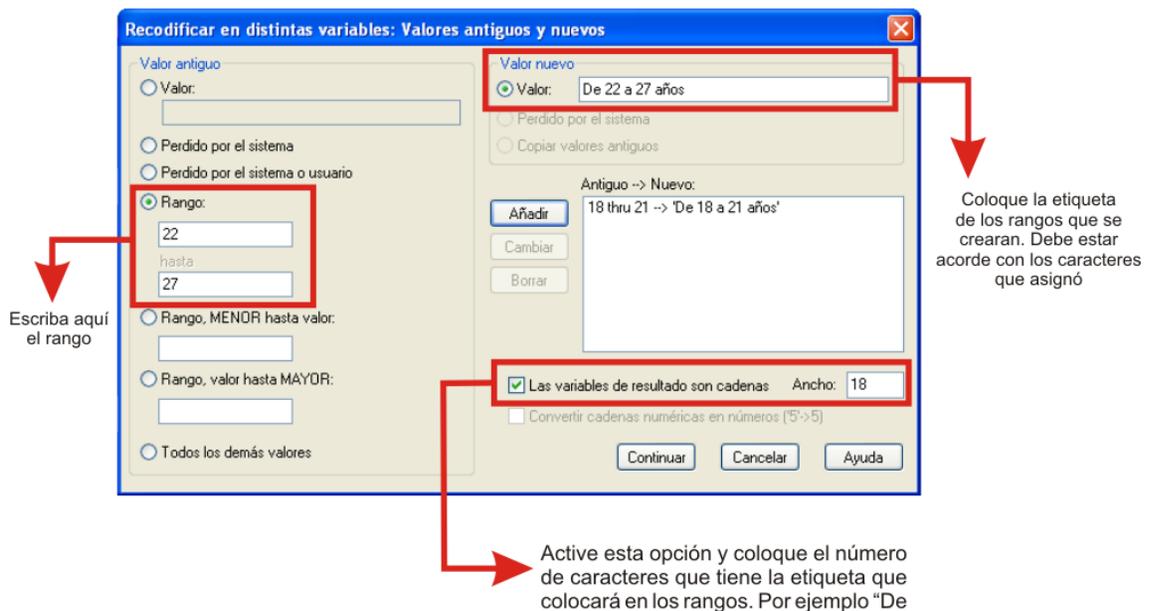
3. En el cuadro de diálogo “**Decodificar en distinta variable**” realice los siguientes pasos:

Figura 5.49. Cuadro de diálogo de recodificar en distinta variable



1. Seleccione la variable Edad y trasládelala a la sección **Var. Numérica - Var. De resultado**.
2. Coloque un nuevo nombre a la variable en la sección de Variable de resultado - Nombre, en este caso se colocó el nombre de **edad2**.
3. Coloque una etiqueta a la variable (**Rango de edades**).
4. Haga clic en la opción de cambiar, automáticamente el programa pasa el nuevo nombre a la sección de **Var. Numérica - Var. De resultado**.
5. Haga clic en la opción de **Valores antiguos y nuevos**, para continuar con el procedimiento, se despliega un nuevo cuadro de diálogo (Figura 5.50).

**Figura 5.50.** Cuadro de dialogo valores antiguos y nuevos



En este nuevo cuadro de diálogo haga los siguientes procedimientos:

1. Active la opción **Rangos**.
2. Active la opción "**Las variables de resultado son cadena**".
3. Fije **Ancho** en 18 caracteres.
4. En la sección de **Valor antiguo**, digite el primer rango 18 hasta 21.
5. En la sección **Valor nuevo** digite "De 18 a 21 años" y haga clic en la opción **Añadir**.
6. Crear dos rangos mas así: De 22 a 27 y de 28 a 36.
7. Haga clic en la opción **Continuar**.
8. Haga clic en la opción **Aceptar**.

De esta forma se crea una nueva variable denominada edad2 (Figura 5.51).

**Figura 5.51.** Editor de datos creación de la nueva variable edad2.

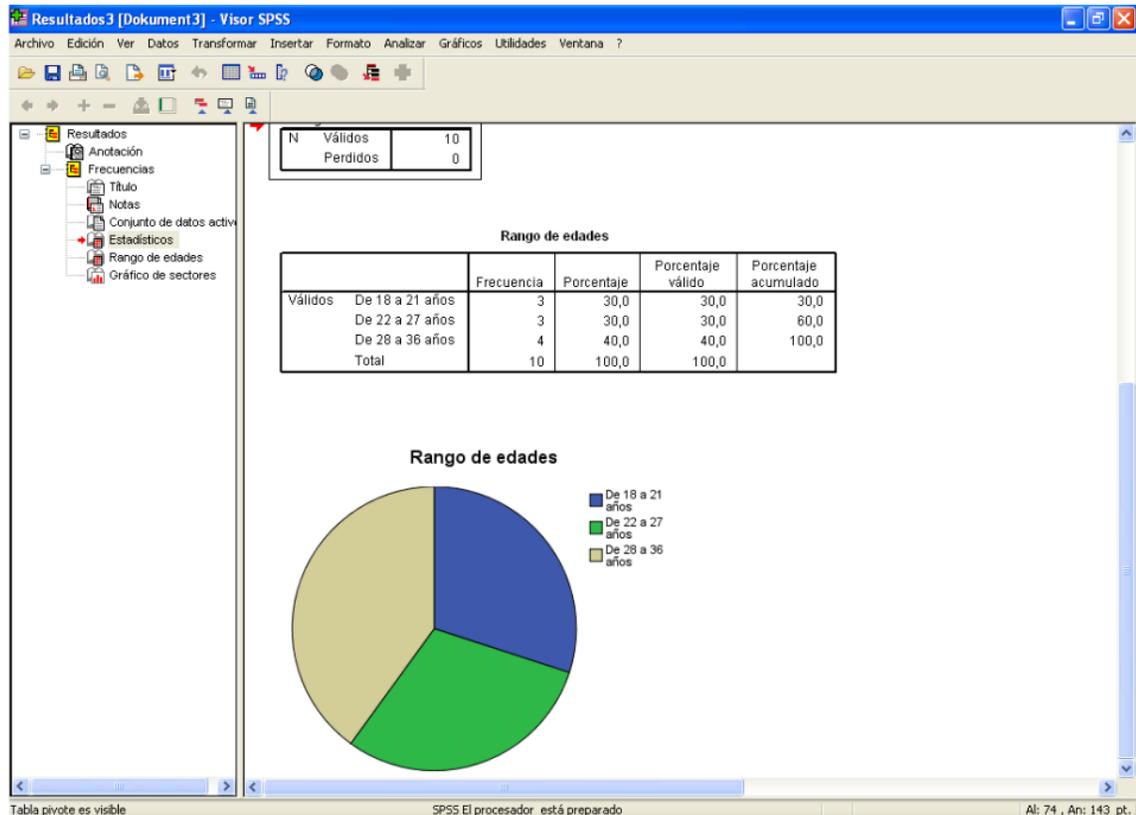
The screenshot shows the SPSS Data Editor window for a file named '\*Ejercicios de aplicación.sav [Conjunto\_de\_datos1]'. The window title bar includes standard OS icons and the text '\*Ejercicios de aplicación.sav [Conjunto\_de\_datos1] - Editor de datos SPSS'. Below the title bar is a menu bar with options: Archivo, Edición, Ver, Datos, Transformar, Analizar, Gráficos, Utilidades, Ventana, and ?. A toolbar with various icons is located below the menu bar. The main area displays a data table with 15 rows and 9 columns. The columns are labeled: 'sexo', 'aspectos', 'aspecto 1', 'trabajo1', 'trabajo2', 'trabajo3', 'edad', 'Edad2', and 'var'. The 'Edad2' column is highlighted with a red border. The data in the 'Edad2' column consists of age ranges: 'De 18 a 21 añ', 'De 18 a 21 añ', 'De 18 a 21 añ', 'De 22 a 27 añ', 'De 22 a 27 añ', 'De 22 a 27 añ', 'De 28 a 36 añ', 'De 28 a 36 añ', 'De 28 a 36 añ', and 'De 28 a 36 añ'. The status bar at the bottom indicates 'Área del procesador' and 'SPSS El procesador está preparado'.

	sexo	aspectos	aspecto 1	trabajo1	trabajo2	trabajo3	edad	Edad2	var
1	1	2	1	1	2	7	18	De 18 a 21 añ	
2	1	2	1	1	3	7	20	De 18 a 21 añ	
3	1	4	1	1	8	12	21	De 18 a 21 añ	
4	2	4	4	5	7	13	23	De 22 a 27 añ	
5	2	4	5	2	4	8	25	De 22 a 27 añ	
6	1	3	5	1	2	3	27	De 22 a 27 añ	
7	2	3	2	6	1	2	30	De 28 a 36 añ	
8	1	1	2	13	1	2	32	De 28 a 36 añ	
9	2	1	2	1	8	12	34	De 28 a 36 añ	
10	2	2	2	1	2	5	36	De 28 a 36 añ	
11									
12									
13									
14									
15									

Procesar esta nueva variable así:

1. Seleccione **Analizar**.
2. **Estadísticos descriptivos**.
3. **Frecuencia**.
4. **Gráficos**. Gráfico de sectores.
5. **Aceptar**. Los resultados se presentan en la figura 5.52.

**Figura 5.52.** Visor de Resultados Variable rango de Edades.



## 5.6. GRAFICAS INTERACTIVAS

Las gráficas interactivas son una alternativa a las gráficas expuestas en las anteriores secciones; de otra parte, la calidad de presentación es mejor que las gráficas tradicionales o normales vistas en las secciones anteriores.

Las gráficas interactivas están disponibles en el programa SPSS para Windows desde la versión 8.

Los tipos de gráficas interactivas son:

1. Barras
2. Puntos
3. Líneas
4. Bandas. Sólo se encuentran con gráficas interactivas.
5. Líneas verticales.
6. Áreas
7. Sectores

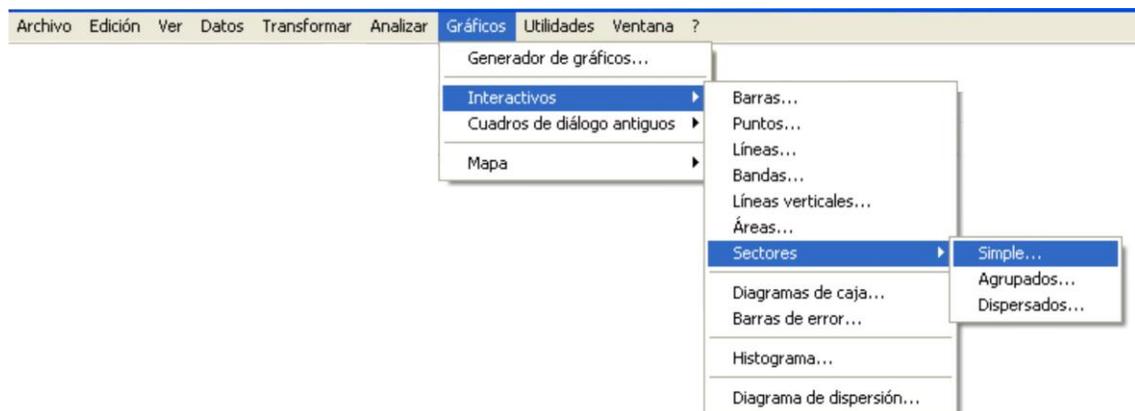
8. Diagrama de cajas
9. Barrar de error
10. Histogramas
11. Diagramas de dispersión. Son las nubes de puntos.

**5.6.1. Cómo crear una grafica interactiva.** Para explicar esta sección, vamos a graficar la variable “Aspectos que más le preocupa”.

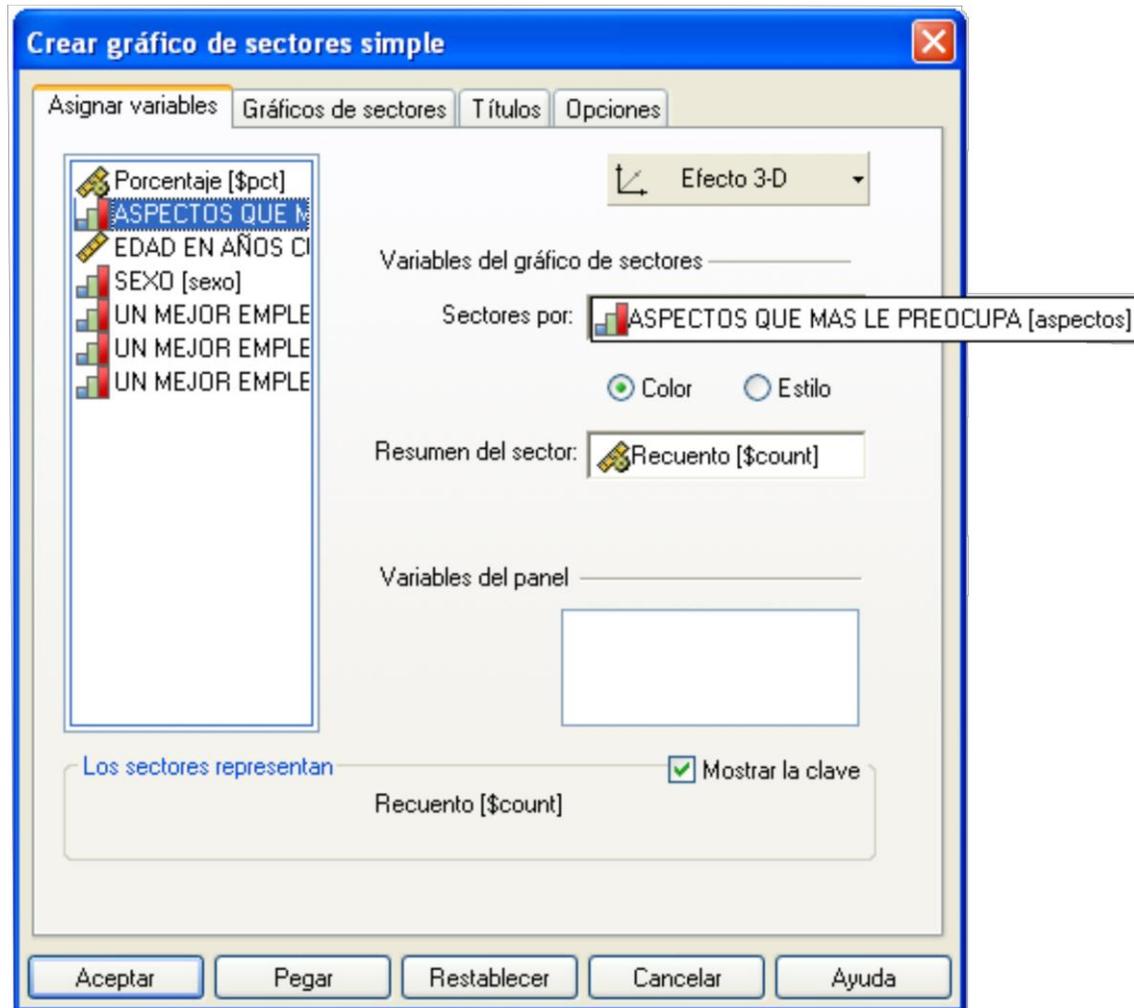
Realice el siguiente procedimiento:

1. Seleccione **Gráficas** del menú principal (Figura 5.53)
2. **Interactivas.**
3. **Sectores.**
4. **Simple.** Al seleccionar esta opción se despliega un cuadro de diálogo. (Figura 5.53).

**Figura 5.53.** Graficas interactivas.



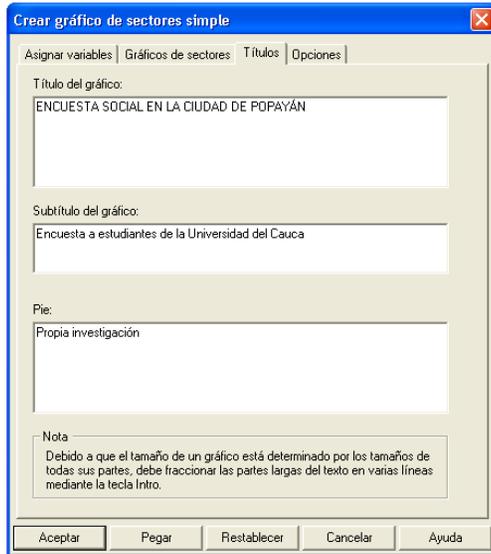
**Figura 5.54.** Cuadro de diálogo Gráficas Interactivas – Sectores Simples



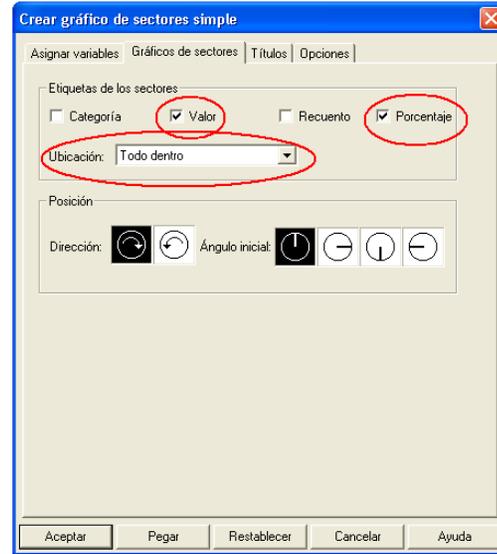
En el cuadro de diálogo proceda así:

1. Seleccione la variable con el ratón, usted notará que el cursor se convierte en una mano , arrastre la variable manteniendo presionado el ratón hasta la sección **Sectores por**.
2. Coloque el título de la gráfica, como aparece en la figura 5.55.
3. Seleccione la pestaña de Gráficos de Sectores, realice los procedimientos según la Figura 5.56.
4. Haga clic en **Aceptar**.
5. Los resultados son como los que aparecen en la Figura 5.57.

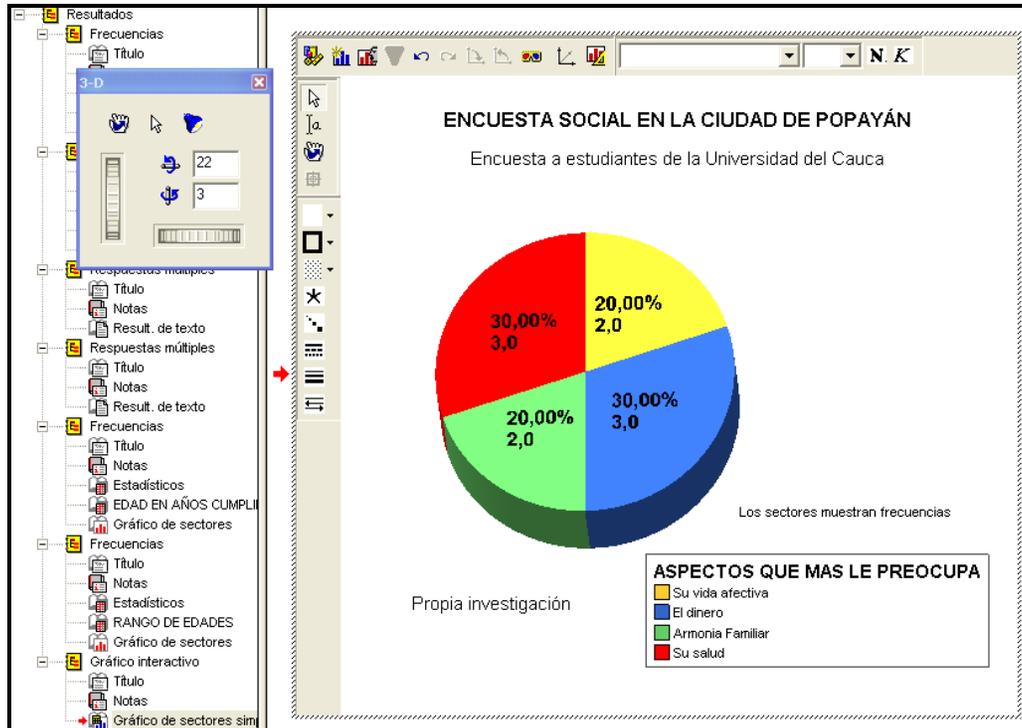
**Figura 5.55.** Título de la gráfica interactiva



**Figura 5.56.** Sección de Gráficos de Sectores de la Gráfica Interactiva.



**Figura 5.57.** Visor de Gráfica Interactiva





# Capítulo seis

## Tipos de escala y análisis de la información

---

### 6. INTERPRETACIÓN Y ANÁLISIS DE LA INFORMACIÓN.

Procesados los datos en el programa SPSS, estamos listos para realizar el análisis de la información. En el capítulo 2 (punto 2.6) se definió los tipos escala; esta clasificación resulta imprescindible para determinar la forma adecuada del análisis de la información.

Para determinar que técnica de análisis de datos se debe emplear es necesario tener en cuenta los siguientes aspectos:

- El número de variables a analizar;
- Si nos interesa hacer inferencia de los datos analizados y;
- Los tipos de escala que se utilizaron en el momento de la recolección de la información.

**6.1. NÚMERO DE VARIABLES A ANALIZAR.** Una primera clasificación de las técnicas de análisis de datos permite distinguir entre investigación univariante, bivalente y multivariante.

- **Análisis univariante:** Es la técnica de análisis de datos que se aplican sobre una variable individualmente o de forma aislada.
- **Análisis bivariado:** Técnica de análisis de datos que aplican sobre dos variables de forma conjunta. El objetivo de la técnica bivalente es la búsqueda de relaciones de asociación, dependencia causa – efecto.
- **Análisis multivariado.** Técnica de análisis de datos que se aplican sobre más de dos variables de forma simultanea. Se busca la relación entre ellas.

 <div style="background-color: #333; color: white; padding: 5px; display: inline-block; border-radius: 10px;"><b>Nota</b></div>	<p>En este texto se explicará en una forma exhaustiva con ejemplos, las técnicas de análisis <i>univariado</i> y <i>bivariado</i>, al igual que la forma de procesar los estadísticos en el programa SPSS. Los estudiantes que están interesados en el análisis <i>multivariado</i>, al final del texto se relacionan algunos libros de investigación de mercados donde se puede consultar y profundizar en este tema.</p>
--	--

## 6.2. INFERENCIA ESTADÍSTICA Y DESCRIPCIÓN DE VARIABLES

En la investigación de mercados se hace necesario aplicar la estadística descriptiva y la estadística de inferencia.

- **La estadística descriptiva.** Es una rama de la estadística que proporciona a los investigadores estimaciones puntuales de los datos de la muestra como: media o promedio, medidas de dispersión y nivel de asociación entre las variables.
- **La estadística Inferencial.** Es una rama de la estadística que permite hacer juicios a cerca de la población total, basados en los resultados generados por la muestra.

## 6.3. TIPOS DE ESCALA DE MEDICIÓN.

El tercer aspecto que se debe analizar es el tipo de escala que se utilizo en la variable o variables a analizar. En la tabla 6.1 se retoma la clasificación de las variables No Métricas y las Métricas según las escalas de medición que se trato en el capítulo 2.

**Tabla 6.1. Clasificación de escalas**

NO MÉTRICAS	NOMINALES	Se identifica la pertenencia a una categoría u otra.
	ORDINALES	Se establece un orden de preferencia
MÉTRICAS	INTERVALO	Se valora una característica o atributo, asignando un valor dentro de un rango previamente establecido.
	DE RELACIÓN	Valor numérico real.

Si el investigador conoce el número de variables que van a analizar simultáneamente, ya sea que el interés esté en la descripción o inferencia, y conoce el nivel de escala de la variable o las variables, entonces puede seleccionar el procedimiento estadístico apropiado.

## 6.4. ANÁLISIS UNIVARIADO DE DATOS.

El análisis univariado permite al investigador realizar un estudio exhaustivo de todos los datos tabulados. El análisis univariado permite hacer una descripción del comportamiento general de la muestra con respecto a los aspectos contenidos en el cuestionario. En la Tabla 6.2 presenta una visión general de las técnicas estadísticas disponibles para el análisis univariado de datos.

**Tabla 6.2.** Análisis univariado de datos.

ESCALA	ESTADÍSTICA DESCRIPTIVA			ESTADÍSTICA INFERENCIAL
	Medidas de tendencia central	Análisis de frecuencia	Medidas de dispersión	
Nominal	Moda	Frecuencias relativas Frecuencias absolutas		Prueba de ji - cuadrado
Ordinal	Moda Mediana	Frecuencias relativas Frecuencias absolutas	Rango intercuartil	Prueba kolmogorov - Smirnov
De intervalo	Moda Mediana Media	Frecuencias relativas Frecuencias absolutas	Varianza Desviación estándar Rango intercuartil	Prueba Z Prueba t
De relación	Media		Varianza Desviación estándar	

**6.4.1. Análisis de frecuencia.** El análisis de la distribución de frecuencias es un método para organizar y resumir datos. Bajo este método los datos se clasifican y ordenan, indicándose el número de veces que se repite. El análisis de frecuencias puede expresarse en valores absolutos o relativos.

- **Valores absolutos.** Número de encuestados que han contestado las diferentes alternativas de respuesta.
- **Valores relativos.** % de encuestados que han contestado las diferentes alternativas de respuesta.

**6.4.2. Medidas de tendencia central.** Las medidas de posición o de tendencia central nos permiten determinar la posición de un valor respecto a un conjunto de datos, el cual consideramos como representativo para el total de las observaciones. Las medidas de tendencia central que se analizarán son: Moda, Mediana y la Media.

- **Moda:** valor que se repite con mayor frecuencia

- **Mediana:** Se define como aquel valor de la variable que supera a no más de la mitad de las observaciones y al mismo tiempo es superado por no más de la otra mitad de las observaciones. La mediana es el valor central.
- **Media:** Es la medida de posición más utilizada, la más conocida y sencilla de calcular, es la suma de los valores de la variable dividida por el número de valores observados.

