



Universidad de Talca  
Centro de Inserción Laboral  
Programa Jóvenes Profesionales

# Manual Básico SPSS

Manual de introducción a SPSS



Elaborado por:

Juan Antonio González González | Ingeniero Comercial | Universidad de Talca | 2009 | [juagonzalez@utalca.cl](mailto:juagonzalez@utalca.cl)

## AGRADECIMIENTOS

Quisiera agradecer a todas las personas que de alguna manera han ayudado en la elaboración de este documento. A todas las personas que trabajan en el Centro de Inserción Laboral: Don Julio, la Anita, La Margarita, Andrés, los Psicólogos, la Sra. Berta y al personal de aseo, por la buena disposición y todo el apoyo prestado durante el periodo en que me desempeñé como joven profesional del CIL.

Quisiera también agradecer a mis compañeros de labores allá en el centro, que siempre han sido una fuente de apoyo bastante grande, aunque quizás, muchos no se daban cuenta... ¡muchas gracias chiquillos(as)!

Finalmente, agradecer a mi niña, María Paz, por el apoyo que siempre me ha prestado en cada uno de los proyectos en los que me embarco, nunca limitando ni mi imaginación ni mis ideas... Así mismo quisiera agradecer a los profesores Marcelo Rodríguez Gallardo y Jorge Zamora por todo el apoyo que me prestaron para la elaboración de este texto, por las bases de datos, libros e ideas para hacer esto lo más simple posible.

**Muchas Gracias.**

## LICENCIA

De manera de facilitar la difusión de este contenido, he decidido licenciar el texto con una licencia Creative Commons del tipo: "Atribución - No Comercial - Licenciar Igual 2.0 Chile", bajo las siguientes condiciones:

### Eres libre de:



Copiar, distribuir, comunicar y ejecutar públicamente la obra.



Hacer obras derivadas.

### Bajo las siguientes condiciones:



**Atribución.** Debes reconocer y citar la obra de la forma especificada por el autor o el licitante.



**No Comercial.** No puedes utilizar esta obra para fines comerciales.



**Licenciar Igual.** Si alteras o transformas esta obra, o generas una obra derivada, solo puedes distribuir la obra derivada bajo una licencia idéntica a ésta.

# INDICE

## Contenido de capítulos



Capítulo 1 - Introducción a SPSS	4
Capítulo 2 - Trabajo con Bases de Datos	7
Capítulo 3 - Análisis de Estadísticos Descriptivos	17
Capítulo 4 - Análisis de Variables Categóricas	26
Capítulo 5 - Pruebas de Hipótesis	29
Capítulo 6 - Análisis de Varianza (ANOVA)	35
Capítulo 7 - Gráficos	38
Capítulo 8 - Análisis de Correlación	46
Capítulo 9 - Análisis de Regresión Lineal	49
Capítulo 10 - Análisis de Regresión Logística	53
Capítulo 11 - Análisis Discriminante	58
Capítulo 12 - Análisis Factorial	63

# CAPÍTULO 1

Introducción a SPSS



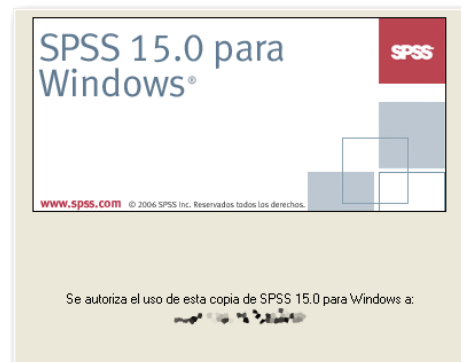
## Introducción a SPSS

Hoy en día resulta obligatorio para los nuevos profesionales poseer herramientas que le permitan desarrollar su trabajo de forma más óptima y eficiente. Independiente del área en que se desempeñen, tener la capacidad de interpretar datos para transformarlos en conocimiento, resulta vital para cualquier profesional y para la posterior toma de decisiones.

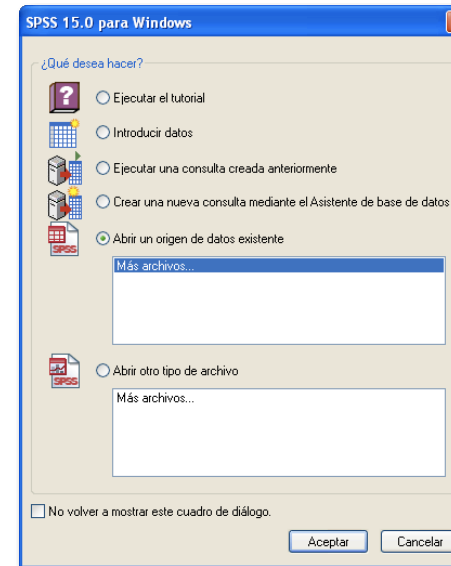
SPSS es uno de los Software más conocidos y utilizados para la realización de investigaciones de mercado e investigaciones relacionadas al área de la sociología y psicología. Una de las características fundamentales de SPSS es su facilidad de uso, junto a la potencia e integridad del software, convirtiéndolo en una de las herramientas más potentes para éste tipo de trabajos.



La imagen que se puede ver al lado izquierdo corresponde al icono de la aplicación, el que por defecto se instala bajo el nombre de "SPSS para Windows".

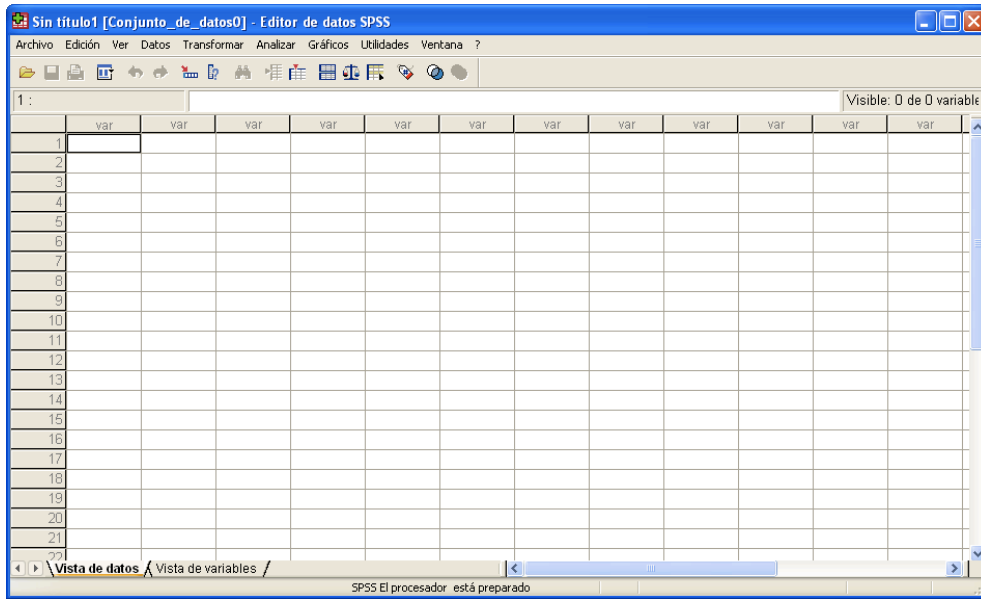


Al iniciar el Software, lo primero que se despliega es una ventana de información (que se observa a la izquierda) la que indica la versión del software, que en este caso corresponde a la versión 15 para Windows y además muestra la información referente al propietario de la licencia.



Una vez abierto el software, se observa una ventana de "Bienvenida", en la que se ofrecen distintas opciones, donde la más importante es la de abrir archivos utilizados recientemente.

Pese a lo anterior, se recomienda no utilizar esta ventana, debido a que generalmente no se puede observar bien la base de datos con la que se está trabajando, pudiendo generar errores al momento de abrir una base de datos particular.

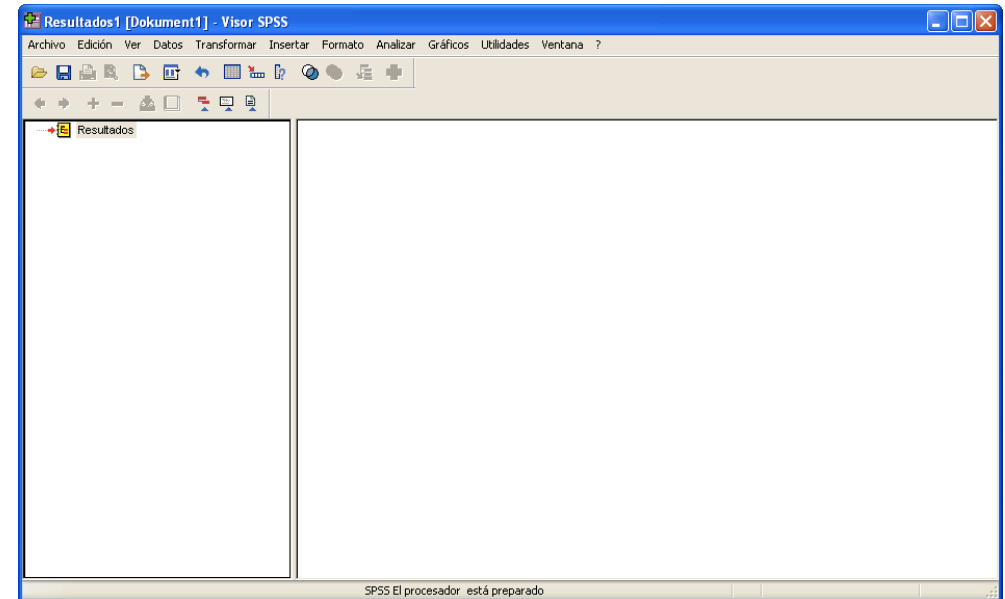


### Editor de Datos:

SPSS se compone de dos ventanas de trabajo. La primera recibe el nombre de Editor de Datos (arriba) y es la ventana donde se ingresan los datos con los que se trabajará. Para los familiarizados con Microsoft Excel®, esta ventana les resultará familiar y su forma de trabajo será bastante similar. En las columnas (hacia la izquierda) se introducirán las distintas variables que compongan la base de datos. En términos prácticos, cada pregunta de un cuestionario realizado corresponde a una variable.

En las filas (hacia abajo) son introducidos los datos, o las respuestas entregadas por los encuestados, o donde sea que se hayan extraído los datos para la realización de la base de datos. Así, si se entrevistó a 100 personas, hacia abajo deberá haber 100 registros.

Este Editor posee dos tipos de vista: “Vista de datos” y “Vista de variables”. En la primera, el investigador observará los datos ingresados, mientras que en la segunda se puede acceder a los atributos de las distintas variables para poder editarlos. Lo anterior será ejemplificado en el capítulo siguiente.



### Visor de Resultados:

La segunda ventana correspondiente al área de trabajo de SPSS se conoce como “Visor de Resultados” y, tal como su nombre lo indica, corresponde a una ventana donde el software despliega los resultados que se obtienen de los distintos análisis realizados. En la sección izquierda, el software mostrará los títulos de los distintos análisis y los desplegará como una diagrama de árbol.

Por otra parte en la sección derecha se muestran los resultados obtenidos, los cuales pueden ser copiados y pegados a software como Microsoft Word® o Excel® para realizar informes u otras acciones que el investigador considere necesarias.

# CAPÍTULO 2

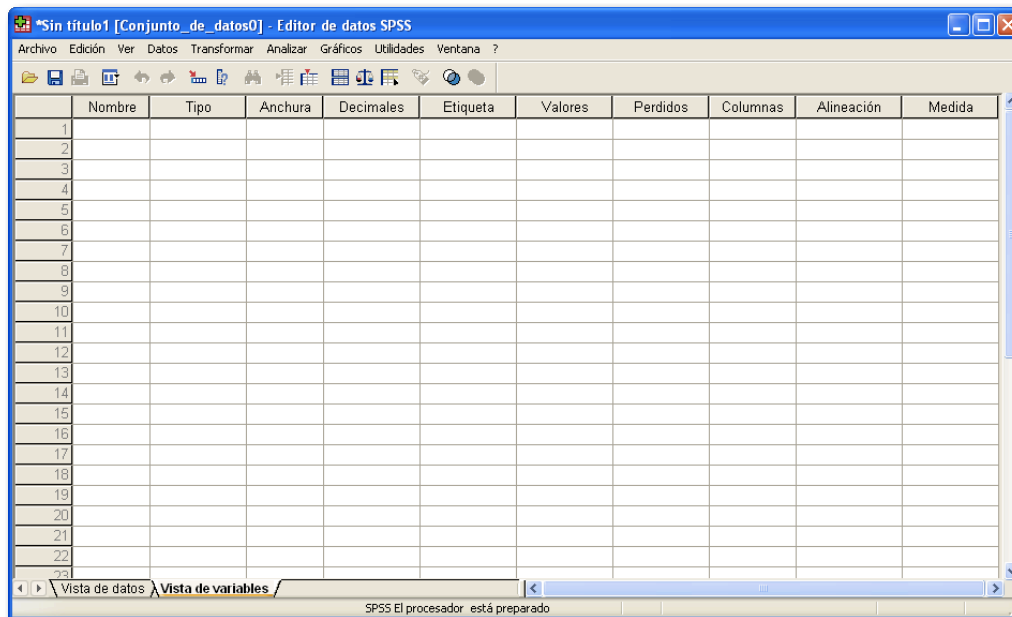
Trabajo con Bases de Datos



## Trabajo con Bases de Datos

En este apartado se explicará la forma de crear y editar las distintas bases de datos, de manera que se encuentren listas para realizar los distintos análisis requeridos por el investigador.

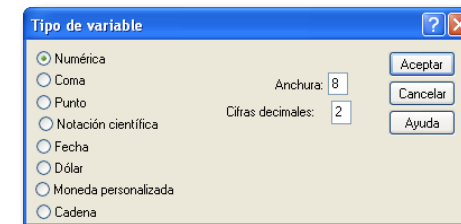
Lo primero que se debe hacer antes de ingresar datos a una base de datos, es prepararla para ello, lo cual debe ser efectuado en el "Editor de datos", utilizando la "Vista de variables", lo que mostrará una ventana como la que se observa a continuación.



Como se puede observar, en la parte superior se muestra una serie de "características" de las variables. Sin embargo, no todas se deben utilizar siempre, por lo que solamente se explicará el funcionamiento de las características que resulten más relevantes para el uso común de SPSS.

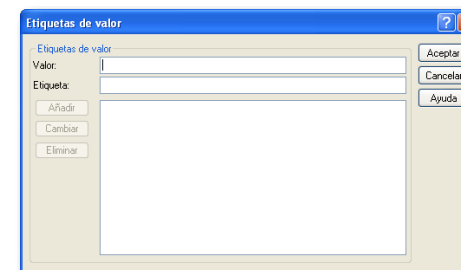
**Nombre:** En esta sección se debe asignar un nombre para la variable. Dicho nombre debe poseer un máximo de 8 letras y se debe evitar el uso de tildes, símbolos y caracteres como la letra "ñ".

**Tipo:** Se ofrece la posibilidad de asignar una característica a la variable, de manera que ésta sea más fácil de trabajar y de forma que el software la interprete de manera correcta. A continuación se muestra la ventana de "Tipos" de variable:



**Etiqueta:** Para que en el visor de resultados muestre correctamente los nombres de las variables (y sin el límite de las 8 letras) se puede establecer una etiqueta para ellas, de manera de en este sector se podrá escribir lo que el investigador considere necesario.

**Valores:** Para cierto tipo de variables, la etiqueta debe ser definida para más de un valor, como por ejemplo cuando estamos hablando del sexo de una persona (que puede ser masculino o femenino), por lo cual hay que decir a SPSS como debe identificar eso. Para ello se utiliza esta sección, la cual despliega una ventana como la siguiente:





**Medida:** Para que SPSS realice correctamente los análisis solicitados, es necesario establecer correctamente las unidades de medida de cada una de las variables. Para ello ofrece tres opciones: Escala, Ordinal y Nominal. A continuación se explicará brevemente cada una de estas unidades de medidas utilizadas por el software.

Medida	Descripción
Escala	Corresponde a variables numéricas sin una característica especial. Por ejemplo, Notas del colegio, Edad, Estatura, Salario, etc.
Ordinal	Corresponde a variables que representan categorías y además dichas categorías poseen un ranking implícito. Ejemplo de este tipo de variable es aquella donde hay que seleccionar que tan de acuerdo o en desacuerdo se está con algo.
Nominal	Corresponde a variables que representan categorías pero que no poseen ningún tipo de ranking. Por ejemplo: Tramo de edad, sexo, etc.

Las características anteriormente definidas son las comúnmente utilizadas por los usuarios, no obstante el investigador puede utilizar el resto de características de manera de personalizar aún más su base de datos. Una vez que estas características se encuentren correctamente definidas, el investigador se encuentra con la capacidad de comenzar a ingresar los datos a la base de datos.

### Libro de Códigos:

Una de las características fundamentales de todo tipo de software de análisis de datos es que estos trabajan con números, por lo que cada una de las preguntas (variables) debe ser definida por un valor o un conjunto de valores numéricos para así ingresarlos al software.

Una recomendación es la utilización de un libro de códigos, el cual no es más que una ficha en que se incluyen todos los valores numéricos asignados a cada una de las variables, de manera que al momento de ingresar los datos a SPSS, el investigador, así como los digitadores sepan que rango de valores debe usar cada una de las variables.

A continuación se mostrará un ejemplo práctico para la creación de un libro de códigos, la edición de las características de las variables y el ingreso de datos a SPSS.

### Ejemplo Práctico:

La Empresa JPP requiere obtener información acerca de la percepción de sus clientes sobre un nuevo servicio ofrecido desde hace unos meses. Para ello se aplicará un cuestionario a 30 de sus mejores clientes para conocer la opinión de éstos. Su misión como investigador será preparar el libro de códigos y la base de datos para que la realización de los análisis sea efectuada de manera correcta y sin errores. En la página siguiente se presenta el cuestionario aplicado por la empresa a sus clientes.

**Marque con una X la alternativa que corresponda:**

1. Sexo:      Masculino               Femenino
2. Edad:      18 a 29     30 a 49     50 o más

**Escriba su respuesta en la zona designada para ello:**

3. ¿Cuál es su renta líquida?  
\_\_\_\_\_
4. ¿Cuántos años de estudio posee?  
\_\_\_\_\_

**5. Marque con una X la alternativa que más se acerque a su percepción del servicio:**

	Muy de Acuerdo	De Acuerdo	Indiferente	En Desacuerdo	Muy en Desacuerdo
El Servicio ofrece una nueva experiencia en la compra.					
El precio del servicio es acorde a la calidad de este					

Muchas Gracias por participar

**Libro de Códigos:**

El libro de códigos es una ficha donde se especifican los valores numéricos con que se codificará cada una de las variables del estudio (en este caso del cuestionario), de manera que el ingreso de los valores a la base de datos se realice uniformemente, evitando así errores derivados de la digitación de los cuestionarios.

Si bien no existe una forma única de hacer un libro de códigos, a continuación se ofrecerá una vista parcial de un libro de códigos estándar.

N°	Nombre de la Variable	Etiqueta de la Variable	Valores
1	Sexo	Sexo	1 : Masculino 2 : Femenino
2	Edad	Edad	1 : 18 a 29 2 : 30 a 49 3 : 50 o más años
3	Ren_Liq	Renta Líquida	
4	Niv_Est	Años de Estudio	
5.1	Exp_Serv	El Servicio ofrece una nueva experiencia en la compra	1 : Muy de Acuerdo 2 : De acuerdo 3 : Indiferente 4 : En Desacuerdo 5 : Muy en Desacuerdo
5.2	Pre_Serv	El precio del servicio es acorde a la calidad de este	1 : Muy de Acuerdo 2 : De acuerdo 3 : Indiferente 4 : En Desacuerdo 5 : Muy en Desacuerdo

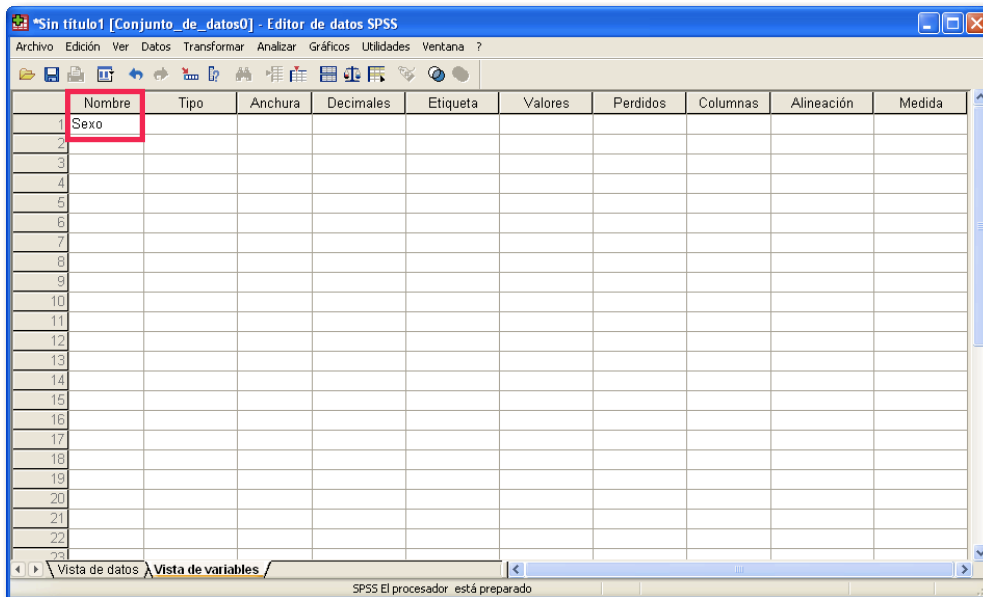
### Edición de las Variables:

Una vez listo el libro de códigos, toca el turno de preparar la base de datos para el ingreso de la información. Lo primero que hay que hacer es definir las características de cada una de las variables, resultando el libro de códigos una herramienta indispensable para esta parte del proceso.

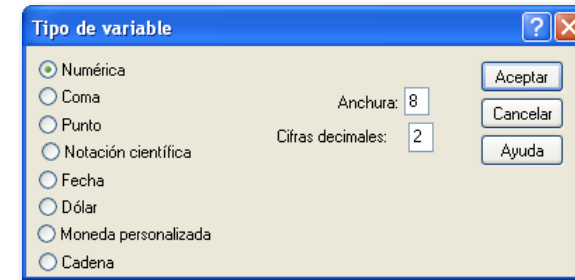
A modo práctico solo se hará la edición de las variables 1, 3 y 5.1 del libro de códigos. En el archivo "BaseDatosC1.sav" se encuentra la base de datos lista para el ingreso de información, con todas las variables especificadas.

### Variable 1 (Sexo):

Lo primero que se debe insertar es el nombre de la variable. En este caso "Sexo", bajo la columna "Nombre", tal como muestra la imagen siguiente (Recuadro rojo).



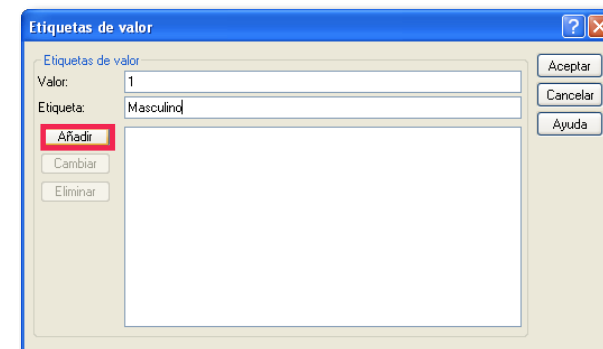
Posteriormente en "Tipo" hay que escoger "Numérica", pues se trata de una variable numérica, sin características especiales, tal como se muestra a continuación:

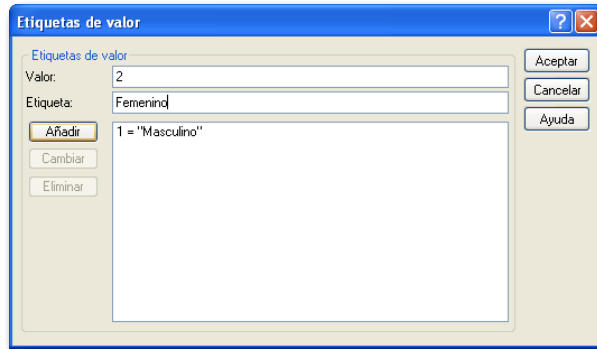


En "Etiqueta" se debe establecer el Nombre completo que recibirá la variable, en este caso se llamará "Sexo".

