



UNIVERSIDAD DE PUERTO RICO
RECINTO DE RIO PIEDRAS
FACULTAD DE ADMINISTRACIÓN DE EMPRESAS

INTRODUCCIÓN A SPSS



Preparado por:
Marisela Santiago Castro, Ph.D

Seminario de Integración: Cómo analizar datos de cuestionarios en SPSS
Título V: Iniciativas de Investigación y Actividad Creativa Subgraduadas (iINAS)

1 de mayo de 2015

CONTENIDO:

INTRODUCCIÓN	3
PARTE I – RESUMEN DE CONCEPTOS BÁSICOS PARA LA INSPECCION ANTES DE UN ANALISIS ESTADISTICO	4
PARTE II – INSTRUCCIONES BÁSICAS PARA EL MANEJO DE SPSS	10
REFERENCIAS	24

INTRODUCCION

Este folleto sirve de referencia para el Seminario de Integración - Cómo analizar datos de cuestionarios en SPSS. Estas páginas proveen una introducción al análisis estadístico a través del uso del programado SPSS, que significa *Statistical Package for Social Sciences*. Este programado permite analizar estadísticamente datos a diferentes niveles. Las funciones de SPSS se pueden clasificar de las siguientes maneras: (1) funciones para describir o resumir los datos, (2) funciones para hacer pruebas de hipótesis y (3) funciones para examinar relaciones entre variables.

Es meritorio advertir que muchos de los términos que se utilizan en el folleto son en inglés dado que la versión de SPSS que se va a utilizar durante el curso es en inglés. Estos términos se escriben en letra especial *italics*, para que sean fáciles de identificar en el texto. Otra limitación de este escrito es que la información que se presenta no representa todas las posibilidades de análisis estadístico que se pueden lograr con SPSS.

Se prohíbe el uso de toda o parte de la información de este folleto sin la previa autorización de la autora.

PARTE I - Resumen de Conceptos Básicos para la Inspección de Datos antes de un Análisis Estadístico.

INSPECCIÓN DE LOS DATOS

La inspección de los datos antes de la aplicación de las técnicas estadísticas, permite al investigador obtener un entendimiento previo de las características de éstos. El investigador obtiene un perfil básico de los datos y las relaciones entre las variables. Este conocimiento puede ayudar a la especificación y refinamiento del modelo y puede proveer una perspectiva inicial para la interpretación de los resultados.

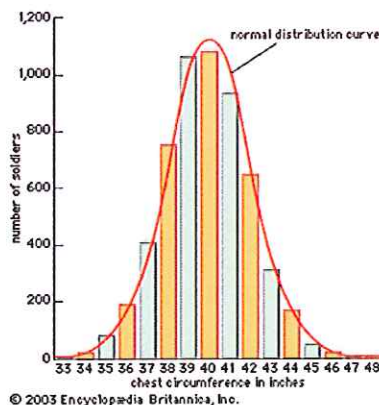
La inspección de los datos puede darse en cuatro etapas:

1. inspección gráfica.
2. detección de datos que faltan (*missing values*)
3. detección de valores extremos (*outliers*)
4. identificación de las suposiciones estadísticas.

Inspección gráfica de los datos

Un modo para entender la naturaleza de los datos es conocer las características básicas de los éstos y las relaciones que puedan existir entre variables. Existen técnicas para inspeccionar las características de la distribución de los datos y variables, las relaciones entre dos variables, y diferencias entre grupos de casos u observaciones.

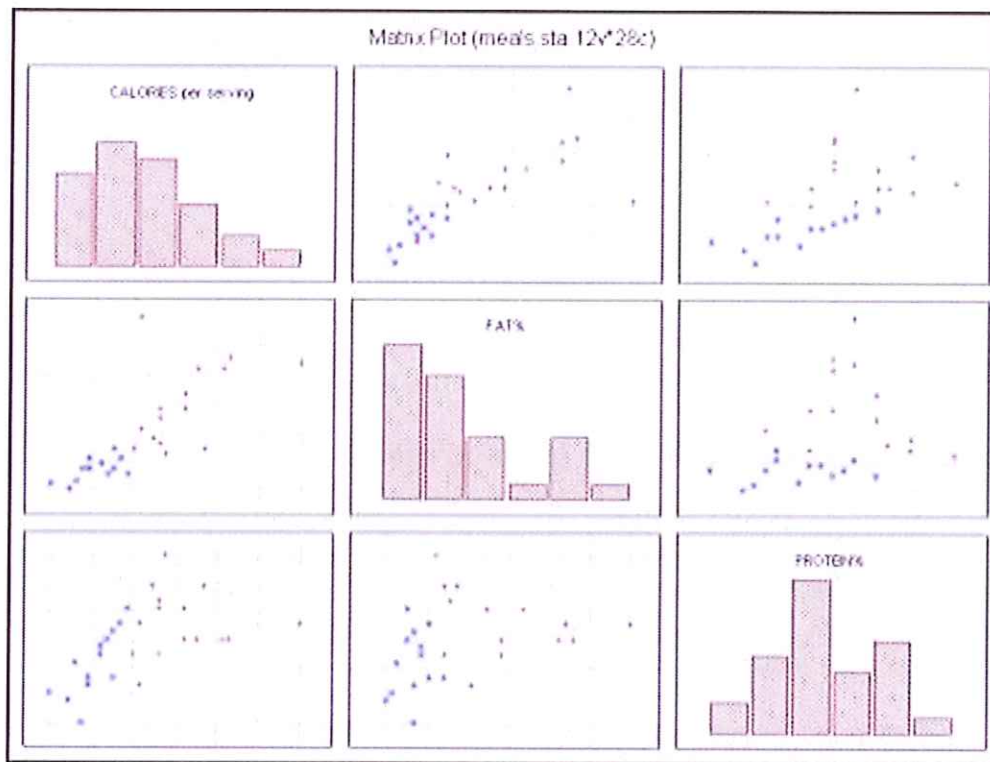
- a. Perfil de la distribución – un histograma permite obtener una perspectiva adecuada de las variables. Un histograma es una representación gráfica de una variable que representa la frecuencia de observaciones o casos dentro de cada categoría.