**6.4.3. Medidas de dispersión.** Son aquellas que nos determinan cómo se agrupan o se dispersan los datos alrededor de un promedio. Cuando se trabaja con variables de intervalo existen una serie de medidas de dispersión que permiten evaluar precisamente esto: el grado de homogeneidad o heterogeneidad en las respuestas de la muestra. Entre las medidas de dispersión que analizaremos están:

- **Rango o recorrido.** Diferencia entre el valor máximo y el valor mínimo otorgado por los encuestados.
- **Recorrido intercuartílico.** Diferencia entre el 3er cuartil y el 1er cuartil.
- **Varianza.** Suma de los cuadrados de las diferencias entre los valores obtenidos y su media, dividiendo el resultado por el número de casos.
- **Desviación estándar.** Raíz cuadrada de la varianza.

**6.4.4. Ejemplos de análisis univariante.** En apéndice A. aparecen los detalles de la base de datos con que se trabajara las interpretaciones de los ejemplos de este capítulo.

 <b>En el CD-ROM</b>	En el CD que encontrara en la contraportada de este libro ubique el archivo COMERCIO.SAV, que esta en la carpeta CAPITULO 7.
---	--

**6.4.4.1. Ejemplo 1 Variables nominales.** Los habitantes de la ciudad de Popayán han manifestado inconformidad con el comercio de la Ciudad, en cuanto a la tención al cliente, precio y variedad en los productos. De los siguientes aspectos que le presento a continuación, Seleccione con el que hasta más inconforme.

1. No hay variedad en los producto ( )
2. Mala atención al cliente ( )
3. Mala calidad en los productos ( )
4. Los precios son muy altos ( )
5. NS/NC ( )

Los resultados aparecen en la Tabla 7.3, con un total de 437 respuestas válidas y 24 perdidos, es decir, que 24 de los encuestados no contestaron esta pregunta. En la tabla de frecuencias 6.3 hay que hacer las siguientes aclaraciones:

- Valores válidos. Hacen referencia a aquellos individuos que contestaron.
- Valores perdidos. Hacen referencia a aquellas personas que no contestaron.
- La diferencia entre “porcentaje” y “porcentaje válido” se refiere a la presencia de valores perdidos.
- En las variables nominales, el porcentaje acumulado no tiene ningún sentido de interpretación.

Tabla 6.3.

**Aspecto de mayor inconformidad**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	No hay variedad en los productos	40	8,7	9,2	9,2
	Mala atención al cliente	136	29,5	31,1	40,3
	Mala calidad en los productos	112	24,3	25,6	65,9
	Los precios son muy altos	149	32,3	34,1	100,0
	Total	437	94,8	100,0	
Perdidos	NS/NC	24	5,2		
Total		461	100,0		

Como se observa en la Tabla 6.3 el 32% de los entrevistados está inconforme con los precios de los productos, siendo esta opción la mayor (moda). El 24% está inconforme con la calidad de los productos, y aproximadamente el 30% está inconforme con la mala atención al cliente. En definitiva, las opiniones de los entrevistados están bastante repartidas. Los datos se presentan en la Figura 6.1.

Figura 6.1.



**6.4.4.2. Ejemplo 2: Variables ordinales.** En este ejemplo utilizamos la pregunta que aparece a continuación, en la que, a través de una escala ordinal, se les ha pedido a los entrevistados que establezcan un orden de los factores que considera más importantes a la hora de comprar vestidos y sus complementos, para un total de siete ítems. Por tanto se crean siete variables, una por cada opción de respuesta. Cada una estas variables se deben analizar, en este caso tan solo se analizará la primera opción las alternativas.

- Cuando usted acude a un almacén a comprar vestidos y sus complementos, cuáles de los siguientes factores considera más importantes, ordénelos, colocando en primer lugar aquel que considera más importante, y en último lugar aquel que considere menos importante.

- B1 Variedad de las marcas
- B2 Los precios de los productos
- B3 Ubicación del almacén
- B4 Que encuentre todo en un mismo lugar
- B5 Los precios de los zapatos y accesorios
- B6 Orientación del comerciante
- B7 Ofertas del almacén

**Tabla 6.4. Opciones de preguntas**

1. Variedad de las marca de la ropa que tiene el almacén	1a	2a	3a	4a	5a	6a	7a
2. Los precios del almacén.	1a	2a	3a	4a	5a	6a	7a
3. Ubicación del almacén.	1a	2a	3a	4a	5a	6a	7a
4. Que encuentre todo en el mismo lugar	1a	2a	3a	4a	5a	6a	7a
5. Los precios de los zapatos y accesorios	1a	2a	3a	4a	5a	6a	7a
6. La orientación que da el comerciante	1a	2a	3a	4a	5a	6a	7a
7. Ofertas que brinda el almacén	1a	2a	3a	4a	5a	6a	7a

Cuando se trabaja con variables ordinales es posible realizar un mayor número de análisis: no sólo frecuencias y moda, sino también mediana y cuartiles.

Pero antes de iniciar con el análisis es preciso hacer las siguientes aclaraciones:

- El % máximo del porcentaje válido nos permite conocer el valor de la moda.
- El porcentaje acumulado nos permite conocer:
  - La mediana: cuando igualamos o superamos el 50% de los datos acumulados.
  - Los cuartiles cuando:
    - Primer cuartil: cuando igualamos o superamos el 25% de los datos acumulados.

- Segundo cuartil: la mediana
- Tercer cuartil: Cuando igualamos o superamos el 75% de los datos acumulados.
- Cuanto mayor sea la diferencia entre el tercer y el primer cuartil, mayor será la variabilidad o dispersión de las respuestas.

Tabla 6.5

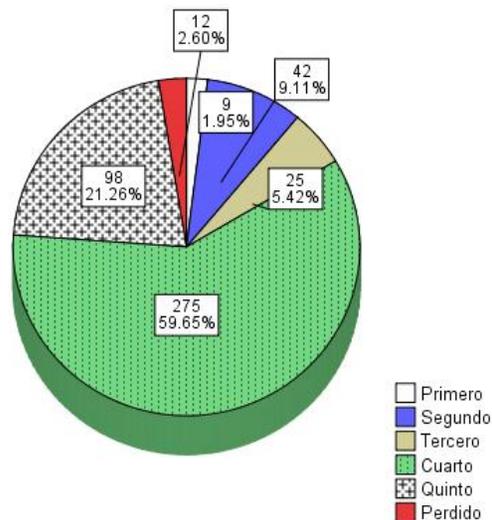
Variedad en marcas

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Primero	9	2,0	2,0	2,0
	Segundo	42	9,1	9,4	11,4
	Tercero	25	5,4	5,6	16,9
	Cuarto	275	59,7	61,2	78,2
	Quinto	98	21,3	21,8	100,0
Total		449	97,4	100,0	
Perdidos	Ns/Nc	12	2,6		
Total		461	100,0		

En la **tabla 6.5** se puede observar que la variedad en marcas no es relevante para los entrevistados, dado que la ubican en el cuarto lugar, es decir, que el 61% de los entrevistados la sitúan en el cuarto lugar.

Figura 6.2.

Variedad en marcas



**6.4.4.3. Ejemplo 3. Variables de intervalo.** En este ejemplo se van a analizar variables de intervalo. Utilizamos una pregunta en la que a través de una escala de intervalo de cinco puntos, se le pide al entrevistado que se le de una puntuación individualizada, para un total de once (11) variables en que las alternativas de respuesta son: 1, 2, 3, 4, y 5. En este caso analizaremos únicamente la dos primeras.

Utilizando esta escala de 5 puntos en la que 1 valora muy poco y 5 valora mucho, Cuando usted realiza sus compras de vestido y complementos ¿qué es lo que más valora?

**Tabla 6.6 Opciones de preguntas.**

	Valora muy poco <span style="float: right;">Valora mucho</span>				
	1	2	3	4	5
B8 Variedad					
B9 Diseño					
B10 Atención					
B11 Amabilidad					
B12 Ambiente					
B13 Mobiliario					
B14 Decoración					
B15 Marca					
B16 Exclusividad					
B17 Limpieza					
B18 Calidad					

Cuando se trabaja con variables de intervalo se pueden realizar un mayor número de análisis estadístico, más que con las variables ordinales y, por supuesto, más que con variables nominales.

Los análisis a efectuar son:

- Frecuencias
- Medidas de tendencia central (moda, mediana y media), y
- Medidas de dispersión (desviación, varianza, rango y recorrido intercuartilico).

**Tabla 6.7. Tabla de frecuencias, valoración de la opción “Variedad”**

		Variedad			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1. Valora muy poco	37	8,0	8,7	8,7
	2	37	8,0	8,7	17,5
	3	79	17,1	18,7	36,2
	4	96	20,8	22,7	58,9
	5. Valora mucho	174	37,7	41,1	100,0
	Total	423	91,8	100,0	
Perdidos	Sistema	38	8,2		
Total		461	100,0		

**Tabla 6.8. Tabla de frecuencias, valoración de la opción “Diseño”**

		Diseño			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1. Valora muy poco	42	9,1	10,0	10,0
	2	54	11,7	12,9	22,9
	3	157	34,1	37,4	60,2
	4	92	20,0	21,9	82,1
	5. Valora mucho	75	16,3	17,9	100,0
	Total	420	91,1	100,0	
Perdidos	Sistema	41	8,9		
Total		461	100,0		

**Tabla 6.8.**

Resultados estadísticos,Ejemplo 3

	Media	Mediana	Moda	Desv. típ.	Varianza	Rango
Variedad	3,79	4,00	5	1,302	1,694	4
Diseño	3,25	3,00	3	1,185	1,404	4

**6.4.4.4. Medidas de tendencia central.** Las medidas de tendencia central para la variable “Variedad” son muy altas. Esto indica que los encuestados le dan mucha importancia a esta variable: la mayoría de ellos valoro esta variable con la máxima puntuación (5 puntos). En el caso de la variable “Diseños” es también una variable importante para los encuestados, pero en menor medida que la “Variedad”: En este caso las opiniones de los individuos consultados son más variadas, aunque la mayoría de ellas se sitúan en las puntuaciones 2,3 y 4. Se puede concluir que la variedad de los productos importa más a los entrevistados que los diseños.

La variabilidad de las respuestas ha sido mayor en el caso de la variable “Variedad”. Así lo demuestran las tres medidas de dispersión analizadas, más bajas para la variable “Diseño”. Es preciso indicar que de las tres medidas de dispersión analizadas, es la *desviación típica la más exacta*, o mejor dicho la más discriminante a la hora de hacer la comparación entre variables.

**6.4.4.5. Medidas de forma de la distribución:** asimetría y curtosis. Cuando se trabaja con variables de tipo métrico un aspecto interesante es comparar la forma en la que se distribuyen los datos en la variable analizada con respecto a una distribución normal.

A continuación se analizará los estadísticos relacionados en la tabla 6.9.

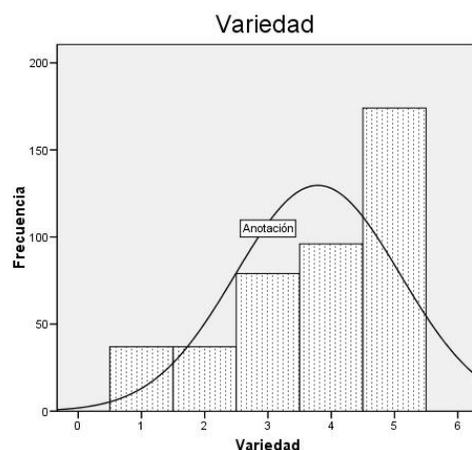
Tabla 6.9. Estadísticos de asimetría y curtosis

	Valoración de la Variedad	Valoración de Diseño
N Válidos	423	420
Perdidos	38	41
Asimetría	-,799	-,204
Error típ. de asimetría	,119	,119
Curtosis	-,494	-,639
Error típ. de curtosis	,237	,238

- **Variable variedad:**

- **El coeficiente de asimetría.** Es negativo -0,799: existe una ligera asimetría hacia la izquierda (Figura 6.3)

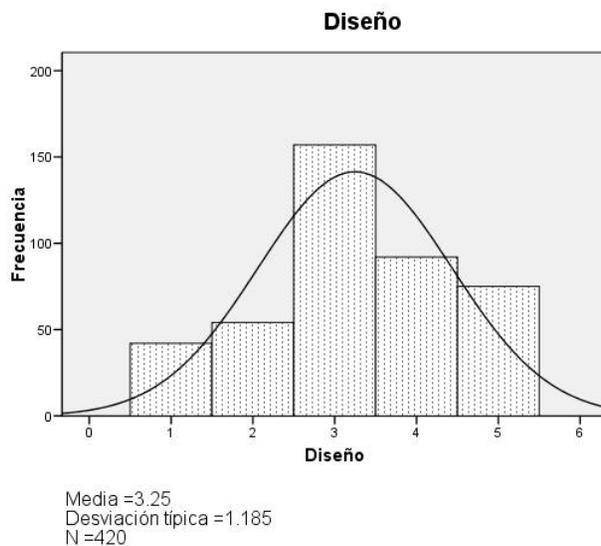
Figura 6.3. Histograma



Media =3,79  
Desviación típica =1.302  
N =423

- **El coeficiente de curtosis.** Es negativo  $-0,494$ , es decir, que la forma de distribución es achatada o platicúrtica.
- **Variable “Diseño”**
  - **El coeficiente de asimetría.** Es negativo  $-0,204$ , este coeficiente es más cercano a cero, por tanto su forma es más simétrica con una pequeña tendencia negativa o a la izquierda. (Figura 6.4.)

Figura 6.4. Histograma



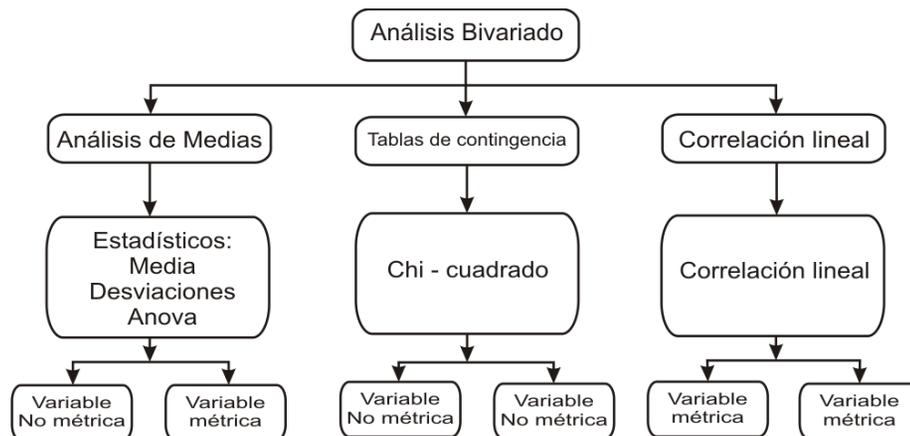
- **El coeficiente de curtosis.** Es negativo  $-0,639$ , es decir, que la forma de la distribución es más apuntada o leptokúrtica, de otra parte hay más concentración de la información.

## 6.5. ANÁLISIS BIVARIANTE.

El análisis bivalente se clasifica de acuerdo al tipo de variable, como aparece en la figura 6.5, dando como resultado tres tipos de análisis.

- **Análisis bivalente de medias:** cuando una de las variables es no métrica y la otra métrica.
- **Análisis de tablas de contingencia.** Cuando las dos variables de análisis son no métricas.
- **Análisis de correlación lineal.** Cuando las dos variables son métricas.

**Figura 6.5. Análisis Bivariado.**



**6.5.1. Análisis bivariante de medias.** El objetivo del análisis bivariante de medias es el de segmentar una muestra con una variable de clasificación, con la finalidad de encontrar diferencias significativas entre los distintos grupos formados. Las diferencias se determinan a partir de la comparación entre las medias de las variables analizadas para los distintos grupos.

Para realizar este análisis las variables empleadas deben tener las siguientes características:

**La variable dependiente:** debe ser **métrica**, ya que se calculará su media.

**La variable independiente:** debe ser no métrica que permita dividir la muestra.

**6.5.1.1. Ejemplo 4.** Con este ejemplo se cruzará la pregunta analizada en la sesión 6.4.4 ejemplo 3 (análisis univariado) con la variable de clasificación “SEXO”. Se analizará si existen diferencias significativas entre el sexo de los entrevistados y las características que se valoran cuando compran vestidos y sus complementos. Las variables se clasificaron de la siguiente forma:

**Variable dependiente:** Características que los entrevistados valoran a la hora de comprar vestidos y sus complementos (Escala de 1 a 5). Variable de la que se va a calcular su media.

**Variable independiente.** Sexo de los consumidores.

Se va a calcular la importancia media concedida por las mujeres a este tipo de características, la importancia concedida por los hombres y la importancia media total (sin distinguir entre hombres y mujeres). Además de considerar la media, resulta

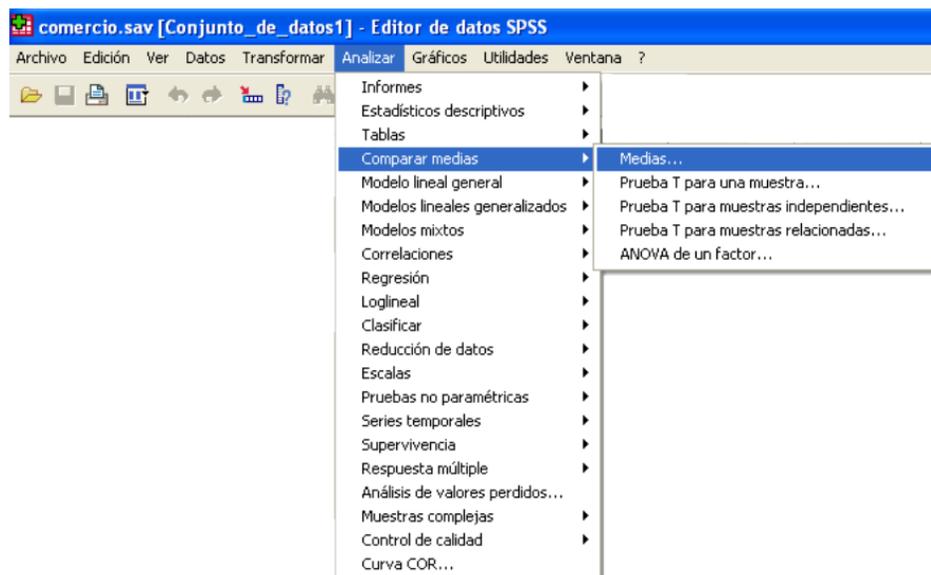
interesante la variabilidad en las respuestas de los hombres, la variabilidad en el grupo de mujeres y la variabilidad total (medida en este caso a partir de la desviación típica). Otro dato a tener en cuenta, especialmente al extrapolar los datos, es el número de encuestas realizadas en cada segmento. A continuación se muestran las tablas de análisis.

**6.5.2. Procedimiento en el programa estadístico SPSS.** Para realizar el análisis bivariado en el programa estadístico SPSS, haga el siguiente procedimiento.

Seleccione en el menú principal

- Analizar
- Comparar medias
- Medias (Figura 6.6)

Figura 6.6. Procedimiento en SPSS para análisis de comparación de medias.



- Seleccione las variables.
- Coloque la variable “Variedad” en la sesión Dependientes
- La variable “Sexo” en Independientes
- Seleccione Opciones (Figura 6.7)

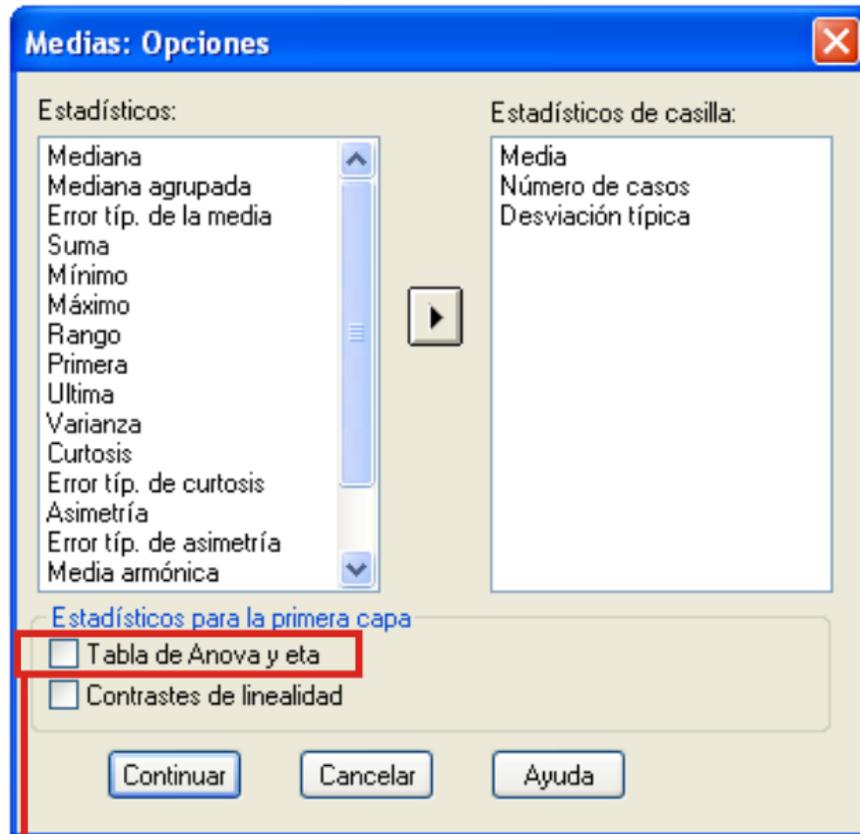
Figura 6.7. Selección de variables.



En la ventana de opciones active:

- Tabla anova y eta;
- **Sesión estadísticos de casilla:** El programa tiene predeterminado los estadísticos: media, número de casos y desviación típica, es decir, que estas opciones están ubicadas en ésta sección. Si no están ubicadas en la sección de Estadísticos de casilla, entonces procesa a realizar esta actividad;
- Presione continuar (Figura 6.8).

Figura 6.8. Ventana de opciones.



La ventana de resultados de su computadora debe tener una apariencia similar a la de la Figura 6.9.

Figura 6.9. Ventana de resultados.

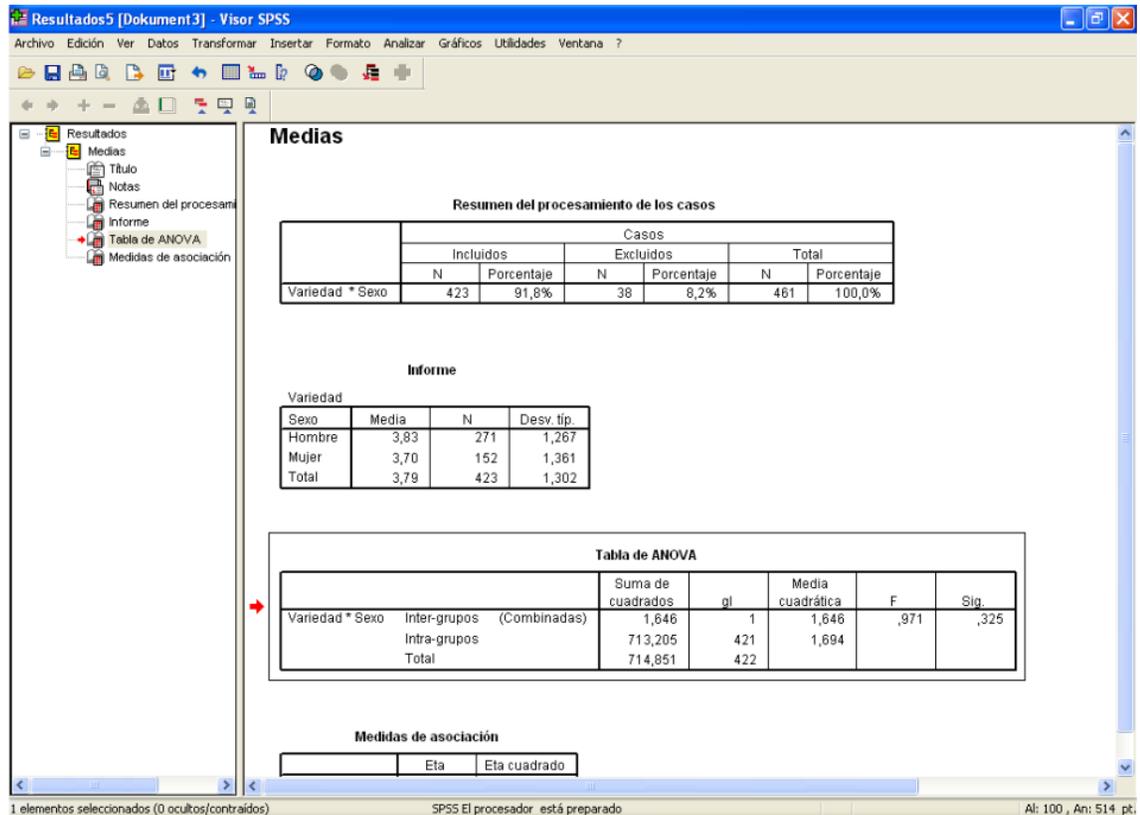


Tabla 6.10.

### Análisis bivalente de medias

Variedad

Sexo	Media	N	Desv. típ.
Hombre	3,8339483395	271	1,267
Mujer	3,7039473684	152	1,361
Total	3,7872340426	423	1,302

Los datos arrojan el siguiente análisis:

- Los hombres dan más importancia a este tipo de acciones tal y como lo demuestra un mayor valor en la media (3,83).
- El grado de acuerdo entre los hombres es inferior con una desviación típica de 1,267, frente al grupo de mujeres con una desviación típica de 1,361.
- Se han realizado un mayor número de encuestas a los hombres frente a la mujeres (cuando mayor es el número de un segmento, más consistencia se tiene en los estadísticos calculados; segmentos muy pequeños pueden distorsionar los resultados).

**6.5.3. Análisis de varianza (la Existencias de diferencias significativas).** El análisis de varianza nos permite identificar las posibles diferencias entre los segmentos utilizados y adicionalmente permite determinar si estas diferencias son estadísticamente significativas.

El análisis de la varianza analiza la variabilidad en las respuestas del colectivo (hombres y mujeres), distinguiendo entre:

- **Variabilidad entre-grupos:** diferencias de opiniones del grupo de hombres frente al de las mujeres.
- **Variabilidad intra-grupos:** diferencia de opiniones dentro del grupo de los hombres (hombres frente a hombres) o dentro del grupo de las mujeres (mujeres frente a mujeres).
- **Variabilidad total.** Variabilidad inter-grupos más variabilidad intra –grupos.