Siguiendo, bajo "valores" se debe ingresar los valores asignados a cada una de las posibles respuestas a la pregunta. En este caso "Masculino" o "Femenino". En este caso los valores ingresados serán 1 para Masculino y 2 para Femenino, de acuerdo al libro de códigos.

Al hacer clic ahí, se abrirá una nueva ventana, donde se debe ingresar el "Valor" y el "Nombre" para cada una de las posibles respuestas. La imagen siguiente muestra los valores que se deben ingresar para la respuesta "Masculino". Una vez listo se debe presionar el botón "Añadir" (En recuadro) y el valor de la etiqueta pasará a la zona inferior. Posteriormente se debe ingresar los valores para "Femenino", tal como lo muestran las imágenes siguientes:





Una vez ingresados ambos valores se debe hacer clic en el botón “Aceptar”.

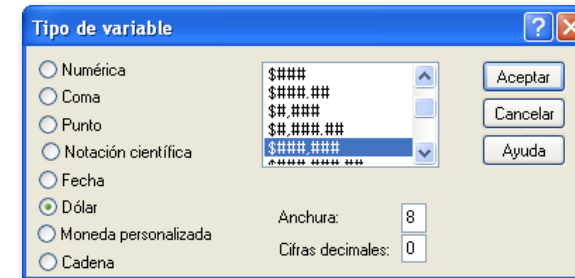
Finalmente se debe establecer la unidad de medida. Debido a que la variable sexo es una variable que representa una categoría (Masculino o Femenino) y que a su vez no se puede establecer un ranking, en la medida que ser Masculino no es superior a ser Femenino o viceversa, se está hablando de una variable con una escala de medida “Nominal”. Finalmente, cuando esté toda la variable definida, el software mostrará lo siguiente:

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	Sexo	Númérico	8	0	Sexo	{1, Masculino}, Ninguno		8	Derecha	Nominal
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										

### Variable 3 (Renta Líquida):

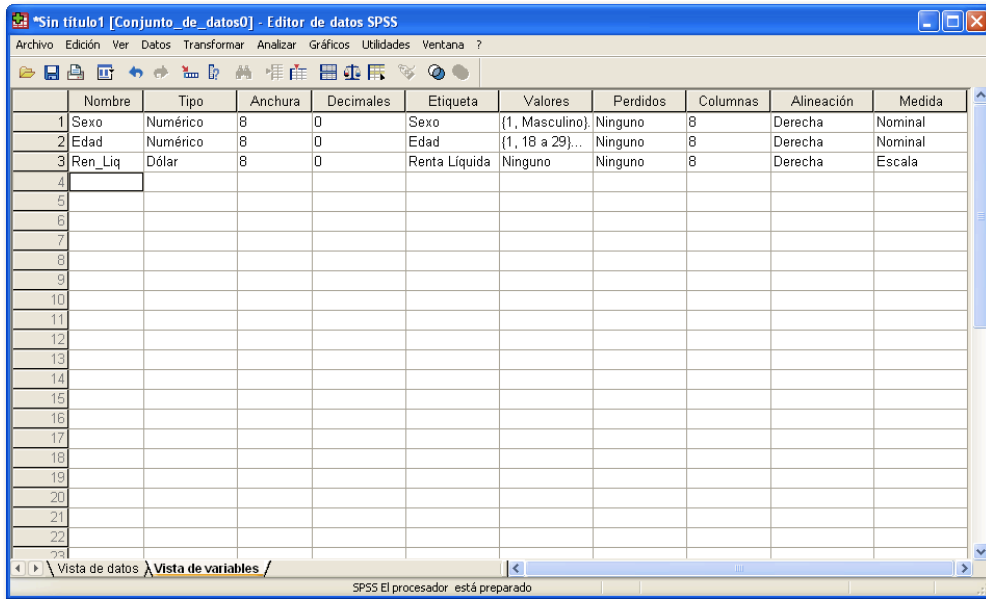
Lo primero que se debe insertar es el nombre de la variable. En este caso “Ren\_Liq”, bajo la columna “Nombre”.

Al definir el “Tipo” de variable, se debe tener en cuenta que esta es una unidad de medida monetaria, por lo que debe ser definida como tal. En el cuadro de diálogo se ofrece una opción denominada “Dólar” que es para definir variables monetarias.

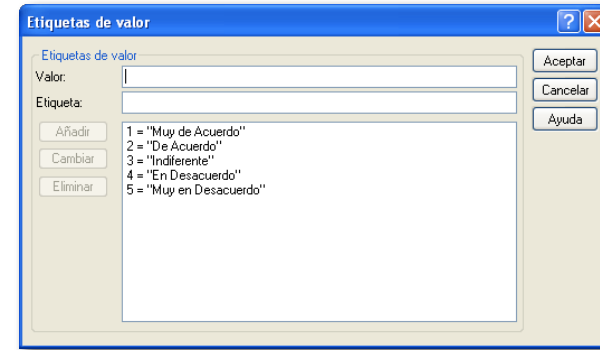


Bajo la columna etiqueta se debe escribir el valor de la etiqueta de acuerdo a lo especificado por el libro de códigos. En este caso “Renta Líquida” y debido a que esta variable no tiene respuestas adicionales, el cuadro “Valores” debe ser dejado sin llenar.

En cuanto a la unidad de medida, ya que se trata de una variable donde el encuestado ingresa un valor numérico (su salario), se está haciendo referencia a una variable de con unidad de medida de “Escala”, por lo que esa debe ser la unidad de medida asignada. La imagen de la página siguiente muestra como debe verse la base de datos con la variable 1, 2 y 3 completas:



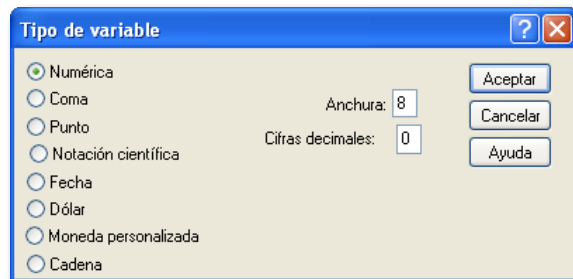
Bajo la columna etiqueta se debe escribir el valor de la etiqueta de acuerdo a lo especificado por el libro de códigos. En este caso “El Servicio ofrece una nueva experiencia en la compra” y posteriormente se debe llenar el cuadro de “Valores” para esta etiqueta, al igual como se hizo con la variable “Sexo”, pero esta vez utilizando las 5 posibles respuestas, tal como lo muestra la imagen siguiente:



**Variable 5.1 (El servicio ofrece una nueva experiencia en la compra):**

Lo primero que se debe insertar es el nombre de la variable. En este caso “Exp\_Serv”, bajo la columna “Nombre”.

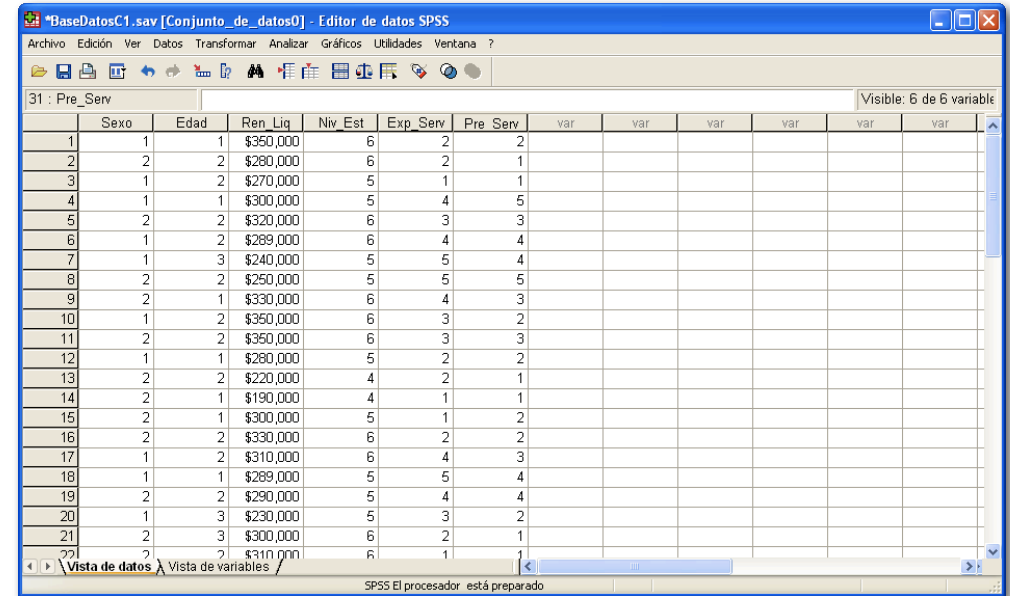
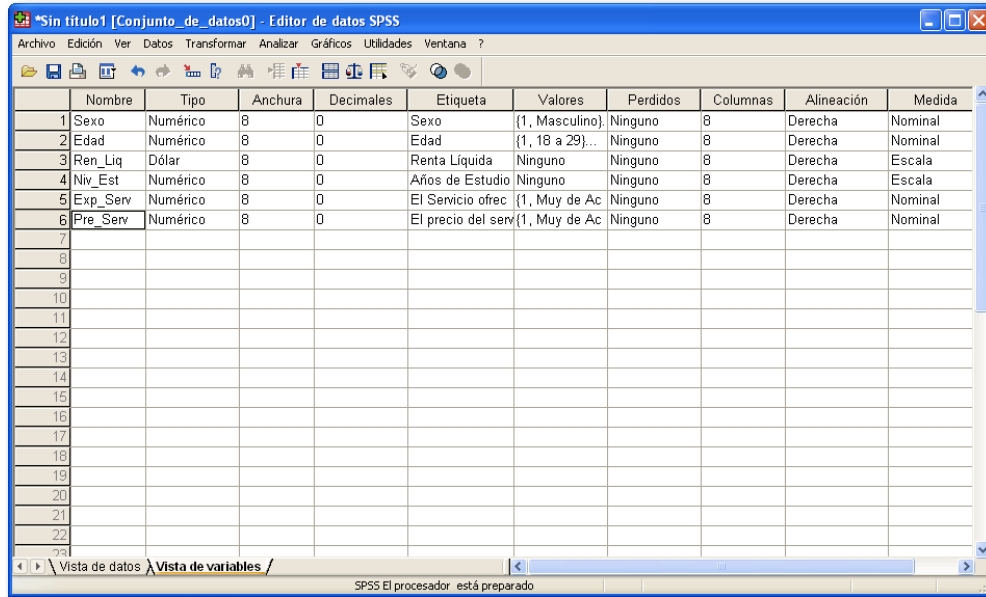
Al definir el “Tipo” de variable, se debe tener en cuenta que esta es una unidad de medida numérica, por lo que debe ser definida como tal:



Finalmente se debe definir una unidad de medida para la variable. Como se puede observar, esta variable también sirve para categorizar la información, sin embargo posee la característica que está definiendo un “ranking” de con la información, de acuerdo a las posibilidades de respuesta, por lo que se trata de una variable con medida “Ordinal”

En la página siguiente se observa la base de datos editada para cada una de las 5 variables del cuestionario, la cual se encuentra disponible en el archivo “BaseDatos1.sav” de manera que pueda ser verificado por ud.

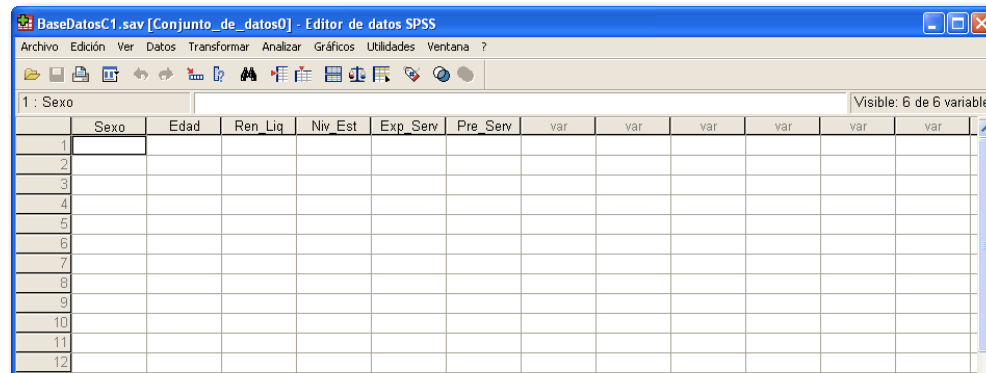
La base de datos, una vez completa debiese lucir como la imagen siguiente:



Al volver a la “Vista de datos” se obtendrá una pantalla como la siguiente. En ese momento SPSS se encuentra listo para que se comience a ingresar los datos en el, de acuerdo a lo especificado por el libro de códigos. El archivo “BaseDatosC1a” contiene una base de datos con la información de los 30 encuestados, de manera que se pueda revisar y comprender la forma en que se ingresa la información en el software.

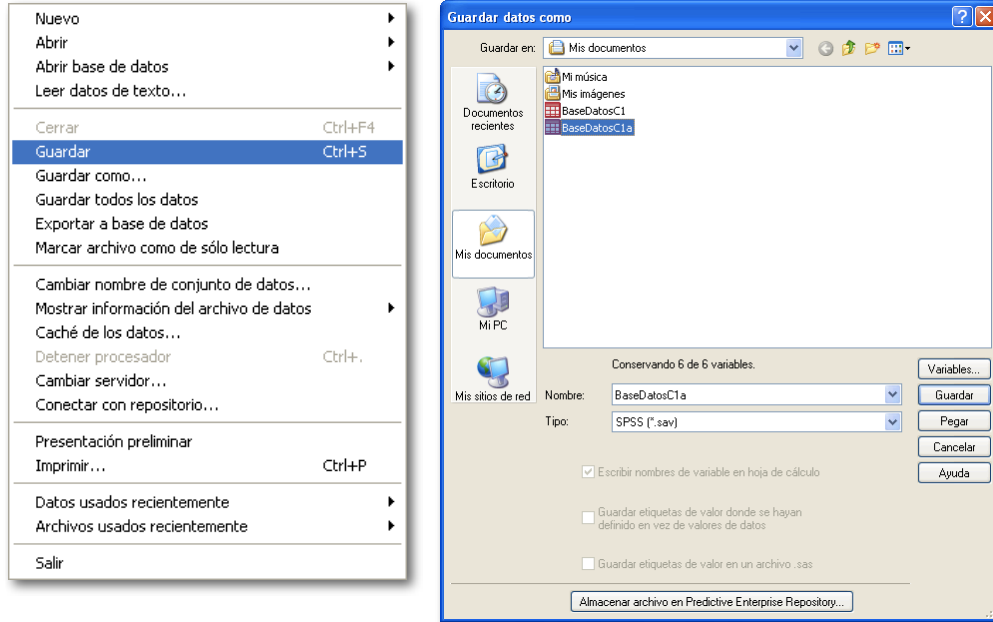
Guardar una Base de Datos:

SPSS tiene una particularidad, por cuanto el archivo de bases de datos, así como el archivo de resultados se guardan como archivos independientes y con extensiones distintas.



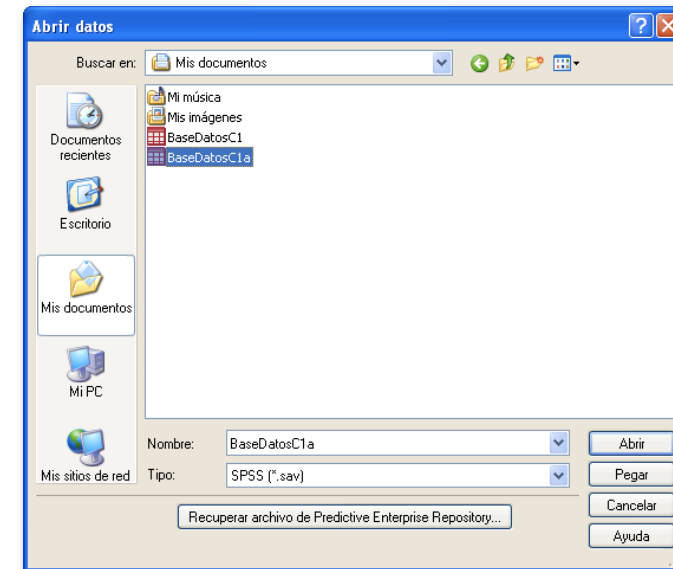
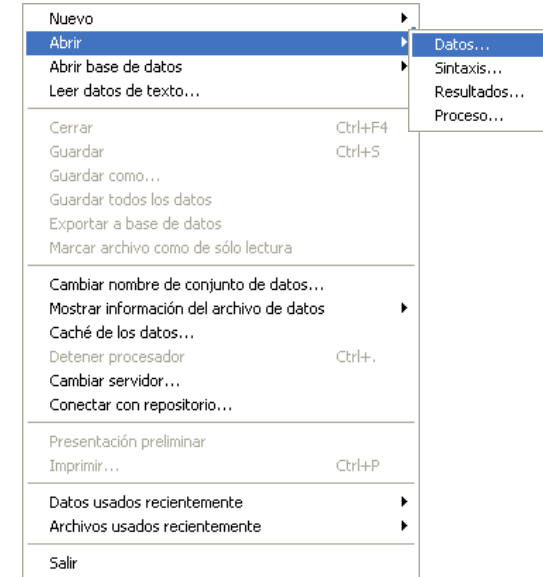
Ítem	Extensión
Guardar Base de Datos	Nombre.sav
Guardar Archivo de Resultados	Nombre.spo

Para Guardar una base de datos, se debe ir al menú “Archivo \ Guardar” y asignar un nombre al archivo. Posteriormente hacer clic en el botón “Guardar”. Tal como se observa en las imágenes siguientes:



### Abrir una Base de Datos:

Para abrir una base de datos se debe hacer clic en el menú “Archivo \ Abrir \ Datos”, y buscar el archivo que deseamos abrir, tal como muestran las imágenes siguientes:

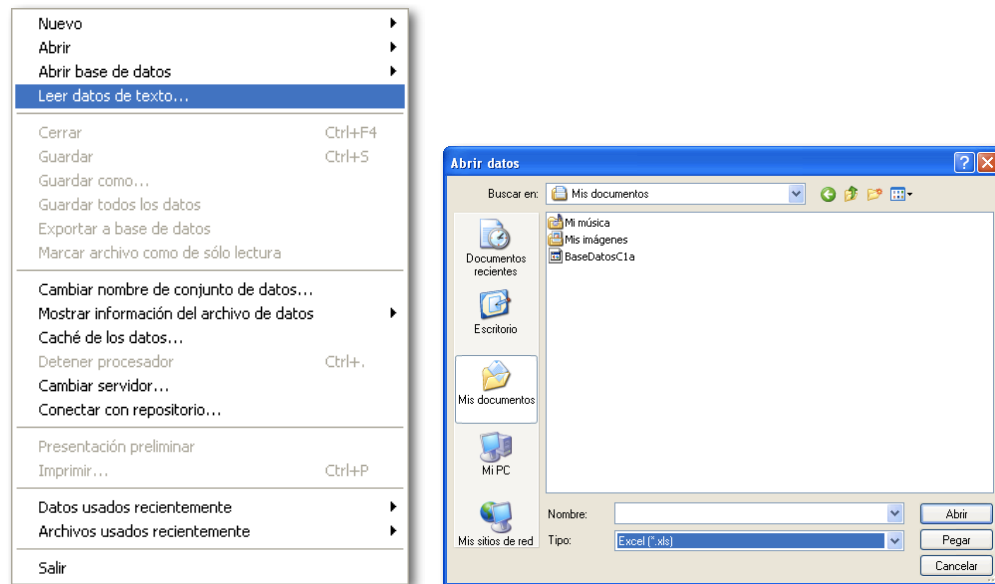


### Importar Datos de Microsoft Excel®

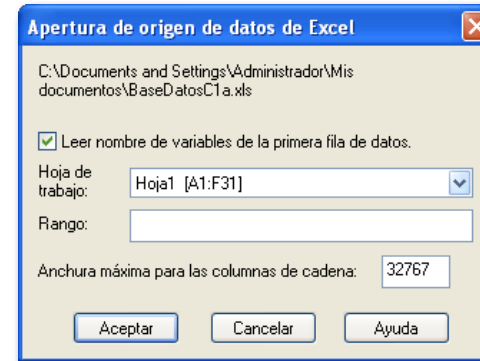
Una forma común de encontrar bases de datos es en formatos distintos a SPSS, como por ejemplo Microsoft Excel (.xls). Para poder trabajar con bases de datos de otro software primero se debe importar de manera de transformarla en datos comprensibles para SPSS.

La forma de hacer esto es la siguiente:

Primero se debe ir al menú "Archivo \ Leer datos de texto". En la ventana que abre debemos seleccionar "Excel [.xls]" donde dice "Tipo" y posteriormente buscar y seleccionar el archivo Excel que se desee importar.



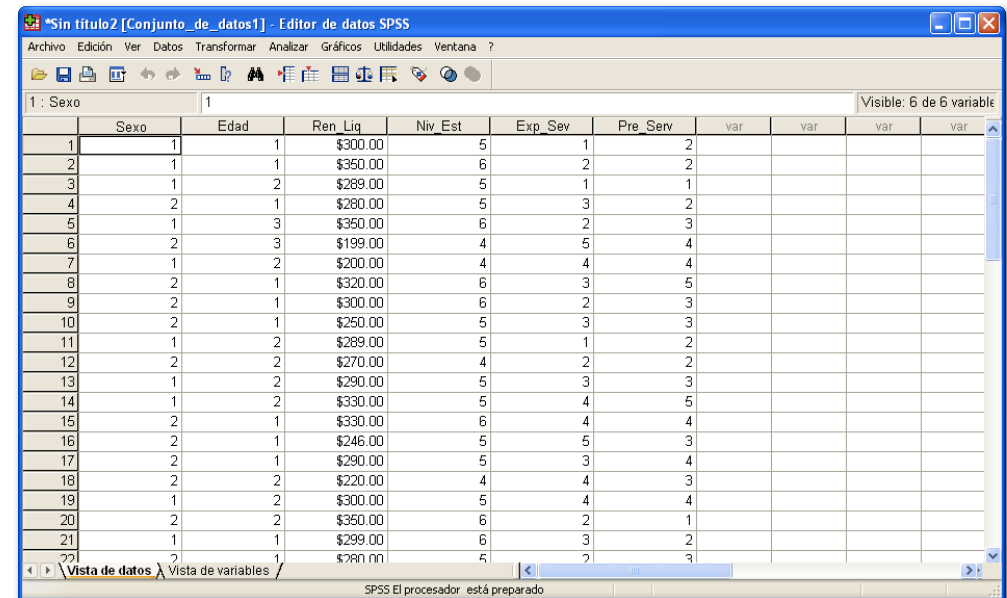
Una vez seleccionado el archivo se debe hacer clic en el botón "Abrir" y se abrirá un nuevo cuadro de diálogo.



Se recomienda dejar este diálogo tal cual aparece en el programa, pues generalmente cambiar esto puede generar problemas al investigador cuando no comprende a cabalidad los software que se están haciendo funcionar (SPSS y Excel).

Hacer clic en el botón "Aceptar" para continuar e importar los datos.