Un histograma con la distribución normal sobrepuesta se pueden obtener en SPSS de varias formas:

1. *ANALYZE* → *DESCRIPTIVE STATISTICS* → *FRECUENCIAS* → *CHARTS*
2. *GRAPHS* → *LEGACY DIALOGS* → *HISTOGRAM*

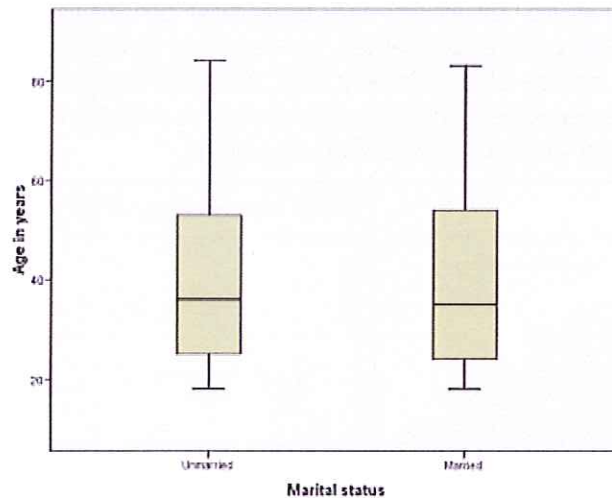
- b. Relación entre variables – un diagrama de puntos (*scatterplot*) es el método más común para examinar la relación entre dos o más variables. Una variable define el eje de X y otra variable el eje de Y. Las variables pueden ser observaciones, valores esperados o residuos. El patrón de los puntos representa la relación entre las variables.



Un diagrama de puntos se puede obtener en SPSS a través del menú de *graphs*. Si se desea el diagrama de puntos de los residuos de una regresión se obtiene:

ANALYZE → *REGRESSION* → *LINEAR* → *PLOTS*

- c. Diferencias entre grupos – se busca entender la extensión y el grado de las diferencias entre dos o más grupos para una o más variables métricas. El investigador necesita entender cómo los valores están repartidos en cada grupo y si hay suficientes diferencias entre los grupos para ser estadísticamente significativas. Se puede utilizar un *boxplot*, que es una representación ilustrada de la distribución de los datos.



En SPSS: ANALYZE → DESCRIPTIVE STATISTICS → EXPLORE → PLOTS

Inspección de datos que faltan (*missing values*)

Missing values son datos específicos no disponibles para un caso o sujeto para el cual se tiene otra información. Usualmente ocurre cuando alguien no responde alguna(s) pregunta(s). Este tipo de situación puede afectar la generalización de los resultados. El objetivo del investigador es determinar las razones para que falte información de alguno de los casos.

Las razones para que datos falten en la muestra pueden ser externas al sujeto bajo estudio (errores al entrar los datos o problemas al recolectar los datos) o pueden ser por acciones del mismo (no querer contestar alguna pregunta). Si las razones para que falten estos datos se desconocen se debe tratar de identificar algún patrón. Cualquier patrón que no sea aleatorio podría asociarse con factores externos al sujeto bajo estudio.

Maneras (remedios) para trabajar con *missing values*:

1. Utilizar solamente las observaciones que contengan los datos completos (Remedio más práctico y simple). Consecuencias de este remedio:
 - Puede reducir considerablemente el tamaño de la muestra, puede convertir la muestra en inapropiada para ciertos tipos de análisis estadísticos.

- Los resultados no se pueden generalizar.
- Este remedio es apropiado si no hay muchos valores faltantes, la muestra es lo suficiente grande, y las relaciones son fuertes que no se afectan por los datos que faltan.

2. Borrar los casos y/o variables.
3. Imputar información - estimar los valores que faltan basándose en los valores válidos de otras variables o casos de la muestra (sólo aplica a variables métricas).
4. Procedimientos basados en modelos estadísticos.

En SPSS bajo el menú de *Transform* (ver páginas 12, 17) existe un comando para reemplazar los *missing values* de diferentes maneras: media o mediana de la muestra, media o mediana de los puntos más cercanos, interpolación o tendencia lineal. Además, en *Variable view* (ver página 10) se puede definir de antemano los *missing values* para cada variable.