Si lo que se desea es comprobar si existen diferencias significativas, motivadas por la pertenencia de las observaciones a uno u otro grupo, la mayor parte de la variabilidad total tiene que ver con los resultados de la varianza inter-grupos, mientras que la variabilidad intra-grupos debe ser un componente residual.

Con el fin de que el estudiante comprenda los resultados arrojados por el programa SPSS, se desarrollara en forma manual los resultados de la matriz ANOVA.

#### **Nomenclatura utilizada:**

$X_{ij}$  = Valores que toma la variable dependiente

$\bar{X}$  = Media de la variable dependiente

$\bar{X}_j$  = Media de la variable dependiente para cada grupo

$n$  = Número total de datos

$g$  = Número de grupos

$n_j$  = Número de datos de cada grupo

$STC$  = Sumatoria total al cuadrado

$SCE$  = Sumatoria al cuadrado entre grupos

$SCI$  = Sumatoria al cuadrado intra grupos

$gl$  = Grados de libertad

$gl_{n-1}$  = Grados de libertad variabilidad total

$gl_{g-1}$  = Grados de libertad variabilidad entre grupos

$gl_{(n-1)-(g-1)}$  = Grados de libertad intra grupos

$S_E^2$  = Varianza entre grupos

$S_I^2$  = Varianza intra grupos

$F = \text{Test F}$

**Tabla 6.11.**

**Análisis de varianza - Tabla de ANOVA**

			Suma de cuadrados	gl	Media cuadrática	F	Sig.
Variedad	Inter-grupos	(Combinadas)	1,646	1	1,646	,971	,325
* Sexo	Intra-grupos		713,205	421	1,694		
Total			714,851	422			

- **Suma total al cuadrado STC**

$$STC = SCE + SCI = \sum_{j=1}^g \sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2 = 714,851$$

- **Suma de cuadrados entre grupos (inter grupos)**

$$SCE = \sum_{j=1}^g n_j (X_{.j} - \bar{X})^2 = 271(3,833948 - 3,787234)^2 + 152(3,703947 - 3,787234)^2 = 1,645758$$

- **Suma de cuadrados intra grupos**

$$SCI = \sum_{j=1}^g \sum_{i=1}^{n_j} (X_{ij} - \bar{X}_j)^2 = STC - SCE = 714,851 - 1,646 = 713,205$$

- **Varianza entre grupos**

$$S_E^2 = \frac{SCE}{g-1} = \frac{1,646}{1} = 1,646$$

- **Varianza intra grupos**

$$S_I^2 = \frac{SCI}{(n-1)-(g-1)} = \frac{713,206}{(2-1)-(421)} = \frac{713,206}{421} = 1,694$$

- **Test F**

$$F = \frac{S_E^2}{S_I^2} = \frac{1,646}{1,694} = 0,971$$

- **El test F.** Partiendo de la hipótesis nula de que la variabilidad de opiniones de los sujetos se debe principalmente a la pertenencia, en este caso, a uno de los grupos (el

grupo de los hombres o el grupo de las mujeres), si el cociente resultante es alto (mayor que 1) y la significatividad asociada es igual o mayor a 0,05.

Como se puede ver el numerador refleja las diferencias **entre grupos** (en este caso hombres frente a mujeres); el denominador representa la diferencia de opiniones **intra grupo** (dentro de los hombres o dentro de las mujeres).

**Se concluye** que *no existen diferencias* significativas entre grupos (hombre y mujeres) a la hora de valorar las variables al comprar vestidos y accesorios, puesto que el F calculado es un valor alto (mayor que 1) y la significatividad es mayor que 0,05.

## **6.6. TABLAS CONTINGENCIA (Tabulaciones cruzadas).**

Este tipo de análisis permite la búsqueda de relaciones de asociación o dependencia entre dos variable, interpretando el sentido de dicha relación. Este análisis es interesante ya que permite trabajar con variables nominales.

**6.6.1. Ejemplo de tabla de contingencia.** Se trabajará con el ejemplo anterior donde se consideran las dos variables:

- **Variable sexo:** variable de clasificación.
- **Variable variedad:** Importancia concedida a la variedad de mercancía en los almacenes de la ciudad de Popayán. Los entrevistados han tenido que indicar si para ellos la variedad de mercancía lo valoran mucho o muy poco en una escala de 1 a 5.

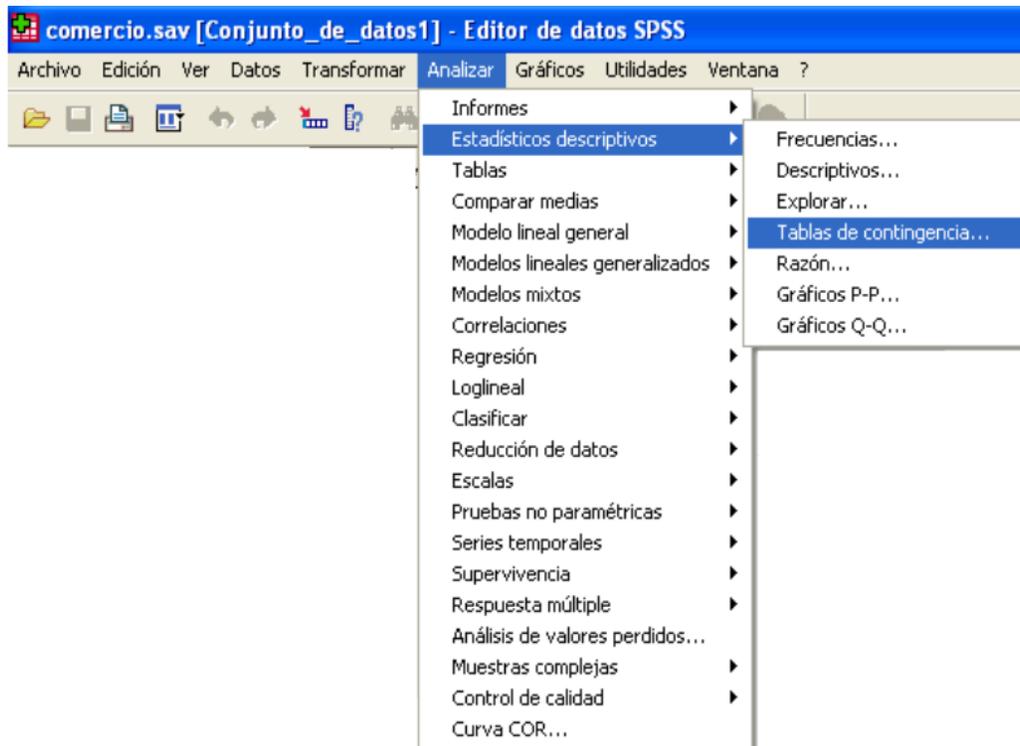
Se parte inicialmente de una posible influencia del sexo sobre la variable variedad. Se analiza por tanto la existencia de una potencial relación de dependencia.

- **Variable Independiente:** Sexo
- **Variable dependiente:** Variedad

**6.6.2. Procedimiento en el programa SPSS.** Para procesar la información en el programa SPSS haga los siguientes pasos.

- Seleccione en el menú principal la opción Analizar
- Estadísticos descriptivos
- Tablas de contingencia (Figura 6.10)

Figura 6.10. Tablas de contingencia.



En la ventana de tablas de contingencia seleccione las variables de la siguiente forma:

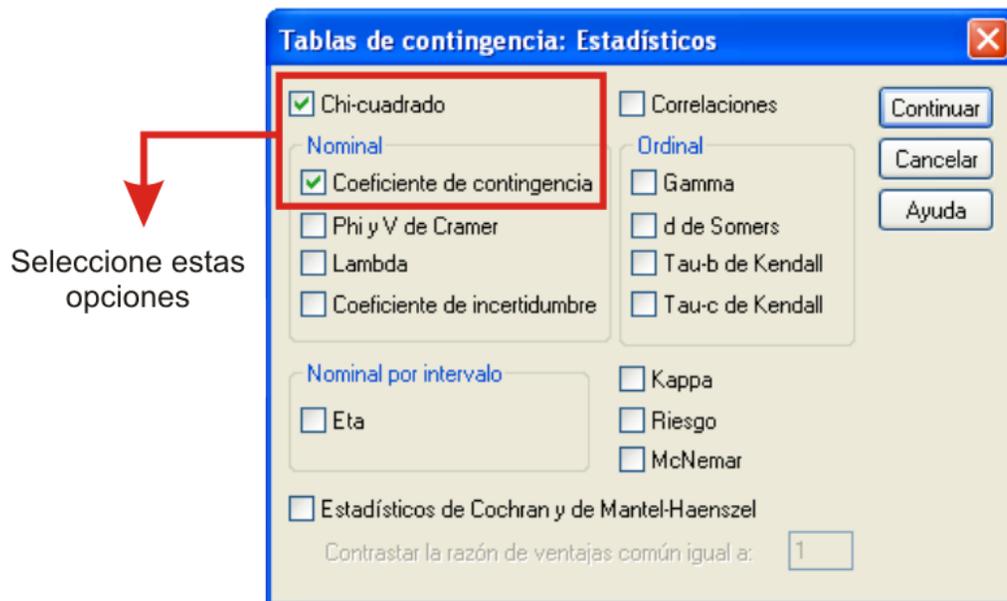
- Traslade la variable “variedad” a la sesión de filas;
- Traslade la variable de clasificación “sexo” a la sesión de columnas;
- Seleccione Estadísticos. (Figura 6.11).

Figura 6.11. Ventana de tablas de contingencia.



Seleccione esta opción

Figura 6.12. Ventana tabla de contingencia: Estadísticos

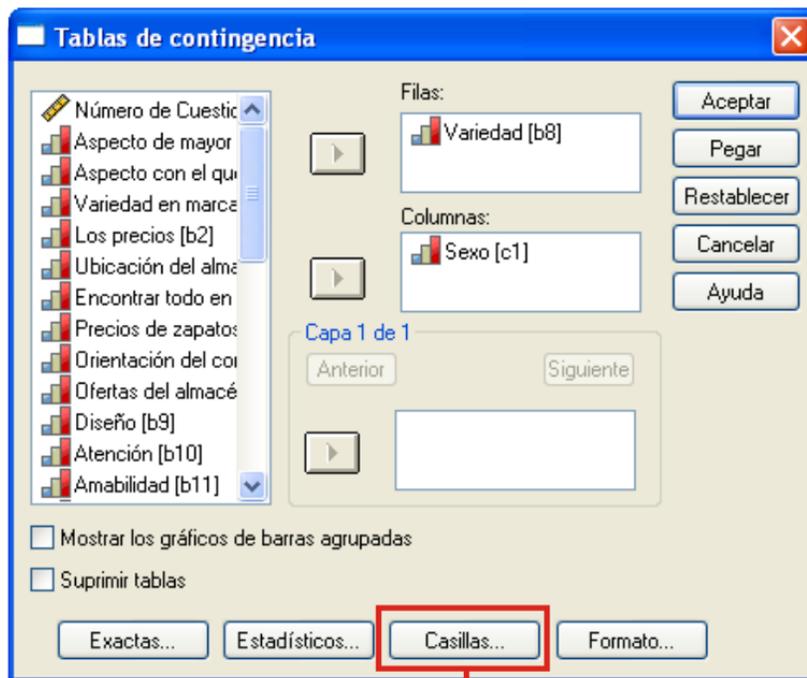


Seleccione estas opciones

En la ventana de estadísticos seleccione:

- Chi – cuadrado
- Coeficiente de contingencia y
- Presione continuar.
- Volvemos a la ventana inicial de tablas de contingencia
- Seleccione Casillas (Figura 6.13)

Figura 6.13. Ventana de tabla de contingencia.

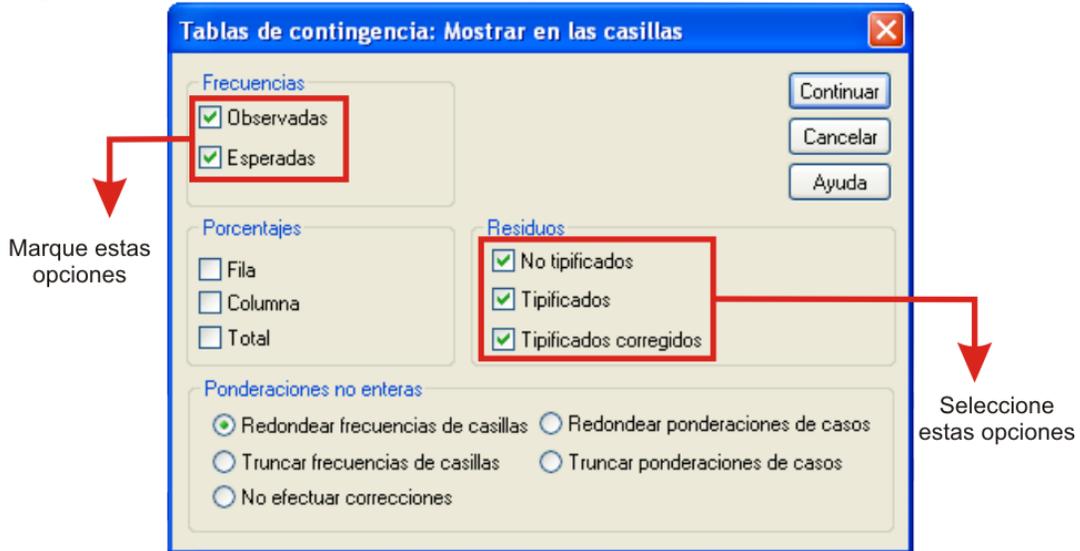


Seleccione esta opción

En la venta de tablas de contingencia: Mostrar en las casillas seleccione:

- Frecuencias observadas
- Frecuencias esperadas
- Residuos no tipificados
- Residuos tipificados
- Residuos tipificados corregidos
- Presione continuar (Figura 6.14)

Figura 6.14. Ventana de tablas de contingencias: Mostrar en las casillas



El visor de resultados debe tener una apariencia como se muestra en la Figura 6.15.

Figura 6.15. Ventana de resultados.

Resultados5 [Documento3] - Visor SPSS

Archivo Edición Ver Datos Transformar Insertar Formato Analizar Gráficos Utilidades Ventana ?

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Variedad * Sexo	423	91,8%	38	8,2%	461	100,0%

Frecuencias observadas - Frecuencias esperadas.  
Tabla de contingencia "Variedad" - "Sexo"

Variedad			Sexo		Total
			Hombre	Mujer	
1. Valora muy poco	Recuento		21	16	37
	Frecuencia esperada		23,7	13,3	37,0
2	Recuento		22	15	37
	Frecuencia esperada		23,7	13,3	37,0
3	Recuento		51	28	79
	Frecuencia esperada		50,6	28,4	79,0
4	Recuento		64	32	96
	Frecuencia esperada		61,5	34,5	96,0
5. Valora mucho	Recuento		113	61	174
	Frecuencia esperada		111,5	62,5	174,0
Total	Recuento		271	152	423
	Frecuencia esperada		271,0	152,0	423,0

Medidas simétricas

		Valor	Sig. aproximada
Nominal por nominal	Coefficiente de contingencia	,060	,818

SPSS El procesador está preparado

Al: 245 , An: 420 pt.

**6.6.3. Análisis de frecuencias.** Una vez se tengan los resultados se inicia con la comparación del recuento o frecuencia observada con la frecuencia esperada, bajo el supuesto de independencia, es decir bajo el supuesto, en este caso, de que no influye el ser hombre o mujer a la hora de valorar las características de un almacén cuando se compara vestidos y accesorios. La frecuencia esperada va a consistir en reparto proporcional entre las celdas existentes, calculándose de la siguiente forma:

$$F_{ij} = \frac{n_i \times n_j}{n}$$

Siendo

$F_{ij}$  = Frecuencia esperada para cada una de las celdas o subgrupos existentes (fila i, columna j)

$n_i$  = Subtotal de la fila i

$n_j$  = Subtotal de la columna j

$n$  = Número total de datos

Por ejemplo la frecuencia esperada de la fila uno (1) (valora muy poco), columna 1 hombres se calcula de la siguiente forma:

$$F_{ij} = \frac{n_i \times n_j}{n} = \frac{37 \times 271}{423} = 23,7$$

Las frecuencias esperadas representan el supuesto de independencia entre las variables, por lo que, cuanto mayor sea la diferencia entre los datos observados y la frecuencias esperadas, mayor será la posibilidad de que exista relación entre las dos variables consideradas.

Tabla 6.12

Frecuencias observadas - Frecuencias esperadas.  
Tabla de contingencia "Variedad" - "Sexo"

			Sexo		Total
			Hombre	Mujer	
Variedad	1. Valora muy poco	Recuento	21	16	37
		Frecuencia esperada	23,7	13,3	37,0
	2	Recuento	22	15	37
		Frecuencia esperada	23,7	13,3	37,0
	3	Recuento	51	28	79
		Frecuencia esperada	50,6	28,4	79,0
	4	Recuento	64	32	96
		Frecuencia esperada	61,5	34,5	96,0
	5. Valora mucho	Recuento	113	61	174
		Frecuencia esperada	111,5	62,5	174,0
Total		Recuento	271	152	423
		Frecuencia esperada	271,0	152,0	423,0

**6.6.4. Análisis de chi cuadrado.** El contraste de chi – cuadrado se convierte en la prueba que determine la existencia de dependencia entre las variables analizadas. Este contraste parte de la diferencia, anteriormente comentada, entre la frecuencia observada y la esperada. Su valor, por tanto, debe ser alto para validar la relación de dependencia entre las dos variables. A continuación se expone la forma de calcular este estadístico así como la tabla de su cálculo (Tabla 6.13)

$$X^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(f_{ij} - F_{ij})^2}{F_{ij}}$$

Donde:

$r$  = Filas

$c$  = Columnas

$f_{ij}$  = Frecuencia observada

$F_{ij}$  = Frecuencia esperada

Tabla 6.13. Cálculo de chi cuadrado

Celda i, j	$f_{ij}$	$F_{ij}$	$f_{ij} - F_{ij}$	$(f_{ij} - F_{ij})^2$	$\frac{(f_{ij} - F_{ij})^2}{F_{ij}}$
1,1	21	23,7	-2,7	7,29	0,31
1,2	16	13,3	2,7	7,29	0,55
2,1	22	23,7	-1,7	2,89	0,12
2,2	15	13,3	1,7	2,89	0,22
3,1	51	50,6	0,4	0,16	0,00
3,2	28	28,4	-0,4	0,16	0,01
4,1	64	61,4	2,6	6,76	0,11
4,2	32	34,5	-2,5	6,25	0,18
5,1	113	111,5	1,5	2,25	0,02
5,2	61	62,5	-1,5	2,25	0,04
				Suma =	1,55

En este caso la chi – cuadrado es  $1,548 \approx 1,55$ . Para determinar si este valor es lo suficientemente alto, se debe comparar con el de las tablas estadísticas de este contraste, teniendo un nivel de confianza del 95% y los grados de libertad correspondiente. Entonces hay dos alternativas para el análisis de este estadístico:

- Si el valor de Chi cuadrado es alto y la significancia asociada es inferior a 0,05: rechazamos la hipótesis nula. Es decir, que las variables están relacionadas.

- Si el valor de chi cuadrado es bajo y las significancia asociada es superior a 0,05 aceptamos la hipótesis nula. Las variables no están relacionadas.

En el ejemplo que estamos analizando, la significatividad asociada es 0,818 por tanto existe una probabilidad baja del 18,2% (1-0,818) de que el sexo de los individuos influya en la valoración de las características de los almacenes a la hora de comprar vestidos y accesorios.

**Tabla 6.14. Contraste de chi cuadrado – prueba de chi cuadrado**

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	1,548 <sup>a</sup>	4	,818
Razón de verosimilitudes	1,524	4	,822
Asociación lineal por lineal	,972	1	,324
N de casos válidos	423		

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 13,30.

Un aspecto importante en el análisis de tabulaciones cruzadas es la posible falta de datos. No hay que olvidar que en el análisis de tabulaciones cruzadas la muestra se divide en subgrupos. La falta de datos en las diferentes casillas puede afectar y sesgar los resultados, impidiendo la extrapolación de la muestra. Cuando el número de celdas con frecuencia esperada es inferior a 5 y supera el 20% del total de celdas, no es posible concluir, por falta de datos, ni a favor de la independencia ni de la dependencia.

En el ejemplo analizado no se tiene este problema. La frecuencia mínima esperada, es del 13,30 como se puede observar en la tabla 6.14. No hay, por tanto, celdas con frecuencias esperadas inferiores a 5.

**6.6.5. Análisis de los residuos y de los porcentajes básicos.** El residuo es igual al valor observado menos el valor esperado bajo el supuesto de independencia. Esta diferencia debe ser grande para que exista relación entre las variables analizadas. El residuo es una medida individualizada que permite analizar el comportamiento de la tabla, celda por celda. Lo primero que hay que comprobar es que el residuo ajustado o corregido sea significativo: es significativo siempre que su valor absoluto supere 1,96. Una vez comprobado, se analizan los residuos significativos interpretándolos en función de su signo. Si el signo es positivo, indica que hay más valores observados que cumplen las características de esa celda que los cabría esperar bajo el supuesto de independencia. Si el signo es negativo, indica que hay menos valores observados que cumplen las características de esa escala que los que cabría esperar bajo el supuesto de independencia.

En el ejemplo analizado son cuatro los residuos significativos:

- Residuo hombre “Valora muy poco”: es negativo. Hay menos hombres que han indicado esta opción de los que corresponderían en caso de independencia.
- Residuo hombre “Valora bastante – 4”: positivo. Hay más hombres que han indicado esta opción de los que corresponderían en caso de independencia.
- Residuo mujer “Valora muy poco”: Positivo. Hay más mujeres que han escogido esta opción que las que cabría esperar bajo el supuesto de independencia.
- Residuo mujer “Valora bastante – 4”: negativo. Hay menos mujeres que han escogido esta opción que las que cabría esperar bajo el supuesto de independencia.

Estos cuatro residuos nos indican que los hombres dan más importancia a la variedad de la mercancía que las mujeres.

**Tabla 6.15. Tabla de contingencia – análisis de residuos.**

			Sexo	
			Hombre	Mujer
Variedad	1. Valora muy poco	Residuo	-2,7	2,7
		Residuos tipificados	-,6	,7
		Residuos corregidos	-1,0	1,0
	2	Residuo	-1,7	1,7
		Residuos tipificados	-,4	,5
		Residuos corregidos	-,6	,6
	3	Residuo	,4	-,4
		Residuos tipificados	,1	-,1
		Residuos corregidos	,1	-,1
4	Residuo	2,5	-2,5	
	Residuos tipificados	,3	-,4	
	Residuos corregidos	,6	-,6	
5. Valora mucho	Residuo	1,5	-1,5	
	Residuos tipificados	,1	-,2	
	Residuos corregidos	,3	-,3	



# *Capítulo siete*

## *Métodos estadísticos de proyección*

---

### **7. ANÁLISIS DE PROYECCIONES MEDIANTE REGRESIÓN Y CORRELACIÓN (CON DOS VARIABLES).**

El pronóstico de la situación futura en el mercado del producto es uno de los aspectos de mayor importancia para todo proyecto. Debido a los continuos cambios que se presentan en el medio y específicamente en variables de tipo económico, social, político, tecnológico y ecológico los resultados proyectados deben ser analizados con cierta reserva. En este sentido, las cantidades demandadas y ofrecidas con los precios esperados de los años futuros deben tomarse como aproximaciones o simples datos de referencia para análisis y toma de decisiones y no como cálculos de absoluta certeza.

Existen variadas alternativas para determinar el comportamiento futuro para un producto, lo cual exige de los proyectistas un análisis detenido, dadas las implicaciones y exigencias de cada método, así como del tipo de proyecto que se esté trabajando, para seleccionar y aplicar correctamente el que se considere más adecuado a la situación objeto de estudio.

A menudo en cualquier estudio se presentan simultáneamente observaciones respecto a dos variables, ligadas por algún tipo de relación, por ejemplo la relación que hay entre el peso y la estatura de una persona, el precio de un producto y su demanda, la publicidad invertida y las ventas. En estos casos se debe hacer un análisis en tal forma que explique el tipo de relación que hay entre las variables observadas; esto se logra por medio de un procedimiento denominado análisis de regresión y de correlación.

Dado lo anterior, se trabajará con un ejemplo para pronosticar la producción de una empresa dedicada a la elaboración y comercialización de prendas de dotación industrial.

## 7.1. ESTUDIO DE CASO.

La empresa de confecciones de prendas de dotación y uniformes de labor “INCOLVES LTDA”, necesita realizar la proyección de la demanda del producto. Para realizar esta tarea se tiene series históricas del personal vinculados a empresas de la ciudad de Popayán. Estos datos fueron suministrados por la oficina de Planeación de la Caja de Compensación Familiar del Cauca “COMFACAUCA”; se escogió esta entidad dado que las empresas deben reportar las nóminas para el pago de parafiscales.

La información se presenta en la siguiente tabla:

**Tabla 7.1.** Trabajadores Reportados a ComfacaUCA.

Año	1993	1994	1995	1996	1997	1980	1999	2000
Trabajadores afiliados	29.448	29.504	31.079	32.194	34.892	34.802	34.702	34.592

*Fuente: Departamento de Planeación – COMFACAUCA*

Adicionalmente se realizó una encuesta en las empresas industriales de la ciudad; según datos arrojados se estima que el 70% de los trabajadores de estas empresas tienen derecho a dotación. De otra parte, los empresarios deben dar cumplimiento a la ley 11 de 1984 del Régimen Laboral Colombiano, la cual dice: “Todo patrono que habitualmente ocupe uno (1) o más trabajadores, deberá suministrar cada cuatro (4) meses en forma gratuita un par de zapatos y un vestido de labor al trabajador cuya remuneración mensual sea hasta de dos veces el salario mínimo legal más alto vigente.....”.

Dado lo anterior se procedió a calcular el número de trabajadores que tienen derecho a dotación.