Una vez importados los datos desde la planilla excel, se abrirá una ventana del "Editor de datos" que lucirá como la siguiente. Si se va a la "Vista de variables", el investigador se dará cuenta que las características de las variables no están ingresadas, por lo que deberá ingresarlas tal cual se explicó anteriormente:





# CAPÍTULO 3

## Análisis de Estadísticos Descriptivos



## Análisis Estadísticos Descriptivos

En el presente apartado se explicará la forma de realizar algunos de los análisis descriptivos más comunes en las investigaciones de mercado, como son las medidas de tendencia central (Media, mediana), medidas de dispersión (Varianza, desviación estándar) y análisis de respuestas de selección Múltiple.

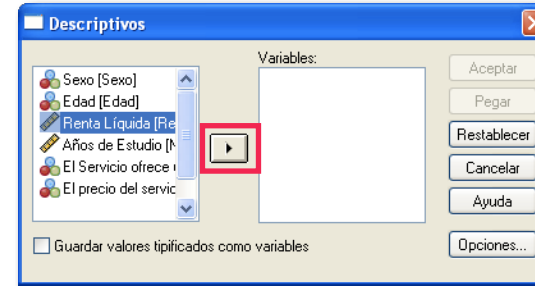
### Medidas de Tendencia Central: La Media.

La media es comúnmente conocida como “promedio” y corresponde a un valor de tendencia central para una variable con medida de Escala. En las variables nominales u ordinales no tiene sentido utilizar este estadístico por su naturaleza.

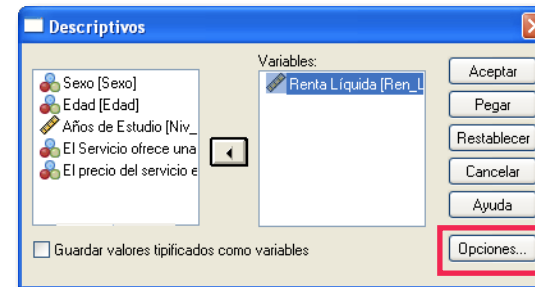
Para el caso práctico se utilizará la misma base de datos creada en el capítulo anterior, debido a que ya se encuentra familiarizado con esta (BaseDatosC1a.sav).

Informes	▶	Frecuencias...
Estadísticos descriptivos	▶	Descriptivos...
Tablas	▶	Explorar...
Comparar medias	▶	Tablas de contingencia...
Modelo lineal general	▶	Razón...
Modelos lineales generalizados	▶	Gráficos P-P...
Modelos mixtos	▶	Gráficos Q-Q...
Correlaciones	▶	
Regresión	▶	
Loglineal	▶	
Clasificar	▶	
Reducción de datos	▶	
Escalas	▶	
Pruebas no paramétricas	▶	
Series temporales	▶	
Supervivencia	▶	
Respuesta múltiple	▶	
Análisis de valores perdidos...	▶	
Muestras complejas	▶	
Control de calidad	▶	
Curva COR...	▶	

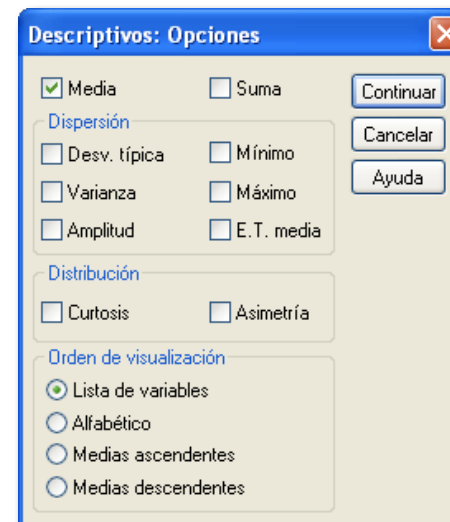
1. Lo primero que se debe hacer es ir al menú **“Analizar”** (que es donde se encuentran todas las opciones de análisis de SPSS) y luego ir al submenú **“Estadísticos descriptivos”** y dentro del dicho menú, escoger **“Descriptivos...”**, Tal cual se observa en la imagen que acompaña este texto.



2. Posteriormente se abrirá un cuadro de diálogo como el de la izquierda. Ahí se debe seleccionar la variable a la que se quiere calcular la media (promedio), en este caso la Renta Líquida y se debe hacer clic en el botón con el triángulo (recuadro rojo).



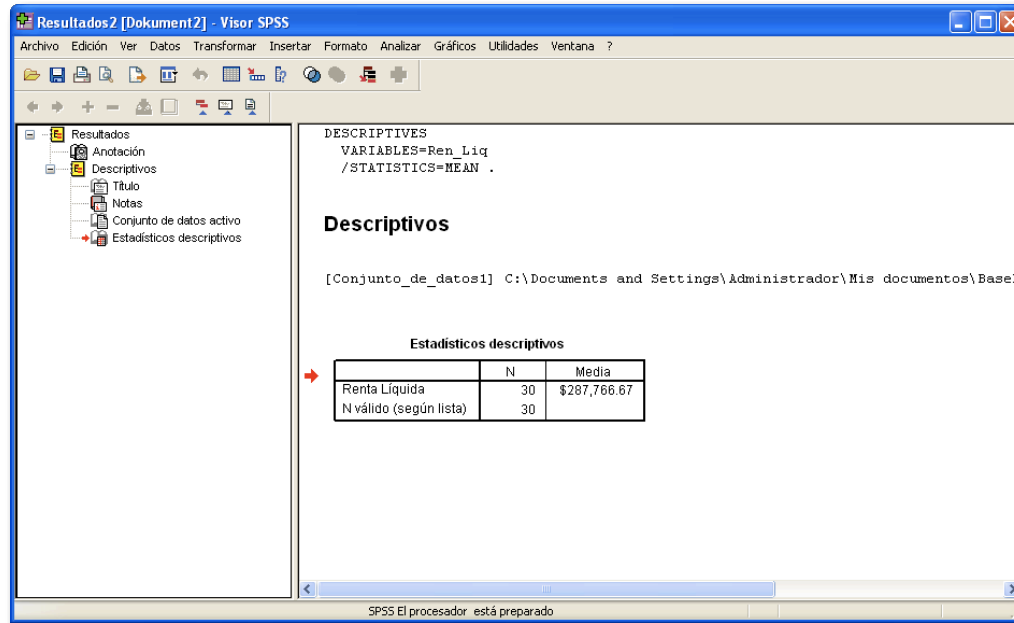
3. Una vez seleccionada la variable esta pasará a ubicarse en la zona de la derecha, tal cual se ilustra en la imagen. Posteriormente se recomienda hacer clic en el botón “Opciones..” para seleccionar los estadísticos que se quiere calcular.



4. El cuadro abierto ofrece distintos análisis a realizar. Por ser este el primer ejemplo se recomienda que solamente se deje seleccionado el análisis “Media” y el resto se deje sin seleccionar.

Es posible realizar más de un análisis al mismo tiempo. Solo basta seleccionar más análisis de los que se ofrecen ahí y SPSS se encargará de mostrar todos los análisis requeridos por el investigador.

Una vez hecho clic en “Continuar” y posteriormente en “Aceptar”, SPSS abrirá una ventana del visor de resultados con la información requerida (en este caso la media). A continuación una captura del visor de resultados:



**Estadísticos descriptivos**

	N	Media
Renta Líquida	30	\$287,766.67
N válido (según lista)	30	

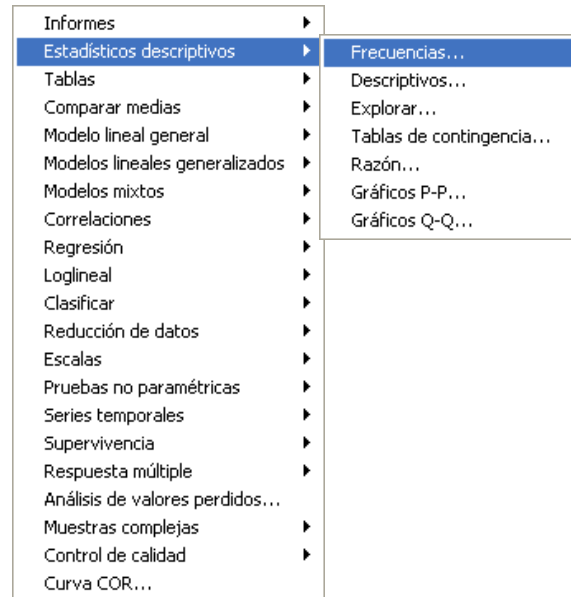
**Interpretación:**

El salario promedio de los 30 Clientes Encuestados es de \$287.766.

**Medidas de Tendencia Central: La Mediana.**

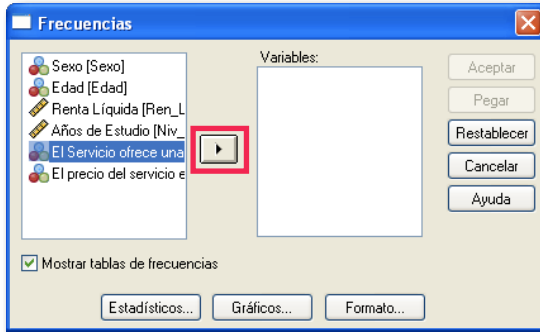
La mediana es similar al promedio, pero es utilizada en variables de tipo Ordinal o Nominal para determinar cual es la tendencia central escogida por parte del encuestado.

Nuevamente se utilizará la base de datos del capítulo anterior y se buscará la Mediana para la variable “El servicio ofrece una nueva experiencia en la compra”

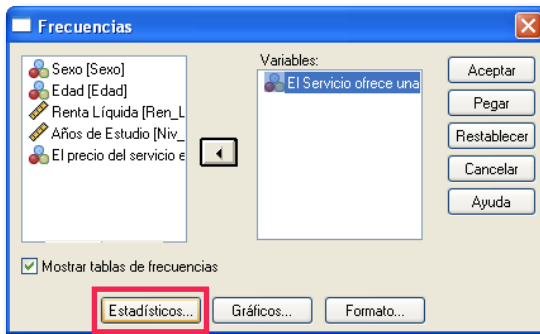


1. Para la determinación de la Mediana se debe ir al menú: **“Analyze \ Estadísticos Descriptivos \ Frecuencias”**.

Una vez abierto el software abrirá un cuadro de diálogo similar al que se mostraba para calcular la media, pero esta vez tendrá algunas variaciones.



2. Una vez abierto el cuadro de diálogo se debe seleccionar la variable que a la que se desea calcular la mediana y luego presionar el botón con el triángulo. En este caso es la variable “El servicio ofrece una...”.



3. Posteriormente se recomienda seleccionar los análisis que se desea realizar. Para ello se debe hacer clic en el botón “Estadísticos”, el que abrirá una nueva ventana con distintos análisis disponibles para el investigador.



4. En el cuadro de dialogo “Estadísticos” se debe seleccionar los análisis que el investigador desee realizar. A modo práctico solamente se seleccionará el análisis de “Mediana”, que es la medida de tendencia central que se está explicando en el presente capítulo.

Una vez seleccionados los análisis se debe hacer clic en el botón “Continuar”.

**Estadísticos**

El Servicio ofrece una nueva experiencia en la compra

N	Válidos	30
	Perdidos	0
Mediana		2,50

**Interpretación:**

Debido a que el valor de la mediana es de 2.5, esto quiere decir que los encuestados tienden a estar entre indiferentes y de acuerdo con que el servicio ofrece una nueva experiencia en la compra

**El Servicio ofrece una nueva experiencia en la compra**

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos				
Muy de Acuerdo	5	16,7	16,7	16,7
De Acuerdo	10	33,3	33,3	50,0
Indiferente	8	26,7	26,7	76,7
En Desacuerdo	5	16,7	16,7	93,3
Muy en Desacuerdo	2	6,7	6,7	100,0
Total	30	100,0	100,0	

**Interpretación:**

El visor de resultados también muestra una tabla que indica las frecuencias (cantidad de veces) que se escogió cada una de las alternativas por parte de los usuarios, así como el porcentaje al que equivale esa cantidad de respuestas sobre el total de encuestados. Si se observa con atención, la tabla muestra claramente que las respuestas tienden a estar agrupadas entre “Indiferente” y “De acuerdo”, tal cual lo determinó el cálculo e interpretación de la mediana.

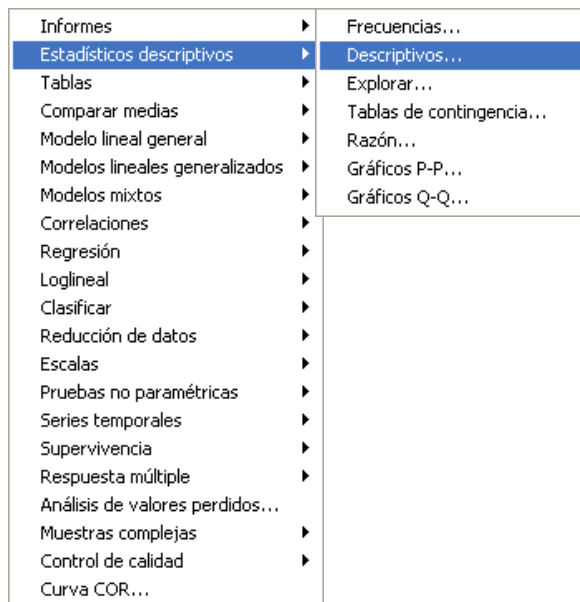
**Medidas de Dispersión: La Varianza y Desviación Estándar.**

Muchas veces el investigador necesita saber que tan dispersos están los datos unos con otros respecto al promedio o en términos generales. Por ejemplo, si tenemos un grupo de dos personas que ganan \$300.000 cada una, no es lo mismo que tener un segundo grupo de dos personas donde una gana \$100.000 y la otra \$500.000, aún cuando en promedio ambos grupos ganan \$300.000.

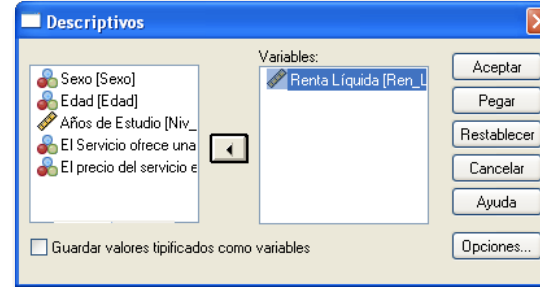
Evidentemente el segundo grupo posee una dispersión más grande que el primero, y es justamente para evitar estos problemas o para advertirlos a tiempo que las llamadas “medidas de dispersión de datos” son necesarias.

Las medidas comúnmente utilizadas en investigación de mercados son la Varianza y la Desviación Estándar, siendo esta última la más utilizada porque es una medida que indica cuan dispersos están los datos respecto del promedio. Obviamente un mayor valor implica una mayor dispersión de los datos.

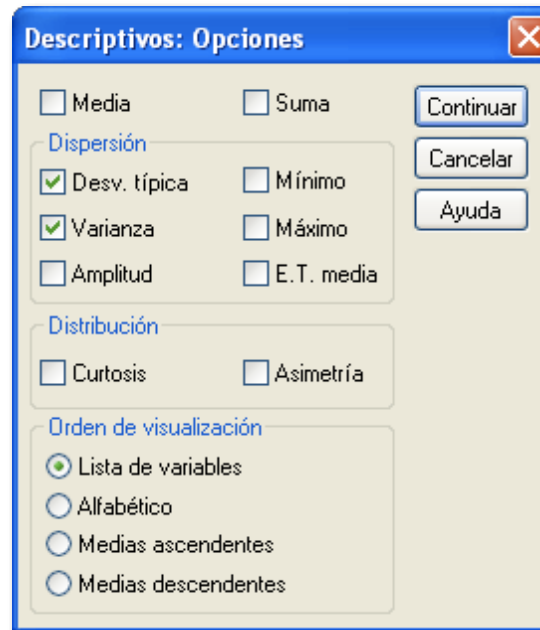
A modo práctico se seguirá utilizando la misma base de datos “BaseDatosC1a.sav” y determinaremos la varianza y desviación estándar para la variable “Sueldo Líquido”.



1. Para calcular la varianza y desviación estándar se debe hacer clic en el menú “**Análisis**” y posteriormente situar el mouse sobre “**Estadísticos Descriptivos**” y una vez allí seleccionar “**Descriptivos**”, tal como lo ilustra la imagen de la izquierda.



Posteriormente se debe seleccionar la variable que se desee analizar y presionar el botón con el triángulo, de la misma forma como se ha hecho anteriormente. En este caso la variable a estudiar será “Renta Líquida”. Posteriormente se debe hacer clic en el botón “Opciones” para escoger los estadísticos que se realizarán.



Una vez dentro del cuadro de diálogo “Opciones” se muestra una serie de análisis posibles. En este caso se seleccionará “Desv. típica” (Desviación Estándar) y “Varianza”, tal como muestra la figura lateral.

Posteriormente se hace clic en el botón “Continuar” y finalmente en el botón “Aceptar” para acceder a la ventana de resultados.

**Estadísticos descriptivos**

	N	Desv. típ.	Varianza
Renta Líquida	30	\$44,276.314	2E+009
N válido (según lista)	30		

**Interpretación:**

Los salarios poseen una desviación estándar de \$44.276.

**Análisis de Respuestas de Selección Múltiple:**

Es común que las investigaciones de mercado o estudios donde se apliquen cuestionarios, posean respuestas de selección múltiple, o en otras palabras, aquellas respuestas donde el encuestado puede responder con más de una opción. Para estos casos SPSS requiere ser trabajado de manera especial, tanto en el ingreso de este tipo de preguntas al Software, como para su posterior análisis.

Para ejemplificar la forma de trabajar este tipo de preguntas se supondrá que una empresa “X” desea ofrecer nuevos servicios a sus clientes. Para ello realizan un pequeño cuestionario a 30 personas, donde se les pide que seleccionen las alternativas por las que estarían dispuestos a pagar. A continuación se muestra las alternativas:

**Seleccione las alternativas por las que estaría dispuesto a pagar (Puede seleccionar más de una):**

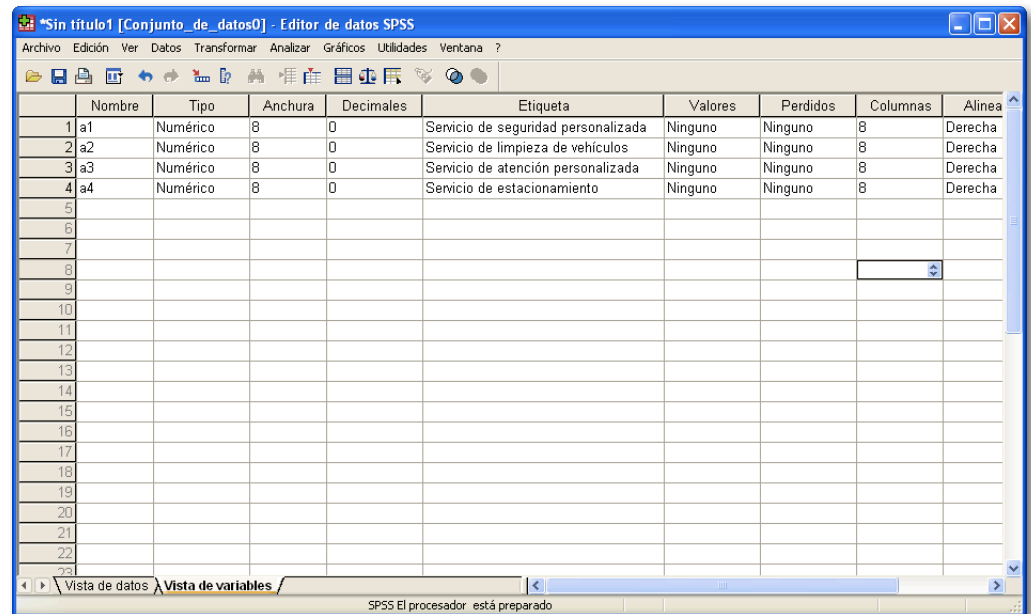
- Servicio de seguridad personalizada.
- Servicio de limpieza de vehículos.
- Servicio de atención personalizada.
- Servicio de estacionamiento.

**Ingreso de datos a la Base de Datos:**

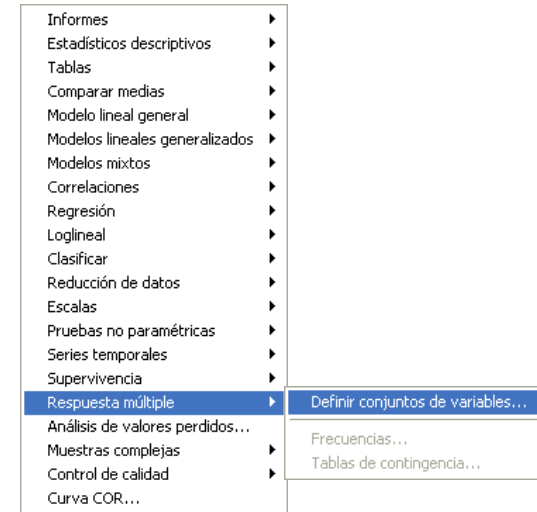
A diferencia de los análisis comunes, cuando se trabaja con respuestas de selección múltiple hay que adaptar la base de datos.

Lo primero que se debe saber, es que cada alternativa debe ser ingresada como una variable independiente en SPSS, es decir, cada alternativa debe ser tratada como si se tratara de una pregunta diferente. En el archivo “BaseDatosMult.sav” se encuentra la base de datos ingresada.

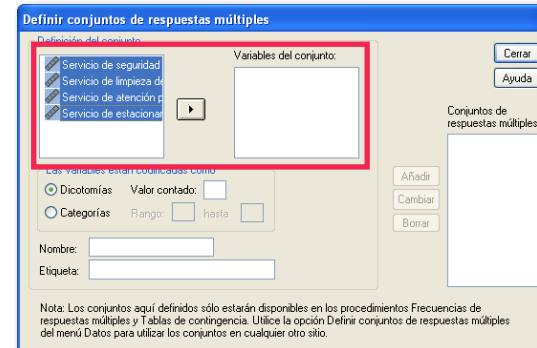
Se debe notar que cuando un encuestado selecciona una alternativa particular, esta debe ser llenada con un valor “1”, mientras que una alternativa que no haya sido seleccionada por el encuestado debe llevar un número “0”. En la base de datos de ejemplo se observa como se debe realizar el procedimiento.



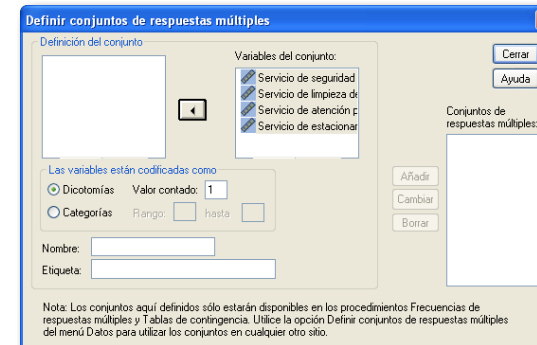
	a1	a2	a3	a4	var	var	var	var	var	var	var	var
1	1	0	1	1								
2	0	1	1	0								
3	1	0	0	0								
4	0	0	0	1								
5	1	0	1	0								
6	1	0	1	1								
7	1	1	0	0								
8	0	0	0	1								
9	0	0	1	0								
10	0	1	1	0								
11	1	1	1	1								
12	0	1	0	0								
13	1	0	0	1								
14	0	1	1	0								
15	0	0	1	0								
16	0	0	0	1								
17	1	1	1	0								
18	0	0	1	0								
19	1	0	1	1								
20	1	0	1	0								
21	0	0	1	0								



1. Lo primero que se debe realizar para definir el conjunto de datos es ir al menú **“Analizar”**, luego posicionarse sobre **“Respuesta múltiple”** y finalmente escoger la opción **“Definir conjuntos de variables...”** (tal como muestra la figura).