Detección de *outliers*

Los *outliers* son una o varias observación(es) que es (son) substancialmente diferente(s) de las otras observaciones (valor extremo). Existen varias maneras para detectar los *outliers*:

1. Inspeccionar la distribución de las observaciones y seleccionar aquellas variables que estén fuera del rango de la distribución.
 - a. Se tiene que designar cuál debe ser ese límite.
 - i. Para muestras pequeñas (80 ó menos observaciones) se recomienda identificar los casos que estén fuera de 2.5 o más desviaciones estándar.
 - ii. Para muestras grandes – 3 a 4 desviaciones estándar.
2. Utilizar un diagrama de puntos. Se designan como valores extremos aquellos casos que caigan extremadamente fuera del límite de las observaciones.
3. *Mahalanobis D²* – medida de distancia en un espacio multidimensional de cada observación desde el centro de las demás observaciones. Se recomienda el nivel de 0.001 para identificar los valores extremos.

En SPSS: *ANALYZE* → *DESCRIPTIVE STATISTICS* → *EXPLORE* → *STATISTICS* → *OUTLIERS*

[Esta instrucción identifica los cinco valores más altos y los cinco valores más bajos de la variable seleccionada]

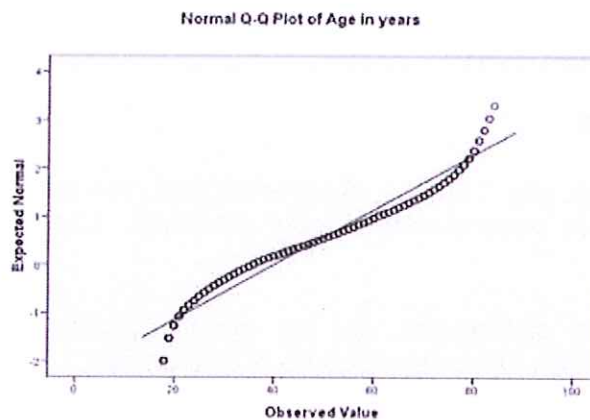
Conformidad con las suposiciones del análisis estadístico

El análisis estadístico con múltiples variables requiere que las suposiciones se prueben para las variables por separado y para el modelo en específico en que se van a emplear las variables.

1. Normalidad en la distribución de las variables

- a. Análisis gráfico - prueba más simple para diagnosticar normalidad.
 - i. *Normal probability Plot* – compara la distribución acumulada de las observaciones con la distribución acumulada de una distribución normal.

En SPSS: *ANALYZE* → *DESCRIPTIVE STATISTICS* → *EXPLORE* → *STATISTICS* → *PLOTS*



- b. Pruebas estadísticas
 - i. Pruebas Z de *Skewness* y *Kurtosis*

En SPSS: *ANALYZE* → *DESCRIPTIVE STATISTICS* → *OPTIONS*

- ii. Prueba *Shapiro-Wilks*
- iii. Prueba *Kolmogorov-Smirnov*

En SPSS: *ANALYZE* → *DESCRIPTIVE STATISTICS* → *EXPLORE* → *PLOT* → *NORMALITY PLOTS WITH TESTS*

2. *Homoscedasticity* – especialmente importante para las regresiones múltiples.

- a. Análisis gráfico
- b. Prueba estadísticas
 - i. Prueba *Levene*

- ii. Prueba *Box's M* – aplicable para cuando se prueban más de una variable métrica a la vez.

En SPSS: *ANALYZE* → *COMPARE MEANS* → *INDEPENDENT SAMPLES T TEST*

3. Linealidad

- a. Análisis gráfico de *scatterplots*.
- b. Análisis de los residuales de una regresión simple

4. Ausencia de errores correlacionados

Transformaciones de datos

Las transformaciones de los datos proveen un medio para modificar variables para corregir violaciones de las suposiciones o para mejorar la relación (correlación) entre variables. Por ejemplo: para transformar una distribución plana a una normal se toma el inverso de la variable; para distribuciones *skewed*: si es *positive skewed* - sacar la raíz cuadrada, si es *negative skewed* - logaritmos o el inverso de la variable.

Una variable se puede transformar para obtener las características deseadas o para tenerla en una base comparable con las demás variables. La necesidad y el tipo específico de las transformaciones puede darse por razones teóricas o razones empíricas.

SPSS provee un menú completo (*Transform*) para hacer transformaciones de los datos y variables (ver páginas 12, 17).

Análisis de confiabilidad

Confiabilidad se puede definir como el grado por el cual alguna medida es libre de errores y resultará en resultados consistentes (Perterson, 1994). Cuando se utilizan en una investigación índices o escalas de múltiples ítems se debe medir la confiabilidad de éstos. Una de las medidas más utilizada en la academia para medir confiabilidad es Cronbach's Alfa.

El coeficiente de Cronbach's Alfa fue desarrollado en 1951 por L. Cronbach y mide consistencia interna. En las ciencias sociales el límite aceptable de confiabilidad es de entre 0.60 a 0.70.