### 7.1.1. Determinación del Número de Trabajadores con Derecho a Dotación.

**Tabla 7.2.** Trabajadores con Derecho a Dotación.

Año	1993	1994	1995	1996	1997	1980	1999	2000
Trabajadores afiliados	29.448	29.504	31.079	32.194	34.892	34.802	34.702	34.592
70% del total de afiliados	20.614	20.653	21.755	22.536	24.424	24.361	24.291	24.214

Los inversionistas están interesados en pronosticar el número de trabajadores con derecho a dotación para los años 2001, 2002, 2003, 2004 y 2005.

Para realizar la proyección haga los siguientes pasos:

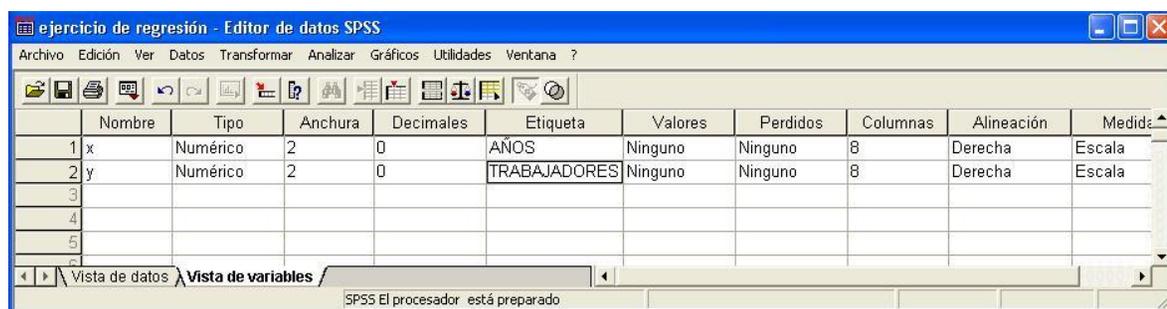
**Crear las variables.** Para crear las variables, siga los procedimientos explicados en capítulo V, punto 5. “DEFINICIÓN DE VARIABLES”, utilice los siguientes parámetros:

**Tabla 7.3.** Definición de las variables X, Y.

Nombre de la Variable	X(*)	Y
Tipo:	Numérica	Numérica
Anchura	2	2
Cifras decimales	0	0
Etiqueta	AÑOS	TRABAJADORES
Etiqueta de valor	Ninguno	Ninguno
Perdidos	Ninguno	Ninguno
Columnas	8	8
Alineación	Derecha	Derecha
Medida	Escala	Escala

(\*) Se reemplazan los valores de los años 1993 a 2000 por 1, 2, 3, ...8, respectivamente.

**Figura 7.1.** Creación de las Variables X y Y



**Digite la información,** como se muestra en la Figura 7.2.

**Figura 7.2.** Digitación de la información.

	x	y	var	var	var	var
1	1	20614				
2	2	20653				
3	3	21653				
4	4	22536				
5	5	24424				
6	6	24361				
7	7	24391				
8	8	24214				
9						

**Determinación del tipo de relación entre las variables.** La relación entre las variables puede ser de tipo lineal, cuadrático, cúbico, exponencial y otros; para determinar que tipo de relación se ajusta más a los datos, realice el siguiente procedimiento, elija:

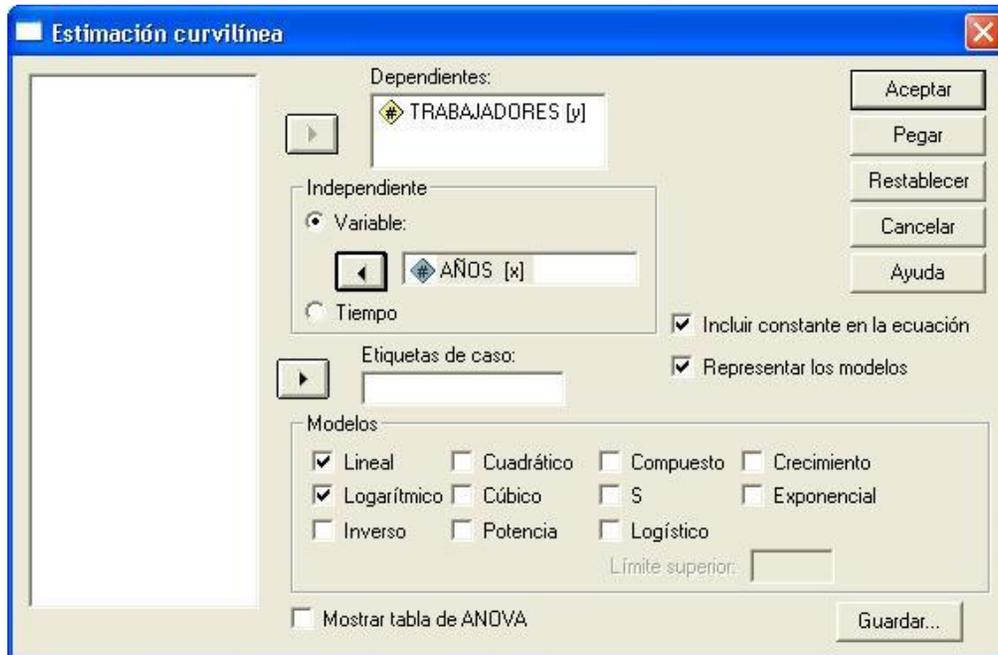
**Analizar → Regresión → Estimación curvilínea**

Como se muestra en la Figura 7.3.

**Figura 7.3.** Estimación Curvilínea de las Variables.



**Figura 7.4.** Definición de la Variable Independiente y la Variable Dependiente.



Después de definir las variables independiente y dependiente, en este caso, se analizará si la función tiene una relación de tipo lineal o logarítmica, es decir, se procede a ver cual de las dos funciones se ajusta mejor a la función que se pretende buscar. Marque lineal y logarítmica, en la sección de Modelos (Figura 7.4) y pulse el botón aceptar. El resultado se produce en modo texto así:

**Figura 7.5.** Ajuste a cada tipo de relación.

Independent: X

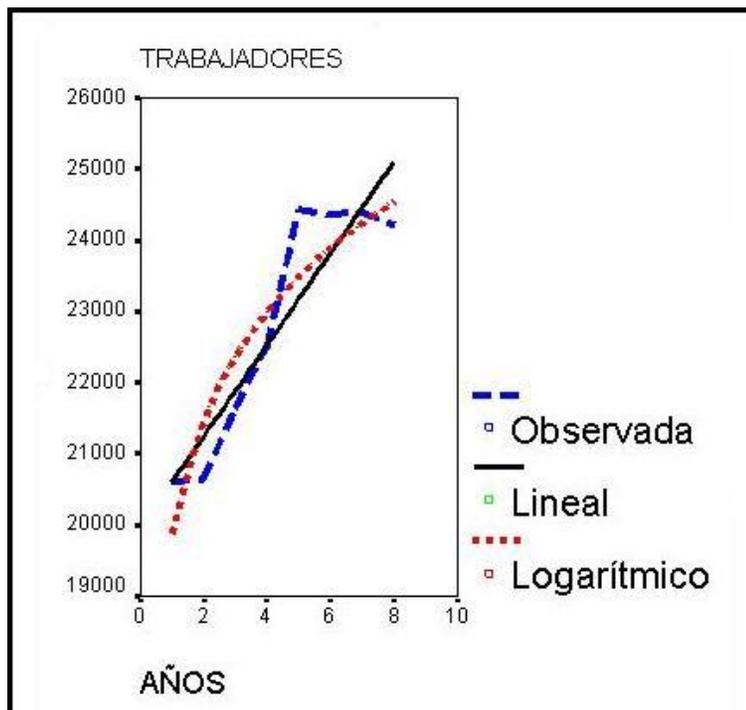
Dependent	Mth	Rsq	d.f.	F	Sigf	b0	b1
Y	LIN	,850	6	33,92	,001	19968,1	641,690
Y	LOG	,850	6	34,05	,001	19892,8	2235,25

Probabilidad  
Coeficiente de determinación  $R^2$   
Parámetros de cada modelo

El ajuste para cada tipo de función se comprueba con el coeficiente de correlación elevado al cuadrado (Rsq.  $R^2$ ). Se ve que los dos modelos tienen un alto coeficiente de determinación,  $R^2 = 0,85$ , es decir, que ambas  $R^2$  son significativas. Cuando los ajustes de varios modelos sean similares es conveniente usar el más parsimonioso, en este caso se trabajará con el modelo de regresión lineal.

La opción ofrece además una gráfica con la que se comprueba cómo los datos observados se ajustan a los modelos.

**Figura 7.6.** Ajuste de variables.



Otros de los datos suministrados por esta opción son los parámetros de las ecuaciones así:

$$\text{Modelo..lineal} \rightarrow Y_i = \beta_0 + \beta_1 X_i$$

$$\text{Modelo..lineal} \rightarrow Y_i = 19.968,1 + 641,69 X_i$$

$$\text{Modelo..log arítmico} \rightarrow Y_i = \beta_0 + \beta_1 \text{Log} X_i$$

$$\text{Modelo..log arítmico} \rightarrow Y_i = 19.892,8 + 2.235,25 \text{Log}(X_i)$$

Antes de continuar es preciso recordar algunos aspectos sobre el análisis de regresión.

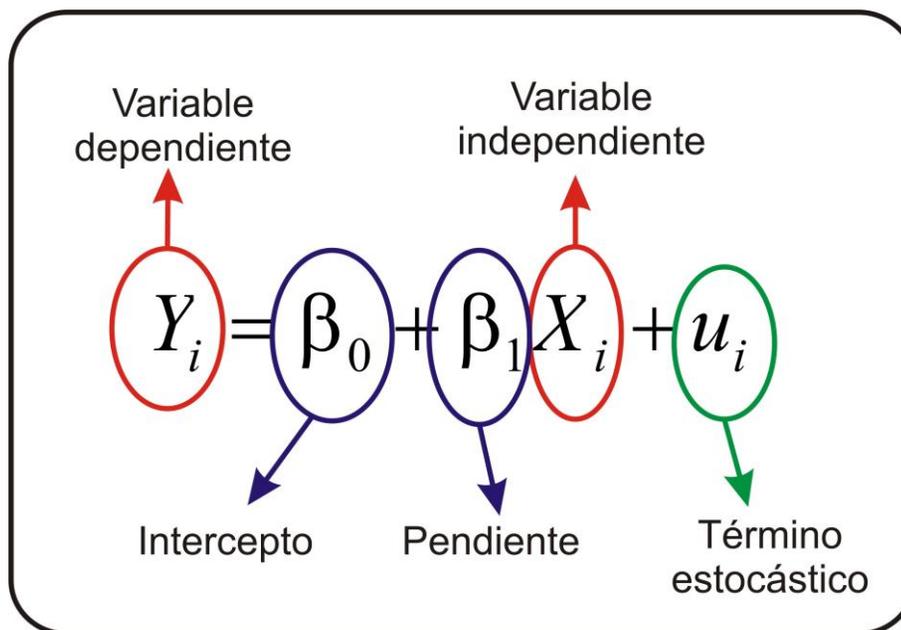
## 7.2. FUNCIÓN DE REGRESIÓN

El análisis de regresión trata del estudio de la dependencia de la variable dependiente, en una o más variables explicativas, con el objetivo de estimar y/o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos de la(s) segunda(s).

El análisis de regresión da lugar a una ecuación matemática que permite describir la relación existente entre dos variables. Es decir, obtener una línea “ideal” conocida como línea de regresión que nos describe la relación o dependencia entre dos variables.

**7.2.1. Significado del Término “Lineal”.** Puesto que la idea de linealidad se concibe con modelos de tipo.

**Figura 7.7.** Modelo de regresión lineal



Es esencial entender lo que significa realmente el término lineal, ya que puede ser interpretado de dos formas diferentes.

**7.2.1.1. Linealidad en las variables.** Geométricamente, la curva de regresión, en este caso, es una línea recta, es decir, que la variable X aparece elevada a la potencia uno (1).

**7.2.1.2. Linealidad en los parámetros.** La segunda interpretación de linealidad se presenta cuando los  $\beta$  están elevados a la potencia uno (1).

Para efectos del método de mínimos cuadrados se requiere que los parámetros estén elevados a la potencia uno, aunque las variables no lo estén. De acuerdo con esta interpretación

$$Y_i = \beta_0 X_i^{\beta_2} e^{u_i} \quad (22)$$

La ecuación (22) es un modelo de regresión lineal (en los parámetros pero no en la variable X), por tanto todos los modelos presentados en la figura 7.7 son modelos de regresión lineal, es decir, modelos lineales en los parámetros.

Las funciones del tipo (22) se pueden transformar en modelos lineales, mediante la manipulación de las propiedades de los logaritmos:

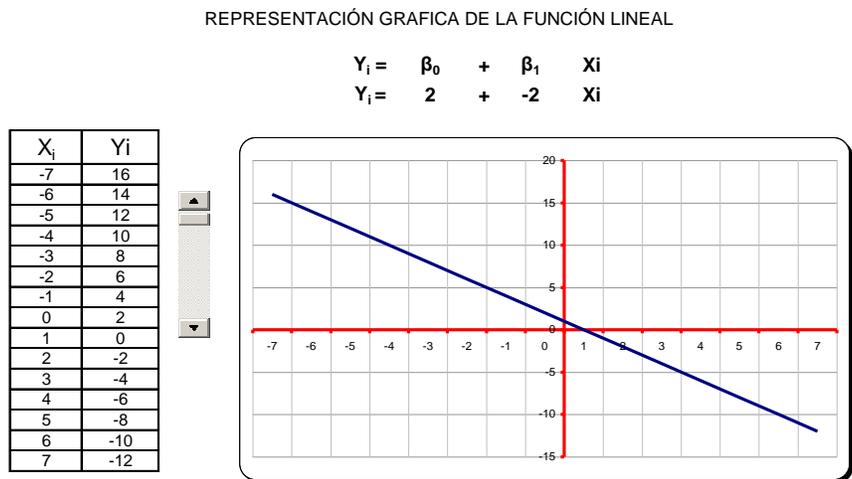
1.  $\text{Log}_b ac = \log_b a + \log_b c$
2.  $\text{Log}_b n/d = \log_b n - \log_b d$
3.  $\text{Log}_b m^p = p \log_b m$
4.  $\text{Log}_b a^x = x \log_b a$

Es decir, que mediante la aplicación de las propiedades de los logaritmos la ecuación (22), quedaría de la siguiente forma:

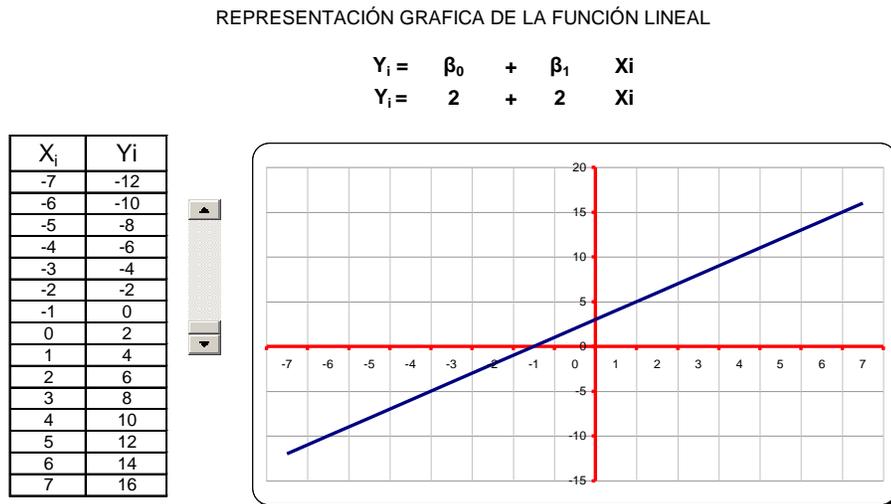
$$\text{Log} Y_i = \log \beta_0 + \beta_1 \log X_i + u_i \quad (23)$$

**7.2.2. Representación gráfica de una función lineal.** Las ecuaciones de primer grado con dos variables se llaman ecuaciones lineales porque representan líneas rectas.

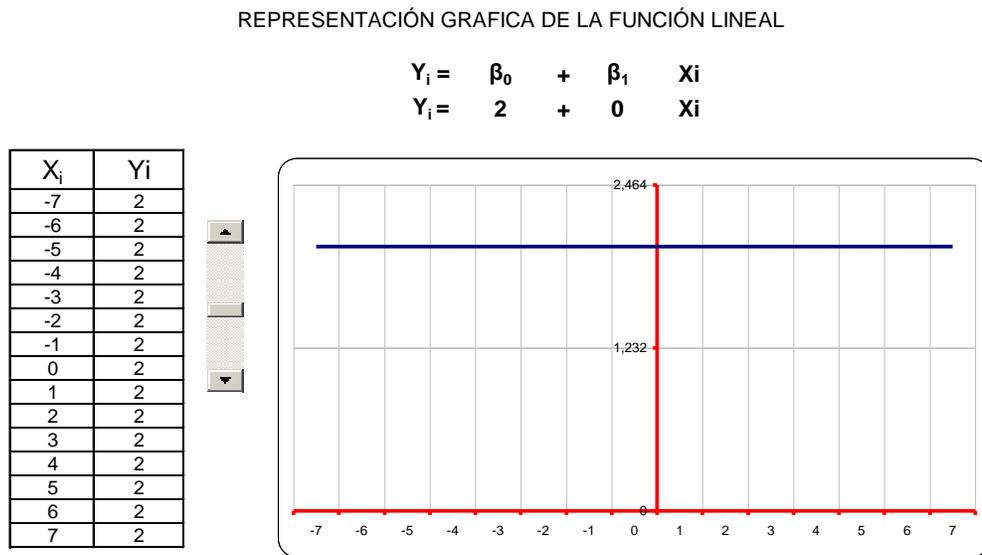
**Figura 7.8.** Representación Gráfica de la Función Lineal con Pendiente Negativa ( $\beta_1 < 0$ )



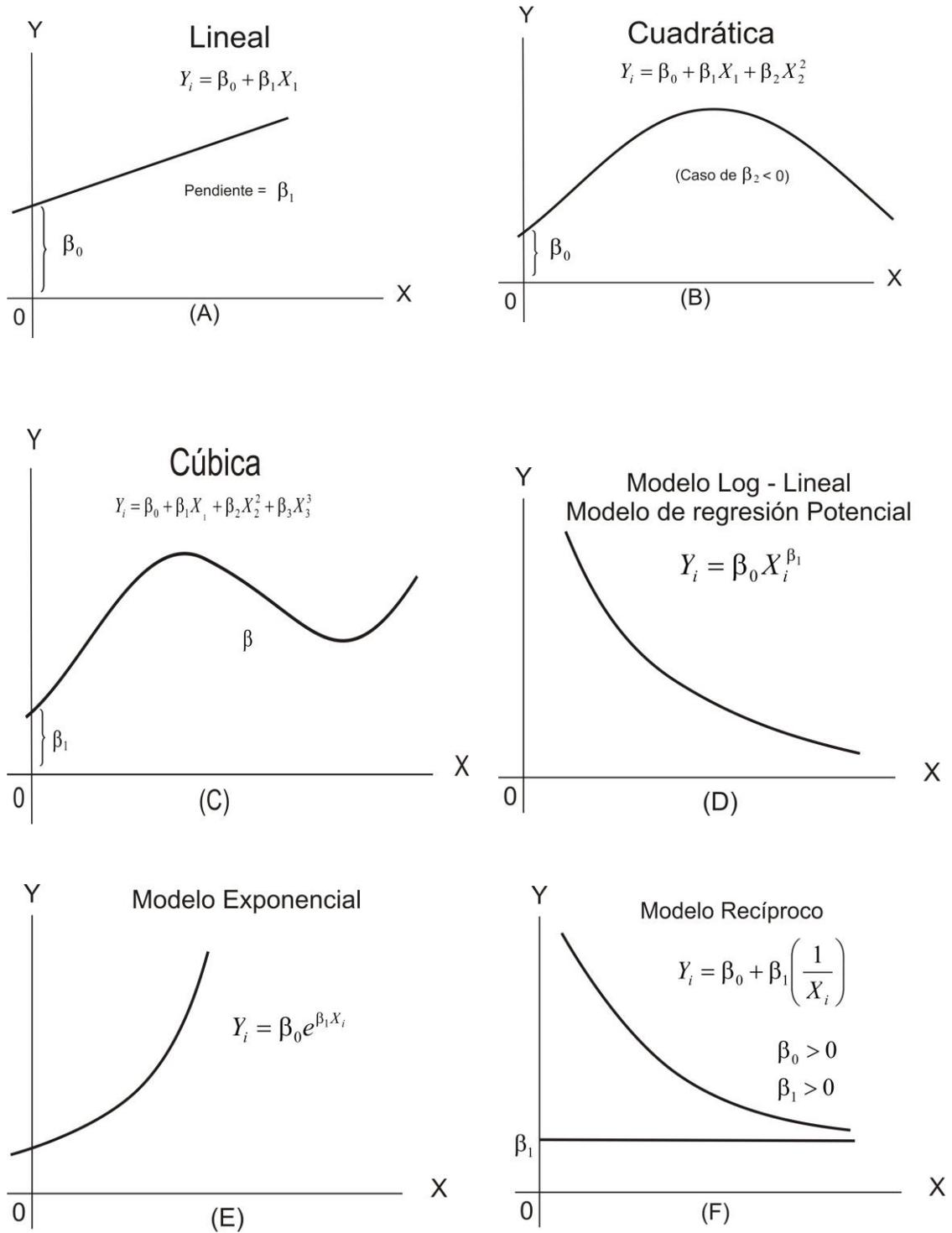
**Figura 7.9.** Representación Gráfica de la Función Lineal con Pendiente Positiva ( $\beta_1 > 0$ )



**Figura 7.10.** Representación Gráfica de la Función Lineal sin Pendiente ( $\beta_1 = 0$ )



**Figura 7.11.** Tipos de modelos de regresión lineal (lineal en los parámetros).



### 7.3. CALCULO DE LOS PARÁMETROS DE LA FUNCIÓN DE REGRESIÓN LINEAL (FRL).

El modelo de regresión simple, presenta la siguiente estructura:

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad (24)$$

Y : Es la variable dependiente.

$\beta_0$  : Es la ordenada en el origen.

$\beta_1$  : Es la pendiente de la recta.

$e_i$  : Es el término de error, es decir, la diferencia entre los valores predichos por la regresión y los valores reales.

$$e_i = Y_i - \beta_0 - \beta_1 X_i \quad (24-A)$$

El objetivo principal es realizar la estimación puntual de los parámetros  $\beta_0$  y  $\beta_1$ , mediante el siguiente procedimiento:

**7.3.1. Método de Mínimos Cuadrados Ordinarios (MCO).** El método de mínimos cuadrados ordinarios se atribuye a Carl Friedrich Gauss, un matemático alemán. Bajo ciertos supuestos este método tiene algunas propiedades estadísticas muy atractivas que lo han convertido en uno de los más eficaces y populares del análisis de regresión. Para entenderlo, se explicará el principio de los mínimos cuadrados.

#### La Función de regresión poblacional (FRP) con dos variables

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (25)$$

Esta función no es observable directamente, dado que por lo general no se dispone de la información, entonces esta función debe ser estimada con una **Función Regresión Muestral (FRM)**, ya que en la mayoría de las investigaciones generalmente se trabaja con muestras, es decir, que se calcula la **Función de Regresión Muestral (FRM)**.

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + u_i \quad (26)$$
$$Y_i = \hat{Y}_i + u_i$$



#### Nota

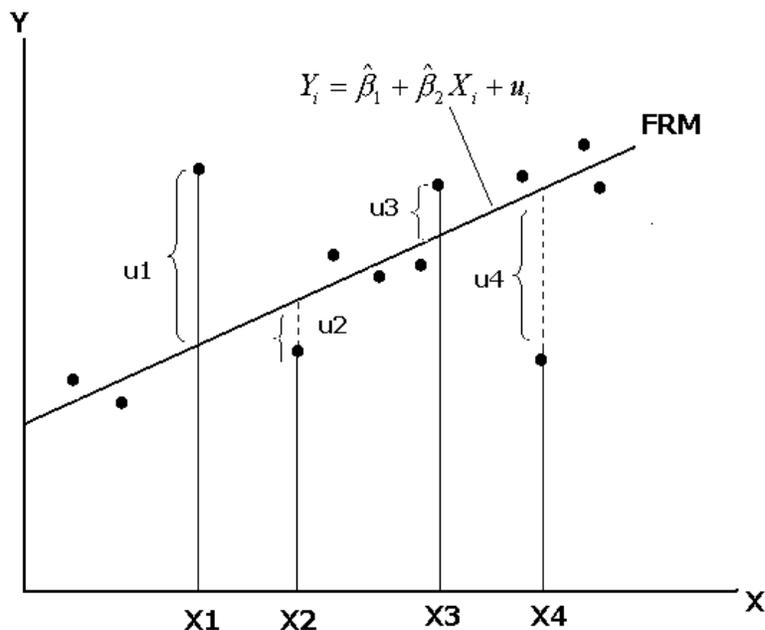
Para efectos de nomenclatura la función de regresión muestral se distingue de la función poblacional, en que los parámetros  $\beta$  se le

coloca un gorro así:  $\hat{\beta}_i$ .

Ahora, dados  $n$  pares de observaciones de  $Y$  y  $X$ , se está interesado en determinar la FRM de tal manera que esté lo más cerca posible a la FRP. Con este fin, se puede adoptar el siguiente criterio: seleccionar la FRM de tal manera que la suma de los residuos sea la menor posible. Este criterio, aunque es intuitivamente atractivo, no es muy bueno, como puede verse en el diagrama de dispersión que aparece en la figura 7.12.

Si se adopta el criterio de minimizar  $\sum u_i$ , la figura 7.12 muestra que los residuos  $u_2$  y  $u_3$  al igual que los residuos  $u_1$  y  $u_4$  reciben el mismo peso en las sumas ( $u_1 + u_2 + u_3 + u_4$ ), aunque los dos primeros están mucho más cerca de la FRM que los dos últimos. En otras palabras, a todos los residuales se les da la misma importancia sin importar qué tan cerca o qué tan lejos estén las observaciones individuales de la FRM. De tal forma que es muy posible que la suma algebraica de los  $u_i$  sea pequeña (aun cero). Se puede evitar este problema si se adopta el criterio de mínimos cuadrados, el cual establece que la FRM pueda determinarse en forma tal que  $\sum \hat{u}_i^2 = \sum (Y_i - \hat{Y}_i)^2$  o  $\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$ , sea lo más pequeña posible, donde  $u_i$  son los residuos elevados al cuadrado. Al elevar al cuadrado  $u_i$  este método da más peso a los residuos tales como  $u_1$  y  $u_4$ . En la figura 7.12 se aprecia mejor los efectos de este método.