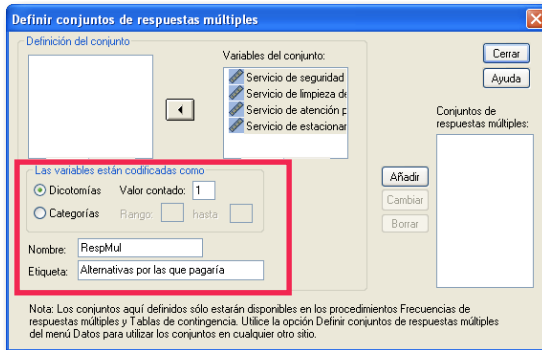


2. Una vez abierto el cuadro de diálogo, se debe escoger las variables que se desee utilizar como alternativas (cada una de las alternativas de la pregunta de selección múltiple y posteriormente hacer clic en el botón con el triángulo negro para que así las variables se ubiquen en la sección de “Variables del Conjunto”, tal como se observa en las figuras de la izquierda.

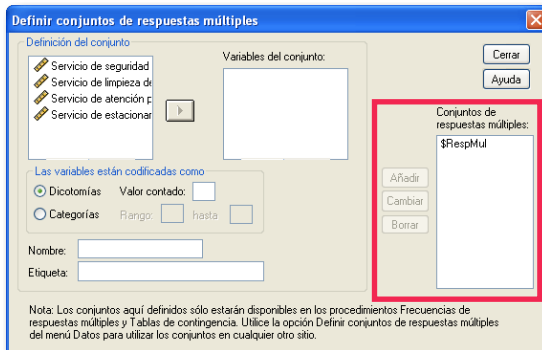


### Definir el conjunto de datos:

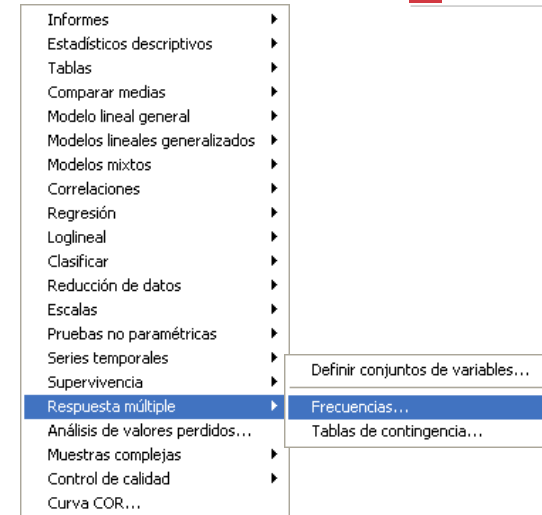
La primera labor que se debe hacer para realizar un análisis de respuestas múltiples es definir el conjunto de datos, lo que consiste básicamente en indicar al software cuales variables corresponden a las alternativas de la pregunta de selección múltiple. El objetivo de esto es que el programa sepa los parámetros sobre los que debe encontrar las respuestas dadas por los encuestados y así evitar malinterpretar los resultados.



4. Posteriormente se debe definir el valor que SPSS identificará como válido para la selección. En este caso será el valor "1", pues con ese se indicó en la base de datos que el encuestado había seleccionado dicha alternativa. Para ello se debe escribir el número "1" donde dice "las variables están codificadas como" y posteriormente se debe agregar un nombre y etiqueta al conjunto de respuesta múltiple.

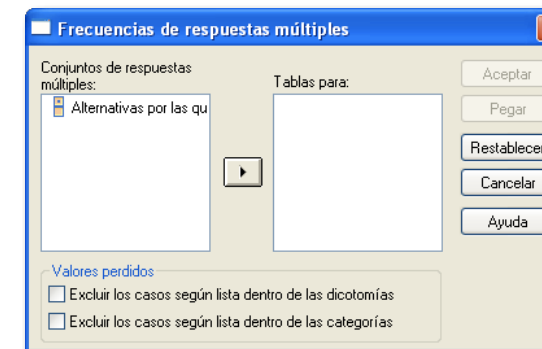


Finalmente se debe hacer clic en el botón "Añadir" y así el conjunto de respuestas estará creado y se podrá realizar los análisis de respuestas múltiples con SPSS



1. Primero se debe ir al menú "Análisis \ Respuesta Múltiple \ Frecuencias...". En caso que se desee realizar una tabla de contingencia, solamente se deberá escoger la segunda opción.

Del uso de tablas de contingencia se hablará en un capítulo posterior.



2. Una vez abierto el cuadro de diálogo se debe escoger el conjunto de respuestas múltiples que se definió anteriormente y posteriormente se debe hacer clic en el botón con el triángulo negro. Una vez listo, se debe presionar el botón "Aceptar" y abrirá una ventana del editor de resultados, con los resultados de los análisis realizados.

### Análisis de conjuntos de Respuestas Múltiples:

Una vez definido el conjunto de datos, es posible realizar los análisis. Resulta necesario mencionar que los análisis que se pueden realizar con este tipo de preguntas son bastante pocos, y se limitan a tablas de contingencia y análisis de frecuencia.

El procedimiento para realizar dichos análisis se describirá a continuación:



**Resumen de los casos**

	Casos					
	Válidos		Perdidos		Total	
	Nº	Porcentaje	Nº	Porcentaje	Nº	Porcentaje
\$RespMul <sup>a</sup>	30	100,0%	0	,0%	30	100,0%

a. Agrupación de dicotomías. Tabulado el valor 1.

El primer cuadro que se muestra es un cuadro resumen que no posee mayor información, salvo indicar que se están analizando 30 datos y que estos equivalen al 100% de la muestra.

**Frecuencias \$RespMul**

		Respuestas		Porcentaje de casos
		Nº	Porcentaje	
Alternativas por las que pagaría	Servicio de seguridad personalizada	14	26,4%	46,7%
	Servicio de limpieza de vehículos	10	18,9%	33,3%
	Servicio de atención personalizada	18	34,0%	60,0%
	Servicio de estacionamiento	11	20,8%	36,7%
Total		53	100,0%	176,7%

a. Agrupación de dicotomías. Tabulado el valor 1.

**Interpretación:**

Se puede deducir de los resultados que el 34% de los encuestados prefieren un servicio de atención personalizada, mientras que un 26,4% pagarían por un servicio de seguridad personalizada, un 20,8% estarían dispuestos a pagar por un servicio de estacionamiento y un 18,9% pagaría por un servicio de limpieza de vehículos.

En la columna "Porcentaje de casos" se indica el porcentaje "Bruto" de cada una de las alternativas, de ahí que sume más del 100%.

# CAPÍTULO 4

## Análisis de Variables Categóricas



## Análisis de Variables Categóricas

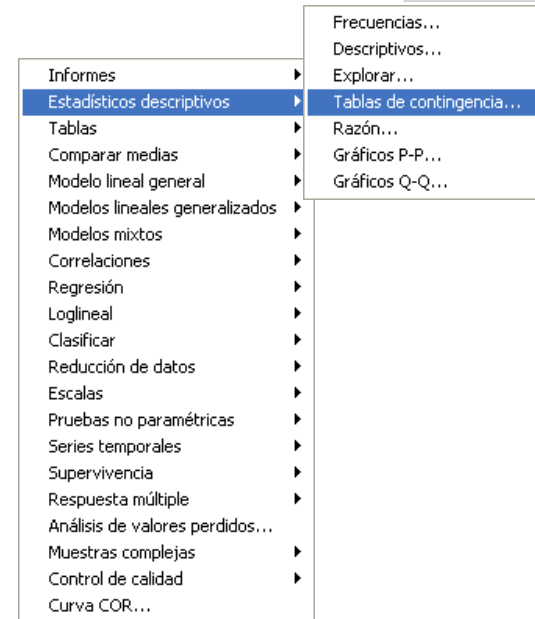
En las investigaciones de mercados y otros trabajos de investigación, en ocasiones, es importante entregar resultados categorizados, es decir, agrupados por alguna categoría o variable que sirva para separar los resultados y dar visiones separadas de estos, como por ejemplo cuando se entregan resultados separados para mujeres y hombres.

Para realizar estas labores SPSS posee una herramienta de análisis que se llama "Tablas de contingencia", que consiste en una tabla que muestra información categorizada de acuerdo a los parámetros que se haya definido previamente.

Para que este tipo de análisis tenga sentido, se necesita al menos de dos variables: una variable que sirva para categorizar y una variable que se desee categorizar. Estas variables deben poseer unidades de medida "Ordinal" o "Nominal", por lo que utilizar variables con medida de Escala no tiene mayor sentido y no ofrecería resultados de los que se pueda extraer mucha información.

### Ejemplo Práctico:

Se utilizará la base de datos "BaseDatosC1a.sav" para categorizar la edad de los encuestados de acuerdo al sexo de estos. Es decir, se buscará definir cuantas personas de cada grupo de edad corresponden a sexo masculino y cuantas al sexo femenino.



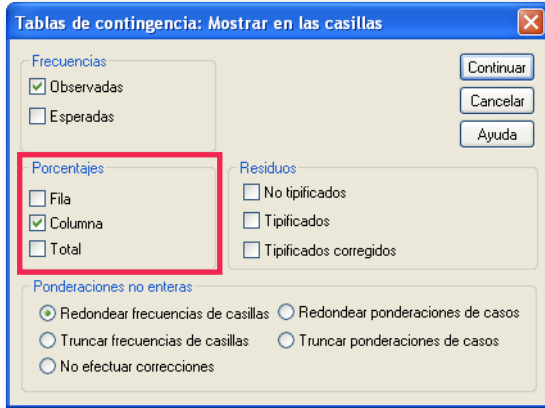
1. Al igual que la mayoría de los análisis, se debe ir al menú "Analizar" y posteriormente a "Estadísticos descriptivos", donde se debe escoger la opción "Tablas de contingencia..."



2. Posteriormente se debe escoger las variables que se desee utilizar. En este caso se utilizarán las variables: "Sexo" y "Edad".

Se debe ubicar una de las variables en la casilla "Filas" y la otra en la casilla "Columnas". Esta ubicación dependerá de la forma en que al investigador resulte más cómodo que se desplieguen los análisis en la tabla de contingencia, por lo que los resultados no variarán si se ubica en otra dirección.

Para mejorar los resultados se recomienda hacer clic en el botón: "Casillas"



3. Dentro de las opciones para mostrar en las casillas, se recomienda al investigador utilizar porcentajes, ya que la tabla de contingencia por si sola, muestra únicamente las frecuencias de los datos, lo que en ocasiones puede resultar insuficiente.

Para mostrar los porcentajes de las frecuencias en la tabla solamente se debe marcar las casillas respectivas. En este caso se seleccionó la opción "Columna" pues se desea que los porcentajes sean mostrados para cada una de las columnas.

Finalmente se debe hacer clic en "Continuar" y posteriormente en "Aceptar" para visualizar los resultados.

**Tabla de contingencia Sexo \* Edad**

		Edad			Total
		18 a 29	30 a 49	50 o más años	
Sexo	Masculino	Recuento 4	8	1	13
		% de Edad 25,0%	72,7%	33,3%	43,3%
	Femenino	Recuento 12	3	2	17
		% de Edad 75,0%	27,3%	66,7%	56,7%
Total		Recuento 16	11	3	30
		% de Edad 100,0%	100,0%	100,0%	100,0%

**Interpretación:**

Como se puede observar en la tabla, dentro de los 16 encuestados del rango de 18 a 29 años, el 75% corresponde a mujeres, mientras que el 25% corresponde a hombres. En el rango de 30 a 49 años sucede lo contrario: El 72,7% de los encuestados son varones y el 27,3% son mujeres, mientras que de los encuestados mayores de 50 años, el 66,7% corresponde a personas del sexo femenino, mientras que el 33,3% corresponde a hombres.

**Resumen del procesamiento de los casos**

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Sexo * Edad	30	100,0%	0	,0%	30	100,0%

Al igual que con cualquier resultado, SPSS muestra un resumen del procesamiento, en el que se indica que se utilizaron 30 casos para el análisis y que estos corresponden al 100% de los datos incluidos.

# CAPÍTULO 5

## Pruebas de Hipótesis



## Pruebas de Hipótesis

Resulta común en el día a día, escuchar afirmaciones que dicen cosas como “Las mujeres ganan menos que los hombres”, “A mayor nivel de educación, mayor nivel de ingresos” y asuntos por el estilo.

Si bien es cierto que la lógica o el sentido común podían indicar que efectivamente lo que sucede es así, cuando se desea comprobar de manera empírica estas afirmaciones es necesario utilizar técnicas estadísticas que apoyen o refuten las teorías del investigador, de manera de tener certeza (desde el punto de vista estadístico), sobre las afirmaciones realizadas.

Para poder realizar este tipo de análisis, SPSS presenta algunas herramientas, sin embargo, se considera prudente interiorizar al lector en el concepto de pruebas de hipótesis, de manera superficial, con la finalidad de comprender su finalidad.

### Pruebas de Hipótesis:

En términos generales, las pruebas de hipótesis se utilizan cuando se quiere determinar, a través de una muestra, que la totalidad de la población posee una característica determinada, además de aquellas ocasiones en que esto no es así.

La prueba de hipótesis consta de dos afirmaciones: Una hipótesis del investigador (que se identifica como  $H_1$ ) y una hipótesis nula (identificada como  $H_0$ ). Cada una de estas afirmaciones puede escribir en términos normales, pero también se debe identificar como notación matemática. Por ejemplo, si deseamos determinar que el promedio de las edad de las mujeres es menor que el de los hombres, se deberá expresar de la siguiente manera:

$H_0$ : El promedio de edad de las mujeres es igual al de los hombres.  
 $H_1$ : El promedio de edad de las mujeres es menor al de los hombres.

O en notación matemática:

$$H_0 : \bar{x}_{edad\ mujeres} = \bar{x}_{edad\ hom\ bres}$$

$$H_1 : \bar{x}_{edad\ mujeres} < \bar{x}_{edad\ hom\ bres}$$

### Valor p:

Una vez establecidas las hipótesis, es necesario conocer el método por el cual el investigador escogerá entre una hipótesis y la otra. Para ello, de manera general, se utiliza un estadístico llamado “valor p”.

Este estadístico indica la probabilidad que existe de rechazar la hipótesis del investigador, cuando esta es verdadera. Por convención, se asume que si esta probabilidad es más de un 5% (0,05) debemos rechazar la hipótesis del investigador ( $H_1$ ) en favor de la hipótesis nula ( $H_0$ ). Entonces:

Criterio	Opción
Si, Valor p < 0,05	Se debe aceptar $H_1$ (Hipótesis del investigador)
Si, Valor p > 0,05	Se debe aceptar $H_0$ (Hipótesis nula)

### Ejemplo Práctico:

La empresa “Doble Equis” está enfrentando una seria disputa entre su personal. Por una parte los profesionales pelean porque de acuerdo a su percepción, el salario recibido por las mujeres es menor que el salario recibido por los hombres.

Además todos reclaman que el salario promedio de la empresa es inferior al salario de mercado, que ronda los \$ 520.000 para empresas similares.

Finalmente, todos concuerdan que en los últimos años el salario no ha aumentado, en promedio, de manera considerable, por lo que piden que se reajuste a todos los trabajadores su nivel de renta.

La empresa decide tomar cartas en el asunto y genera una base de datos (BaseDatosC5.sav) en la que incluye información como: Sexo, Salario inicial, Salario actual y años de escolaridad para cada uno de sus 35 trabajadores.

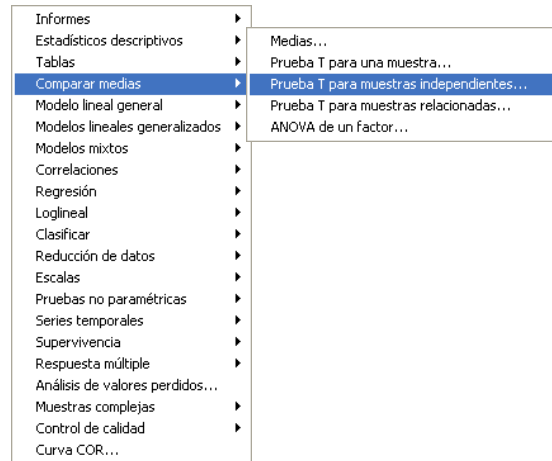
Se le pide analizar la base de datos y determinar la veracidad de cada una de las situaciones anteriormente descritas.

**Prueba de medias para muestras independientes:**

La primera situación a la que se enfrenta el investigador es a determinar si efectivamente existe discriminación en el salario entre las mujeres y los hombres. Para ello se debe realizar una prueba de medias para muestras independientes. A continuación se explica como realizarlo con SPSS, pero primero se debe establecer la prueba de hipótesis:

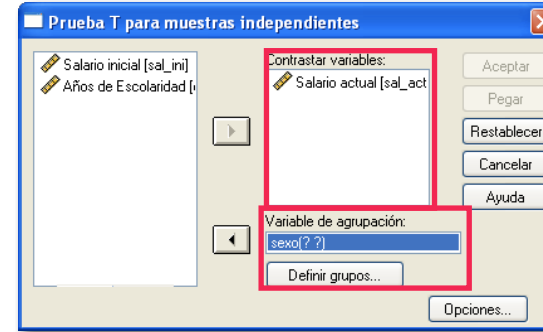
- H<sub>0</sub>: En promedio, el salario de las mujeres es igual al salario de los hombres.**
- H<sub>1</sub>: En promedio, el salario de las mujeres es menor al salario de los hombres.**

De lo anterior, será responsabilidad del investigador, determinar la veracidad de la hipótesis H<sub>1</sub>.



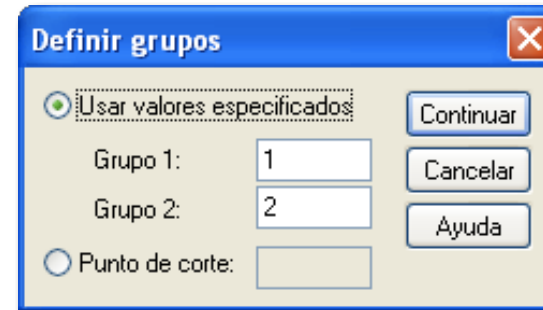
1. Lo primero que se debe hacer es ir al menú **"Análisis"**, posteriormente ir al apartado **"Comparar medias"** y finalmente escoger **"Prueba T para muestras independientes..."**.

Esto abrirá una ventana de diálogo donde se deberán definir los parámetros para la realización de la prueba de contraste de medias.



2. Una vez en el cuadro de diálogo, se debe escoger la variable que se desea medir, en este caso "Salario actual" y debe ser llevada hasta donde dice "Contrastar variables", mientras que la variable por la que se desee agrupar el estadístico (en este caso, el sexo del trabajador) debe ir donde dice "Variable de agrupación".

Posteriormente se deben definir los grupos. Para ello hay que hacer clic en el botón "Definir grupos"



3. Al hacer clic en el botón "Definir Grupos", se abrirá una pequeña ventana como la que se ve a la izquierda. En ella se debe ingresar los valores correspondientes a las etiquetas de sexo. De acuerdo a la información de la base de datos: 1=Masculino y 2=Femenino. A continuación hay que escribir los números en las casillas que se denominan "Grupo 1" y "Grupo 2", tal cual se observa en la figura. Finalmente se debe hacer clic en el botón "Continuar" y para finalizar, en el botón "Aceptar".

Estadísticos de grupo

	Sexo	N	Media	Desviación tip.	Error típ. de la media
Salario actual	Masculino	18	\$488,888.89	\$55,401.092	\$13,058.163
	Femenino	17	\$480,000.00	\$46,502.688	\$11,278.559

**Interpretación:**

Los resultados indican que el salario promedio de los hombres es de \$488.888, mientras que el salario de las mujeres es de \$480.000. Como primera aproximación se observa una diferencia de salario a favor de los hombres de \$8.000 en promedio. Sin embargo esto necesita ser contrastado por pruebas más específicas.

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error tip. de la diferencia	95% Intervalo de confianza para la diferencia	
									Inferior	Superior
Salario actual	Se han asumido varianzas iguales	2,489	,124	,513	33	,612	\$8,888.889	\$17,342.855	-\$26,395.416	\$44,173.193
	No se han asumido varianzas iguales			,515	32,568	,610	\$8,888.889	\$17,254.608	-\$26,233.588	\$44,011.366

**Interpretación:**

SPSS utiliza el criterio del valor p para determinar la validez de los resultados de la prueba de hipótesis. Sin embargo, en este software, el valor p, viene escrito como “Sig (bilateral)”. Por ende, para medir el valor p, solamente se debe ver donde dice “Sig (bilateral)”.

Utilizando el criterio del valor p, se puede afirmar que, estadísticamente, no hay diferencias entre el salario de las mujeres en relación al de los hombres (valor p = 0,612), por lo que esa acusación resulta falsa (Se debe aceptar H<sub>0</sub>).

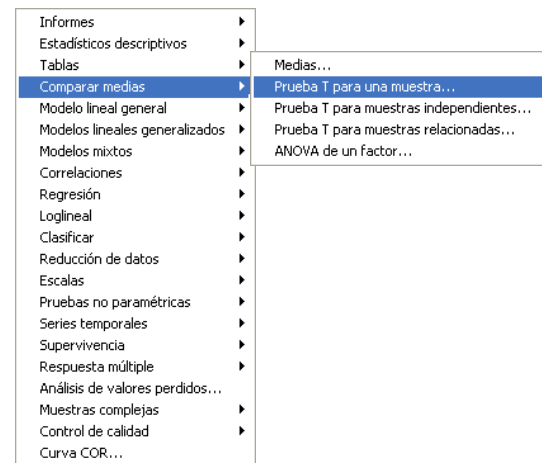
**Prueba de medias para una muestra:**

La segunda situación que debe resolver el investigador es si efectivamente la empresa ofrece un salario promedio inferior al del mercado, que ronda los \$520.000. Lo primero que se debe realizar es establecer la prueba de hipótesis correspondiente:

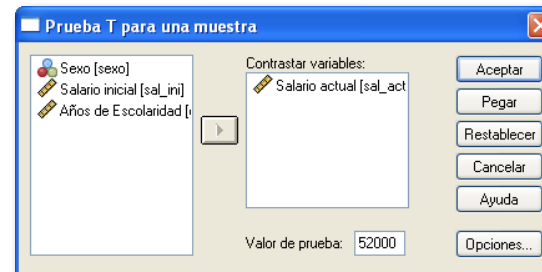
**H<sub>0</sub>: El salario promedio pagado por la empresa es igual al salario de mercado (\$520.000).**

**H<sub>1</sub>: El salario promedio pagado por la empresa es menor que el salario de mercado (\$520.000).**

Para determinar cual hipótesis es la correcta, el procedimiento a seguir es el siguiente:



1. Para realizar este análisis se debe ir al menú “**Analizar**” posteriormente seleccionar donde dice “**Comparar medias**” y finalmente escoger “**Prueba T para una muestra**”.



2. Una vez abierta la ventana, se debe escoger la variable que se desea contrastar (en este caso “Salario actual” y se debe llevar hacia el cuadro. Donde dice “Valor de prueba” se debe ingresar el valor contra el que se quiere contrastar (en este caso el salario de mercado de \$ 520.000).



Estadísticos para una muestra

	N	Media	Desviación tjp.	Error tjp. de la media
Salario actual	35	\$484,571.43	\$50,720.854	\$8,573.389

Interpretación:

La primera parte de los resultados indican que el salario promedio pagado por la empresa es de \$ 484.571, lo que está indicando que en términos brutos, el salario efectivamente es más bajo que el de mercado.

Prueba para una muestra

	Valor de prueba = 520000					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
Salario actual	-4,132	34	,000	-\$35428.57	-\$52,851.79	-\$18,005.35

Interpretación:

Utilizando el criterio del valor p, y dado que el valor entregado por el software es de "0,000" se puede asumir que el investigador tiene razón al afirmar que la empresa paga un salario más bajo que el promedio de mercado.