SPSS provee una prueba directa para medir el Cronbach's Alfa:

ANALYZE → *SCALE* → *REALIBILITY ANALYSIS*

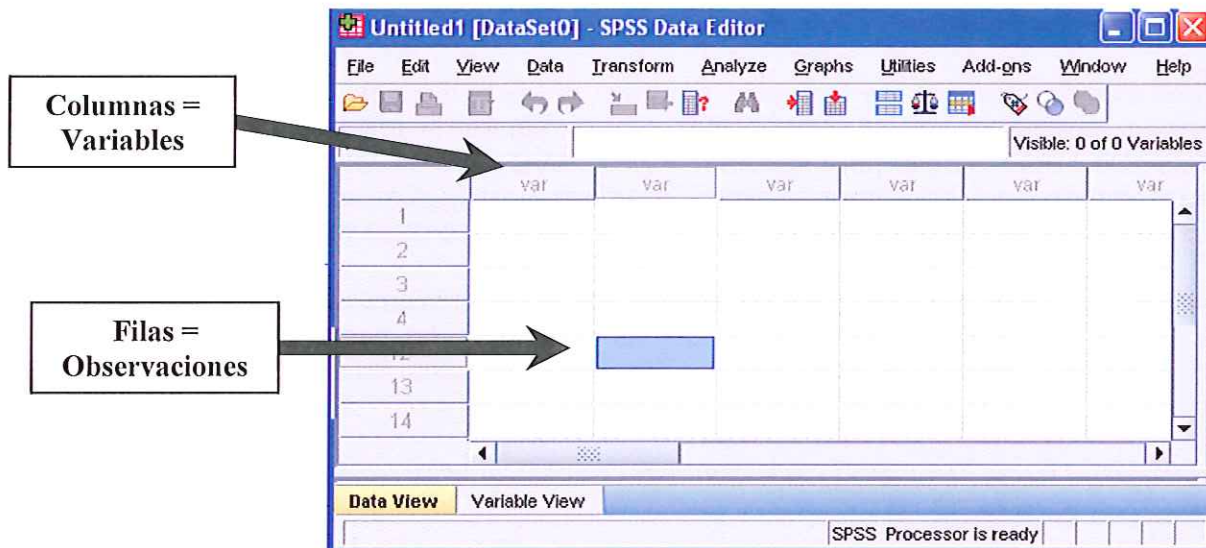
PARTE II – Instrucciones Básicas para el Manejo de SPSS.

INTRODUCCIÓN

Pantallas básicas

SPSS se compone de tres pantallas básicas:

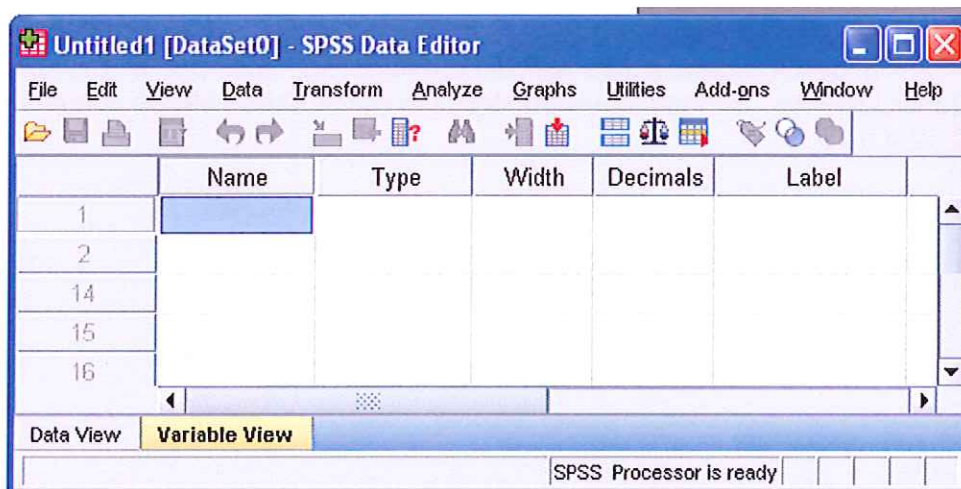
1. *Data view* – primera pantalla que aparece al abrir el programa. Despliega la información de las observaciones que se va a analizar (una vez creado el archivo de los datos).