**Figura 7.12.** Función de regresión muestral



El principio de mínimos cuadrados escoge  $\beta_1$  y  $\beta_2$  de tal manera que para una muestra dada o conjunto de datos,  $\sum u_i^2$  sea la más pequeña posible. En otras palabras, una muestra dada proporciona valores estimados únicos de  $\beta_1$  y de  $\beta_2$  que producen el valor más pequeño o reducido posible de  $\sum u_i^2$ .

Diferenciando la ecuación **24-A**, parcialmente con respecto a  $\beta_1$  y  $\beta_2$ , se obtiene:

$$\begin{aligned} \frac{\partial(\sum \hat{u}_j^2)}{\partial \hat{\beta}_1} &= -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = -2\sum \hat{u}_i \\ -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) &= 0 \\ \sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) &= 0 \end{aligned} \quad (27)$$

$$\text{PRIMERA ECUACIÓN NORMAL: } \sum Y = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \quad (28)$$

$$\begin{aligned} \frac{\partial(\sum \hat{u}_j^2)}{\partial \hat{\beta}_2} &= -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)X_i = -2\sum \hat{u}_i X_i \\ -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)X_i &= 0 \\ \sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)X_i &= 0 \end{aligned}$$

$$\text{SEGUNDA ECUACIÓN NORMAL } \sum X_i Y_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \quad (29)$$

Donde **n** es el tamaño de la muestra. Estas ecuaciones simultáneas se conocen como ecuaciones normales. Resolviendo las ecuaciones normales simultáneamente, se obtiene:

$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} \quad (30)$$

Reemplazando este valor en la primera ecuación obtenemos  $\beta_1$ , así:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \quad (31)$$

O por el sistema de ecuaciones se tiene:

$$\hat{\beta}_1 = \frac{\sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2} \quad (32)$$

**7.3.2. Procesamiento de la información en SPSS.** Ahora bien, el programa SPSS, facilita el cálculo de estos parámetros haciendo el siguiente procedimiento:

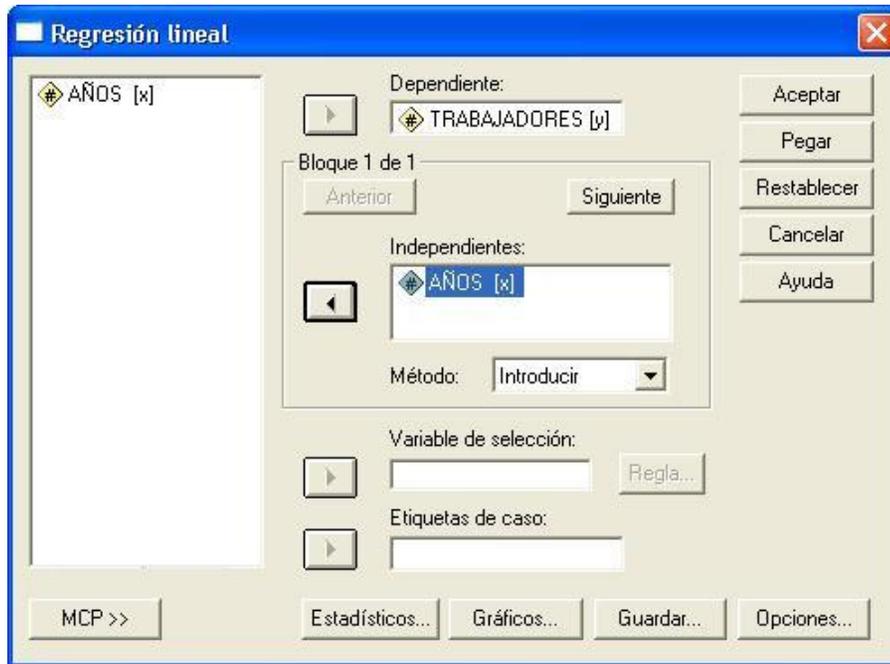
1. **Analizar** (figura 7.13)
2. **Regresión.**
3. **Lineal.** Cuando se selecciona esta opción se despliega el cuadro de diálogo (Figura 7.14)

**Figura 7.13.** Selección de la Opción Regresión.



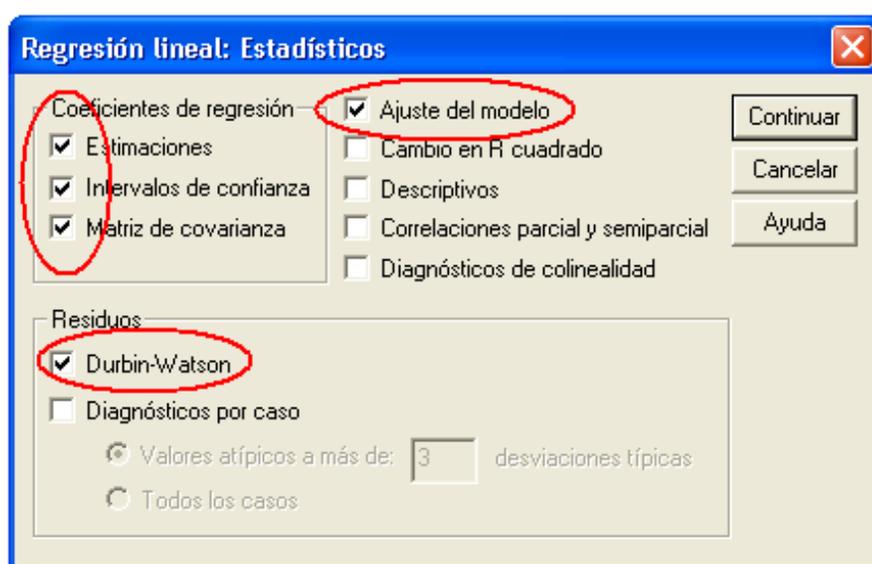
Coloque las variables como se muestran en la Figura 7.14.

**Figura 7.14.** Definición de las Variables.



Haga clic en la opción de **Estadísticos** y marque las opciones como se ven en la Figura 7.15.

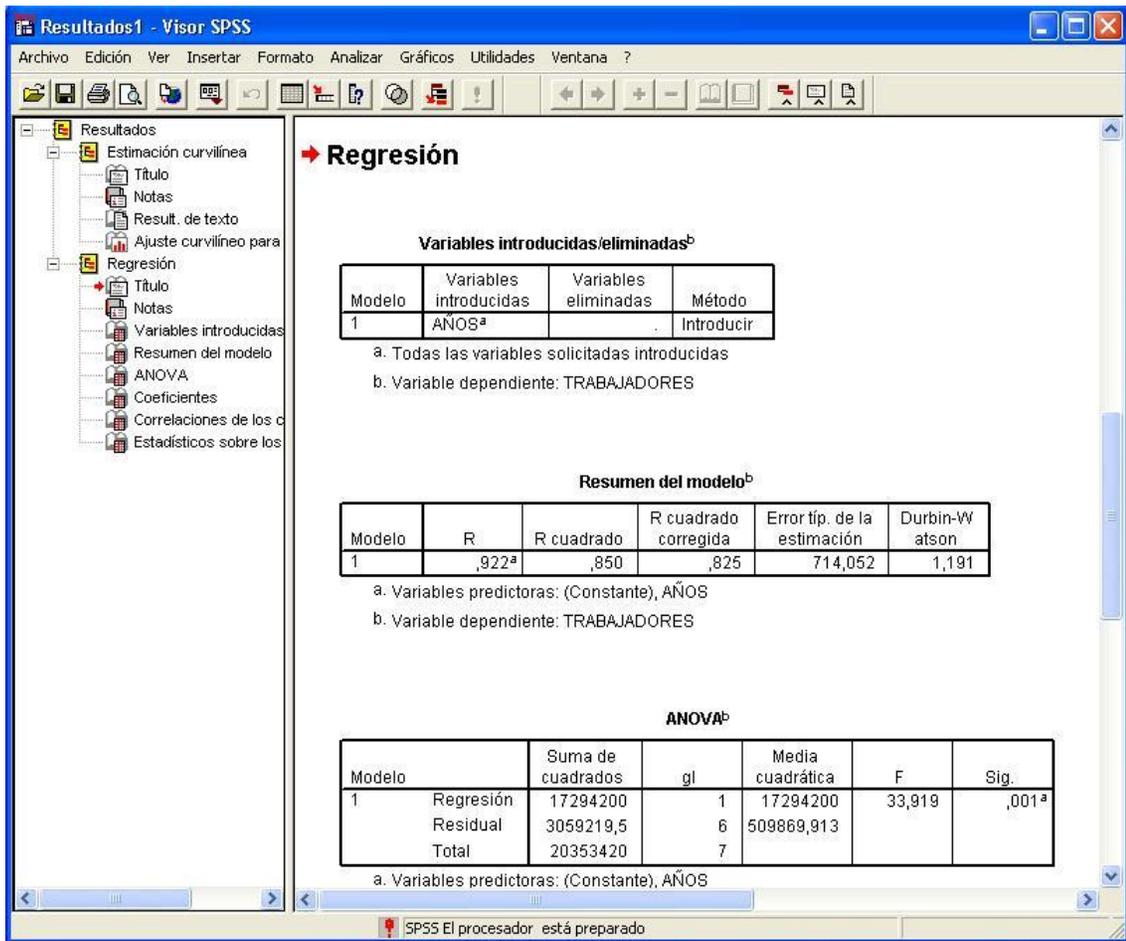
**Figura 7.15.** Regresión lineal: Estadísticos.



Haga clic en **Continuar**.  
 Haga clic en **Aceptar**.

El visor de resultados se presenta en la figura 7.16.

**Figura 7.16.** Visor de Resultados de la Regresión.



Los resultados se presentan en modo de tablas así:

**Figura 7.17.** Coeficientes de correlación y determinación

Resumen del modelo <sup>b</sup>				
Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,922 <sup>a</sup>	,850	,825	714,052

a. Variables predictoras: (Constante), AÑOS  
b. Variable dependiente: TRABAJADORES

Este primer cuadro muestra los cálculos de los coeficientes de correlación R y determinación  $R^2$ . Para interpretar estos coeficientes es preciso realizar un repaso relacionado con los conceptos de los mismos.

#### 7.4. COEFICIENTE DE DETERMINACIÓN $R^2$ : MEDIDA DE LA BONDAD DEL AJUSTE.

El coeficiente de determinación mide la bondad del ajuste, es decir, se ve que tan bien se ajusta la línea de regresión a los datos. Es claro que si todas las observaciones caen en la línea de regresión, se obtendrá un ajuste perfecto, pero rara vez se presenta este caso. Generalmente hay algunas  $u_i$  positivas y algunas  $u_i$  negativas, se espera que estos residuos alrededor de la muestra sean lo más pequeños posibles. El coeficiente de determinación  $R^2$  es una medida de resumen que nos dice qué tan bien se ajusta la línea de regresión muestral a los datos; en otras palabras *el coeficiente de determinación mide la proporción o el porcentaje de la variabilidad en Y explicada por el modelo de regresión.*

##### 7.4.1. Propiedades del Coeficiente de Determinación $R^2$ .

1. Es una cantidad no negativa.
2. Sus límites son  $0 \leq r \leq 1$ . Un  $R^2$  de 1 significa un ajuste perfecto, es decir,  $\hat{Y}_i = Y_i$  para cada i. Por otra parte, un R de cero significa que no hay relación alguna entre las variables dependiente y la variable explicativa (es decir,  $\beta_2 = 0$ ). En este caso, como lo indica la ecuación  $\hat{Y}_i = \beta_1 = \bar{Y}$ , es decir, la mejor predicción de cualquier valor de Y es simplemente el valor de su media. Por consiguiente, en esta situación, la línea de regresión será horizontal al eje X.

Su valor se determina así:

$$r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} \quad (33)$$

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} \quad (34)$$

### 7.5. COEFICIENTE DE CORRELACIÓN R.

Una cantidad estrechamente relacionada con  $R^2$ , pero conceptualmente muy diferente de éste, es una medida de asociación entre dos variables. Se calcula a partir de

$$r = \pm \sqrt{r^2}$$

o a partir de su definición

$$r = \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}}$$

$$r = \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2] \times [n \sum Y_i^2 - (\sum Y_i)^2]}} \quad (35)$$

En el contexto de la regresión,  $R^2$  es una medida con más significado que R. El grado de correlación lo podemos clasificar tentativamente, evitando un tanto la rigidez de sus límites

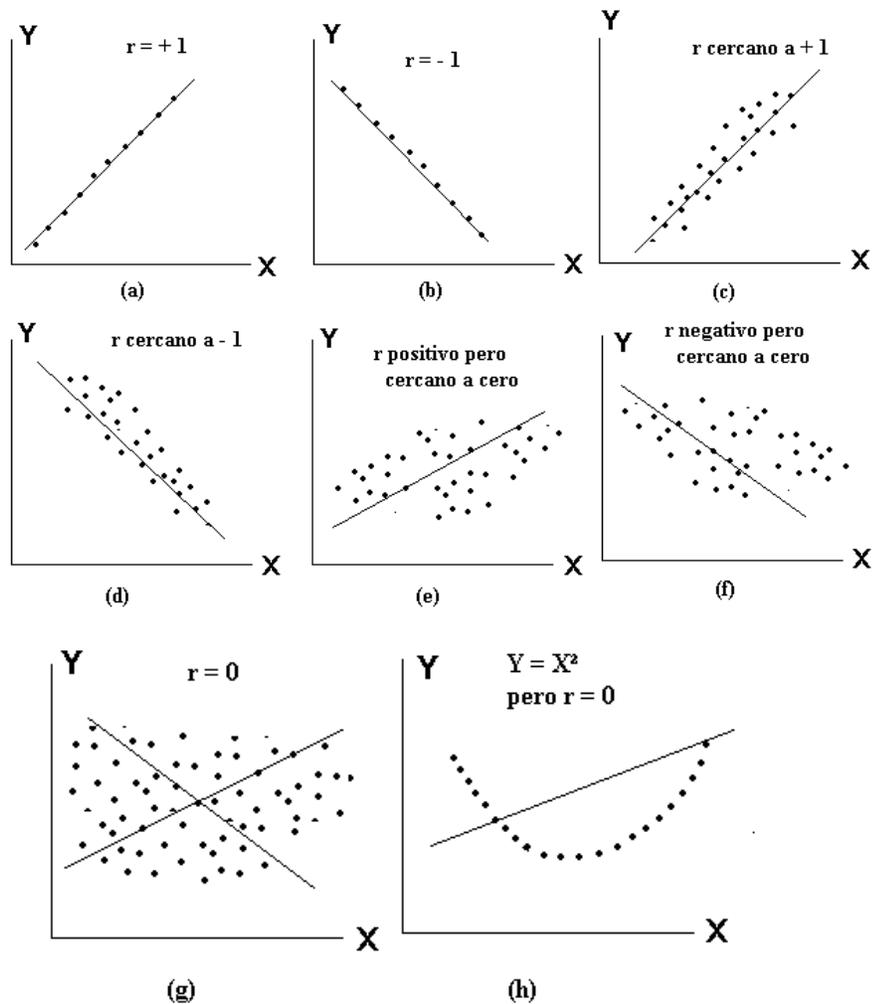
- a) Correlación perfecta, cuando  $(r = 1 \text{ o } -1)$
- b) Correlación excelente, cuando  $(-1 < r < -0,90)$
- c) Correlación aceptable, cuando  $(-0,90 < r < -0,80)$
- d) Correlación regular, cuando  $(-0,80 < r < -0,60)$
- e) Correlación mínima, cuando  $(-0,60 < r < -0,30)$
- f) No hay correlación, cuando  $(-0,30 < r < 0)$

#### 7.5.1. Propiedades del Coeficiente de Correlación R.

1. Puede tener signo positivo o negativo, dependiendo del signo del término en el numerador, el cual mide la covariación muestral de dos variables.
2. Cae entre los límites de  $-1$  y  $+1$ ; es decir,  $-1 \leq R \leq 1$ .
3. Es simétrico por naturaleza; es decir, el coeficiente de correlación entre X y Y es el mismo que entre Y y X.

4. Es independiente del origen y de la escala.
5. Es una medida de asociación lineal o dependencia lineal solamente; su uso en la descripción de relaciones no lineales no tiene significado. Así en la figura 76 (h),  $Y=X^2$  es una relación exacta y aún  $r$  es cero.
6. Aunque es una medida de asociación lineal entre dos variables, esto no implica necesariamente alguna relación causa-efecto.

**Figura 7.18.** Tipos de Correlación.



### 7.6. CALCULO DE LOS PARÁMETROS $\beta$ 's .

Uno de los objetivos es calcular los parámetros  $\beta$ 's , como se observa en la figura 7.19.

**Figura 7.19.** Calculo de los Parámetros.

		Coeficientes no estandarizados		Coeficientes estandarizados			Intervalo de confianza para B al 95%	
Modelo		B	Error típ.	Beta	t	Sig.	Límite inferior	Límite superior
1	(Constante)	19968,143	556,385		35,889	,000	18606,719	21329,567
	AÑOS	641,690	110,181	,922	5,824	,001	372,088	911,293

a. Variable dependiente: TRABAJADORES

## 7.7. PRESENTACIÓN DE LA INFORMACIÓN.

En este texto se presenta la información de siguiente forma:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$Y_i = 19.968,14 + 641,69 X_i + u_i$$

$$t = (35,889) \quad (5,824) \quad (36)$$

$$\rho = (0,000) \quad (0,001)$$

$$R^2 = 85\%$$

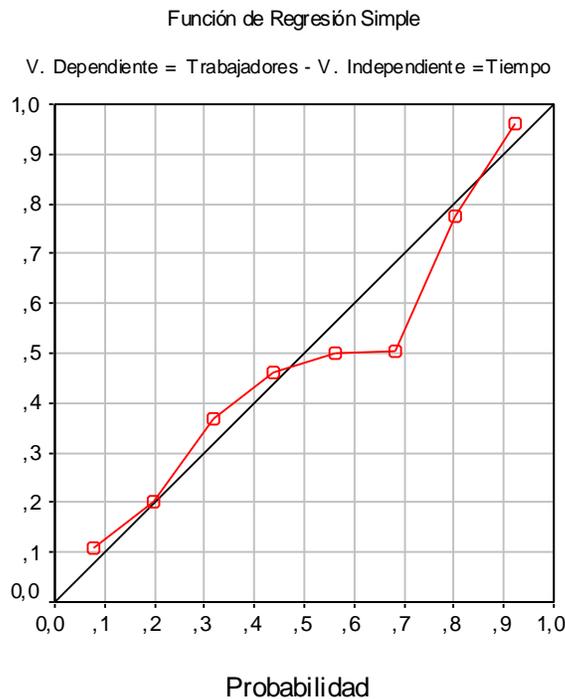
- La primera fila corresponde a la función de regresión lineal;
- la cifras de la segunda fila hacen referencia a los parámetros de la función;
- Las cifras del tercer fila corresponde a los valores de t estimado, bajo la hipótesis nula de que el verdadero valor poblacional de cada coeficiente de regresión lineal es cero;
- Las cifras de la cuarta fila corresponde a la probabilidad de que los parámetros estimados sean iguales a cero.
- La cifra de la quinta fila corresponde al Coeficiente de Determinación  $R^2$ .

### 7.7.1. Interpretación de los resultados

- **Interpretación del  $\beta_0 = 641,69$ .** Significa que anualmente se reportan en promedio 642 trabajadores con derecho a dotación a COMFACAUCA.
- **Interpretación de  $\rho = 0,001$ .** La probabilidad que el parámetro  $\hat{\beta}_1$  sea igual cero es aproximadamente de 0%.
- **Interpretación de  $R^2 = 85\%$ .** Significa que el 85% de las variaciones del número de trabajadores reportados a COMFACAUCA están explicados por la variable tiempo. Considerando que  $R^2$  puede tomar un valor máximo de uno (1), se puede afirmar que los datos se ajustan bastante bien a la ecuación encontrada.

 <p><b>Véase también</b></p>	<p>Compare los resultados de esta opción, con los datos de la figura 7.5. “Ajustes a cada tipo de relación”</p>
---	---

**Figura 7.20.** Representación del Modelo de Regresión Lineal Simple.



### 7.8. PROYECCIÓN DE DATOS.

Uno de los objetivos del proyecto de inversión, es estimar la demanda futura del mismo. Los datos suministrados por la ecuación de regresión son una buena fuente para realizar este trabajo.

En el ejemplo que se está realizando se requiere proyectar el número de trabajadores con derecho a dotación para los años 9, 10, 11, 12, 13 y 14, es decir, los años 2001, 2002, 2003, 2004 y 2005.

Para realizar la proyección reemplace en la ecuación (36) el valor de X por 9, 10, 11, 12, 13 y 14, así:

**Tabla 7.4.** Proyección de datos  
(Demanda futura del número de trabajadores con derecho a dotación).

DETALLE	FUNCIÓN	VALORES PROYECTADOS
Función de regresión	$Y_i = \beta_0 + \beta_1 X_i$	
Parámetros de la función	$Y_i = 19.968,14 + 641,69 X_i$	
Estimación para el año 2001	$Y_i = 19.968,14 + [641,69 \times (9)]$	25.743
Estimación para el año 2002	$Y_i = 19.968,14 + [641,69 \times (10)]$	26.385
Estimación para el año 2003	$Y_i = 19.968,14 + [641,69 \times (11)]$	27.027
Estimación para el año 2004	$Y_i = 19.968,14 + [641,69 \times (12)]$	27.668
Estimación para el año 2005	$Y_i = 19.968,14 + [641,69 \times (13)]$	28.310

# Capítulo ocho

## Formas funcionales de los modelos de regresión

---

### 8. FORMAS FUNCIONALES DE LOS MODELOS DE REGRESIÓN.

En el capítulo anterior se analizó el modelo lineal en los parámetros y en las variables. A continuación se considera dos tipos de modelos de regresión que son lineales en los parámetros pero no en las variables. Estas funciones se pueden transformar en modelos lineales mediante transformaciones apropiadas. En particular se analizará los modelos de regresión:

- El modelo log-lineal
- Modelo semilogarítmico

Se analizará las características especiales de cada modelo, los casos en los cuales su uso es apropiado y la forma como son estimados mediante el programa SPSS. Cada modelo es ilustrado con un ejemplo apropiado.

#### 8.1. MODELO LOG – LINEAL (POTENCIAL).

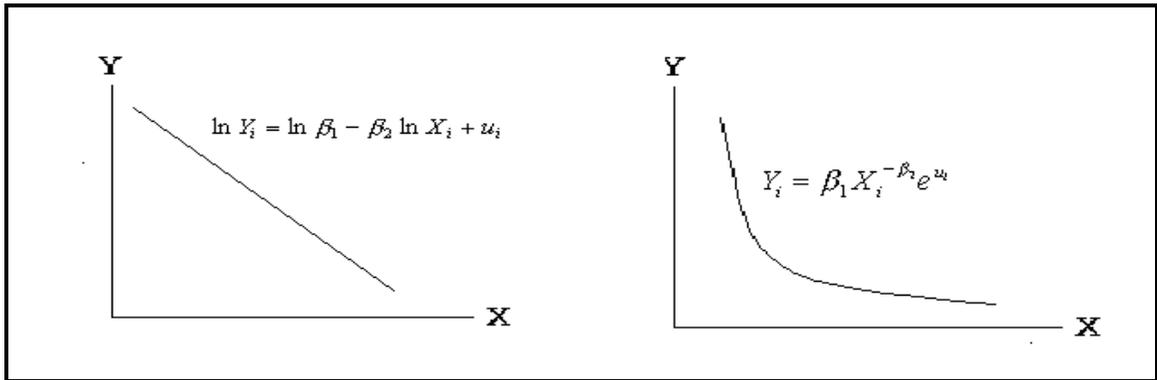
El modelo de regresión potencial representado en la siguiente ecuación

$$Y_i = \beta_1 X_i^{-\beta_2} e^{u_i} \quad (37)$$

Este modelo es lineal en los parámetros pero no en las variables, sin embargo, puede ser expresado en forma lineal mediante algunas transformaciones haciendo uso de las propiedades de los logaritmos, así:

$$\ln Y_i = \ln \beta_1 - \beta_2 \ln X_i + u_i \quad (38)$$

**Figura 8.1.** Representación Gráfica del Modelo Log – Lineal.



### 8.1.1. Características del Modelo Log – Lineal

- **Mide la elasticidad.** Una característica importante de este modelo es que el coeficiente de la pendiente  $\beta_2$  mide la elasticidad de Y con respecto a X, es decir, el cambio porcentual en Y ante un cambio porcentual en X dado.
- **El coeficiente de elasticidad del modelo permanece constante.**

El coeficiente de la elasticidad, en la notación de cálculo, se define como:

$$\left[ \frac{\frac{\partial Y}{Y}}{\frac{\partial X}{X}} \right] = \left[ \frac{\partial Y}{\partial X} \times \frac{X}{Y} \right] \quad (38)$$

Ahora bien, se tiene que:  $\left( \frac{\partial \text{Log} X}{\partial X} \right) = \frac{1}{X} = \frac{\partial X}{X}$  (39)

Es decir, para cambios infinitesimales pequeños (obsérvese el operador diferencial  $\partial$ ), un cambio en  $\log X$  es igual al cambio relativo o proporcional en X. Sin embargo, si el cambio en X es pequeño, esta relación puede escribirse como: cambio en  $\log X \approx$  cambio relativo en X, donde  $\approx$  significa aproximadamente. Así, para cambios pequeños, se tiene:

- **Cambio relativo en X**  $\approx (\log X_t - \log X_{t-1}) = \frac{(X_t - X_{t-1})}{X_{t-1}}$  (40)
- **Cambio absoluto en X**  $.(X_t - X_{t-1})$

- **Cambio porcentual o tasa de crecimiento porcentual**  $\left[ \frac{(X_t - X_{t-1})}{X_{t-1}} \right] \times 100$

Donde:  $X_t$  = Valor actual y  $X_{t-1}$  valor anterior.