Recordemos que cuando el valor p es menor que 0,05 se puede aceptar como verdadera la hipótesis del investigador (H<sub>1</sub>).

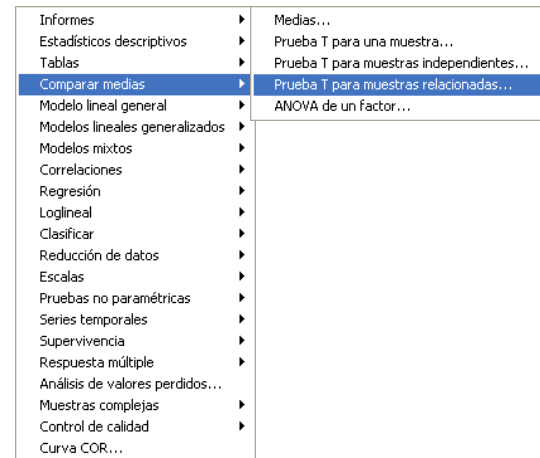
Prueba de medias para muestras relacionadas:

La última situación presentada por los trabajadores de la empresa indica que en promedio sus sueldos no han subido significativamente conforme ha pasado el tiempo. Para resolver lo anterior, el investigador debe realizar una prueba que se denomina: "Prueba T para muestras relacionadas".

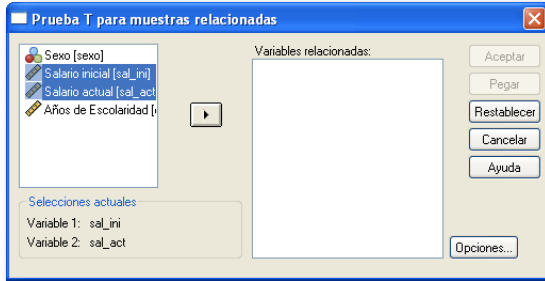
Las hipótesis detrás de esta prueba son las siguientes:

- H<sub>0</sub>: El salario promedio actual de la empresa es igual al salario inicial.**
- H<sub>1</sub>: El salario promedio actual de la empresa es mayor al salario inicial.**

Para determinar cual hipótesis es la correcta, el procedimiento a seguir es el siguiente:



1. Primero hay que ir al menú "Análisis", posteriormente ir al submenú "Comparar medias" y ahí escoger "Prueba T para muestras relacionadas..." tal como muestra la figura de la izquierda.



2. En la ventana que se abre, se debe escoger las variables que se desea contrastar y agregarlas al sector de “Variables seleccionadas”. Para el desarrollo de este ejemplo, las variables son “Salario inicial” y “Salario actual”.

Una vez seleccionadas y ubicadas en la sección “variables Seleccionadas, se debe hacer clic en el botón “Aceptar”.

**Estadísticos de muestras relacionadas**

		Media	N	Desviación tip.	Error típ. de la media
Par 1	Salario inicial	\$473,714.29	35	\$51,454.475	\$8,697.394
	Salario actual	\$484,571.43	35	\$50,720.854	\$8,573.389

**Interpretación:**

Los estadísticos descriptivos indican que inicialmente el salario inicial promedio de los trabajadores era de \$473.715, mientras que el salario promedio actual de los trabajadores es de \$484.571.

**Prueba de muestras relacionadas**

		Diferencias relacionadas				t	gl	Sig. (bilateral)	
		Media	Desviación tip.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
					Inferior	Superior			
Par 1	Salario inicial - Salario actual	-\$10,857.143	\$10,674.717	\$1,804.356	-\$14524	-\$7190.25	-6,017	34	,000

**Interpretación:**

De acuerdo al análisis del valor p, se puede determinar que en promedio, el salario actual de la empresa si es estadísticamente mayor que el salario que se pagaba inicialmente, debido a que como el valor  $p < 0,05$ , se debe aceptar la hipótesis del investigador ( $H_1$ ), que dice que el salario actual es mayor que el salario inicial.

# CAPÍTULO 6

## Análisis de Varianza



## Análisis de Varianza

En el apartado anterior, se introdujo el concepto de pruebas de hipótesis y pruebas de diferencia de tres tipos: para muestras independientes, para una muestra y para muestras relacionadas. Sin embargo, estas pruebas de diferencias de media solo era posible realizarlas entre dos grupos de datos. ¿Qué sucede entonces, cuando se desea contrastar el promedio entre más de dos grupos de datos?. Obviamente, el uso de las técnicas anteriores no logra cumplir ese cometido, por lo que resulta necesario utilizar una técnica diferente, que se denomina Análisis de Varianza (ANOVA, por sus siglas en inglés).

La ANOVA se basa en el análisis de las varianzas de los grupos de datos presentados y se utiliza comúnmente para demostrar que el promedio de un grupo de datos es diferente al de los demás. Por ende la prueba de hipótesis asociada a esta prueba es la siguiente:

$$H_0 : \mu_1 = \mu_2 = \mu_3 \dots = \mu_n = 0$$

$$H_1 : \text{al menos uno es } \neq 0$$

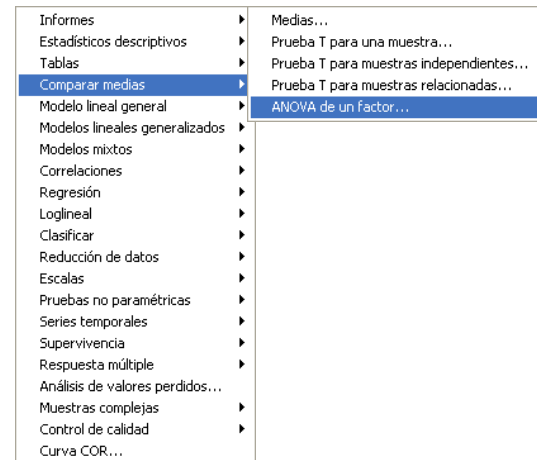
La prueba anterior indica bajo la hipótesis nula, que el promedio para cada uno de los conjuntos de datos es igual entre ellos y a su vez es igual a cero, mientras que la hipótesis del investigador indica que al menos un promedio es distinto de cero.

### Ejemplo Práctico:

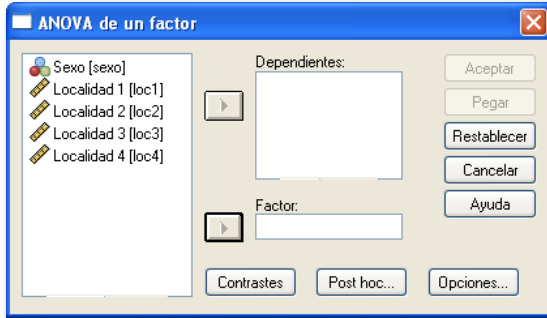
La empresa “OMG” fabrica mesas y sillas, las cuales son iguales para cada lugar al que distribuyen. De acuerdo a los reportes de la fuerza de ventas, existe descontento por parte de los clientes de algunas de las cuatro localidades a las que se distribuye los productos.

Para verificar esta información, la empresa decide preguntar a sus clientes de las cuatro localidades sobre el grado de satisfacción del producto. Para ello se les pide que evalúen el producto con una nota del 1 al 10, permitiendo usar cifras decimales, donde 1 sería muy insatisfactorio y 10 sería completamente satisfecho.

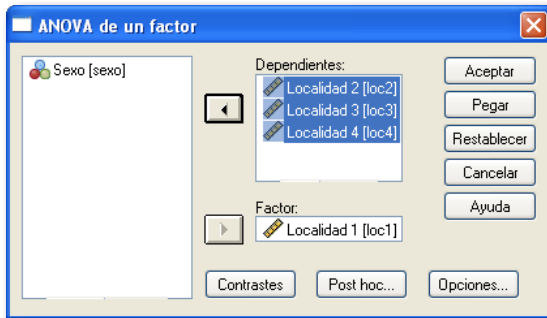
Una vez con la base de datos en su poder (BaseDatosC6.sav), la empresa le pide a usted que determine si efectivamente existen diferencias en la satisfacción de los clientes de la Localidad 1 con respecto a los de las otras localidades.



1. Debido a que el análisis de varianza (ANOVA) es utilizado para contrastar medias, se debe ir al menú “**Analizar**”, luego al apartado “**Comparar medias**” y finalmente se debe escoger la opción “**ANOVA de un factor...**”.



2. Una vez abierto el cuadro de diálogo, se debe escoger las variables dependientes y el factor. Para esclarecer este punto, el “Factor” es la variable que se tomará como referencia para comparar: en este caso, la localidad 1 (de acuerdo a las instrucciones del ejercicio), mientras que las variables dependientes serán: Localidad 2, Localidad 3 y Localidad 4.



La segunda figura muestra como deben ser ubicadas las distintas variables para la realización del análisis.

Una vez listo, se debe hacer clic en el botón “Aceptar”.

**Interpretación:**

De acuerdo a los resultados obtenidos por el análisis ANOVA, no existe evidencia estadística (de acuerdo al criterio del valor p - Sig) para rechazar la hipótesis nula, por ende se puede asegurar que NO existe diferencias en la satisfacción percibida por los clientes de ninguna localidad.

Como se puede observar, los distintos valores p (Sig) que entrega SPSS son todos mayores que el valor de referencia que es de 0,05, por lo que el investigador debe aceptar la hipótesis nula que indica que no hay diferencias entre la satisfacción percibida por el cliente en ninguna de las localidades.

Todo lo anterior derriba la hipótesis de la fuerza de venta que decía que existían diferencias en la percepción de la satisfacción producida por los productos comercializados.

**ANOVA**

		Suma de cuadrados	gl	Media cuadrática	F	Sig.
Localidad 2	Inter-grupos	51,531	16	3,221	1,114	,459
	Intra-grupos	23,124	8	2,890		
	Total	74,654	24			
Localidad 3	Inter-grupos	32,496	16	2,031	1,774	,208
	Intra-grupos	9,158	8	1,145		
	Total	41,654	24			
Localidad 4	Inter-grupos	23,901	16	1,494	,323	,974
	Intra-grupos	36,985	8	4,623		
	Total	60,886	24			

# CAPÍTULO 7

## Gráficos



## Gráficos

En numerosas ocasiones, el investigador se enfrenta con lectores que difícilmente podrán comprender gran parte de los resultados obtenidos en el marco de una investigación. Sin embargo, esto puede mejorar en la medida que el investigador busca nuevas formas de presentar la información. Una de estas formas es la utilización de gráficos, los que resultan un complemento visual importante para el desarrollo de una correcta investigación y ofrecen herramientas que permiten al lector hacerse una mejor idea de lo que visualizarían si solamente mirasen los resultados numéricos.

Si bien es cierto que hay numerosos estilos de gráficos que se pueden utilizar, en este apartado solamente se utilizarán los más comunes: Gráficos de torta, Gráficos de Barra, Gráficos de dispersión (diagramas de dispersión).

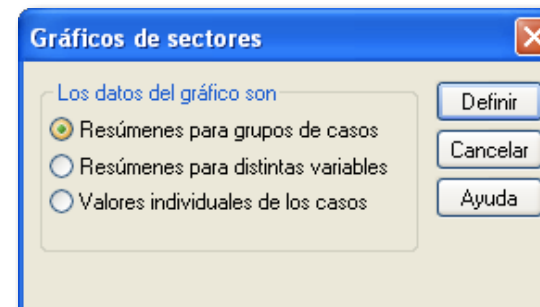
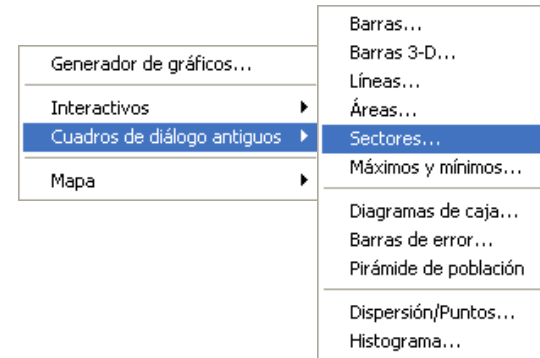
### Gráficos de Torta:

Este tipo de gráficos es muy común para visualizar información de carácter cualitativo y categórico, ya que ofrece una forma muy gráfica para visualizar el peso (o cuantía) de algún aspecto de una variable en relación a otros.

A modo práctico se utilizará la base de datos del capítulo 5 (BaseDatosC5.sav).

En este caso, se realizará un gráfico de torta para representar la segmentación de los encuestados según su sexo (algo muy común en cualquier estudio).

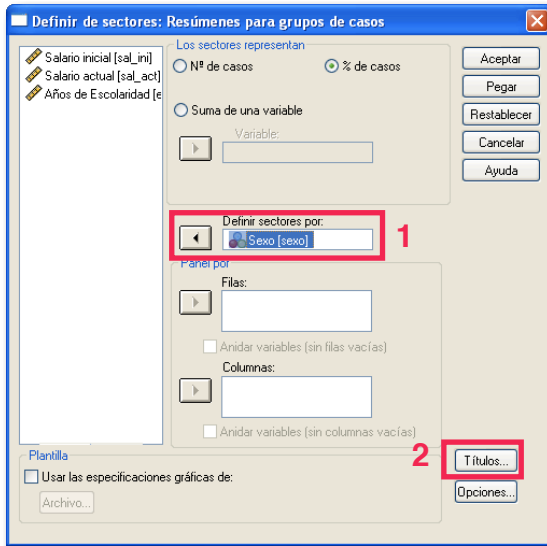
Es importante mencionar que para que este tipo de gráficos tenga sentido, la unidad de medida de los datos debe ser "Nominal" u "Ordinal". Con variables medidas en Escala no tiene mucho sentido realizar este tipo de gráficos, pues entregarán un dibujo muy complejo de leer y sin lógica aparente.



1. Para generar un gráfico cualquiera, se debe ir al Menú "Gráficos" (Al lado del menú "Analizar") y se debe escoger alguna de las opciones. Se recomienda que se escoja la opción "Cuadros de diálogo antiguos" (por asuntos de comodidad). Posteriormente se debe escoger el tipo de gráfico que se desea, en este caso, un gráfico de "Sectores..." (que es comúnmente llamado "Gráfico de Torta").

2. Posteriormente ofrece la opción de como se desea obtener los datos desde la base de datos. Como este es un manual de orden básico, se recomienda dejar tal cual lo ofrece el software, es decir, se debe seleccionar la opción "Resúmenes para grupos de casos".

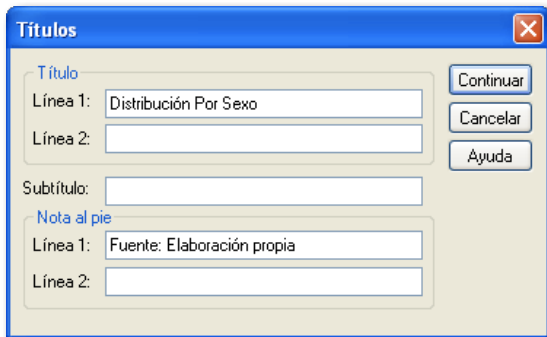
Con lo anterior, el investigador se asegura que se utilizará todos los datos de la base de datos de acuerdo a los grupos en los que se encuentren definidos.



3. Una vez abierto el cuadro de diálogo, se debe definir los sectores del gráfico. Para ello, se debe seleccionar la variable que se desee graficar (en este caso "Sexo") y se debe ubicar en el sector que dice: "Definir sectores por:"

Con esto, el investigador le está diciendo al software que defina los sectores del gráfico de acuerdo a las etiquetas de la variable seleccionada.

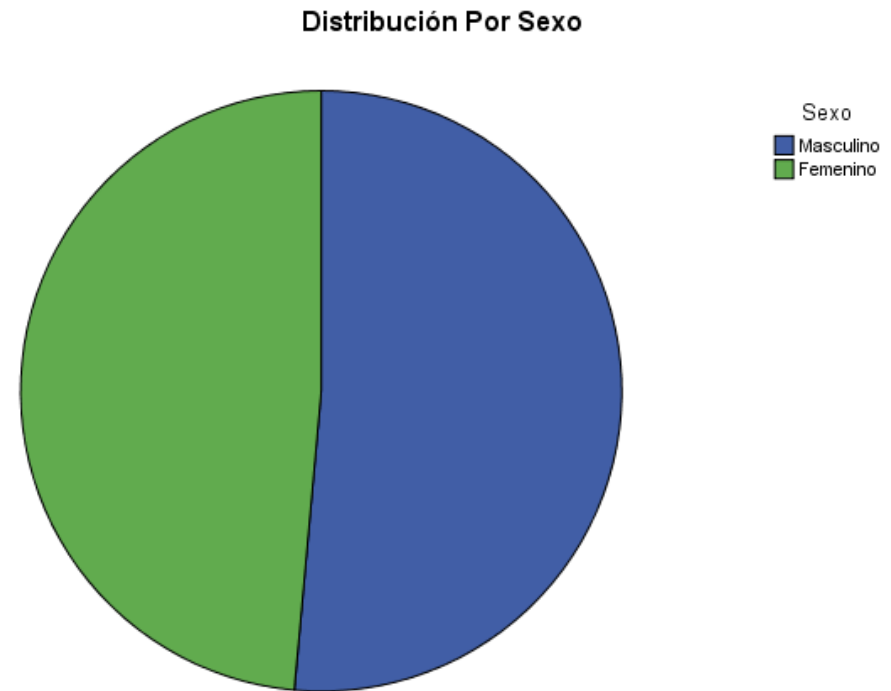
Posteriormente se recomienda agregar títulos al gráfico. Para ello se debe hacer clic en el botón "Títulos...".



4. En el cuadro de diálogo de títulos, se debe agregar el título principal (Línea 1 y Línea 2), el Subtítulo y las notas al pie (Línea 1 y Línea 2 de la zona inferior). Sin embargo es posible que el investigador decida solo colocar un título y dejar el resto en blanco, o simplemente colocar nada.

No es obligación llenar cada uno de los espacios disponibles en esta ventana.

Finalmente se debe hacer clic en el botón "Continuar" y posteriormente en el botón "Aceptar".



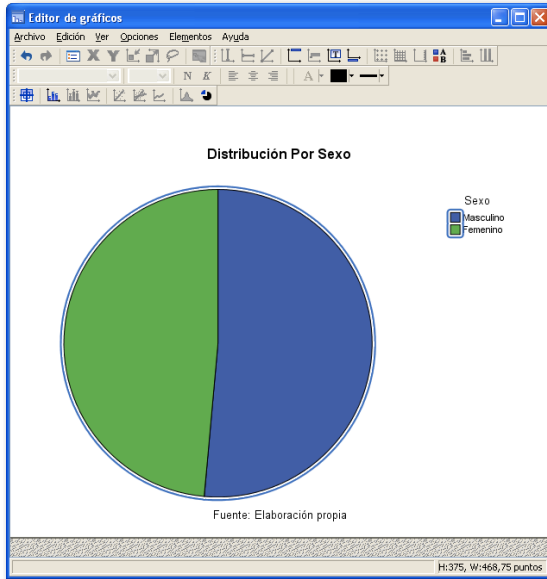
Fuente: Elaboración propia

Una vez hecho clic en el botón aceptar, el software abrirá una ventana de resultados con un gráfico similar al de la imagen superior, en el cual se puede observar el título y la nota al pie ingresada, y la representación gráfica de la distribución por sexo de los encuestados.

No obstante lo anterior, el gráfico aún se puede encontrar muy incompleto o falto de identificadores. Para ello SPSS ofrece herramientas para editar gráficos. Solo basta con hacer doble clic sobre el gráfico en el visor de resultados y abrirá una nueva ventana con el editor de gráficos, donde el investigador podrá editar el gráfico a gusto.

Como no es el objetivo de este manual, queda a criterio del investigador aprender a utilizar dicho editor de la forma que estime conveniente. Más adelante se ofrece una vista a la ventana del editor de gráficos.

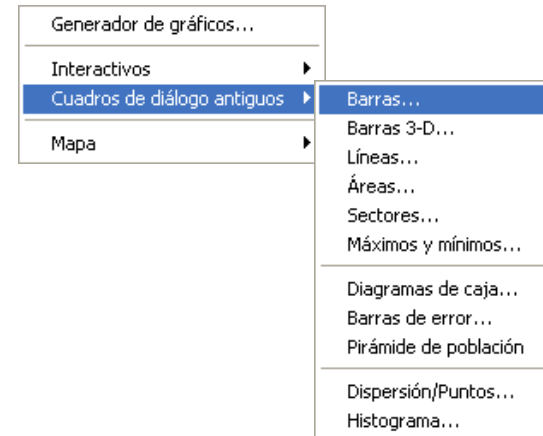




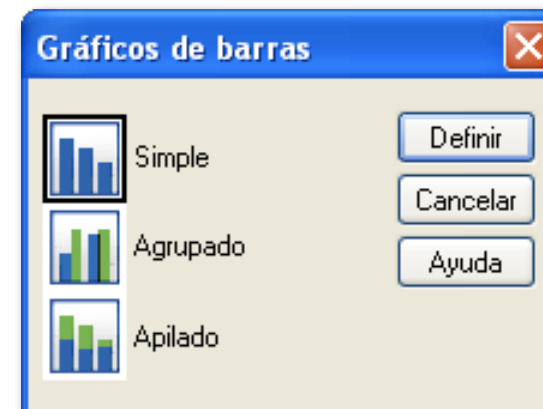
La imagen de la izquierda muestra la composición del editor de gráficos provisto por SPSS. En el, el investigador podrá cambiar los colores, formas, textos, etc, del gráfico, de manera de acomodarlo a las necesidades de su investigación.

### Ejemplo Gráfico de Barra Simple:

Este es el tipo más común de gráfico de barra, en que solamente se grafica una variable. En este caso se graficará la variable “Rango de edad”, de manera de observar la forma en que este tipo de gráfico despliega la información:



1. Al igual que con el gráfico de sectores, para realizar este, es necesario ir al menú “**Gráficos**”, posteriormente se recomienda elegir la opción “**Cuadros de diálogo antiguos**” y dentro de ella, seleccionar donde dice “**Barras**” para poder dibujar un gráfico de barras con la información disponible.



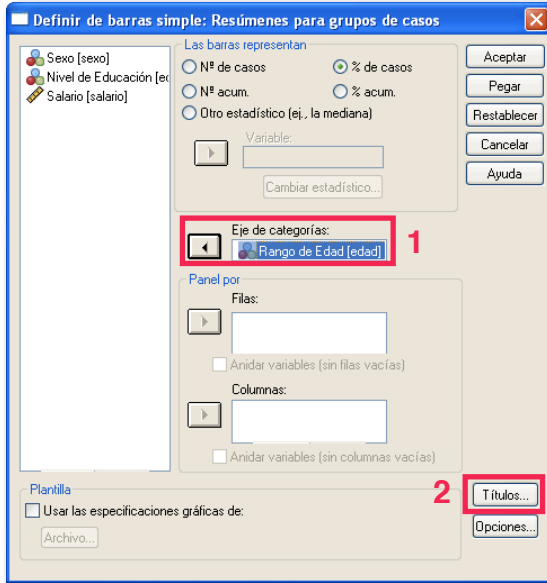
2. Nuevamente se abre un cuadro de diálogo para definir que tipo de gráfico se quiere realizar. En este caso será un gráfico “simple” pues solamente considera una variable. Más adelante se realizará este procedimiento para un gráfico agrupado.

Una vez definido el gráfico “Simple” hay que hacer clic en el botón “Definir”.