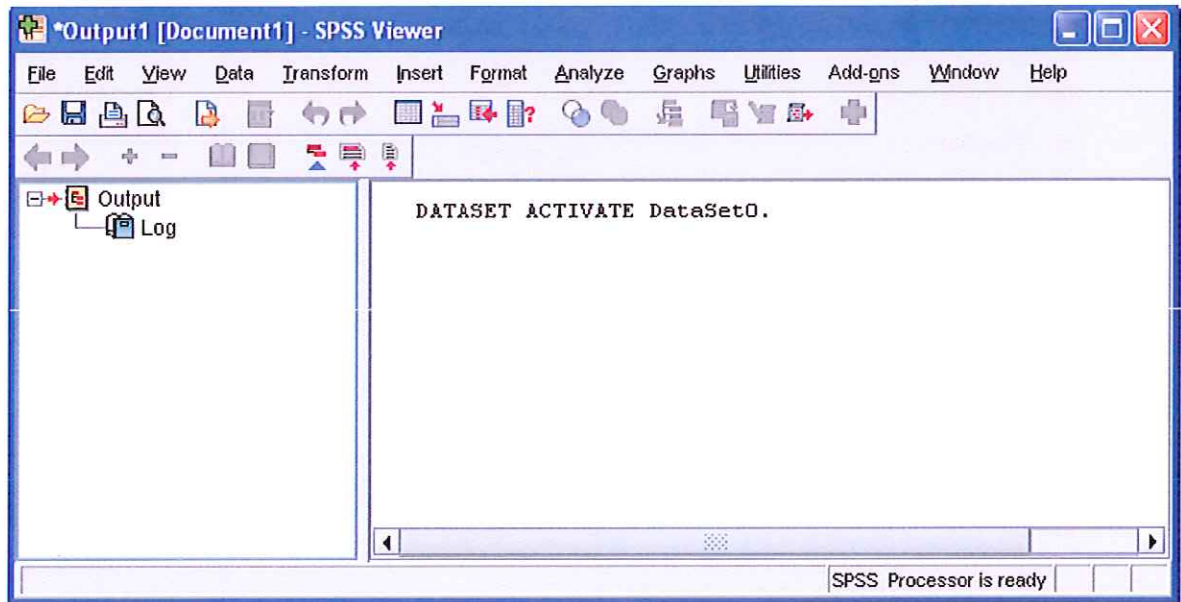
Para poder hacer cualquier análisis, es necesario que este mensaje aparezca en esta parte de la pantalla.



2. *Variable view* – pantalla para definir las características de las variables y las observaciones. Se recomienda que se definan los campos necesarios y aplicables al estudio antes de comenzar a entrar los datos de la muestra.

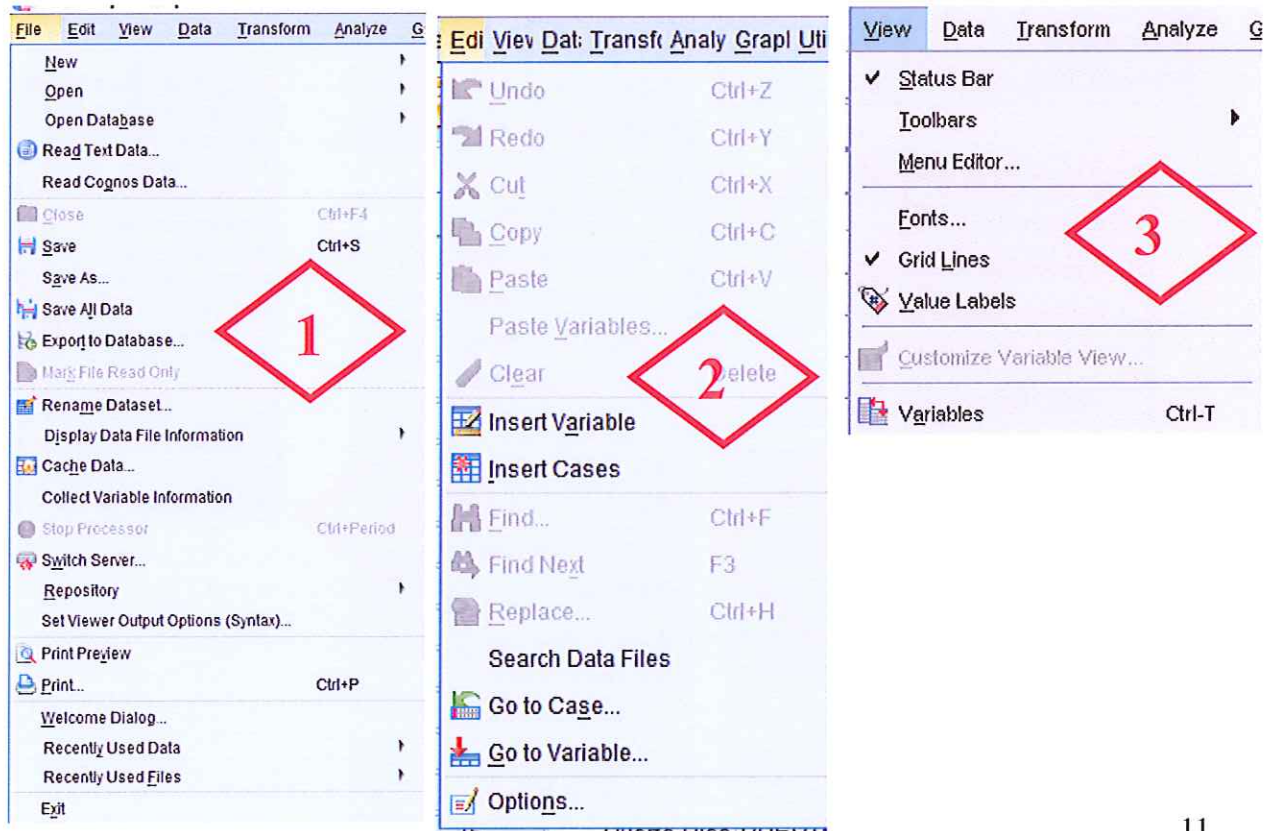


3. Pantalla de resultados – provee los resultados de análisis estadísticos. Provee en el lado izquierdo un registro (*log*) de todos los resultados que se presentan.