## 8.2. ESTUDIO DE CASO<sup>16</sup>.

En la tabla 8.1 se da la información del consumo per cápita por libras de carne de pollo en los Estados Unidos (1960 a 1982) y el precio real al menudeo del pollo por libras, se requiere estimar la función de elasticidad constante y con los parámetros proyectar el consumo para los años 1983, 1984, 1985, 1986.

**Tabla 8.1.** Consumo de Pollo y Precio por Libra.

Años	Y Consumo per cápita de carne de Pollo	X Precio US\$, al menudeo de carne de Pollo por libras
1960	27.9	42.2
1961	29.9	38.1
1962	29.8	40.3
1963	30.8	35.9
1964	31.2	37.3
1965	33.3	38.1
1966	35.6	39.3
1967	36.4	37.8
1968	36.7	38.4
1969	38.4	40.1
1970	40.4	38.6
1971	40.3	39.8
1972	41.8	39.7
1973	40.4	52.1
1974	40.7	48.9
1975	40.1	58.3
1976	42.7	57.9
1977	44.1	56.5
1978	46.7	63.7
1979	50.6	61.6
1980	50.1	58.9
1981	51.7	66.4
1982	52.9	70.4

<sup>16</sup> GUJARATI, Damodar. Econometría. 2003, p. 229 y 230.

### 8.2.1. Creación de las Variables en el Programa SPSS.



Véase también

Para la creación de las variables realice el procedimiento explicado en el capítulo V, punto 5. “Definición de Variables”.

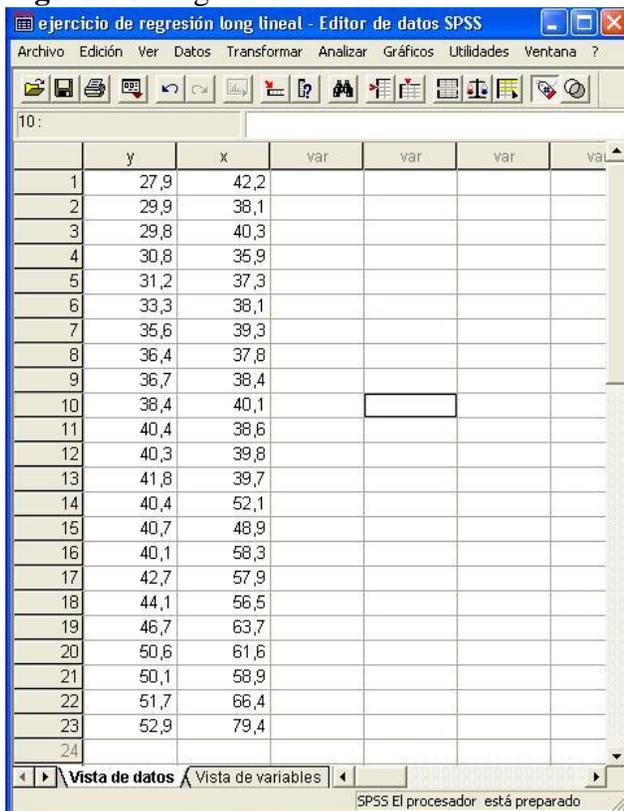
Figura 8.2. Creación de las Variables.



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	y	Numérico	8	1	Consumo per cápita de pollo	Ninguno	Ninguno	8	Derecha	Escala
2	x	Numérico	8	1	Precio raal en US\$ por libra de pollo	Ninguno	Ninguno	8	Derecha	Escala
3										
4										
5										
6										
7										

Digite los datos como se muestra en la Figura 8.3.

Figura 8.3. Digitación de los datos



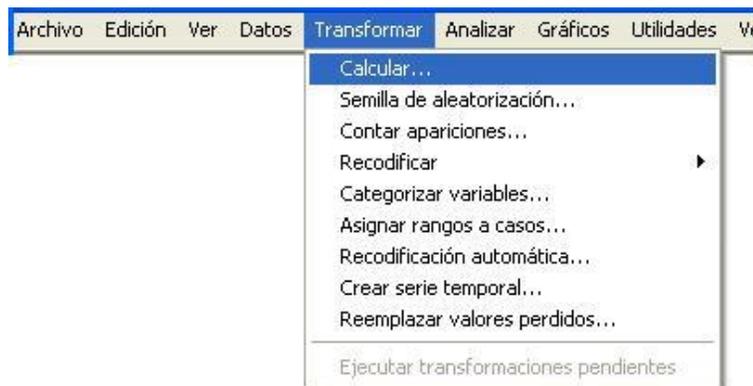
	y	x	var	var	var	val
1	27,9	42,2				
2	29,9	38,1				
3	29,8	40,3				
4	30,8	35,9				
5	31,2	37,3				
6	33,3	38,1				
7	35,6	39,3				
8	36,4	37,8				
9	36,7	38,4				
10	38,4	40,1				
11	40,4	38,6				
12	40,3	39,8				
13	41,8	39,7				
14	40,4	52,1				
15	40,7	48,9				
16	40,1	58,3				
17	42,7	57,9				
18	44,1	56,5				
19	46,7	63,7				
20	50,6	61,6				
21	50,1	58,9				
22	51,7	66,4				
23	52,9	79,4				
24						

### 8.2.2. Creación y Cálculo de los Logaritmos de las Variables.

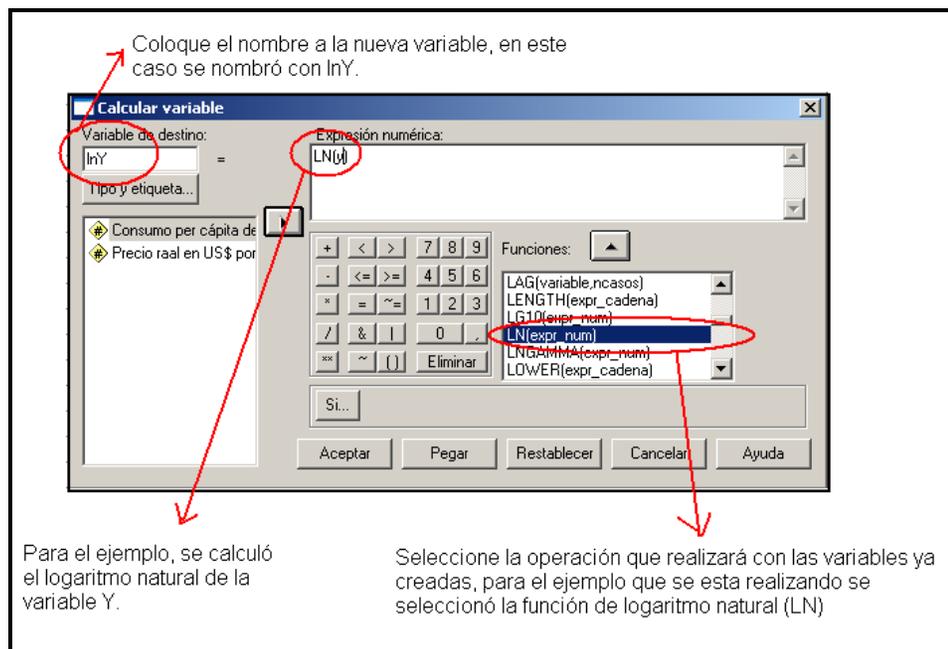
Para crear los logaritmos de las variables, realice los siguientes pasos:

- Seleccione la opción Transformar del menú principal
- Seleccione el submenú Calcular (Figura 8.4).
- Una vez seleccionado esta opción aparece un cuadro de diálogo como se muestra en la figura 8.4.

**Figura 8.4.** Uso de la opción transformar para la creación y cálculo de los logaritmos de las variables.



**Figura 8.5.** Uso de la Opción Calcular Variable.

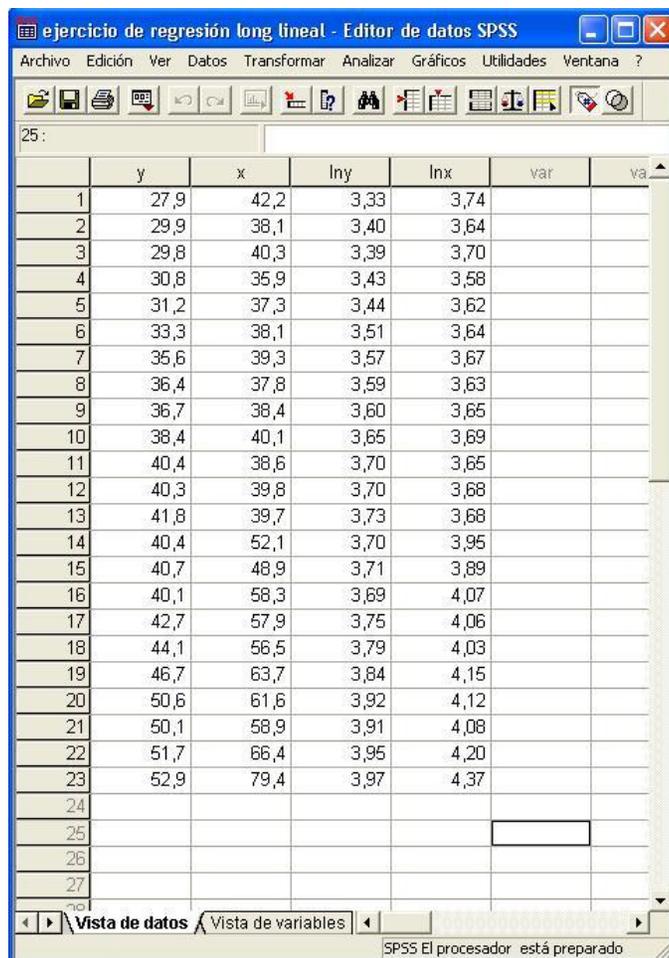


Continúe con los siguientes procedimientos:

- En el cuadro de diálogo **Calcular Variable**, escriba en la casilla **Variable de Destino** el nombre de la nueva variable: LNY
- En la casilla Expresión numérica se calcula la variable, se puede escribir directamente en ella o se puede emplear la lista de las variables, los operadores aritméticos, los números y las funciones. En este caso, se utilizó los elementos del cuadro de diálogo así:

- Seleccione la función **LN(Exp. Num)** y haga clic en el botón 
- Seleccione la variable Y y haga clic en el botón 
- Haga clic en el botón **Aceptar**, para crear la variable Lny, la ventana de datos debe tener una apariencia similar a la figura 8.6.

**Figura 8.6.** Creación de la Variable lny.



	y	x	lny	ln x	var	va
1	27,9	42,2	3,33	3,74		
2	29,9	38,1	3,40	3,64		
3	29,8	40,3	3,39	3,70		
4	30,8	35,9	3,43	3,58		
5	31,2	37,3	3,44	3,62		
6	33,3	38,1	3,51	3,64		
7	35,6	39,3	3,57	3,67		
8	36,4	37,8	3,59	3,63		
9	36,7	38,4	3,60	3,65		
10	38,4	40,1	3,65	3,69		
11	40,4	38,6	3,70	3,65		
12	40,3	39,8	3,70	3,68		
13	41,8	39,7	3,73	3,68		
14	40,4	52,1	3,70	3,95		
15	40,7	48,9	3,71	3,89		
16	40,1	58,3	3,69	4,07		
17	42,7	57,9	3,75	4,06		
18	44,1	56,5	3,79	4,03		
19	46,7	63,7	3,84	4,15		
20	50,6	61,6	3,92	4,12		
21	50,1	58,9	3,91	4,08		
22	51,7	66,4	3,95	4,20		
23	52,9	79,4	3,97	4,37		
24						
25						
26						
27						

Realice el mismo procedimiento para crear la variable **lnx**.

**8.2.3. Estimación del modelo log – lineal.** Para estimar el modelo se realiza el mismo procedimiento que se explicó para estimar el modelo lineal así:

- Seleccione Analizar del menú principal.
- Seleccione la opción Regresión.
- Seleccione la opción Lineal del submenú, como se muestra en la Figura 8.7.

**Figura 8.7.** Opción Analizar → Regresión → lineal.



Una vez seleccionada la opción lineal aparece una ventana, como se muestra en la Figura 8.8. Realice el siguiente procedimiento:

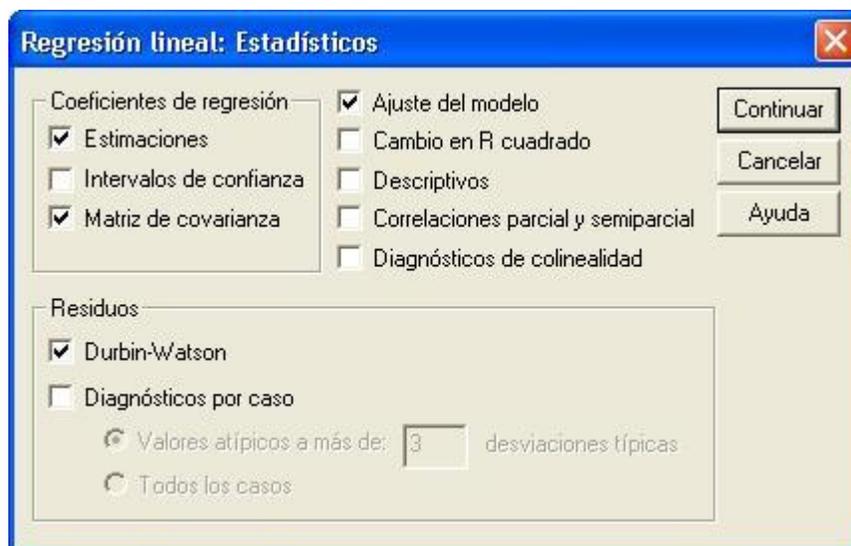
- Seleccione la variable Lny y haga clic en el botón , y pase la variable a la casilla **Dependiente**.
- Seleccione la variable Lnx y haga clic en el botón , y pase la variable a la casilla **Independientes**.

Figura 8.8. Selección de variables.



- Haga clic en el botón **Estadísticos**, aparecerá un cuadro de diálogo similar al mostrado en la Figura 8.9.

Figura 8.9. Opción Estadísticos.

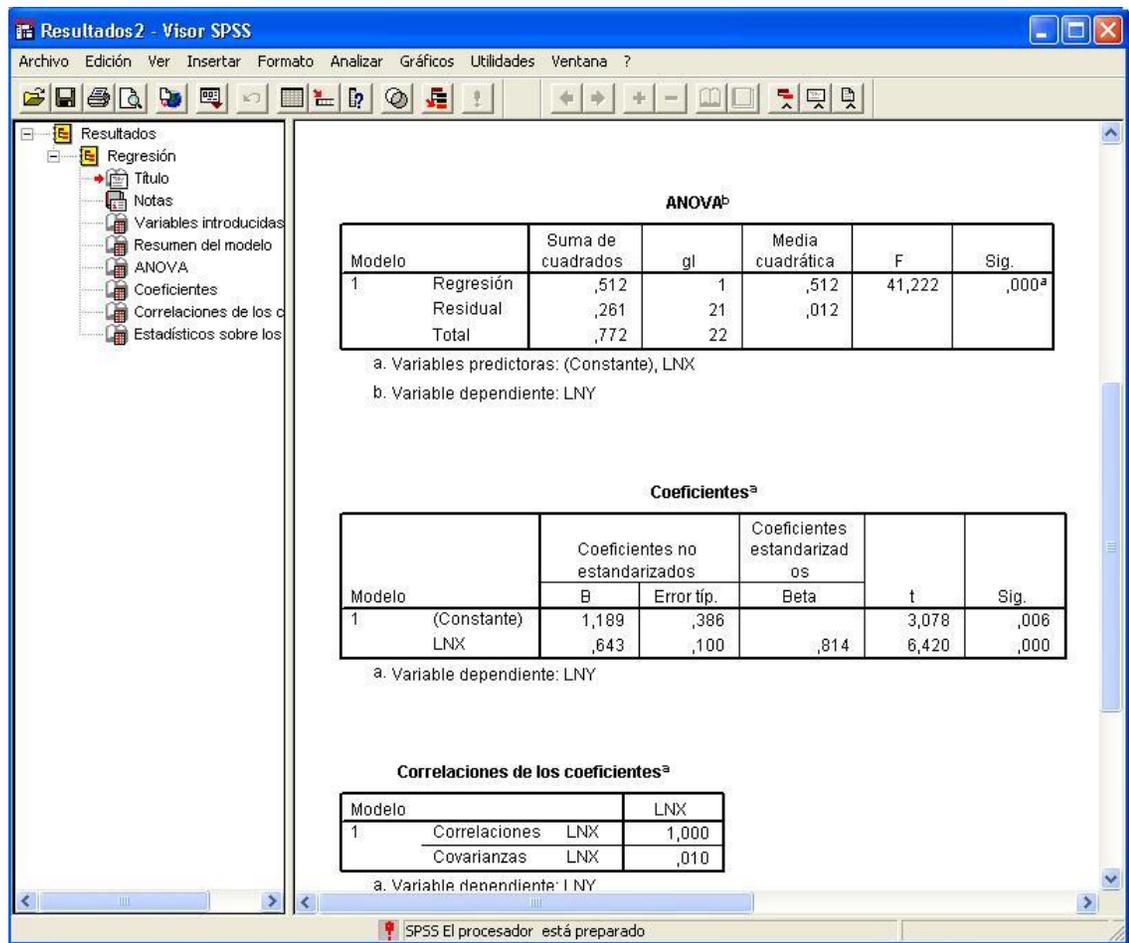


Seleccione las opciones como se muestra en la grafica 8.9.

- Haga clic en continuar.
- Haga clic en Aceptar para ver los resultados. Figura 8.10.

**8.2.4. Análisis de los Parámetros del Modelo.** El visor de los resultados presenta la información que se solicitó en esta opción, como se muestra en la figura 8.10.

**Figura 8.10.** Resultados de la Regresión.



**8.2.5. Presentación de resultados.** Los resultados que se presentan en las figuras 8.11 y 8.12.

**Figura 8.11.** Estimación de los Coeficientes del Modelo Log – Lineal

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	1,189	,386		3,078	,006
	LNx	,643	,100	,814	6,420	,000

a. Variable dependiente: LNY

**Figura 8.12.** Estimación de los coeficientes de correlación y determinación del modelo log – lineal.

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación	Durbin-Watson
1	,814 <sup>a</sup>	,662	,646	,11141	,527

a. Variables predictoras: (Constante), LNx

b. Variable dependiente: LNY

Los resultados se presentan de la siguiente forma:

$$\begin{aligned}
 \ln Y_i &= \beta_0 + \beta_1 \ln X_i + u_i \\
 \ln Y_i &= 1,189 + 0,643 \ln X_i + u_i \\
 t &= (3,078) \quad (6,420) \\
 \rho &= (0,006) \quad (0,000) \\
 R^2 &= 66,2\%
 \end{aligned}
 \tag{41}$$

- La primera fila corresponde a la función de regresión log -lineal;
- la cifras de la segunda fila hacen referencia a los parámetros de la función;
- Las cifras de la tercera fila corresponde a los valores de t estimado, bajo la hipótesis nula de que el verdadero valor poblacional de cada coeficiente de regresión sean iguales a cero;
- Las cifras de la cuarta fila corresponde a la probabilidad de que los parámetros estimados sean iguales a cero.
- La cifra de la quinta fila corresponde al Coeficiente de Determinación  $R^2$ .

### 8.2.6. Interpretación de los Resultados.

- **Interpretación del  $\beta_1 = 0,643$ .** Significa que si el precio al menudeo de la libra de pollo aumenta en un 1%, en promedio, el consumo de pollo se incrementa casi en un 0,64%. En consecuencia el consumo de carne de pollo no es tan sensible al cambio en el precio. Es decir que, la carne de pollo se clasifica como bien inelástico, o de primera necesidad, ya que la elasticidad precio de la demanda es inferior a 1.
- **Interpretación de  $\rho = 0,000$ .** La probabilidad del que el parámetro  $\hat{\beta}_1$  sea igual cero es de 0%, es decir, que en muestreo repetido la probabilidad de que una muestra similar a la tomada en este ejemplo de parámetros iguales a cero es nula.
- **Interpretación de  $R^2 = 66,2\%$ .** Significa que las variaciones del consumo de pollo están explicadas por la variable precio. Considerando que  $R^2$  puede tomar un valor máximo de uno (1), se puede afirmar que los datos se ajustan bastante bien a la ecuación encontrada.

**8.2.7. Proyección de los Resultados.** En el ejemplo que se está realizando se requiere proyectar el consumo per cápita de carne de pollo, cuando los precios se incrementan por la inflación esperada del 6% así:

El último precio reportado se incrementa por la inflación esperada, US\$ 79,4 x (1,06) = US\$ 84,16.

**Tabla 8.2.** Cálculos de Proyección de Datos.

DETALLE	FUNCIÓN	CONSUMO PROYECTADO
Función de regresión	$\ln Y_i = \ln \beta_0 + \beta_1 \ln X_i$	
Parámetros de la función	$\ln Y_i = \ln(1,189) + 0,643 \ln X_i$	
Para un precio = 84,16	$\ln Y_i = \ln(1,189) + 0,643 \times \ln(84,16)$ $Y_i = \text{Ant } \ln(1,189) + [0,643 \times 84,16]$ $Y_i = 3,2838 + [0,643 \times 84,16]$ $Y_i = 57,3987$	57,3987

### 8.3. MODELO LIN – LOG (SEMILOGARÍTMICO).

En algunos casos se está interesado en encontrar la tasa de crecimiento de las variables de interés para el proyecto, como ventas, precios de los productos, costo de materias primas, etc.

Supongamos que deseamos conocer la tasa de crecimiento de los trabajadores con derecho a dotación que son reportados a COMFACAUCA. Para realizar los cálculos se debe utilizar la fórmula de interés compuesto.

$$Y_t = Y_0(1 + r)^t \quad (42)$$

Donde:

- $r$  es la tasa de crecimiento compuesta de  $Y$  (es decir, a través del tiempo).
- $t$  es la tendencia del tiempo (variable tiempo).

Tomando el logaritmo natural se puede expresar de la siguiente manera:

$$\begin{aligned} \ln Y_t &= \ln Y_0 + t \ln(1 + r) + u_t \\ \text{donde: } \beta_1 &= \ln Y_0 \\ \text{y } \beta_2 &= \ln(1 + r) \end{aligned} \quad (43)$$

Quedando la ecuación de la siguiente forma:

$$\ln Y_t = \beta_1 + \beta_2 t + u_t \quad (44)$$

Este modelo es igual a cualquier otro modelo de regresión lineal, en el sentido de que los parámetros  $\beta_1$  y  $\beta_2$  son lineales. La particularidad está en que la variable dependiente o regresada es el logaritmo de  $Y$  y la variable independiente es el tiempo, que adquiere valores de 1,2,3, etc.

Antes de realizar los cálculos veamos las propiedades de este modelo:

- En este modelo el coeficiente de la pendiente mide el cambio proporcional constante o relativo en  $Y$  para un cambio absoluto dado en el valor de la variable independiente, en este caso  $t$ ; es decir, si se multiplica el cambio relativo en  $Y$  por 100 nos dará entonces el cambio porcentual, o la tasa de crecimiento, en  $Y$  ocasionada por un cambio absoluto en  $X$ .
- Este modelo es particularmente útil en situaciones en las cuales la variable  $X$  es el tiempo, el modelo describe la tasa de crecimiento constante relativa ( $\beta_2$ ) o porcentual ( $100 \times \beta_2$ ), si  $\beta_2 > 0$  será una tasa creciente y si  $\beta_2 < 0$  será una tasa decreciente, por esta razón este tipo de regresión se denomina *modelos de crecimiento constante*.

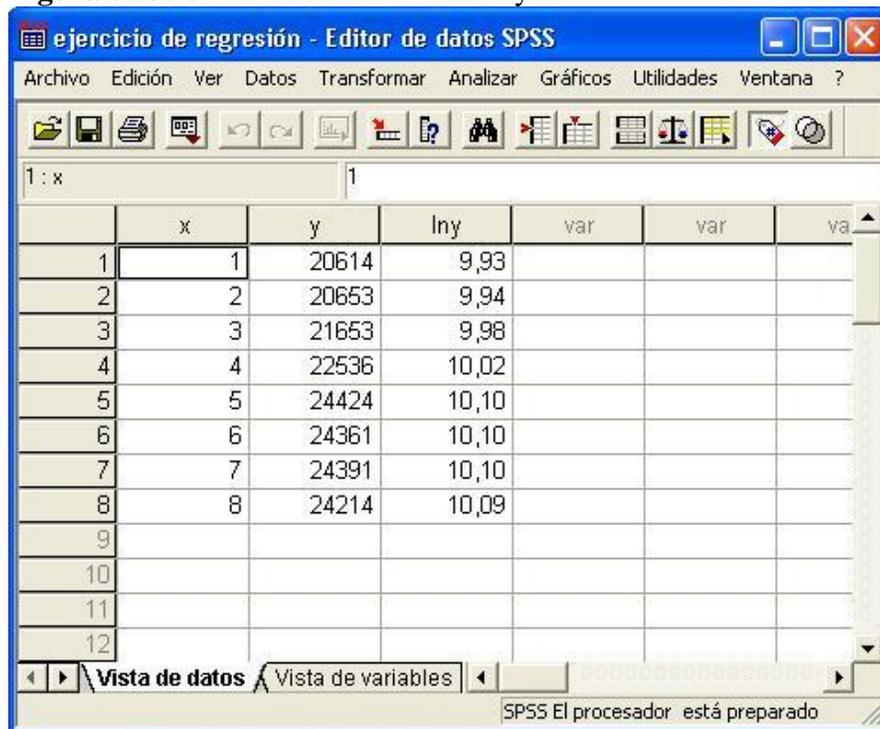
**8.3.1. Crear las variables en el programa SPSS.** Utilizamos los datos digitados con la elaboración del modelo de regresión lineal.