### Gráficos de Barra:

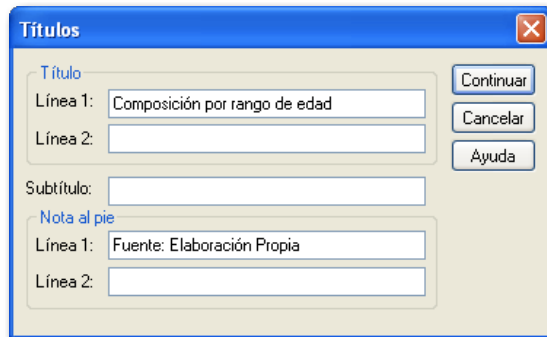
Los gráficos de barra son, junto con los de torta, los más utilizados al momento de desplegar información gráficamente en un estudio. Al igual que los gráficos de torta, generalmente tienen sentido solamente con variables medidas en unidades nominales u ordinales.

Para ejemplificar la utilización de este tipo de gráficos, se tiene una base de datos que contiene la siguiente información: Sexo del encuestado, Rango de edad, Nivel educacional y salario. La base de datos está disponible en el archivo “BaseDatosC7.sav”.



3. En la ventana que abre, se debe especificar a SPSS la variable que se desea graficar. Para ello se debe situar la variable que se desee donde dice “Eje de categorías” (que corresponde al “piso” del gráfico de barras. En este caso se selecciona la variable “Rango de Edad”.

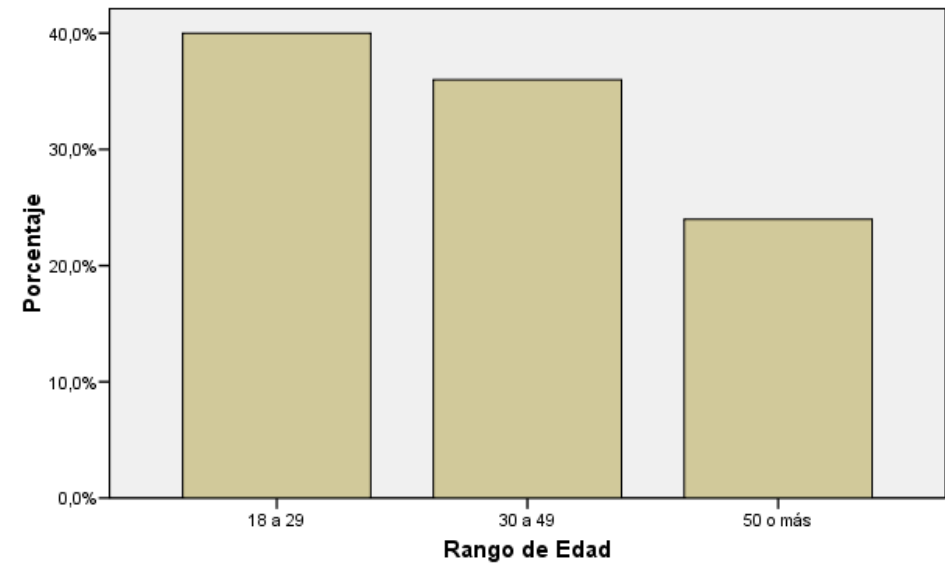
Posteriormente se recomienda hacer clic en el botón “Títulos” y establecer los títulos necesarios para el gráfico



4. En el cuadro de diálogo de los Títulos, se debe (al igual que en el caso anterior) especificar los títulos más idóneos para definir la variable. Esta decisión de títulos corresponde plenamente al investigador.

Cuando esté listo, se debe hacer clic en el botón “Continuar” y para finalizar hay que hacer clic en el botón “Aceptar”

**Composición por rango de edad**



Fuente: Elaboración Propia

El resultado del gráfico de barras desplegado es similar al que se muestra en la parte superior. En dicho gráfico se observan columnas que representan a los 3 grupos de edad definidos en la base de datos y hacia arriba se puede observar el porcentaje de participación de cada una de ellas.

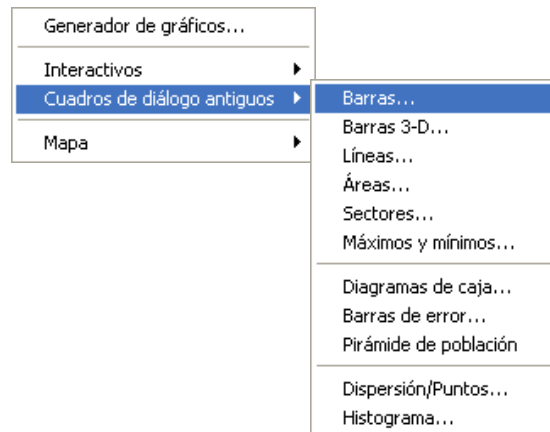
De lo anterior se puede desprender que cerca del 40% de los encuestados posee entre 18 y 29 años, una cifra cercana al 35% de los encuestados están en el rango de 30 a 49 años. mientras que los encuestados con 50 o más años están en el rango del 25%.

Si el investigador desea editar el gráfico, los colores u otros aspectos, solamente debe hacer doble clic sobre el gráfico y SPSS abrirá una ventana con el editor de gráficos, donde podrá hacer algunas correcciones a este.

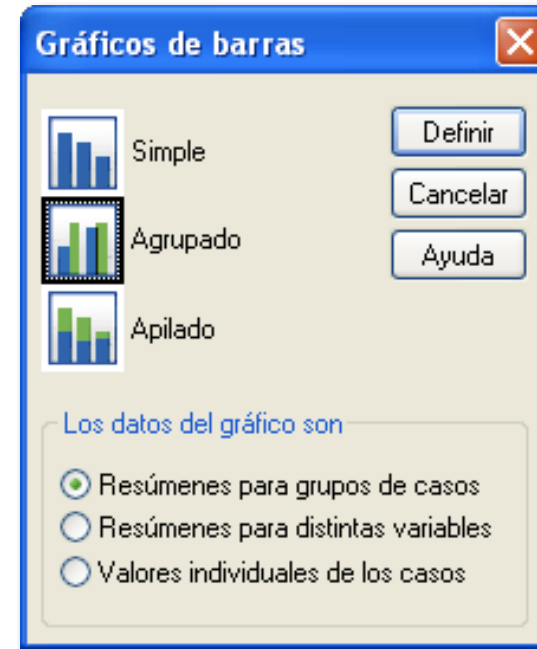
### Ejemplo Gráfico de Barra Agrupado:

Este es un tipo de gráfico de barras en que la variable ubicada en el eje de categorías (en el ejemplo la variable "Rango de edad") es categorizada por otra variable, de manera de entregar resultados segmentados por otra variable que sirva como segmentadora. Todo esto dentro del mismo gráfico.

Utilizando la base de datos "BaseDatosC7.sav" se pretenderá elaborar un gráfico de barras agrupado, en que se categorizará la variable "Rango de edad" de acuerdo al "nivel de escolaridad" alcanzado. esto proporcionará al lector una idea sobre el nivel de escolaridad existente para cada uno de los rangos de edad que se utilizaron en el estudio.

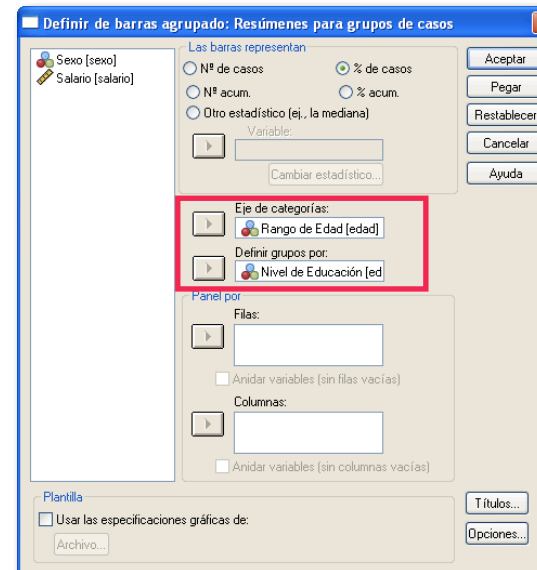


1. Al tratarse de un gráfico de barras, hay que ir al menú "Gráficos" y ahí escoger la opción "Cuadro de diálogo antiguos" y finalmente la opción de "Barras..." tal cual se hizo la vez anterior (ver imagen de la izquierda).



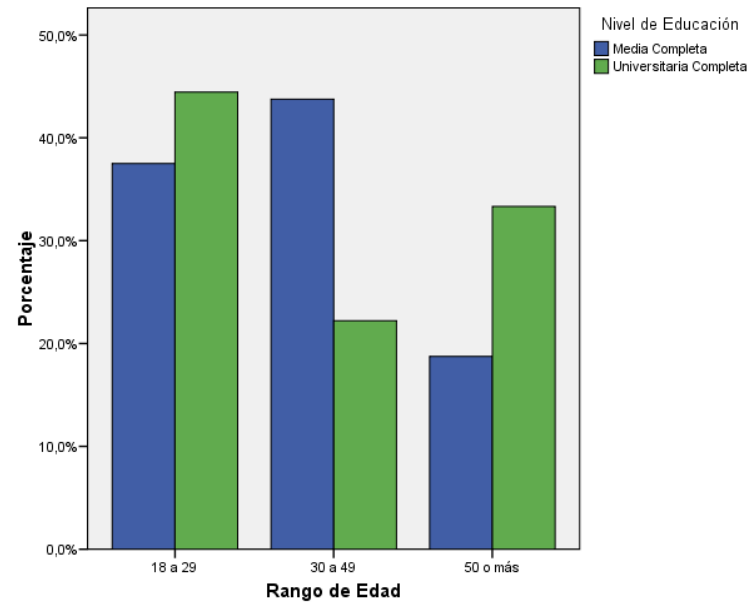
2. En la ventana que se abre a continuación se debe escoger la segunda opción (Agrupado) de manera de generar un gráfico con datos agrupados. Donde dice "Los datos del gráfico son" el investigador debe cerciorarse que la opción "Resúmenes para grupos de casos" esté seleccionada".

Luego que se haga clic en el botón definir, se abrirá una ventana para establecer los parámetros del gráfico.



3. La variable que se desea categorizar (en este caso el rango de edad) se debe colocar donde dice "Eje de Categorías", mientras que la variable por la que se definirán los grupos (Nivel de educación), deberá ser agregada donde dice "Definir grupos por", tal como se observa en la figura de la izquierda.

Además se deben agregar los títulos y una vez listo eso, se debe hacer clic en el botón "Aceptar" para que posteriormente el software nos muestre el gráfico requerido.



**Interpretación:**

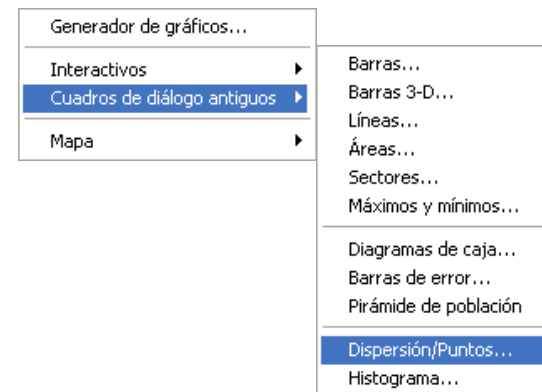
En el rango de edad de 18 a 29 años, se observa que más del 40% de los encuestados posee educación universitaria completa, no así en el rango de 30 a 49 años, en que es aproximadamente de un 20%. En cuanto a los encuestados mayores de 50 años, es posible establecer que cerca del 30% posee educación Universitaria completa.

**Gráfico de dispersión (Tendencia):**

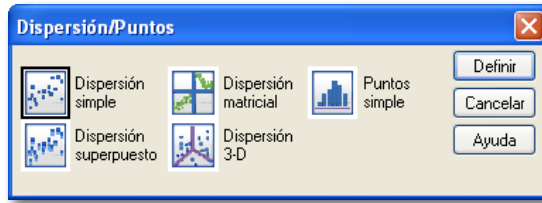
En ocasiones el investigador necesita determinar la tendencia seguida por cierta variable con el paso del tiempo o bajo otras situaciones. Para ello es posible ejemplificar de manera gráfica lo que está ocurriendo, de manera que el lector pueda visualizar de mejor manera la tendencia que existe (o no existe) con determinada variable). Un punto que se debe tener en cuenta con este tipo de gráficos es que se debe trabajar con, al menos, una variable de tipo Escala.

Para ejemplificar este tipo de gráfico, se utilizará la base de datos “BaseDatosC7a.sav”, la que consiste básicamente en una base de datos con meses y las respectivas cantidades vendidas por la empresa durante cada uno de los meses, para su producto “XYZ”.

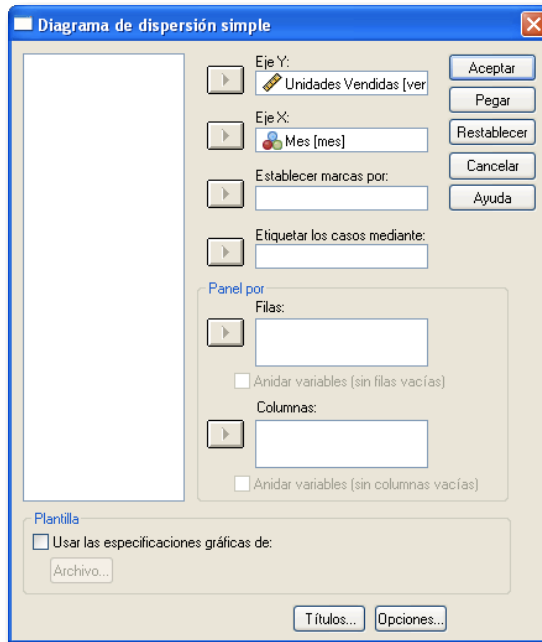
Del gráfico se pretenderá determinar la tendencia de las ventas del producto XYZ.



1. Lo primero que se debe hacer es ir al Menú “**Gráficos**” y una vez ahí situarse donde dice “**Cuadros de diálogo antiguos**” y posteriormente seleccionar la opción “**Dispersión/Puntos**”, que es el tipo de gráfico que se utiliza para visualizar tendencias de variables.



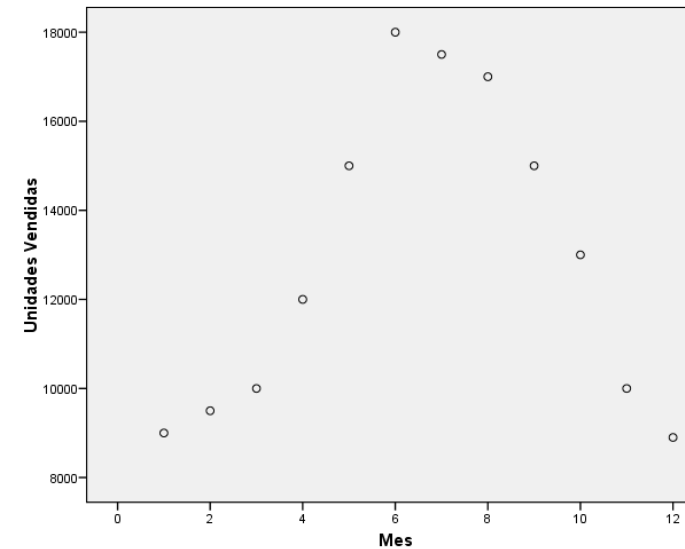
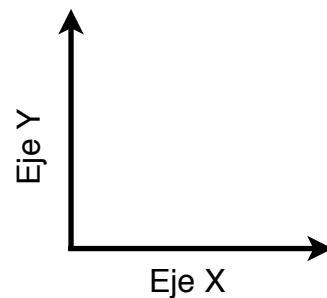
2. Una vez que se muestran las opciones de gráficos, se debe seleccionar la opción “Dispersión Simple”, tal como se observa en la figura y posteriormente se debe hacer clic en el botón “Definir”.



3. Una vez abierto el cuadro de diálogo se debe ingresar las variables. En el “eje X” se debe seleccionar la variable que se desee que se encuentre en la parte inferior del gráfico, mientras que en el “Eje Y”, se debe seleccionar la variable que estará hacia arriba en el gráfico (abajo se observa una figura con la disposición del eje X e Y).

Además se puede agregar títulos al gráfico, tal cual se hace con cualquier tipo de gráficos de SPSS.

Una vez listo lo anterior, se debe hacer clic en el botón “Aceptar” para acceder al gráfico.



**Interpretación:**

Como se puede observar en el gráfico, las ventas comienzan a ascender desde el mes 1 y alcanzan su máximo en el mes de Junio (mes 6), para luego comenzar a decaer nuevamente, hasta llegar a un nuevo mínimo en el mes de Diciembre (12).

# CAPÍTULO 8

## Análisis de Correlación



## Correlación

Con frecuencia el investigador necesita determinar si existe relación entre un conjunto de variables, como por ejemplo, la existencia de una relación entre el salario y los años de estudio de una persona, u otra.

La correlación lineal mide el grado de relación lineal entre dos variables o conjuntos de variables. Los resultados entregan al investigador nociones sobre la dirección de esta relación (positiva, negativa), la fuerza (correlación fuerte, correlación débil, sin correlación).

La medida de correlación más utilizada en investigaciones de mercado es la denominada "Correlación lineal de Pearson", que indica el grado de relación lineal de las variables que se desea medir.

El valor de la correlación fluctúa entre -1 y 1, donde valores cercanos a -1 indican una fuerte correlación negativa, mientras que valores cercanos a 1 indican una fuerte relación positiva. Si el valor de la correlación es cercano a cero, se puede decir que el grado de relación lineal entre las variables es cero o muy débil.

Si bien es criterio del investigador el determinar que tan fuerte es la correlación de las variables, a continuación se dejan rangos de valores que a menudo se utilizan para determinar la fuerza de la relación entre las variables.

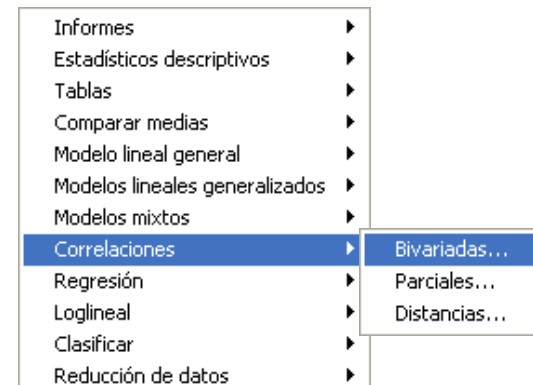
Rango Correlación	Significado
$r = 1$	Correlación Perfecta
$0,8 < r < 1$	Correlación muy Alta
$0,6 < r < 0,8$	Correlación Alta
$0,4 < r < 0,6$	Correlación Moderada

Rango Correlación	Significado
$0,2 < r < 0,4$	Correlación Baja
$0 < r < 0,2$	Correlación Muy Baja
$r = 0$	Correlación Nula.

### Ejemplo Práctico:

Comúnmente se cree que existe una fuerte relación entre los años de estudio de una persona con el salario que recibe en su trabajo. Por lo mismo se le pide a usted que determine si es efectiva esta afirmación y si realmente más años de estudio implican un mejor salario.

Para ello se utilizará una base de datos que contiene información sobre salario actual, salario inicial, años de educación y años de experiencia previa de 30 trabajadores de distintas empresas. La base de datos se encuentra en el archivo "BaseDatosC8.sav"



1. En el menú "Analyze", se debe ubicar el mouse sobre el menú "Correlaciones" y posteriormente se debe escoger el tipo "Bivariadas", que corresponde a correlaciones para dos variables. Una nota importante es que para que este tipo de análisis tenga sentido se debe tratar de utilizar variables con unidad de medida de Escala u Ordinales (aunque este último tipo podría generar problemas).



2. Una vez abierto el cuadro de diálogo, se debe escoger las variables entre las que se desea determinar la correlación: en este caso se trata de las variables: “Años de educación” y “Salario actual”. Una vez seleccionadas las variables, se deben posicionar donde dice “Variables” (como muestra la figura) y se debe cerciorar que se encuentra marcada la casilla “Pearson” dentro de “Coeficientes de correlación”.

Finalmente se debe hacer clic en el botón “Aceptar”.

**Correlaciones**

		Años de Educación	Salario Actual
Años de Educación	Correlación de Pearson	1	,932**
	Sig. (bilateral)		,000
	N	30	30
Salario Actual	Correlación de Pearson	,932**	1
	Sig. (bilateral)	,000	
	N	30	30

\*\* La correlación es significativa al nivel 0,01 (bilateral).

### Interpretación:

Una vez hecho clic en el botón “Aceptar”, se abrirá una ventana del visor de resultados donde se indicará el valor de la correlación de Pearson. En este caso la correlación es de 0,932, lo que significa que existe una relación lineal muy alta entre los años de escolaridad con el salario actual. Es decir, en la medida que se posee una mayor educación (expresada en años), mayor es el salario que se recibirá por parte del trabajador.



# CAPÍTULO 9

Análisis de Regresión Lineal



## Regresión Lineal

Generalmente en investigaciones relacionadas con economía y otras, el investigador necesita determinar si existe alguna relación entre distintas variables, de que tipo es esta relación y si es posible predecir de valor de una de ellas en función de las otras.

Para realizar esto, es que el investigador puede utilizar la herramienta conocida como "Regresión Lineal". El sentido de esta técnica estadística, es el de predecir una medida basándose en el conocimiento de otras.

La regresión lineal consta de dos tipos de variables:

**Variable Dependiente:** Es la variable por predecir (o modelar) y se denota con la letra Y.

**Variable Independiente:** Son las variables que se utilizan para predecir (o modelar) y se denotan con los símbolos  $x_1, x_2, x_3, \dots, x_n$ .

**Error Aleatorio:** Corresponde a las desviaciones de los valores verdaderos de Y con respecto a los valores esperados de Y.

El modelo se plantea de la siguiente manera:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Donde los valores  $\beta_0, \beta_1, \beta_2, \dots, \beta_n$  son llamados parámetros del modelo y  $\varepsilon$  corresponde al término de error aleatorio.

El software SPSS se encarga de entregar los valores para cada uno de los parámetros junto con sus respectivas pruebas de hipótesis, de manera que el investigador puede generar un modelo de regresión lineal y a su vez, verificar la validez de este.

A continuación se explicará brevemente las pruebas de hipótesis de este tipo de modelo estadístico, sin embargo, se recomienda recurrir a algún libro que entregará una idea más completa respecto a este tema.

## Pruebas de Hipótesis:

Para verificar la validez de este tipo de modelos, existen dos pruebas de hipótesis que generalmente se utilizan. La primera busca determinar si el modelo completo tiene sentido o si no funciona para lo que se quiere determinar. Para ello existe una prueba de hipótesis que se basa en el uso del estadístico F de Fisher (el que se obtiene de un análisis de varianza) y que busca determinar si los parámetros en conjunto son iguales o distintos de "cero". La prueba en cuestión es la siguiente:

$$H_0 : \beta_1 = \beta_2 = \beta_n = 0$$

$$H_1 : \beta_j \neq 0 \text{ para algún } j = 1, 2, \dots, n$$

La segunda prueba de hipótesis que se utiliza con este tipo de modelos, busca determinar si alguno de los parámetros de manera individual es estadísticamente distinto de "cero". Para ello se utiliza la siguiente prueba de hipótesis.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 \text{ para } j = 1, 2, \dots, n$$

## R Cuadrado:

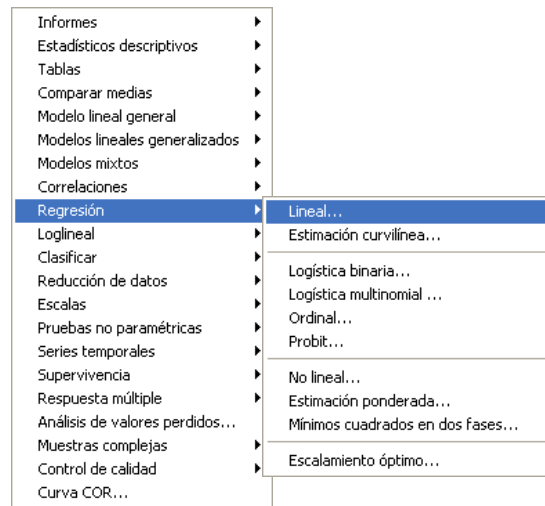
El R cuadrado, o conocido comúnmente como "Coeficiente de determinación" permite determinar el grado de ajuste del modelo, o dicho de otra forma, que tanto de la variable dependiente es explicada por las variables independientes (o por el modelo).