	<b>Véase también</b>	<p>Como se observa en las Figuras 7.1 y 7.2 Punto 7.1.1. “Crear las Variables en el Programa”.</p>
---	----------------------	--

**8.3.1.1. Calcular el logaritmo de la variable Y.** Para calcular el logaritmo de la variable Y, realice el mismo procedimiento explicado en el punto 8.2.2 “Calcule las variables”, “Crear y calcular los logaritmos de las variables”.

Una vez creada la variable, el visor de datos debe tener la apariencia similar a la Figura 8.13.

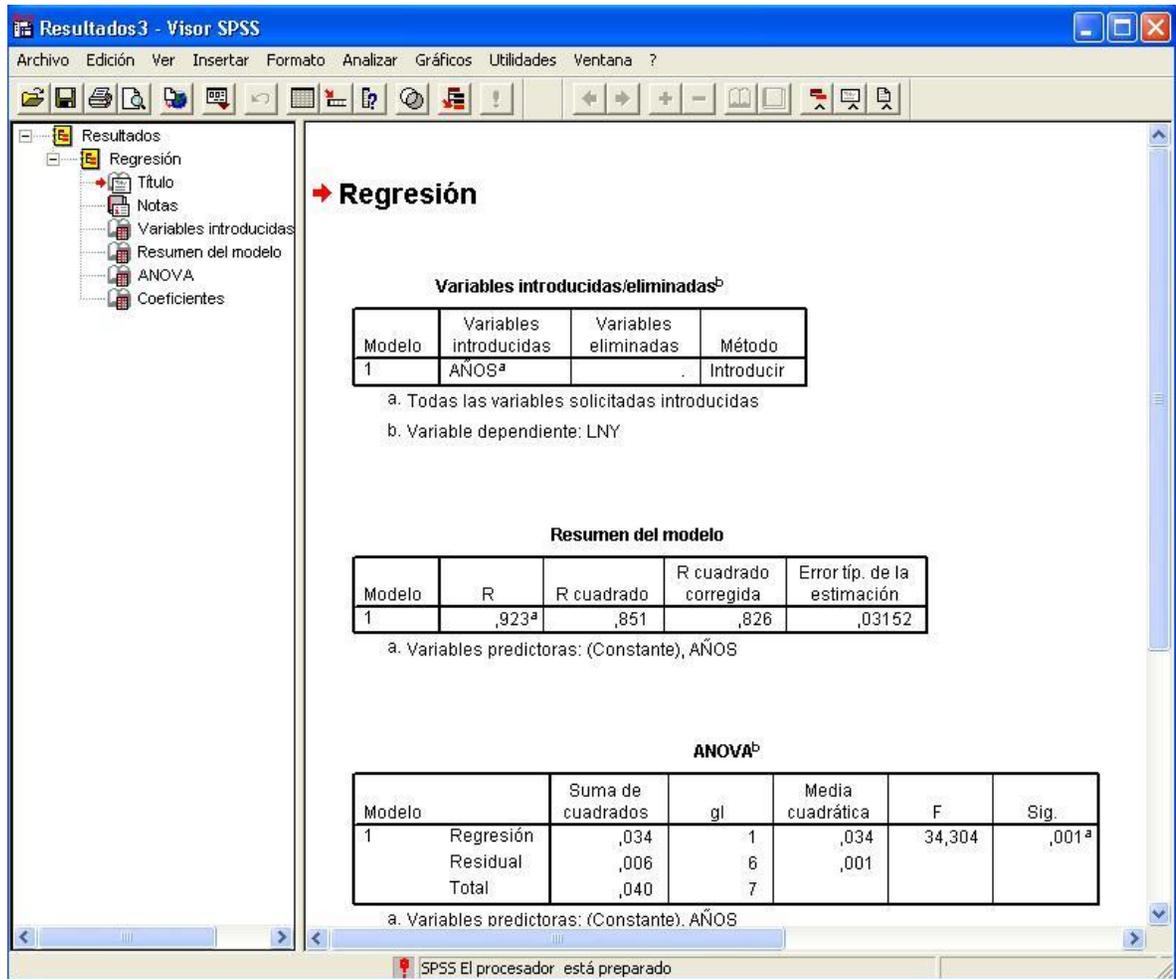
**Figura 8.13.** Creación de la Variable lny.



Creada la variable lny, se procesa la información,

	<b>Véase también</b>	<p>Para más detalles vea el punto 8.1. “Estimación del modelo log-lineal”, Figuras 8.2, 8.3, 8.4 y 8.5. La apariencia de su pantalla debe ser similar a la figura 8.10.</p>
---	----------------------	---

Figura 8.14. Resultados de la Regresión log – lin



### 8.3.2. Presentación de la información y análisis de datos.

Los resultados se presentan de la siguiente forma:

$$\ln Y_i = \beta_0 + \beta_1 X_i + u_i$$

$$\ln Y_i = 9,906 + 0,028 X_i + u_i$$

$$t = (403,41) \quad (5.657) \tag{45}$$

$$\rho = (0,000) \quad (0,001)$$

$$R^2 = 85,1\%$$

- La primera fila corresponde a la función de regresión log -lin;
- La cifras de la segunda fila hacen referencia a los parámetros de la función;

- Las cifras de la tercera fila corresponde a los valores de t estimado, bajo la hipótesis nula de que el verdadero valor poblacional de cada coeficiente de regresión sean iguales a cero;
- Las cifras de la cuarta fila corresponde a la probabilidad de que los parámetros estimados sean iguales a cero.

### 8.3.3. Interpretación de los Resultados.

- **Interpretación del  $\beta_1 = 0,028$ .** Significa que durante el primer año se incrementó en aproximadamente 2.8% las afiliaciones a COMFACAUCA de los trabajadores con derecho a dotación. Este porcentaje se refiere a la tasa de crecimiento instantánea (en un punto del tiempo), pero la tasa de crecimiento compuesta (durante los periodos de análisis) se calcula de la siguiente forma: Obténgase el antilogaritmo de la  $\beta_1$  estimada, réstele 1 y multiplique la diferencia por 100, así:

$$\text{Antilogaritmo de } 0,028 = 1,0284 - 1 = 0.0284 * 100 = 2,84\%$$

**Es decir, que la tasa compuesta de crecimiento de los Reportes a COMFACAUCA fue casi de 2,84%.**

- **Interpretación de  $\rho = 0,001$ .** La probabilidad del que el parámetro  $\beta_1$  sea igual cero es de 0%, es decir, que en muestreo repetido, la probabilidad de que una muestra similar a la tomada en este ejemplo de parámetros iguales a cero es un caso entre mil.
- **Interpretación de  $R^2 = 85,1\%$ .** Significa que el 85,1% de las variaciones de los reportes a COMFACAUCA están explicadas por la variable tiempo. Considerando que  $R^2$  puede tomar un valor máximo de uno (1), se puede afirmar que los datos se ajustan bastante bien a la ecuación encontrada.

**8.3.4. Proyección de Datos.** En el ejemplo que se retomó de los trabajadores con derecho a dotación, para realizar la proyección de la demanda, es necesario realizar el siguiente procedimiento:

**Tabla 8.3.** Proyección de los Parámetros del Modelo log - lin

DETALLE	FUNCIÓN	CONSUMO PROYECTADO
Función de regresión	$\ln Y_i = \beta_0 + \beta X_i$	
Parámetros de la función	$\ln Y_i = 9,906 + 0.028X_i$	
Estimación para el año 2001	$LnY_{ii} = 9.906 + 0,028(9)$ $Ant \ln Y_i = 10,158$ $Y_i = 25.796$	25.796
Estimación para el año 2002	$LnY_{ii} = 9.906 + 0,028(10)$ $Ant \ln Y_i = 10,186$ $Y_i = 26.529$	26.529
Estimación para el año 2003	$LnY_{ii} = 9.906 + 0,028(11)$ $Ant \ln Y_i = 10,21404$ $Y_i = 27.282$	27.282
Estimación para el año 2004	$LnY_{ii} = 9.906 + 0,028(12)$ $Ant \ln Y_i = 10,2420$ $Y_i = 28.057$	28.057
Estimación para el año 2005	$LnY_{ii} = 9.906 + 0,028(13)$ $Ant \ln Y_i = 10,2700$ $Y_i = 28.854$	28.854

## Apéndice A.

### DESCRIPCIÓN GENERAL DEL ARCHIVO COMERCIO.SAV

Los datos de este archivo hacen referencia a una encuesta realizada por estudiantes de la facultad de Ciencias Contables, Económicas y Administrativas de la Universidad del Cauca del programa de Administración de empresas en la asignatura de Investigación de mercado en el año 2005. El tema general de la encuesta es “Compre lo nuestro”, lema que se promocionó las agremiaciones de comerciantes, como Cámara de Comercio, FENALCO, ANDI y ACOPI, para incentivar a los Payaneses para que compren en el comercio de la Ciudad. A continuación se describe la ficha técnica.

### FICHA TÉCNICA

Elemento	:	Habitantes de la ciudad de Popayán mayores de 18 años.
Unidad de muestreo	:	Habitantes de la ciudad de Popayán mayores de 18 años.
Alcance	:	Ciudad de Popayán.
Tiempo	:	Enero de 2005.
Nivel de confianza	:	95% (Z=1,96).
Ocurrencia de evento P	:	0,5
No ocurrencia de evento Q	:	0,5
Error	:	4,6%
N	:	120.000
Tamaño de la muestra	:	452 se realizaron 461 encuestas

La muestra de 461 encuestas, de uno y otro sexo, mayores de 18 años, residentes en Popayán, el trabajo de campo se realizó en el mes de enero de 2005, mediante encuesta personal.

La muestra fue ponderada por sexo y edades de los individuos de acuerdo con la siguiente distribución.

**Tabla A.1.**

Distribución de la muestra por sexo y edades

Recuento		Sexo		Total
		Hombre	Mujer	
Rango de edades	de 18 a 20	14	2	16
	de 21 a 30	71	62	133
	de 31 a 40	85	49	134
	de 41 a 50	72	24	96
	de 51 a 60	33	26	59
	de 61 a 64	4	8	12
	Mas de 65	9	2	11
Total		288	173	461

El archivo analiza un total de 31 variables, divididas en dos (2) grupos:

- a) Aspectos relacionados con el comercio
- b) Variables de clasificación

A continuación se lleva a cabo una breve descripción de cada una de las variables, la escala utilizada, la codificación realizada y la pregunta.

1. Los habitantes de la ciudad de Popayán han manifestado inconformidad con el comercio de la Ciudad, en cuanto a la tención al cliente, precio y variedad en los productos. De los siguientes aspectos que le presento a continuación, Seleccione con el que hasta más inconforme.

**A1. Aspecto de mayor inconformidad.**

- 6. No hay variedad en los producto (     )
- 7. Mala atención al cliente (     )
- 8. Mala calidad en los productos (     )
- 9. Los precios son muy altos (     )
- 10. NS/NC (     )

2. También se ha manifestado que no todo en el comercio es malo, hay algunos aspectos en que los Payaneses se encuentran satisfechos. ¿Con cual de los siguientes aspectos usted se encuentra más satisfecho?

**A2. Aspecto con el que se siente más satisfecho.**

- 1. El comercio esta ubicada en una sola parte (     )
- 2. Puedo ubicar los productos en menos tiempo (     )
- 3. Puedo encargar la mercancía que no encuentro en la ciudad (     )
- 4. Los encargos los traen en el menor tiempo posible (     )
- 5. NS/NC (     )

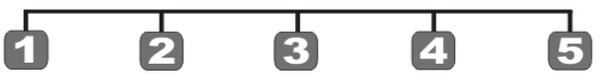
3. Cuando usted acude a un almacén a comprar vestidos y sus complementos, cuáles de los siguientes factores considera más importantes, ordénelos, colocando en primer lugar aquel que considera más importante, y en último lugar aquel que considere menos importante. (Pregunta de la pagina 142)

- B1 Primero**
- B2 Segundo**
- B3 Tercero**
- B4 Cuarto**
- B5 Quinto**
- B6 Sexto**

### B7 Séptimo

1. Variedad de las marca de la ropa que tiene el almacén	1a	2a	3a	4a	5a	6a	7a
2. Los precios del almacén.	1a	2a	3a	4a	5a	6a	7a
3. Ubicación del almacén.	1a	2a	3a	4a	5a	6a	7a
4. Que encuentre todo en el mismo lugar	1a	2a	3a	4a	5a	6a	7a
5. Los precios de los zapatos y accesorios	1a	2a	3a	4a	5a	6a	7a
6. La orientación que da el comerciante	1a	2a	3a	4a	5a	6a	7a
7. Ofertas que brinda el almacén	1a	2a	3a	4a	5a	6a	7a

4. Utilizando esta escala de 6 puntos en la que 1 valora muy poco y 6 valora mucho, Cuando usted realiza sus compras de vestido o complementos ¿qué es lo que más valora?

	Valora muy poco <span style="float: right;">Valora mucho</span> 				
B8 Variedad					
B9 Diseño					
B10 Atención					
B11 Amabilidad					
B12 Ambiente					
B13 Mobiliario					
B14 Decoración					
B15 Marca					
B16 Exclusividad					
B17 Limpieza					
B18 Calidad					

### Datos de clasificación.

C1. Sexo

1. Hombre (     )     2. Mujer (     )

C2. Edad en años cumplidos

C3. Estado civil

1. Soltero (a)                    (     )

- 2. Casado (a) ( )
- 3. Viviendo en pareja ( )
- 4. Separado (a) ( )
- 5. Divorciado (a) ( )
- 6. Viudo (a) ( )
- 8. NC ( )

**C4 Relación con el cabeza de familia**

- 1. Cabeza de familia ( )
- 2. Esposo (a), compañero (a) ( )
- 3. Hijo (a) ( )
- 4. Padre, madre, suegro (a) ( )
- 5. Hermano (a), cuñado (a) ( )
- 6. Otro ( )
- 9. NS ( )

**C5: Número de miembros del hogar**

- 1. Uno ( )
- 2. Dos ( )
- 3. Tres ( )
- 4. Cuatro ( )
- 5. Cinco ( )
- 6. Seis ( )
- 7. Siete ( )
- 8. Ocho ( )
- 9. Nueve ( )
- 0. NS/NC ( )

**C6. Estudios del entrevistado**

- 0. Ninguno. ( )
- 1. Universitarios ( )
- 2. Bachiller ( )
- 3. Media vocacional ( )
- 4. Estudios primarios completos ( )
- 5. Técnico ( )
- 6. Tecnólogo ( )
- 7. Estudios primarios incompletos ( )
- 8. Profesional ( )
- 9. Postgrado ( )

**C7. Lugar de nacimiento.**

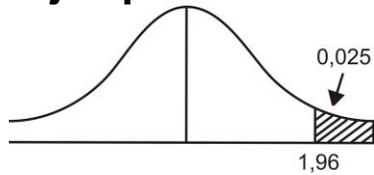
01.	Almaguer	( )
02.	Arauca	( )
03.	Argelia	( )
04.	Armenia	( )
05.	Armenia	( )
06.	Barranquilla	( )
07.	Bogotá	( )
08.	Popayán	( )
09.	Bucaramanga	( )
10.	Cajamarca	( )
11.	Calcará	( )
12.	Cali	( )
13.	Candelaria	( )
14.	Cartagena	( )
15.	Cartago	( )
16.	Cartegena del Chairá	( )
17.	Cúcuta	( )
18.	Florencia	( )
19.	Guapi	( )
20.	Ibagué	( )
21.	Ipiales	( )
22.	Jamundí	( )
23.	La Cruz	( )
24.	La Unión	( )
25.	Leticia	( )
26.	Medellín	( )
27.	Mercaderes	( )
28.	Miranda	( )
29.	Mocoa	( )
30.	Montería	( )
31.	Neiva	( )
32.	Palmira	( )
33.	Pasto	( )
34.	Patía	( )
35.	Paujil	( )
36.	Piendamó	( )
37.	Bolívar	( )
38.	Pradera	( )
39.	Puerto Rico	( )
40.	Puerto Tejada	( )
41.	Quibdó	( )
42.	Riohacha	( )
43.	Rosas	( )
44.	San Pablo	( )
45.	Santamartha	( )

- 46. Santander ( )
- 47. Sevilla ( )
- 48. Silvia ( )
- 49. Sincelejo ( )
- 50. Tulúa ( )
- 51. Tunja ( )
- 52. Valledupar ( )
- 53. Villavicencio ( )
- 54. Yopal ( )

C8. Marque la opción que mejor describe sus ingresos mensuales por todos los conceptos que ingresan en su hogar.

- 1. Hasta 381.500 ( )
- 2. De 381.501 a 500.000 ( )
- 3. 500.001 a 550.000 ( )
- 4. 550.001 a 600.000 ( )
- 5. 600.001 a 650.000 ( )
- 6. 650.001 a 700.000 ( )
- 7. 700.001 a 750.000 ( )
- 8. 750.001 s 800.000 ( )
- 9. Más de 800.001 ( )
- 0. NC ( )

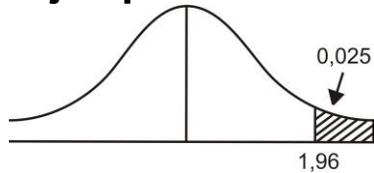
**Apéndice B.**  
**Ejemplo:**



**TABLA DE DISTRIBUCIÓN NORMAL**

<b>Z</b>	<b>0,00</b>	<b>0,01</b>	<b>0,02</b>	<b>0,03</b>	<b>0,04</b>	<b>0,05</b>	<b>0,06</b>	<b>0,07</b>	<b>0,08</b>	<b>0,09</b>
<b>0,00</b>	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
<b>0,10</b>	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
<b>0,20</b>	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
<b>0,30</b>	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
<b>0,40</b>	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
<b>0,50</b>	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
<b>0,60</b>	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
<b>0,70</b>	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
<b>0,80</b>	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
<b>0,90</b>	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
<b>1,00</b>	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
<b>1,10</b>	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
<b>1,20</b>	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
<b>1,30</b>	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
<b>1,40</b>	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
<b>1,50</b>	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
<b>1,60</b>	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
<b>1,70</b>	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
<b>1,80</b>	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
<b>1,90</b>	<b>0,0287</b>	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

**Ejemplo:**



**TABLA DE DISTRIBUCIÓN NORMAL**

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
<b>2,00</b>	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
<b>2,10</b>	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
<b>2,20</b>	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
<b>2,30</b>	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
<b>2,40</b>	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
<b>2,50</b>	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
<b>2,60</b>	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
<b>2,70</b>	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
<b>2,80</b>	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
<b>2,90</b>	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
<b>3,00</b>	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
<b>3,10</b>	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
<b>3,20</b>	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
<b>3,30</b>	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
<b>3,40</b>	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
<b>3,50</b>	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002
<b>3,60</b>	0,0002	0,0002	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
<b>3,70</b>	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
<b>3,80</b>	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001
<b>3,90</b>	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

**Fuente:** Esta tabla fue generada en la hoja de cálculo Excel mediante la función DISTR.NORM.ESTAND.

## Apéndice C.

### TABLA PARA LA DETERMINACIÓN DE UNA MUESTRA

Margen de confianza del 95%

Z = 1,96

Hipótesis  $\rho = 50\%$

POBLACIÓN N	TAMAÑO DE LA MUESTRA SEGÚN MÁRGENES DE ERROR					
	1%	2%	3%	4%	5%	10%
500					217	81
1.000				375	278	88
1.500			624	429	306	90
2.000			696	462	322	92
2.500		1.225	748	484	333	92
3.000		1.334	787	500	341	93
3.500		1.424	818	512	346	93
4.000		1.500	842	522	350	94
4.500		1.566	863	530	354	94
5.000		1.622	879	536	357	94
5.500		1.671	894	541	359	94
6.000		1.715	906	546	361	95
6.500		1.753	917	550	363	95
10.000	4.899	1.936	964	566	370	95
15.000	5.855	2.070	996	577	375	95
20.000	6.488	2.144	1.013	583	377	96
25.000	6.939	2.191	1.023	586	378	96
30.000	7.275	2.223	1.030	588	379	96
35.000	7.536	2.247	1.036	590	380	96
40.000	7.745	2.265	1.039	591	381	96
60.000	8.279	2.309	1.048	594	382	96
65.000	8.368	2.315	1.050	595	382	96
70.000	8.445	2.321	1.051	595	382	96
100.000	8.762	2.345	1.056	597	383	96
200.000	9.164	2.373	1.061	598	383	96
300.000	9.306	2.382	1.063	599	384	96
900.000	9.503	2.395	1.066	600	384	96
1.000.000	9.513	2.395	1.066	600	384	96

## Apéndice C.

### TABLA PARA LA DETERMINACIÓN DE UNA MUESTRA

Margen de confianza del 99%

Z = 2,58

Hipótesis  $\rho = 50\%$

POBLACIÓN N	TAMAÑO DE LA MUESTRA SEGÚN MÁRGENES DE ERROR					
	1%	2%	3%	4%	5%	10%
500					286	125
1.000				510	400	143
1.500			828	614	461	150
2.000			961	684	499	154
2.500		1.562	1.063	734	526	156
3.000		1.743	1.144	772	545	158
3.500		1.901	1.210	802	559	159
4.000		2.039	1.264	825	571	160
4.500		2.162	1.311	845	580	160
5.000		2.271	1.350	861	587	161
5.500		2.369	1.384	875	594	162
6.000		2.457	1.413	886	599	162
6.500		2.537	1.440	897	604	162
10.000	6.246	2.938	1.560	942	624	164
15.000	7.889	3.257	1.646	973	637	165
20.000	9.083	3.444	1.693	989	644	165
25.000	9.991	3.567	1.722	999	648	165
30.000	10.704	3.654	1.742	1.005	651	165
35.000	11.279	3.718	1.756	1.010	653	166
40.000	11.752	3.768	1.767	1.014	655	166
60.000	13.028	3.890	1.794	1.022	658	166
65.000	13.249	3.910	1.798	1.024	659	166
70.000	13.445	3.927	1.801	1.025	659	166
100.000	14.267	3.994	1.815	1.029	661	166
200.000	15.363	4.075	1.832	1.035	663	166
300.000	15.766	4.103	1.838	1.036	664	166
900.000	16.339	4.141	1.845	1.039	665	166
1.000.000	16.369	4.143	1.846	1.039	665	166

## Índice

### A

**ACEPTAR**., 124  
Actitudes, 13, 14  
American Marketing Association, 5  
**ANÁLISI BIVARIANTE**, 182  
Análisis bivariado, 172  
Análisis bivalente de medias, 182, 183  
*Análisis de chi cuadrado*, 196  
Análisis de contenido, 15  
Análisis de correlación lineal, 182  
*Análisis de frecuencia*, 174  
*Análisis de frecuencias*, 195  
*Análisis de los residuos*, 197  
Análisis de tablas de contingencia, 182  
*Análisis de varianza*, 188  
Análisis de vestigios, 15  
Análisis multivariado, 172  
**ANÁLISIS UNIVARIADO DE DATOS**, 174  
Análisis univariante, 172  
Analizar, 103  
Archivo, 103  
*asignación proporcional*, 90  
Auditorias, 15  
**AYUDA**, 124  
Ayuda?, **103**

### C

**CANCELAR**, 124  
Clasificación de escalas, 35  
Codificación, 17  
**COEFICIENTE DE CORRELACIÓN R**, 217  
**COEFICIENTE DE DETERMINACIÓN R<sup>2</sup>**, 216  
componente cognoscitivo, 14  
Componentes del cuestionario, 20  
Comportamiento anterior, 13  
concluyente, 7  
**Conglomerados**, 107  
Contrastes de hipótesis, 75  
Copiar una Variable, 131  
Creación del plan de muestreo, 107

### D

**Datos**, 103  
Datos de clasificación, 20  
**DATOS DE CLASIFICACIÓN**, 28

Datos de identificación, 20  
Datos primarios, 10  
Datos secundarios, 10  
**Desventajas de realizar un censo**, 61  
**Desviación estándar**, 175  
**Diseño de estudios transversales**, 9  
**DISEÑO DE FORMULARIOS**, 20  
diseño de investigación, 6, 7  
**Diseño del cuestionario**, 17  
**Diseño longitudinal**, 9  
**DISTRIBUCIONES MUESTRALES**, 65

### E

Edición, 103  
*Edición de gráficos*, 137  
Editor de datos, 101, 102  
Editor de gráficos, 101  
Editor de procesos, 102  
Editor de propiedades del grafico, 139  
Editor de resultados de texto, 101  
Editor de sintaxis, 102  
Editor de tablas pivote, 101  
El coeficiente de asimetría, 181  
El coeficiente de curtosis, 182  
El coeficiente de elasticidad, 223  
El Nivel de Confianza, 72  
El test F, 189  
**ELEMENTO**, 62  
Encuesta personal, 15  
Encuesta por correo, 15  
*Encuesta postal*, 19  
Encuesta telefónica, 15, 19  
**Entrevistas en profundidad**, 12  
Error de tipo I, 75  
Error de tipo II, 75  
Escala de calificación Q y otros procedimientos, 38  
*Escala de caras sonrientes*, 44  
Escala de clasificación continúa, 40  
Escala de clasificación de partidas, 41  
Escala de comparación apareada, 36  
**Escala de diferencial semántico**, 42  
Escala de estimación de la magnitud, 39  
*ESCALA DE INTENCIÓN DE COMPRA*, 46  
**Escala de Likert**, 41  
**Escala de Stapel**, 43  
Escala de suma de constantes, 37  
*Escala del termómetro.*, 45  
*Escala nominal*, 31