A continuación se presenta un ejemplo práctico acerca de la utilización del modelo.

**Ejemplo Práctico:**

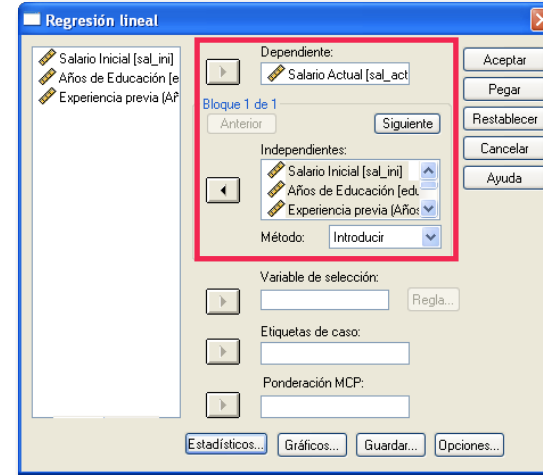
El investigador cree que el salario actual que recibe una persona depende de diversos factores, entre ellos: el salario que recibía en un comienzo, los años de estudio y la experiencia previa del trabajador.

Para ello se genera una base de datos con información de 30 trabajadores para determinar la existencia de esta relación.



1. Para poder generar un modelo de regresión, el investigador debe ir al menú **“Analizar”** y posteriormente ubicarse sobre el menú **“Regresión”** y ahí seleccionar donde dice **“Lineal...”**.

Una vez hecho clic sobre el menú, se abrirá una nueva ventana donde se introducirán las variables.



2. En la ventana que se abre, se debe seleccionar la variable dependiente así como las independientes. En este caso, la variable dependiente corresponde al salario actual, mientras que las variables independientes corresponde a las variables: Salario inicial, Años de educación y Experiencia previa. Una vez seleccionadas y ubicadas como muestra la figura de la izquierda, se debe hacer clic en el botón aceptar, para abrir una nueva ventana con los resultados

**Variables introducidas/eliminadas<sup>b</sup>**

Modelo	Variables introducidas	Variables eliminadas	Método
1	Experiencia previa (Años), Años de Educación, Salario Inicial <sup>a</sup>		Introducir

a. Todas las variables solicitadas introducidas

b. Variable dependiente: Salario Actual

El visor de resultados proporciona al investigador una serie de tablas y cuadros que indican toda la información que se necesita. El primero de ellos (en la parte superior) indica cuales de las variables ingresadas, finalmente han sido introducidas al modelo. En este caso particular, todas las variables fueron ingresadas y además indica que la variable dependiente corresponde al “Salario Actual”:

**Resumen del modelo**

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,995 <sup>a</sup>	,991	,990	11001,269

a. Variables predictoras: (Constante), Experiencia previa (Años), Años de Educación, Salario Inicial

Un segundo cuadro, ofrece al investigador información básica del modelo. En este caso el valor que más resulta útil es el "R cuadrado", que indica básicamente que el modelo explica un 99,1% (0,991) del salario actual. En otras palabras, el salario actual se explica en un 99,1% por el salario inicial, los años de estudio y la experiencia previa.

**ANOVA<sup>b</sup>**

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	3,44E+011	3	1,1E+011	947,215	,000 <sup>a</sup>
	Residual	3,15E+009	26	1,2E+008		
	Total	3,47E+011	29			

a. Variables predictoras: (Constante), Experiencia previa (Años), Años de Educación, Salario Inicial

b. Variable dependiente: Salario Actual

Posteriormente, SPSS ofrece una tabla ANOVA, la que se utiliza para determinar la validez del modelo completo (la prueba F que se mencionó anteriormente). Para ello el investigador debe fijarse en el valor p (Sig.) y como este es menor que 0,05 se puede aceptar la hipótesis del investigador, lo que significa que al menos uno de los parámetros es estadísticamente distinto de "cero", por ende el modelo es válido en su conjunto.

**Coefficientes<sup>a</sup>**

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados		Sig.
		B	Error típ.	Beta	t	
1	(Constante)	-89,956	8568,297		-,010	,992
	Salario Inicial	,863	,048	,858	17,904	,000
	Años de Educación	13732,300	4603,251	,139	2,983	,006
	Experiencia previa (Años)	1341,143	2306,890	,014	,581	,566

a. Variable dependiente: Salario Actual

Finalmente, la tabla que indica los parámetros y las pruebas de hipótesis para ellos. En cuanto a los parámetros, el modelo se puede escribir de la siguiente manera:

$$Y = -89,956 + 0,863X_1 + 13732X_2 + 1341X_3 + \epsilon$$

En otras palabras:

$$\text{Salario Actual} = -89,95 + 0,863 * \text{Salario Inicial} + 13732 * \text{Años Experiencia} + 1341 * \text{Experiencia Previa}.$$

Ahora, revisando las pruebas de hipótesis y mediante el criterio del valor p (Sig.), se puede decir que tanto el salario inicial como los años de educación son estadísticamente distintos de cero (Sig < 0,05), por lo que pueden permanecer en el modelo, mientras que la experiencia previa no es distinta de cero (Sig. > 0,05), por lo que no resulta relevante para explicar el salario actual.

Por lo anterior, el modelo finalmente quedará de la siguiente manera:

$$\text{Salario Actual} = -89,95 + 0,863 * \text{Salario Inicial} + 13732 * \text{Años Experiencia} + \epsilon$$

# CAPÍTULO 10

## Análisis de Regresión Logística



## Regresión Logística (Logit)

La regresión logística resulta útil para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de predictores. Es similar a un modelo de regresión lineal pero está adaptado para modelos en los que la variable dependiente es dicotómica (esto quiere decir que toma los valores 0 ó 1). Los coeficientes de regresión logística pueden utilizarse para estimar la razón de las ventajas de cada variable independiente del modelo.

En su forma funcional, el modelo de regresión logística se presenta de la siguiente manera:

$$\hat{p} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x}}$$

La ecuación anteriormente presentada, representa la probabilidad (estimada) que un sujeto elija la opción 1 (el valor de la variable dicotómica), dado un determinado valor de x.

En este tipo de ecuaciones, también se puede definir la ventaja de elegir la opción 1, por sobre la opción 0, mediante el siguiente cálculo:

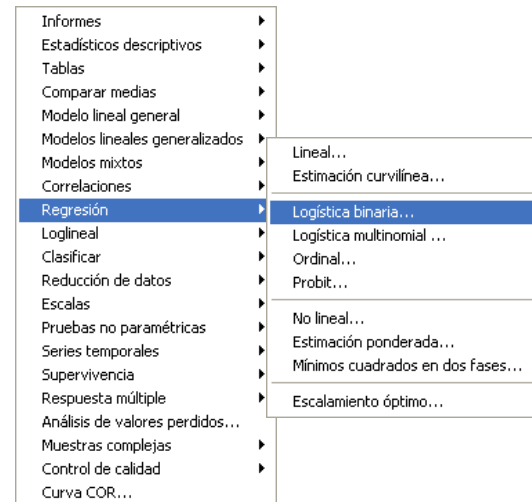
$$\hat{\Omega}_i = \frac{P_i}{1 - P_i} = e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}$$

### Ejemplo Práctico:

Para comprender de mejor manera la utilidad de los modelos de regresión logística, se aplicará el siguiente ejemplo práctico. Si bien es bastante trivial, se pretende dar al lector una visión sobre las posibilidades de este método.

Una entidad financiera prepara el lanzamiento de un nuevo producto. Con el fin de diseñar una adecuada política de promoción, el departamento de marketing estaría interesado en conocer hasta qué punto la aceptación del producto está relacionada con el nivel ingresos de sus clientes. Con esta intención se realiza una encuesta, en la cual la variable dependiente sería: ¿adquiriría usted el producto A?, la que ha sido codificada con el valor 1 en caso de respuesta afirmativa y con valor 0 en caso contrario (pregunta dicotómica). Por otra parte, la variable independiente representa el nivel de ingresos de cada encuestado.

Para efectos prácticos se encuestó a 42 personas y la base de datos se guardó en el archivo "BaseDatosC10.sav".



1. Como el modelo de regresión logística es similar al modelo de regresión lineal (visto en el capítulo anterior), para acceder a él se debe ir al menú **"Análisis"** y posteriormente ir donde dice **"Regresión"** y escoger la opción **"Logística binaria..."** tal cual se ilustra en la imagen de la derecha.



2. Se debe seleccionar las variables que se incluirán en el modelo. En este caso, la variable dependiente (dicotómica) es ¿adquiriría usted el ...?, mientras que las covariables sería el nivel de ingresos, Se deben ubicar en las respectivas casillas, tal cual se muestra en la imagen.

Finalmente se debe hacer clic en el botón "Aceptar". Si no está familiarizado con este modelo, se sugiere no cambiar nada en la ventana.

**Resumen del procesamiento de los casos**

Casos no ponderados <sup>a</sup>		N	Porcentaje
Casos seleccionados	Incluidos en el análisis	42	100,0
	Casos perdidos	0	,0
	Total	42	100,0
Casos no seleccionados		0	,0
Total		42	100,0

a. Si está activada la ponderación, consulte la tabla de clasificación para ver el número total de casos.

**Codificación de la variable dependiente**

Valor original	Valor interno
No	0
Si	1

**Interpretación:**

Cuando se realiza un modelo de regresión logística, SPSS provee un cuadro de resumen en que se indica la cantidad de casos incluidos en el cálculo. En este caso, corresponde a 42, que es el número de personas que contestaron el cuestionario y que además estas 42 personas corresponden al 100% de la muestra. Además muestra un cuadro donde se indica los valores asignados a la variable dependiente (0 y 1).

**Tabla de clasificación<sup>a,b</sup>**

Observado			Pronosticado		Porcentaje correcto
			¿Adquiriría Usted el Producto A?		
			No	Si	
Paso 0	¿Adquiriría Usted el Producto A?	No	0	20	,0
		Si	0	22	100,0
Porcentaje global					52,4

- a. En el modelo se incluye una constante.
- b. El valor de corte es ,500

**Variables en la ecuación**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 0 Constante	,095	,309	,095	1	,758	1,100

**Variables que no están en la ecuación**

				Puntuación	gl	Sig.
Paso 0	Variables	ingresos		22,024	1	,000
Estadísticos globales				22,024	1	,000

**Interpretación:**

Posteriormente, SPSS realiza un modelo en que no se tiene en cuenta la covariable asignada (Nivel de ingresos). En este punto, solamente se está tomando en cuenta la existencia de una variable constante, que es propia de este tipo de modelos.

En la medida que se vayan agregando variables (covariables) a un modelo de regresión logística es que el software irá agregando un paso adicional por cada una de las variables que se agreguen.

En este caso, los valores entregados para la constante son absolutamente inútiles, por lo que habría que dejarlos de lado.

**Pruebas omnibus sobre los coeficientes del modelo**

	Chi-cuadrado	gl	Sig.
Paso 1	29,392	1	,000
Bloque	29,392	1	,000
Modelo	29,392	1	,000

**Interpretación:**

Seguido al análisis con la constante, SPSS comienza a incluir (de una en una) las variables asignadas al modelo. En este caso solamente la variable “ingresos”.

La tabla que se muestra arriba indica (de acuerdo al criterio del valor p) que la variable incluida es válida (Paso), y que además el modelo es válido.

**Resumen de los modelos**

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	28,737 <sup>a</sup>	,503	,672

a. La estimación ha finalizado en el número de iteración 6 porque las estimaciones de los parámetros han cambiado en menos de ,001.

**Interpretación:**

Luego de realizar los análisis para determinar que el modelo es válido, entrega un cuadro de resumen, el que básicamente indica algunos estimadores para verificar el modelo, sin embargo, para el usuario principiante no debiese ser información de cuidado, pues no se utiliza en el propósito general del modelo.

**Tabla de clasificación<sup>a</sup>**

Observado		Pronosticado		Porcentaje correcto
		¿Adquiriría Usted el Producto A?		
		No	Si	
Paso 1	¿Adquiriría Usted el Producto A?	No	Si	
		17	3	85,0
		2	20	90,9
Porcentaje global				88,1

a. El valor de corte es ,500

**Interpretación:**

En el siguiente apartado se ofrece al investigador información acerca de la forma en que los encuestados contestaron a la pregunta.

Además la tabla indica que el modelo es capaz de predecir el 88,1% de las veces, acerca de las intenciones de compra del usuario.

**Variables en la ecuación**

		B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1	ingresos	,237	,070	11,556	1	,001	1,268
	Constante	-11,449	3,380	11,476	1	,001	,000

a. Variable(s) introducida(s) en el paso 1: ingresos.

**Interpretación:**

Finalmente se entregan los valores de los  $\beta$  para poder predecir el modelo. En este caso el valor constante es -11,449 y el valor del  $\beta$  es de 0,237. Por lo tanto el modelo queda expresado de la siguiente manera:

$$\hat{p} = \frac{e^{-11,449+0,237x}}{1 + e^{-11,449+0,237x}}$$



El modelo se traduce, a modo práctico, en lo siguiente:

Si se quiere determinar, por ejemplo, la probabilidad de que un cliente con ingresos de 56,2 adquiera (compre) el producto A, esta sería de:

$$\hat{p} = \frac{e^{-11,449+0,237 \cdot 56,2}}{1 + e^{-11,449+0,237 \cdot 56,2}} = 0,867$$

Es decir, la probabilidad de que un cliente con ingresos de 56,2 adquiera el producto A es de un 86,7%.

Además se puede determinar “la ventaja” entre adquirir el producto y no adquirirlo para la misma persona (ingresos de 56,2), y eso vendría dado por:

$$\hat{\Omega} = e^{-11,449+0,237 \cdot 56,2} = 6,527$$

Lo anterior significa que un cliente con ingresos de 56,2 tiene una preferencia por comprar el producto 6,527 veces mayor que por no comprarlo.

**Variables en la ecuación**

	B	E.T.	Wald	gl	Sig.	Exp(B)
Paso 1 ingresos	,237	,070	11,556	1	,001	1,268
1 Constante	-11,449	3,380	11,476	1	,001	,000

a. Variable(s) introducida(s) en el paso 1: ingresos.

**Interpretación:**

De la misma tabla final se puede utilizar el criterio del test de Wald para determinar que las variables utilizadas en el modelo (constante y nivel de ingresos), son estadísticamente distintas de cero. Para ello existen dos criterios. El primero es ver si el valor Wald es mayor que 4 y si es así, se puede asumir que las variables son estadísticamente diferentes de cero. La segunda opción es utilizando el criterio del valor p, que en este caso (0,01) es menor que 0,05, por lo que también se puede asumir que las variables son distintas de cero (o estadísticamente válidas).

# CAPÍTULO 11

## Análisis Discriminante



## Análisis Discriminante

El análisis discriminante es una técnica que permite encontrar funciones capaces de separar dos o más grupos de individuos, tomando como base un conjunto de medidas sobre los mismos representadas por una serie de variables. Dichas funciones discriminan o identifican grupos, definidos por una variable dependiente.

Básicamente se trata de una técnica de reducción de datos que indica cuando un dato debe pertenecer a un grupo y cuando debe pertenecer a otro grupo.

Como se indicó antes, al análisis discriminante permite encontrar una función que ayuda a discriminar los datos, para ello se utiliza la siguiente función:

$$D = a_0 + a_1X_1 + a_2X_2 + \dots + a_pX_p$$

Donde:

X son los valores de las variables independientes y  $a_p$  son los coeficientes estimados en la base de datos.

Si bien es cierto que lo anterior puede parecer un tanto engorroso, a continuación se explicará en términos prácticos (mediante la resolución de un ejemplo), la utilización de este tipo de análisis.

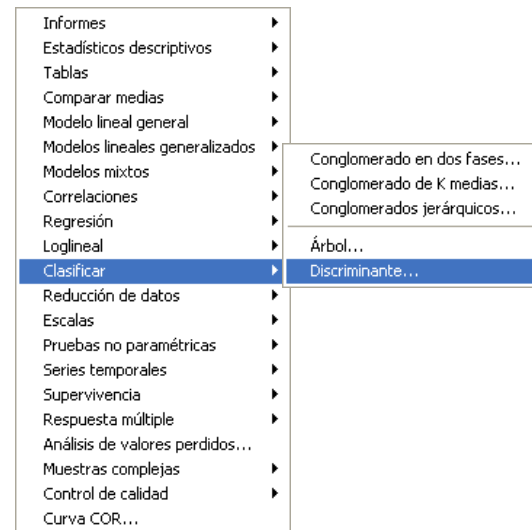
**NOTA:** Este tipo de análisis conlleva la utilización de cierto tipo de cálculo (anexo a los resultados entregados por el Software), el cual será explicado en la medida que vayan apareciendo los resultados. Esto principalmente para evitar confundir al lector con tanta fórmula en los encabezados.

### Ejemplo Práctico:

Los responsables de la empresa "A" s.a. estudian la planificación de su estrategia comunicacional para el año 2009. El jefe de comunicación sabe que hay una relación muy estrecha entre los resultados en términos monetarios que consigue una empresa y el manejo de su estrategia de comunicación. Este directivo desea saber más información sobre los factores, derivados de su estrategia de comunicación, que sirvan para discriminar entre las empresas según sus resultados.

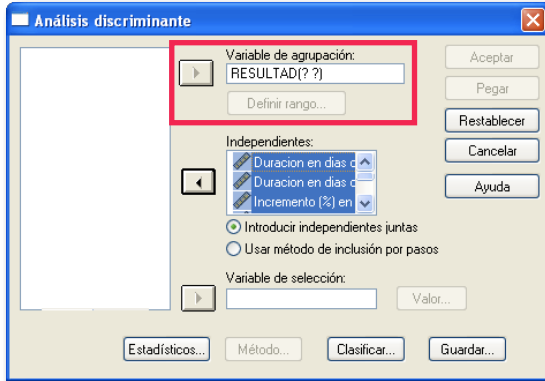
Se dispone de un estudio en el que figura información relativa a 35 empresas competidoras, 18 de las cuales tuvieron resultados que se pueden catalogar como bajos y 17 tuvieron resultados que se puede considerar como medios. La base de datos contiene información acerca de: Duración en días de las actividades publicitarias de la empresa, duración en días de las actividades promocionales de la empresa, incremento (%) en el presupuesto publicitario respecto al periodo anterior, incremento (%) en el presupuesto promocional respecto al periodo anterior, inversión (millones de euros) en actividades de patrocinio en el año 2007 y resultados monetarios obtenidos.

La base de datos se encuentra disponible en el archivo "BaseDatosC11.sav"



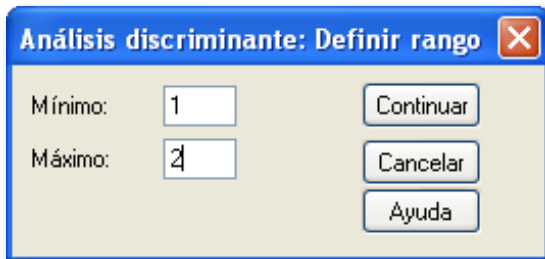
1. Al tratarse de una técnica de clasificación de datos, para acceder al análisis discriminante se debe ir al menú "**Análisis**" y posteriormente seleccionar "**Clasificar**" y finalmente escoger la opción "**Discriminante**".

La ventana de la izquierda muestra gráficamente la forma de acceder a dicho análisis.



2. Una vez abierto el cuadro de diálogo del análisis discriminante, se debe ingresar las variables que se utilizarán. En este caso, la variable dependiente (RESULTAD) debe ir donde dice “Variable de agrupación”, mientras que las variables independientes, van bajo la etiqueta “Independiente”, tal como se observa en la figura.

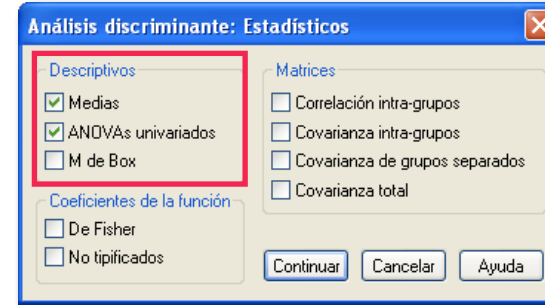
Si se observa con detalle, la variable dependiente presenta unos signos de interrogación (??), para ello debemos definir su rango de acción. Así que hay que presionar donde dice “Definir rango...”.



3. Debido a que la variable “Resultado” podía tomar los valores 1 y 2 (revisar en la base de datos), esos valores deben ser ingresados como “Mínimo” y “Máximo” respectivamente.

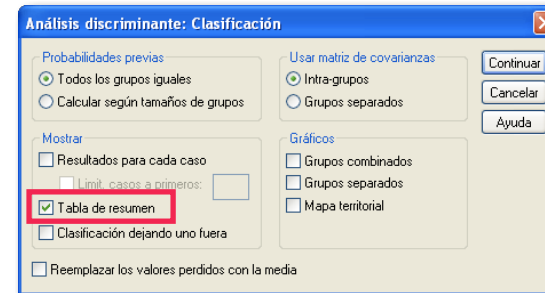
Ojo que los valores ingresados aquí dependen de los valores que posea la variable dependiente en la base de datos (en este caso son 1 y 2, que corresponden a resultados Bajos y Resultados Medios).

Una vez que los valores estén ingresados se debe hacer clic en el botón “Continuar”.



4. De manera de obtener mejores resultados, se recomienda, en la ventana principal, hacer clic en el botón “Estadísticos”, lo que abrirá una ventana como la que se observa a la izquierda. Ahí se debe seleccionar donde dice “Medias” y “ANOVAs univariados”, de manera que aparezcan en los resultados.

Una vez listo, hacer clic en “Continuar”.



5. Una vez hecho clic en el botón continuar, se volverá a la ventana inicial. Una vez ahí, se sugiere hacer clic en el botón “Clasificar”. Cuando abra el cuadro de diálogo, se debe seleccionar donde dice “Tabla de resumen” que entregará un cuadro que será útil para la correcta visualización de los resultados.

Una vez listo se debe hacer clic en el botón Continuar y ya se encuentra listo para realizar el análisis

**IMPORTANTE:**

Antes de proceder a analizar los resultados, se debe saber que este tipo de análisis arroja una gran cantidad de tablas, las que a menudo no resultan útiles para extraer información de ellas. Para efectos prácticos, solamente se analizarán aquellas tablas que ofrezcan información relevante para la toma de decisiones, mientras que otro tipo de tablas serán omitidas de análisis.

**Pruebas de igualdad de las medias de los grupos**

	Lambda de Wilks	F	gl1	gl2	Sig.
Duración en días de las actividades promocionales de la empresa	,414	46,714	1	33	,000
Duración en días de las actividades publicitarias de la empresa	,585	23,393	1	33	,000
Incremento (%) en el presupuesto promocional con respecto al ejercicio anterior	,837	6,410	1	33	,016
Incremento (%) en el presupuesto publicitario con respecto al ejercicio anterior	,862	5,301	1	33	,028
Inversión (en millones de Euros) en actividades de patrocinio en el año 2007	,974	,892	1	33	,352

**Interpretación:**

La tabla que se muestra arriba, lo que hace es indicar al investigador el poder discriminante de cada variable por separado. Para ello usa una prueba F y mediante el criterio del valor p, se puede asumir que solamente una variable no tiene poder discriminante por si sola. Esta variable es "Inversión (En millones de euros) en actividades de patrocinio en el año 2007" (Sig. = 0,352 > 0,05). Por ende esta variable no debiese ser utilizada en el modelo.

**Lambda de Wilks**

Contraste de las funciones	Lambda de Wilks	Chi-cuadrado	gl	Sig.
1	,288	37,969	5	,000

**Interpretación:**

El criterio del Lambda de Wilks permite al investigador determinar si el modelo es válido. Para ello utiliza el estadístico conocido como "Lambda de Wilks". De acuerdo al criterio del valor p, se puede asumir que el modelo es estadísticamente válido, debido a que Sig. = 0,000, lo que es menor que 0,05 (que es el valor generalmente usado para contrastar y que corresponde a un 96% de confianza).

**Coefficientes estandarizados de las funciones discriminantes canónicas**

	Función
	1
Duración en días de las actividades promocionales de la empresa	,975
Duración en días de las actividades publicitarias de la empresa	,586
Incremento (%) en el presupuesto promocional con respecto al ejercicio anterior	,197
Incremento (%) en el presupuesto publicitario con respecto al ejercicio anterior	-,225
Inversión (en millones de Euros) en actividades de patrocinio en el año 2007	-,471

**Interpretación:**

Los coeficientes mencionados en la tabla de la izquierda corresponden a los coeficientes que se deben utilizar en la ecuación discriminante, la que fue definida en forma teórica al comienzo de este capítulo.

Utilizando esta información, la función discriminante quedaría de la siguiente manera:

$$D = 0,675X_1 + 0,586X_2 + 0,197X_3 - 0,225X_4$$

Donde:

- X<sub>1</sub>: Duración de las actividades promocionales de la empresa.
- X<sub>2</sub>: Duración en días de las actividades publicitarias de la empresa.
- X<sub>3</sub>: Incremento (%) en el presupuesto promocional con respecto al ejercicio anterior.
- X<sub>4</sub>: Incremento (%) en el presupuesto publicitario con respecto al ejercicio anterior.

Se debe recordad que la variable “Inversión (en millones de euros) en actividades de patrocinio” no debe ser incluida en el modelo por no poseer poder discriminante por si sola, de acuerdo a lo que se mencionó anteriormente.

**Funciones en los centroides de los grupos**

Resultados monetarios obtenidos	Función
Bajos	-1,484
Medios	1,571

Funciones discriminantes canónicas no tipificadas evaluadas en las medias de los grupos

**Interpretación:**

Para determinar el valor de corte (el que dirá al investigador cuando clasificar en un grupo y cuando hacerlo en el otro), se debe utilizar la siguiente ecuación, la que a su vez debe utilizar los valores de los centroides (indicados en la tabla superior):

$$P_0 = \frac{n_a C_a + n_b C_b}{n_a + n_b}$$

Donde n<sub>a</sub> y n<sub>b</sub> corresponde a la cantidad de datos que se encuentran actualmente en uno u otro grupo de clasificación (en este caso con resultados medios o bajos. Para ello se debe revisar el encabezado del problema, Mientras que C<sub>a</sub> y C<sub>b</sub> corresponde al valor de los centroides.

Reemplazando se tiene que:

$$P_0 = \frac{[(18 \cdot (-1,484)) + (17 \cdot 1,571)]}{18 + 17} \approx 0$$

**Interpretación:**

Como el valor de corte tiende a ser “Cero”, se puede decir entonces que se clasificará a un individuo en el grupo 1 (resultados bajos), si su resultado discriminante (si el valor D obtenido de procesar sus datos por la función discriminante) fuese menor que cero, y se clasificaría dentro del grupo 2 (resultados medios), si su resultado discriminante fuese mayor que cero.

**Resultados de la clasificación<sup>a</sup>**

Original	Recuento	Resultados monetarios obtenidos	Grupo de pertenencia pronosticado		Total
			Bajos	Medios	
		Bajos	18	0	18
		Medios	2	15	17
	%	Bajos	100,0	,0	100,0
		Medios	11,8	88,2	100,0

a. Clasificados correctamente el 94,3% de los casos agrupados originales.

**Interpretación:**

Finalmente, el software entrega un recuento, el que solamente indica que la función discriminante anteriormente elaborada es capaz de clasificar correctamente un 94,3% de los casos.

# CAPÍTULO 12

## Análisis Factorial



## Análisis Factorial

Hay circunstancias en que el investigador posee una gran cantidad de variables que debe analizar y agrupar, como por ejemplo el conjuntos de características comunes a ciertos productos o lugares que sean propios de estos.

Para poder hacer este tipo de ordenamiento o reducción de variables, existen métodos como el análisis factorial, en el cual se parte de una gran número de variables y se trata de averiguar si estas tienen algún pequeño número de factores en común que expliquen (básicamente) lo mismo que las variables observables al inicio.

Dicho de otra manera, el análisis factorial sirve como ayuda al investigador para comprender mejor la composición de determinados fenómenos complejos, mediante la separación de factores característicos para cada uno de estos fenómenos.

En análisis factorial cuenta de 4 fases características:

- \* Cálculo de una matriz capaz de expresar la variabilidad (varianza) conjunta de las variables.
- \* La extracción de un número óptimo de factores.
- \* La rotación de la solución para facilitar su interpretación.
- \* La estimación de las puntuaciones de los sujetos en las nuevas dimensiones.

Un aspecto importante de este tipo de técnicas de reducción de datos es que la decisión de incluir ciertos factores dentro del análisis dependerá únicamente de sus criterios personales, por lo que este tipo de análisis solo sirve para orientar al investigador en la forma en que ordena la información.

En general, las grandes aplicaciones del análisis factorial se resumen en:

- \* Reducción de Información.
- \* Identificación de estructuras subyacentes.
- \* Como paso intermedio en la obtención de nuevas variables intercorrelacionadas para utilizar en posteriores análisis.

En el campo del Marketing el análisis Factorial se puede aplicar en:

- \* Investigaciones sobre productos y atributos determinantes.
- \* Investigaciones sobre imagen corporativa de productos o marcas.
- \* Identificación de grupos de consumidores (Segmentación).
- \* Estudios sobre precios.

A continuación se presentará un ejemplo práctico que indica la forma de aplicar, utilizar e interpretar este tipo de análisis.

**Ejemplo Práctico (Adaptado del libro “Técnicas de Investigación de Mercado” de Teodoro Luque):**

Para la toma de determinadas decisiones comerciales a nivel nacional puede interesar sintetizar las principales características socioeconómicas del conjunto de capitales de provincia españolas. Se trata de identificar las dimensiones relevantes que describen a las capitales de provincia españolas. De esta forma conseguiremos un número más reducido y operativo de variables, y para ello parece oportuno plantear un análisis factorial.

Las variables se han seleccionado de la publicación “Anuario del Mercado Español” del año 1992. Están relacionadas con la población, situación económica e indicadores comerciales y económicos. Las variables se detallan a continuación:

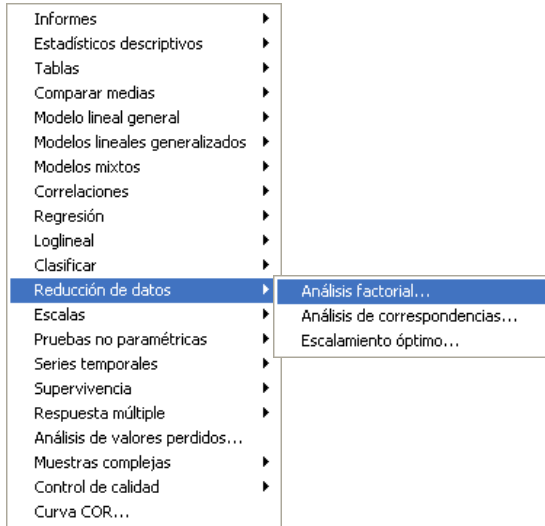
Variable	Descripción
TURISMO	Índice turístico calculado sobre una base nacional de 100.000 unidades.
CUOTA	Cuota de mercado referida a una base nacional de 100.000 unidades



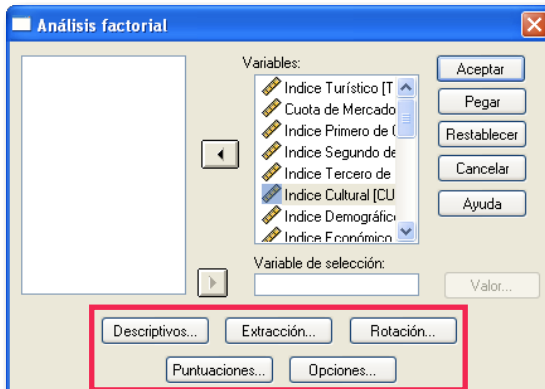
<b>Variable</b>	<b>Descripción</b>
ICC1	Índice primario de capacidad de compra
ICC2	Índice segundo de capacidad de compra
ICC3	Índice tercero de capacidad de compra
CULTURAL	Índice cultural calculado sobre una base nacional de 100.000 unidades
DEMOGRAF	Índice demográfico, referido a una base de 100.000 unidades
ECONOMIC	Índice económico
RACTIVA	Índice de riqueza activa
TLFHABIT	Número de teléfonos por cada 100.000 Háb.
EGBHABIT	Número de alumnos matriculados en EGB por cada 100.000 habitantes
TEXTILHA	Número de licencias fiscales calificables dentro del grupo 2 de actividades comerciales
MADERHAB	Idem calificables dentro del grupo 3
QUIMHAB	Idem calificables dentro del grupo 4
CONSTHAB	Idem calificables dentro del grupo 5

METALHAB	Idem calificables dentro del grupo 6
MAQUINHA	Idem calificables dentro del grupo 7

La base de datos de este ejemplo se encuentra en el archivo "BaseDatosC12.sav".



1. El análisis factorial es una técnica de reducción de datos, por ende, para acceder a el, es necesario ir al menú “**Analizar**” y posteriormente “**Reducción de datos**”. Finalmente se debe hacer clic donde dice “**Análisis factorial...**”.



2. Debido a que el análisis factorial no es una técnica de dependencia (es decir que no hay variables dependientes), solo se debe agregar las variables que se desee introducir, las que en este caso corresponde a la totalidad de las variables de la base de datos.

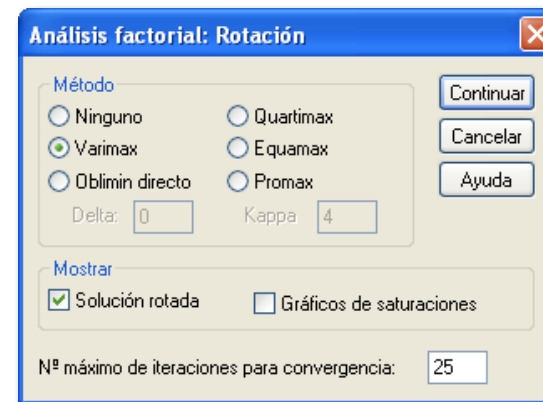
Una vez ingresadas las variables (como muestra la figura de la izquierda), se debe agregar algunas opciones a los resultados de manera que sea más fácil comprender los resultados.

Las opciones están disponibles en los botones de la zona inferior del cuadro de diálogo (recuadro rojo).



3. Primero se debe hacer clic donde dice “**Descriptivos...**” y seleccionar las opciones “Solución Inicial”, “KMO y prueba de esfericidad de Bartlett” y “Anti-imagen”.

Una vez seleccionadas todas las opciones descritas (como muestra la figura), se debe hacer clic en el botón “Continuar”



4. Posteriormente, se debe hacer clic en el botón “**Rotación...**” (en la ventana principal) y se debe seleccionar la opción “Varimax” y se debe pedir mostrar la solución rotada.

Lo anterior responde a que en ocasiones SPSS no muestra claramente la distribución de las características en los factores, por lo que se debe rotar el análisis de manera de obtener resultados más claros.

**NOTA IMPORTANTE:**

Al igual que lo sucedido con el análisis discriminante, la ventana de resultados despliega mucha información que puede no resultar relevante para el estudio que se desee realizar. A continuación solamente se presentarán las tablas de respuesta que resultan útiles para la realización de análisis y para el correcto ordenamiento de los factores.

**KMO y prueba de Bartlett**

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		,774
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	1613,817
	gl	153
	Sig.	,000

**Interpretación:**

El análisis del valor KMO (0,774) y el test de esfericidad de Bartlett (1.613), muestra una cierta estructura de correlación entre las variables, la que puede ser verificada con en análisis de las matrices de correlaciones antiimagen, que es una tabla que no se presentará en este documento, pero que el investigador puede observar.

**Varianza total explicada**

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	8,321	46,228	46,228	8,321	46,228	46,228	6,127	34,041	34,041
2	4,666	25,922	72,151	4,666	25,922	72,151	5,421	30,115	64,157
3	1,928	10,712	82,863	1,928	10,712	82,863	3,283	18,242	82,398
4	1,108	6,158	89,021	1,108	6,158	89,021	1,192	6,623	89,021
5	,506	2,812	91,832						
6	,370	2,054	93,887						
7	,298	1,653	95,540						
8	,228	1,266	96,806						
9	,176	,978	97,784						
10	,131	,726	98,510						
11	,104	,578	99,088						
12	,066	,368	99,456						
13	,054	,302	99,759						
14	,019	,107	99,865						
15	,018	,100	99,965						
16	,003	,015	99,980						
17	,002	,013	99,993						
18	,001	,007	100,000						

Método de extracción: Análisis de Componentes principales.

**Interpretación:**

La tabla que se presenta en la parte superior, se puede observar la descomposición de la varianza por parte de los distintos factores. Como se puede observar, en la parte superior se muestra que hay 4 factores que explican el 89,021% de la varianza total, mientras que el resto de los factores solamente aportan con el 11%, por lo que debiesen ser desechados.

En el cuadro rojo se observa cuales son los cuatro factores en los que se puede simplificar el análisis.

Matriz de componentes<sup>a</sup>

	Componente			
	1	2	3	4
Índice Turístico	,731	-,568	-,054	,048
Cuota de Mercado	,734	-,628	-,162	-,001
Índice Primero de Capacidad de Compra	,579	,080	,747	,184
Índice Segundo de Capacidad de Compra	,665	,181	,710	-,025
Índice Tercero de Capacidad de Compra	,733	,192	,576	-,191
Índice Cultural	,723	-,850	-,155	,124
Índice Demográfico	,743	-,631	-,168	,101
Índice Económico	,799	-,561	-,151	,063
Índice de Riqueza Activa	,791	-,544	-,154	,084
Número de Teléfonos por cada 100000 hab	,852	,063	-,024	-,195
Número de Instrumentos Notariales por cada 100000 hab	,775	,245	,227	,124
Número de Alumnos Matriculados en EGB por cada 100000 hab	-,129	,281	,026	,934
Número de licencias para industria textil cada 100000 hab	,579	,652	-,366	,005
Número de licencias para industria Madera cada 100000 hab	,745	,422	-,218	-,065
Número de licencias para industria Química cada 100000 hab	,531	,710	-,209	-,016
Número de licencias para industria Construcción cada 100000 hab	,487	,700	-,256	,217
Número de licencias para industria Metálica cada 100000 hab	,614	,587	-,155	-,120
Número de licencias para Maquinarias cada 100000 hab	,681	,598	-,197	-,079

Método de extracción: Análisis de componentes principales.  
a. 4 componentes extraídos

Interpretación:

La matriz de componentes muestra la forma en que se debe descomponer cada una de las variables (en otras palabras, indica que variable debe ser asignada a cada factor). Sin embargo, se puede notar que el factor 1 (Componente 1) posee un gran número de cargas elevadas (con un valor alto), por lo que para obtener información más clara se debe utilizar una matriz de componentes rotada (rotación varimax).

A priori, esta tabla no sirve de mucho al investigador, solamente debe ser analizada utilizando la matriz de componentes rotada.

Matriz de componentes rotados<sup>a</sup>

	Componente			
	1	2	3	4
Índice Turístico	,902	,035	,208	-,070
Cuota de Mercado	,965	,038	,105	-,127
Índice Primero de Capacidad de Compra	,178	,081	,932	,160
Índice Segundo de Capacidad de Compra	,150	,227	,950	-,040
Índice Tercero de Capacidad de Compra	,201	,335	,864	-,209
Índice Cultural	,987	,008	,104	-,005
Índice Demográfico	,987	,040	,104	-,027
Índice Económico	,969	,117	,154	-,060
Índice de Riqueza Activa	,956	,124	,150	-,036
Número de Teléfonos por cada 100000 hab	,517	,552	,371	-,241
Número de Instrumentos Notariales por cada 100000 hab	,321	,522	,584	,104
Número de Alumnos Matriculados en EGB por cada 100000 hab	-,162	,072	,008	,968
Número de licencias para industria textil cada 100000 hab	,043	,943	,025	,043
Número de licencias para industria Madera cada 100000 hab	,266	,819	,200	-,063
Número de licencias para industria Química cada 100000 hab	-,071	,896	,150	,034
Número de licencias para industria Construcción cada 100000 hab	-,052	,871	,089	,266
Número de licencias para industria Metálica cada 100000 hab	,042	,838	,218	-,089
Número de licencias para Maquinarias cada 100000 hab	,096	,900	,213	-,051

Método de extracción: Análisis de componentes principales.  
Método de rotación: Normalización Varimax con Kaiser.  
a. La rotación ha convergido en 5 iteraciones.

Interpretación:

Una vez con la matriz de componentes rotados, se puede determinar con mayor facilidad la forma de incluir cada variable en cada factor (componente). Para ello hay que ver para cada uno de los componentes, cuales son las variables que presentan valores más altos.

El factor 1, posee las siguientes variables (nótese que todas las seleccionadas poseen un valor alto):

- \* Índice turístico.
- \* Cuota de mercado.
- \* Índice cultural.
- \* Índice demográfico.
- \* Índice económico.
- \* Índice de riqueza activa.

Matriz de componentes rotados<sup>a</sup>

	Componente			
	1	2	3	4
Índice Turístico	,902	,035	,208	-,070
Cuota de Mercado	,965	,038	,105	-,127
Índice Primero de Capacidad de Compra	,178	,081	,932	,160
Índice Segundo de Capacidad de Compra	,150	,227	,950	-,040
Índice Tercero de Capacidad de Compra	,201	,335	,864	-,209
Índice Cultural	,987	,008	,104	-,005
Índice Demográfico	,987	,040	,104	-,027
Índice Económico	,969	,117	,154	-,060
Índice de Riqueza Activa	,956	,124	,150	-,036
Número de Teléfonos por cada 100000 háb	,517	,552	,371	-,241
Número de Instrumentos Notariales por cada 100000 háb	,321	,522	,584	,104
Número de Alumnos Matriculados en EGB por cada 100000 háb	-,162	,072	,008	,968
Número de licencias para industria textil cada 100000 háb	,043	,943	,025	,043
Número de licencias para industria Madera cada 100000 háb	,266	,819	,200	-,063
Número de licencias para industria Química cada 100000 háb	-,071	,896	,150	,034
Número de licencias para industria Construcción cada 100000 háb	-,052	,871	,089	,266
Número de licencias para industria Metálica cada 100000 háb	,042	,838	,218	-,089
Número de licencias para Maquinarias cada 100000 háb	,096	,900	,213	-,051

Método de extracción: Análisis de componentes principales.  
 Método de rotación: Normalización Varimax con Kaiser.  
 a. La rotación ha convergido en 5 iteraciones.

Interpretación:

Nuevamente, en la matriz de componentes rotados, se debe observar el resto de los factores. En este caso, se verán las variables correspondientes al Factor 2 (Componente 2). Utilizando el mismo criterio que en el anterior caso, se debe escoger las variables que posean los valores altos. De acuerdo a lo anterior, las variables seleccionadas para el segundo factor son las siguientes:

- \* Número de licencias para industria textil (cada 100.000 háb.).
- \* Número de licencias para industria madera (cada 100.000 háb.).
- \* Número de licencias para industria química (cada 100.000 háb.).
- \* Número de licencias para industria construcción (cada 100.000 háb.).
- \* Número de licencias para industria metálica (cada 100.000 háb.).
- \* Número de licencias para maquinarias (cada 100.000 háb.).

Matriz de componentes rotados<sup>a</sup>

	Componente			
	1	2	3	4
Índice Turístico	,902	,035	,208	-,070
Cuota de Mercado	,965	,038	,105	-,127
Índice Primero de Capacidad de Compra	,178	,081	,932	,160
Índice Segundo de Capacidad de Compra	,150	,227	,950	-,040
Índice Tercero de Capacidad de Compra	,201	,335	,864	-,209
Índice Cultural	,987	,008	,104	-,005
Índice Demográfico	,987	,040	,104	-,027
Índice Económico	,969	,117	,154	-,060
Índice de Riqueza Activa	,956	,124	,150	-,036
Número de Teléfonos por cada 100000 háb	,517	,552	,371	-,241
Número de Instrumentos Notariales por cada 100000 háb	,321	,522	,584	,104
Número de Alumnos Matriculados en EGB por cada 100000 háb	-,162	,072	,008	,968
Número de licencias para industria textil cada 100000 háb	,043	,943	,025	,043
Número de licencias para industria Madera cada 100000 háb	,266	,819	,200	-,063
Número de licencias para industria Química cada 100000 háb	-,071	,896	,150	,034
Número de licencias para industria Construcción cada 100000 háb	-,052	,871	,089	,266
Número de licencias para industria Metálica cada 100000 háb	,042	,838	,218	-,089
Número de licencias para Maquinarias cada 100000 háb	,096	,900	,213	-,051

Método de extracción: Análisis de componentes principales.  
 Método de rotación: Normalización Varimax con Kaiser.  
 a. La rotación ha convergido en 5 iteraciones.

Interpretación:

Para seleccionar las variables correspondientes al factor 3, se vuelve a utilizar el criterio del valor más alto. De acuerdo a esto, las variables seleccionadas para el factor 3 (Componente 3) son:

- \* Índice primero de capacidad de compra.
- \* Índice segundo de capacidad de compra.
- \* Índice tercero de capacidad de compra.

Matriz de componentes rotados<sup>a</sup>

	Componente			
	1	2	3	4
Índice Turístico	,902	,035	,208	-,070
Cuota de Mercado	,965	,038	,105	-,127
Índice Primero de Capacidad de Compra	,178	,081	,932	,180
Índice Segundo de Capacidad de Compra	,150	,227	,950	-,040
Índice Tercero de Capacidad de Compra	,201	,335	,864	-,209
Índice Cultural	,987	,008	,104	-,005
Índice Demográfico	,987	,040	,104	-,027
Índice Económico	,989	,117	,154	-,060
Índice de Riqueza Activa	,956	,124	,150	-,036
Número de Teléfonos por cada 100000 háb	,517	,552	,371	-,241
Número de Instrumentos Notariales por cada 100000 háb	,321	,522	,584	,104
Número de Alumnos Matriculados en EGB por cada 100000 háb	-,162	,072	,008	<b>,988</b>
Número de licencias para industria textil cada 100000 háb	,043	,943	,025	,043
Número de licencias para industria Madera cada 100000 háb	,266	,819	,200	-,063
Número de licencias para industria Química cada 100000 háb	-,071	,896	,150	,034
Número de licencias para industria Construcción cada 100000 háb	-,052	,871	,089	,266
Número de licencias para industria Metálica cada 100000 háb	,042	,838	,218	-,089
Número de licencias para Maquinarias cada 100000 háb	,096	,900	,213	-,051

Método de extracción: Análisis de componentes principales.  
Método de rotación: Normalización Varimax con Kaiser.  
a. La rotación ha convergido en 5 iteraciones.

Interpretación:

Finalmente, el cuatro factor esta compuesto solamente por una variable, esta es el número de alumnos matriculados en EGB.

Analizando la composición de los cuatro factores detectados, se puede asumir que el factor 1 corresponde a aspectos “socio-económicos”, el factor 2 corresponde a aspectos relacionados con las “licencias industriales”, el factor 2 corresponde a “capacidad de compras” y el factor 2 corresponde a aspectos “educacionales”.

Los nombres asignados a cada uno de los factores deben ser asignados por el propio investigador de acuerdo a sus propios criterios personales y profesionales, en vista de las variables que incluye cada uno de los factores determinados.