Escala por orden de clasificación, 37

*Escalas de relación*, 33

*Escalas no comparativas*, 40

*Escalas ordinales*, 32

Estadístico, 65

Estratificación, 106

ESTRUCTURA GENERALES DEL PROGRAMA SPSS, 97

Etiquetas, 126

Etiquetas de Valores, 127

exploratoria, 7

## F

FACTOR DE CORRECCIÓN, 70

FICHA TÉCNICA, 84

flexibilidad, 8

*FORMA FÍSICA DE LA ESCALA*, 44

Formulación del diseño de investigación, 8

FUENTES DE INVESTIGACIÓN, 10

Función de regresión poblacional (FRP), 210

## G

GRAFICAS INTERACTIVAS, 167

Gráficos, 103

Grupo de programas de SPSS, 99

## H

Hipótesis alternativa, 75

Hipótesis nula, 74

*HOMOGENEIDAD DE LA INFORMACIÓN*, 89

## I

identificación, 6

Información solicitada, 20

Iniciar el programa SPSS, 99

Instrucciones, 20

*INTRODUCCIÓN*, 2

Investigación causal, 9

investigación de mercados, 5

Investigación descriptiva, 8

## L

La comunicación, 14, 15

La información solicitada, 21

La observación, 14

La Varianza, 72

Las instrucciones, 21

Lenguaje de comandos (Sintaxis), 144

Linealidad en las variables, 206

Linealidad en los parámetros, 206

## M

MARCO MUESTRAL, 63

Media, 175

Mediana, 175

*Medidas de dispersión*, 175

*Medidas de tendencia central*, 174, 180

Menú Contextual, 130

MÉTRICAS, 173

Mide la elasticidad, 223

Mínimos Cuadrados Ordinarios (MCO), 210

Moda, 174

modelo log-lineal, 222

Modelo semilogarítmico, 222

*Mostrar los porcentajes*, 139

muestra representativa individual, 9

muestra representativa múltiple, 9

*muestreo aleatorio Simple*, 112

*MUESTREO ALEATORIO SIMPLE*, 80

*muestreo con probabilidad proporcional*, 112

*MUESTREO ESTRATIFICADO*, 90

Muestreo no restringido, 107

*Muestreo secuencial proporcional*, 112

*Muestreo secuencial simple*, 112

*Muestreo sistemático proporcional*, 112

*muestreo Sistemático Simple*, 112

Muestro no aleatorio, 107

Múltiples etapas, 107

## N

Nivel de significación, 75

Niveles de confianza, 74

NO MÉTRICAS, 173

## O

Observación mecánica, 15

Observación personal, 15

## P

PEGAR, 124

Perdidos, 127

PLAN DE MUESTREO, 61

*POBLACIÓN*, 62

Ponderaciones muestrales, 107

Precisión de la Estimación, 72

*Preguntas abiertas*, 24

*Preguntas cerradas*, 24

*Preguntas de profundización*, 24

Preguntas de recuerdo respuesta espontánea, 26

Preguntas de recuerdo sugerido, 27  
**Preguntas de respuesta múltiple, 25**  
Preguntas de respuestas múltiples, 24  
**Preguntas dicótomas, 25**  
Preguntas filtro, 27  
**PREPARACIÓN DE LOS INFORMES, 10**  
**PREPARACIÓN Y ANÁLISIS DE DATOS, 10**  
**Probabilidad de selección desigual, 107**  
problema, 6  
**Procedimiento no Probabilístico, 64**  
**Procedimiento Probabilístico, 64**  
**PROCEDIMIENTOS DE MUESTREO, 64**

## R

**Rango o recorrido, 175**  
**Re codificar en distinta variable, 163**  
recopilación, 6  
**Recorrido intercuartílico, 175**  
**relación entre las variables, 203**  
**RESTABLECER, 124**

## S

**scalas de intervalo, 32**  
**Sesiones de grupo, 11**  
Solicitud de cooperación, 20, 21

## T

**tabla de contingencia, 190**  
**Tablas de Contingencia Respuestas Múltiples, 149**  
tamaño de la muestra, 80  
**TAMAÑO DE LA MUESTRA, 72**  
**TAMAÑO OPTIMO EN POBLACIONES FINITAS, 79**  
Técnicas de asociación., 12  
Técnicas de construcción, 12  
**Técnicas de escala comparativas, 35**  
**TÉCNICAS DE ESCALAS, 35**  
Técnicas de terminación, 12  
Técnicas expresivas, 13

técnicas proyectivas, 12  
Teorema, 66  
**TEOREMA DEL LÍMITE CENTRAL, 68**  
**Término "Lineal"., 206**  
Test F, 189  
**Tipo de Variable, 125**  
**TIPOS DE ESCALA, 31**  
**TIPOS DE ESCALA DE MEDICIÓN, 173**  
**TIPOS DE PREGUNTAS, 24**  
Trabajo de campo, 6  
**TRABAJO DE CAMPO, 9**  
**Transformar, 103, 163**  
Transformar el grafico, 141

## U

**UNIDAD DE MUESTREO, 63**  
**Utilidades, 103**

## V

**Valores absolutos, 174**  
**Valores Perdidos, 130**  
**Valores relativos, 174**  
**Variabilidad entre-grupos, 188**  
**Variabilidad intra-grupos, 188**  
**Variabilidad total, 188**  
**Variable dependiente, 183**  
**Variable independiente, 183**  
**Variables de intervalo, 179**  
Variables ordinales, 177  
**Varianza, 175**  
**Varianza entre grupos, 189**  
**Varianza intra grupos, 189**  
**Ventajas de realizar una muestra, 61**  
**Ventana, 103**  
**VENTANAS DEL SPSS, 101**  
**Ver, 103**  
versatilidad, 8  
**Visor de borrador, 101**  
**Visor., 101**



## BIBLIOGRAFÍA

---

- ALEGRE MARTIN, Joaquín et al. Ejercicios de econometría. 1 ed. Madrid. Thomson. 1995.
- CAMACHO ROSALES, Juan. Estadística con SPSS para Windows. 11 ed. México D.F. ALFAOMEGA GRUPO EDITOR, S.A. 2003.
- CARRASCAL, Ursicino; GONZÁLEZ, Yolanda y RODRÍGUEZ, Beatriz. Análisis econométrico con Eviews. 1 ed. México D.F. Alfaomega. 2001.
- CHIANG, Alpha C. Métodos fundamentales de economía matemática. 3 ed. México D.F. 1996.
- DIEBOLD, Francis X. Elementos de pronóstico. 2 ed. México D.F. Thomson. 1998.
- GREENE, William H. Análisis econométrico. 3 ed. Madrid. PRENTICE HALL IBERIA S.R.L. 1999.
- GUJARATI, Damodar N. Econometría. 4 ed. México, D. F. McGraw - Hill. 2003.
- KAZMIER, Leonard y DIAZ MATA, Alfredo. Estadística : Aplicada a la administración y a la economía. 2 ed. México D.F. McGraw – Hill, 1993.
- KINNEAR, Thomas y TAYLOR, James R. Investigación de mercados : un enfoque aplicado, 5 ed. Santa Fé de Bogotá, D.C. McGraw – Hill. 1998.
- KOHLER, Heinz. Estadística para negocios y economía. 1 ed. México D. F. CECSA. 1996.
- MADDAL, G.S. Introducción a la econometría. 2 ed. PRENTICE – HALL. 1996.
- MALHOTRA, Narres K. Investigación de mercados un enfoque práctico. 2 ed. México D.F. Prentice – Hall. 1997.
- MARTINEZ BENCARDINO, Ciro. Estadística y muestreo.9 ed. Santa Fé de Bogotá D.C. Ecoe Ediciones. 2002.
- NOVALES CINCA, Alfonso. Econometría. 2 ed. Madrid. McGraw – Hill. 1993.
- PÉREZ LÓPEZ, César. Métodos estadísticos avanzados con SPSS. 1 ed. Madrid.

Thomson. 2005.

PÉREZ LÓPEZ, César. Muestreo estadístico: Conceptos y problemas resueltos. 1 ed. PEARSON EDUCACIÓN, S.A. 2005.

POPE, Effrey L. Investigación de mercados. 5 ed. New York, N.Y. Norma S.A. 1981.

SPIEGEL. Murray R. Estadística: Schaum. 2 ed. Madrid. McGrawHill.1991.

VISAUTA VINACUA, Bienvenido y MARTORI I CAÑAS. Joan Carles. Análisis estadístico con SPSS para Windows : Estadística multivariante. 2 ed. Madrid. McGraw – Hill. 2003. v. 2.

VISAUTA VINACUA, Bienvenido. Análisis estadístico con SPSS para Windows. : Estadística básica. 1 ed. Madrid. McGraw-Hill. 1997.

WOOLDRIDGE, Jeffrey M. Introducción a la econometría. : Un enfoque moderno. 1 ed. THOMSON. México, D.F. 2000.

## TABLA DE CONTENIDO

	Página
CAPITULO UNO .....	5
<b>1. ESTRUCTURA DE INVESTIGACIÓN.....</b>	<b>5</b>
• <b>¿CÓMO INICIAR EL PROYECTO? .....</b>	<b>5</b>
<b>1.1. DEFINICIÓN DEL PROBLEMA. ....</b>	<b>6</b>
<b>1.1.1. PREGUNTAS FRENTE A LA COMPETENCIA.....</b>	<b>6</b>
<b>1.2. FORMULACIÓN DEL DISEÑO DE INVESTIGACIÓN. ....</b>	<b>7</b>
<b>1.2.1. Investigación exploratoria. ....</b>	<b>8</b>
<b>1.2.2. Investigación concluyente. ....</b>	<b>8</b>
<b>3.2.2.1. Investigación descriptiva. ....</b>	<b>8</b>
<b>1.3. TRABAJO DE CAMPO O RECOPIACIÓN DE DATOS. ....</b>	<b>9</b>
<b>1.4. PREPARACIÓN Y ANÁLISIS DE DATOS.....</b>	<b>10</b>
<b>1.5. PREPARACIÓN DE LOS INFORMES.....</b>	<b>10</b>
<b>1.6. FUENTES DE INVESTIGACIÓN. ....</b>	<b>10</b>
<b>1.6.1. Datos de los encuestados o fuentes de información primarias. ....</b>	<b>10</b>
<b>1.6.2. Datos secundarios o fuentes secundarias. ....</b>	<b>11</b>
<b>1.7. PROCEDIMIENTOS DE INVESTIGACIÓN EXPLORATORIA .....</b>	<b>11</b>
<b>1.7.1. Sesiones de grupo. ....</b>	<b>11</b>
<b>1.7.1.1. Características principales. ....</b>	<b>12</b>
<b>1.7.2. Entrevistas en profundidad. . ....</b>	<b>12</b>
<b>1.7.3. técnicas proyectivas. ....</b>	<b>12</b>
<b>1.8. PROCEDIMIENTOS DE INVESTIGACIÓN CONCLUYENTE .....</b>	<b>13</b>
<b>1.9. DEFINICIÓN DE LA INFORMACIÓN NECESARIA. ....</b>	<b>13</b>
<b>1.10. MÉTODOS DE RECOLECCIÓN DE INFORMACIÓN.....</b>	<b>14</b>
<b>1.10.1. La observación.....</b>	<b>14</b>
<b>1.10.2. La comunicación.....</b>	<b>15</b>
CAPITULO DOS .....	17

<b>INTRODUCCIÓN</b> .....	17
<b>2. ETAPAS PARA ELABORAR UNA ENCUESTA</b> .....	17
2.1. DETERMINACIÓN DE LOS OBJETIVOS.....	18
<b>2.2. DETERMINACIÓN DEL TIPO DE ENCUESTA</b> .....	18
2.2.1. Encuesta personal.....	18
2.2.2. Encuesta telefónica. . . . .	19
2.2.3. Encuesta postal. . . . .	19
<b>2.3. DISEÑO DE FORMULARIOS</b> .....	20
• COMPONENTES DEL CUESTIONARIO . . . . .	20
• Datos de identificación.....	20
• Solicitud de cooperación. . . . .	21
• Las instrucciones. . . . .	21
• La información solicitada. . . . .	21
<b>2.4. TIPOS DE PREGUNTAS</b> .....	24
2.4.1. Preguntas abiertas.....	24
2.4.2. Preguntas de profundización.....	24
2.4.2.1. Preguntas de clarificación.....	24
2.4.3. Preguntas cerradas.....	24
2.4.3.1. Preguntas de respuestas múltiples. . . . .	24
2.4.3.2. Preguntas de recuerdo sugerido. . . . .	27
2.4.3.3. Preguntas filtro.....	27
<b>2.5. DATOS DE CLASIFICACIÓN DEL ENTREVISTADO</b> .....	28
<b>SE RELACIONA CON LAS CARACTERÍSTICAS DEL ENCUESTADO</b> .....	28
<b>2.6. TIPOS DE ESCALA</b> .....	31
2.6.1. Escala nominal.....	31
2.6.2. Escalas ordinales. . . . .	32
2.6.3. Escalas de intervalo. . . . .	32
2.6.4. Escalas de relación. . . . .	33
<b>2.7. TÉCNICAS DE ESCALAS</b> .....	35
2.7.1. Técnicas de escala comparativas. . . . .	35
2.7.1.1. Escala de comparación apareada. . . . .	36
2.7.1.2. Escala por orden de clasificación. . . . .	37
2.7.1.3. Escala de suma de constantes. . . . .	37
2.7.1.4. Escala de calificación Q y otros procedimientos. . . . .	38
2.7.1.5. Escala de estimación de la magnitud. . . . .	39

<b>2.7.2. ESCALAS NO COMPARATIVAS.</b>	40
2.7.2.1. Escala de clasificación continúa.	40
2.7.2.2. Escala de clasificación de partidas.	41
<b>2.8. FORMA FÍSICA DE LA ESCALA</b>	44
<b>2.8.1. Escala de caras sonrientes.</b>	44
<b>2.8.2. Escala del termómetro.</b>	45
<b>2.9. ESCALA DE INTENCIÓN DE COMPRA.</b>	46
<b>2.10. ESTUDIO DE CASO 1.</b>	47
2.10.1. Descripción del problema.	47
2.10.2. Formulación del problema.	48
2.10.3. Objetivo general.	48
2.10.5. Diseño de la encuesta.	48
<b>2.11. ESTUDIO DE CASO 2.</b>	54
2.11.1. Objetivo General.	54
2.11.2. Objetivos Específicos:	54
2.11.3. Diseño de la encuesta.	54
<b>CAPITULO TRES</b>	61
<b>3. PLAN DE MUESTREO</b>	61
3.1. ELEMENTO.	62
3.2. POBLACIÓN.	62
3.3. UNIDAD DE MUESTREO.	63
3.4. MARCO MUESTRAL.	63
3.5. POBLACIÓN DE ESTUDIO.	64
3.6. PROCEDIMIENTOS DE MUESTREO	64
3.6.1. Procedimiento Probabilístico.	64
3.6.2. Procedimiento no Probabilístico.	64
3.6.3. Parámetro.	65
3.6.4. Estadístico.	65
3.7. NOMENCLATURA UTILIZADA.	65
3.8. DISTRIBUCIONES MUESTRALES.	65
3.9. TEOREMA DEL LÍMITE CENTRAL.	68
3.10. FACTOR DE CORRECCIÓN PARA POBLACIONES FINITAS.	70
3.11. DISTRIBUCIÓN MUESTRAL DE UNA PROPORCIÓN.	70
3.12. TAMAÑO DE LA MUESTRA.	72
3.12.1. La Varianza.	72
3.12.2. El Nivel de Confianza.	72

3.12.3. Precisión de la Estimación.....	72
<b>3.13. ESTIMACIÓN DE INTERVALOS DE CONFIANZA PARA PARÁMETROS DE POBLACIÓN.....</b>	<b>72</b>
3.14. CÁLCULO DE LOS COEFICIENTES DE CONFIANZA.....	73
<b>3.13. TEORÍA ESTADÍSTICA DE LAS DECISIONES.....</b>	<b>74</b>
3.13.1. Hipótesis nula ( $H_0$ ).....	74
3.13.2. Hipótesis alternativa ( $H_1$ ).....	75
3.13.3. Contrastes de hipótesis y significación, o reglas de decisión..	75
3.13.4. Errores de tipo I y de tipo II.....	75
3.13.5. Nivel de significación.....	75
3.13.6. Contraste mediante la distribución normal.....	76
3.13.7. Curva de operación características.....	77
<b>3.14. TAMAÑO OPTIMO EN POBLACIONES FINITAS.....</b>	<b>79</b>
3.14.1. Formulas utilizadas con variables discretas (atributos).....	80
<b>3.15. MUESTREO ALEATORIO SIMPLE (M.A.S.).....</b>	<b>80</b>
3.15.1. Cálculo del tamaño de la muestra.....	80
3.15.2. Tamaño de la muestra con la variable discreta (cualitativa).....	85
<b>3.16. TAMAÑO OPTIMO DE MUESTRA CON VARIABLES DISCRETAS.....</b>	<b>88</b>
<b>3.17. HOMOGENEIDAD DE LA INFORMACIÓN.....</b>	<b>89</b>
<b>3.18. RELACIÓN DEL TAMAÑO DE LA MUESTRA CON EL NIVEL DE CONFIANZA Y EL ERROR.....</b>	<b>89</b>
<b>3.19. MUESTREO ESTRATIFICADO.....</b>	<b>90</b>
3.19.1. Muestreo estratificado - asignación proporcional.....	90
3.19.2. Cálculo del tamaño de la muestra.....	93
CAPITULO CUATRO.....	97
INTRODUCCIÓN.....	97
<b>4. ESTRUCTURA GENERALES DEL PROGRAMA SPSS.....</b>	<b>97</b>
4.1. ARCHIVOS QUE GENERA EL SPSS.....	101
4.2. VENTANAS DEL SPSS.....	101
4.3. MENÚ PRINCIPAL.....	103
4.4. BARRA DE HERRAMIENTAS.....	103
4.5. HERRAMIENTAS DE EDITOR DE DATOS.....	104
4.6. VENTANA DE EDICIÓN DE DATOS.....	105
4.7. SPSS Y EL MUESTREO.....	106
4.7.1. Creación del plan de muestreo y almacenamiento de los datos de la muestra.....	107

<b>FIGURA 4.19. VARIABLES DE RESULTADO</b> .....	114
<b>FIGURA 4.22. EXTRAER LA MUESTRA</b> .....	117
<b>CAPITULO CINCO</b> .....	123
<b>5. DEFINICIÓN DE VARIABLES</b> .....	123
<b>5.1. VISTA DE VARIABLES</b> .....	123
<b>5.1.1. Cuadros de Diálogo.</b> .....	124
<b>5.2. PREGUNTA CERRADA – DICÓTOMA.</b> .....	125
<b>5.2.1. Nombre de la Variable.</b> .....	125
<b>5.2.1.1. Características de los nombres de las variables.</b> .....	125
<b>5.2.2. Tipo de Variable.</b> .....	125
<b>5.2.3. Etiquetas.</b> .....	126
<b>5.2.4. Etiquetas de Valores.</b> .....	127
<b>5.2.5. Perdidos.</b> .....	127
<b>5.2.6. Columnas – Alineación – Medida.</b> .....	128
<b>5.3. PREGUNTA DE SELECCIÓN MÚLTIPLE.</b> .....	128
<b>5.3.1. Valores Perdidos.</b> .....	130
<b>5.3.2. Menú Contextual</b> .....	130
<b>Copiar una Variable.</b> .....	131
<b>5.3.3. Digite los datos.</b> .....	133
<b>5.3.4. Análisis de la información.</b> .....	135
<b>5.3.5. Edición de gráficos.</b> .....	137
<b>5.3.5.1. Mostrar los porcentajes y valores de las variables graficadas.</b> .....	139
<b>5.3.5.2. Editor de propiedades del grafico de sectores.</b> .....	139
<b>5.3.5.3. Transformar el grafico a tres dimensiones.</b> .....	141
<b>5.3.6. Lenguaje de comandos (Sintaxis).</b> .....	144
<b>5.3.7. Relación entre variables -Tablas de contingencia.</b> .....	146
<b>5.3.8. Lenguaje de Comandos (Sintaxis).</b> .....	148
<b>5.3.9. Tablas de Contingencia Respuestas Múltiples.</b> .....	149
<b>5.3.10. Procesamiento de pregunta de selección múltiple.</b> .....	153
<b>5.3.11. Opción de Tablas de Contingencia con Respuesta Múltiple.</b> .....	156
<b>5.3.12. Lenguaje de Comandos (Sintaxis), de Pregunta de Respuesta Múltiple.</b> .....	158
<b>5.4. PREGUNTA ABIERTA.</b> .....	159

<b>5.4.1. Procesamiento de la Información - Tabla de Frecuencias.</b> .....	160
<b>5.5. TRANSFORMACIÓN DE LA VARIABLE - EDAD EN RANGOS.</b> .....	163
<b>5.6. GRAFICAS INTERACTIVAS</b> .....	167
<b>5.6.1. Cómo crear una grafica interactiva.</b> .....	168
<b>CAPÍTULO SEIS</b> .....	<b>172</b>
<b>6. INTERPRETACIÓN Y ANÁLISIS DE LA INFORMACIÓN.</b> .....	<b>172</b>
6.1. NÚMERO DE VARIABLES A ANALIZAR. . . . .	172
6.2. INFERENCIA ESTADÍSTICA Y DESCRIPCIÓN DE VARIABLES .....	173
6.3. TIPOS DE ESCALA DE MEDICIÓN. ....	173
6.4. ANÁLISIS UNIVARIADO DE DATOS. ....	174
<b>6.4.1. Análisis de frecuencia.</b> .....	<b>174</b>
<b>6.4.2. Medidas de tendencia central.</b> .....	<b>174</b>
<b>6.4.3. Medidas de dispersión.</b> .....	<b>175</b>
<b>6.4.4. Ejemplos de análisis univariante.</b> .....	<b>175</b>
<b>6.4.4.3. Ejemplo 3. Variables de intervalo.</b> .....	<b>179</b>
6.5. ANÁLISI BIVARIANTE.....	182
<b>6.5.1. Análisis bivalente de medias.</b> .....	<b>183</b>
<b>6.5.2. Procedimiento en el programa estadístico SPSS.</b> .....	<b>184</b>
<b>6.5.3. Análisis de varianza (la Existencias de diferencias significativas).</b> . . . . .	<b>188</b>
6.6. TABLAS CONTINGENCIA (TABULACIONES CRUZADAS). ....	190
<b>6.6.1. Ejemplo de tabla de contingencia.</b> .....	<b>190</b>
<b>6.6.2. Procedimiento en el programa SPSS.</b> .....	<b>190</b>
<b>6.6.3. Análisis de frecuencias.</b> .....	<b>195</b>
<b>6.6.4. Análisis de chi cuadrado.</b> .....	<b>196</b>
<b>6.6.5. Análisis de los residuos y de los porcentajes básicos.</b> .....	<b>197</b>
<b>CAPÍTULO SIETE</b> .....	<b>200</b>
<b>7. ANÁLISIS DE PROYECCIONES MEDIANTE REGRESIÓN Y CORRELACIÓN (CON DOS VARIABLES).</b> .....	<b>200</b>
7.1. ESTUDIO DE CASO. ....	201
<b>7.1.1. Determinación del Número de Trabajadores con Derecho a Dotación.</b> .....	<b>201</b>
7.2. FUNCIÓN DE REGRESIÓN .....	206

<b>7.2.1. Significado del Término “Lineal”.</b> .....	<b>206</b>
<b>7.2.2. Representación gráfica de una función lineal.</b> .....	<b>207</b>
<b>7.3. CALCULO DE LOS PARÁMETROS DE LA FRL.</b> .....	<b>210</b>
<b>7.3.1. Método de Mínimos Cuadrados Ordinarios (MCO)..</b> .....	<b>210</b>
<b>7.3.2. Procesamiento de la información en SPSS.</b> .....	<b>213</b>
<b>7.4. COEFICIENTE DE DETERMINACIÓN <math>R^2</math></b> .....	<b>216</b>
<b>7.4.1. Propiedades del Coeficiente de Determinación <math>R^2</math>.</b> .....	<b>216</b>
<b>7.5. COEFICIENTE DE CORRELACIÓN R.</b> .....	<b>217</b>
<b>7.5.1. Propiedades del Coeficiente de Correlación R.</b> .....	<b>217</b>
<b>7.6. CALCULO DE LOS PARÁMETROS <math>\beta</math>'s</b> .....	<b>218</b>
<b>7.7. PRESENTACIÓN DE LA INFORMACIÓN.</b> .....	<b>219</b>
<b>7.7.1. Interpretación de los resultados</b> .....	<b>219</b>
<b>7.8. PROYECCIÓN DE DATOS.</b> .....	<b>220</b>
<b>CAPITULO OCHO</b> .....	<b>222</b>
<b>8. FORMAS FUNCIONALES DE LOS MODELOS DE REGRESIÓN.</b> .....	<b>222</b>
<b>8.1. MODELO LOG – LINEAL (POTENCIAL).</b> .....	<b>222</b>
<b>8.1.1. Características del Modelo Log – Lineal</b> .....	<b>223</b>
<b>8.2. ESTUDIO DE CASO.</b> .....	<b>224</b>
<b>8.2.1. Creación de las Variables en el Programa SPSS.</b> .....	<b>225</b>
<b>8.2.2. Creación y Cálculo de los Logaritmos de las Variables.</b> .....	<b>226</b>
<b>8.2.3. Estimación del modelo log – lineal.</b> .....	<b>228</b>
<b>8.2.4. Análisis de los Parámetros del Modelo.</b> .....	<b>230</b>
<b>8.2.5. Presentación de resultados.</b> .....	<b>231</b>
<b>8.2.6. Interpretación de los Resultados.</b> .....	<b>232</b>
<b>8.2.7. Proyección de los Resultados.</b> .....	<b>232</b>
<b>8.3. MODELO LIN – LOG (SEMILOGARÍTMICO).</b> .....	<b>232</b>
<b>8.3.1. Crear las variables en el programa SPSS.</b> .....	<b>234</b>
<b>8.3.2. Presentación de la información y análisis de datos.</b> .....	<b>235</b>
<b>8.3.3. Interpretación de los Resultados.</b> .....	<b>236</b>
<b>8.3.4. Proyección de Datos.</b> .....	<b>236</b>
<b>APÉNDICE A.</b> .....	<b>238</b>

<b>DESCRIPCIÓN GENERAL DEL ARCHIVO COMERCIO.SAV.....</b>	<b>238</b>
<b>FICHA TÉCNICA .....</b>	<b>238</b>
<b>APÉNDICE B. ....</b>	<b>244</b>
<b>TABLA DE DISTRIBUCIÓN NORMAL .....</b>	<b>244</b>
<b>APÉNDICE C. ....</b>	<b>246</b>
<b>TABLA PARA LA DETERMINACIÓN DE UNA MUESTRA.....</b>	<b>246</b>
<b>BIBLIOGRAFÍA.....</b>	<b>252</b>