Menús básicos

El menú básico de las pantallas de *Data view* y *Variable view* contiene 11 sub-menús, a saber:



Dat: Transfr Analy Grapl Utiliti Add-o Windr Helj

- Define Variable Properties...
- Set Measurement Level for Unknown...
- Copy Data Properties...
- New Custom Attribute...
- Define Dates...
- Define Multiple Response Sets...
- Identify Duplicate Cases...
- Compare Datasets...
- Sort Cases...
- Sort Variables...
- Transpose...
- Merge Files
- Restructure...
- Aggregate...
- Rake Weights...
- Propensity Score Matching...
- Case Control Matching...
- Copy Dataset
- Split into Files
- Split File...
- Select Cases...
- Weight Cases...



nsform Analyze Grapls Utiliti

- Compute Variable...
- Count Values within Cases...
- Shift Values...
- Recode into Same Variables...
- Recode into Different Variables...
- Automatic Recode...
- Visual Binning...
- Rank Cases...
- Date and Time Wizard...
- Create Time Series...
- Replace Missing Values...
- Random Number Generators...



Analy Grapl Utiliti Add-o Windr Helj

- Reports
- Descriptive Statistics
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests
- Forecasting
- Survival
- Multiple Response
- Simulation...
- Quality Control
- ROC Curve...



Graph Utiliti Add-o Windr Helj

- Chart Builder...
- Graphboard Template Chooser...
- Compare Subgroups
- Regression Variable Plots
- Legacy Dialogs



Utiliti Add-o Windr Helj

- Variables...
- QMS Control Panel...
- QMS Identifiers...
- Scoring Wizard...
- Merge Model XML...
- Calculate with Pivot Table
- Data File Comments...
- Merge Viewer Tables...
- Define Variable Sets...
- Censor Table
- Use Variable Sets...
- Show All Variables...
- Process Data Files
- Spelling...
- Run Script...
- Production Facility...
- Map Conversion Utility...
- Custom Dialogs
- Extension Bundles



Add-o Windr Helj

- IBM SPSS Bootstrapping
- IBM SPSS Categories
- IBM SPSS Conjoint
- IBM SPSS Forecasting
- IBM SPSS Custom Tables
- IBM SPSS Exact Tests
- IBM SPSS Complex Samples
- IBM SPSS Missing Values
- IBM SPSS Decision Trees
- IBM SPSS Data Preparation
- IBM SPSS Direct Marketing
- IBM SPSS Neural Networks
- Applications
- Services
- Programmability Extension



Window Help

- Split
- Minimize All Windows

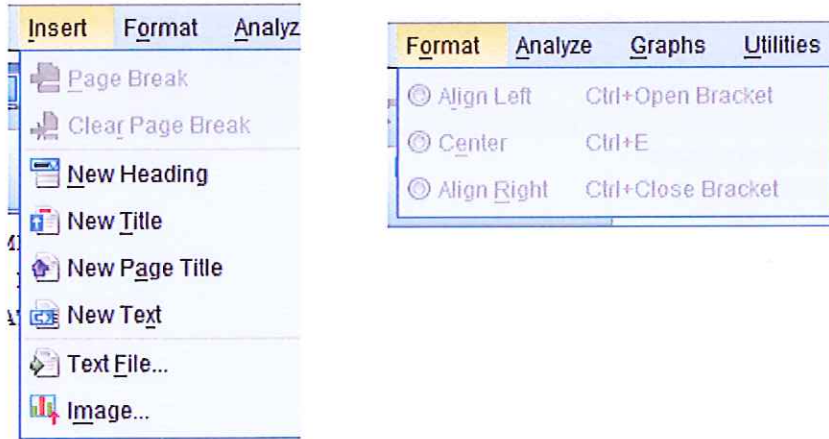


Help

- Topics
- Tutorial
- Case Studies
- Statistics Coach
- Command Syntax Reference
- SPSS Developer Central
- About...
- Algorithms
- SPSS Home
- Check for Updates
- Product Registration



La pantalla de resultados contiene dos sub-menús adicionales:



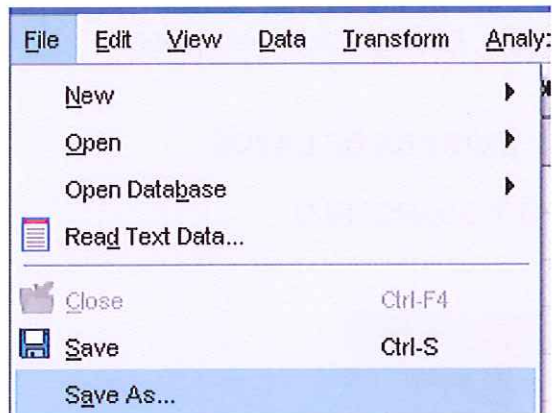
Archivos

Existen cuatro tipos de archivos en SPSS:

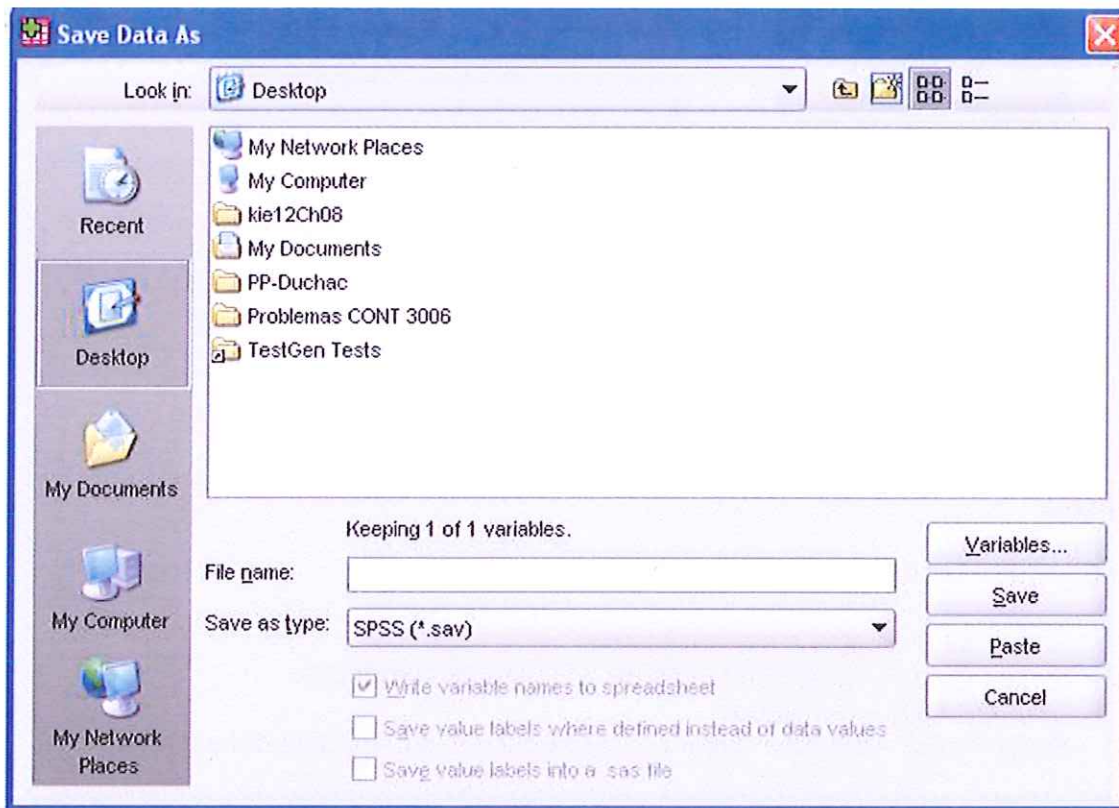
1. *Data* (*.sav) – datos a ser analizados.
2. *Syntax* (*.sps) – procedimientos en lenguaje de comandos.
3. *Output* (*.spv) – resultados.
4. *Script* (*.sbs) – ayuda a hacer procesos hechos a la medida del usuario.

Para trasladar archivos de un lugar a otro se deben guardar los archivos primero. Para esto debe utilizar "Save as", que se encuentra en el menú de *File*.

PASO #1



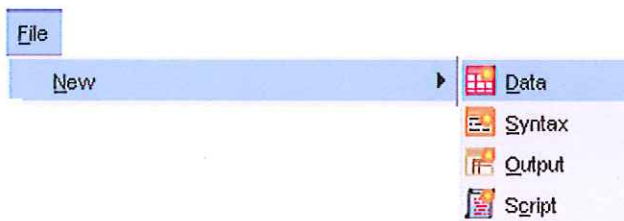
PASO #2 – ELEGIR NOMBRE PARA EL ARCHIVO.



SPSS permite guardar los archivos en diferentes formatos para poder abrirlos en diferentes programados, por ejemplo: EXCEL, SAS, Text y Stata.

ENTRADA DE DATOS

PASO #1 – CREAR EL ARCHIVO Y GUARDARLO.



PASO #2 – DEFINIR EN LA PANTALLA VARIABLE VIEW LAS VARIABLES ANTES DE ENTRAR LAS OBSERVACIONES DE LA MUESTRA.

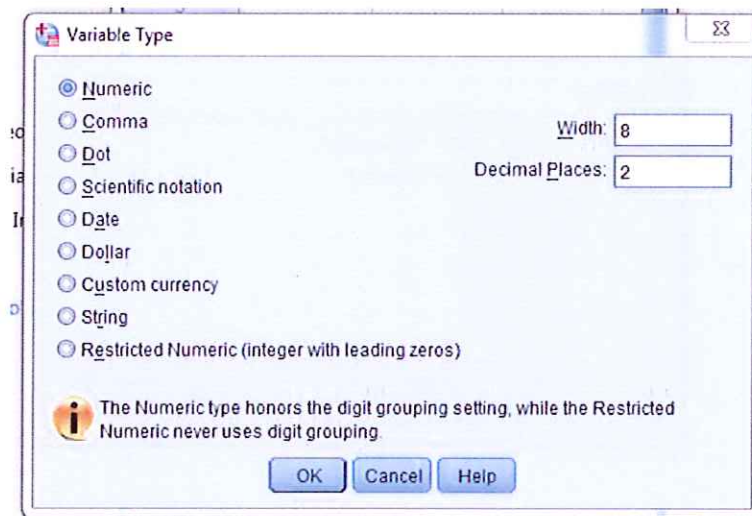
Campos de la pantalla *Variable view*:

1. *Name* – nombre de las variables. Ejemplo:

Reglas más importantes a recordar sobre los nombres de las variables:

- Cada variable debe tener un nombre único, no se pueden repetir.
- Los nombres pueden ser de hasta 64 letras o símbolos de largo.
- El primer carácter de cada nombre debe ser una letra.
- Los nombres no pueden contener espacios entre las letras o símbolos.
- Los siguientes caracteres se pueden utilizar entre letras: _, \$, #, y @.

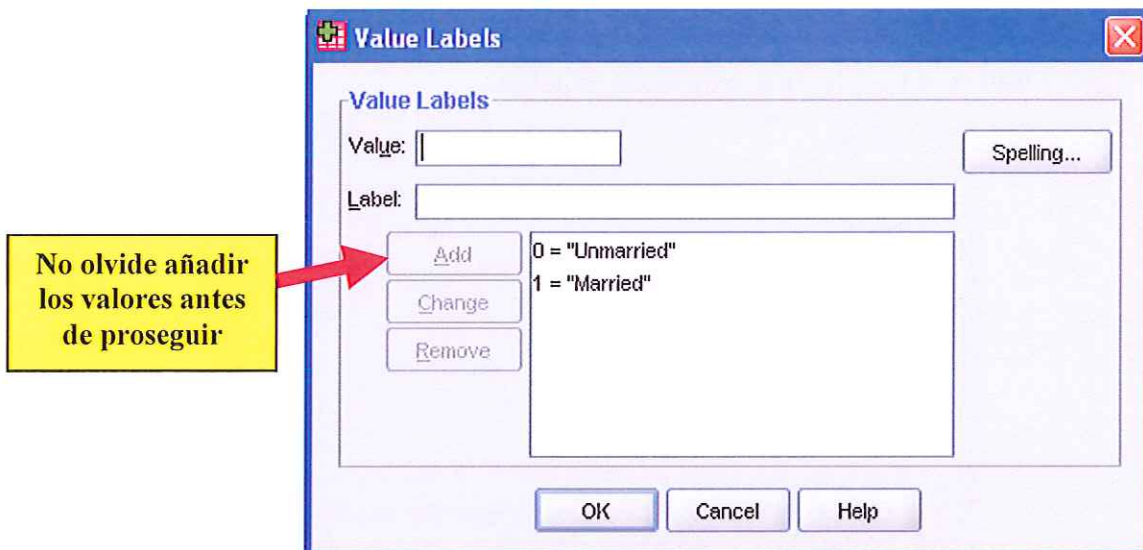
2. *Type* – define el tipo de data para cada variable.



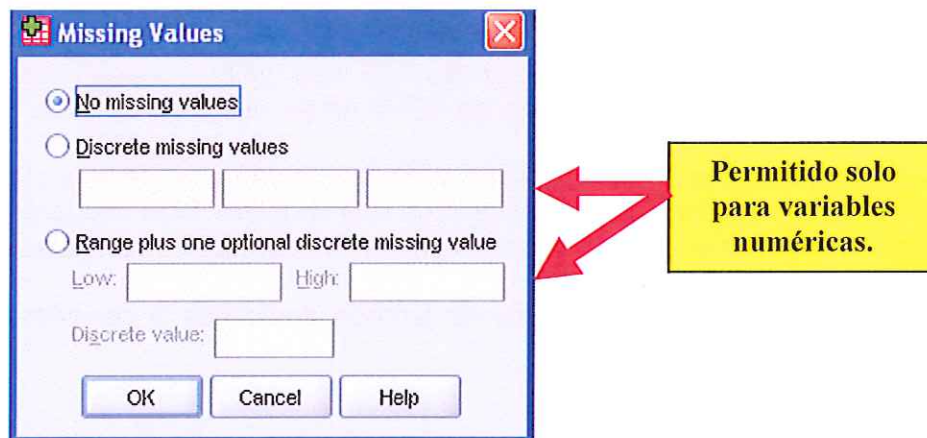
A continuación se definen los tipos de variables más pertinentes a la Administración de Empresas:

- *Numeric* – variable cuyas observaciones son valores numéricos que pueden entrarse en el formato estándar o en notación científica.
 - *Dollar* – valores de dólares. El valor se presenta con un signo de \$, comas y un punto para delimitar los lugares decimales.
 - *String* – variable cuyas observaciones NO son valores numéricos (valores alfanuméricos). Una vez se definan así SPSS no las utiliza en cálculos.
3. *Width* – permite definir el tamaño de las columnas que van a aparecer en la pantalla *Data view*. Esto no afecta el ancho definido de una variable. Si el ancho definido de una variable es menor que el actual, un asterisco (*) aparecerá en el *Data view*.
4. *Decimals* – permite definir el número de lugares decimales a considerar en las variables.

5. *Labels* – define el nombre (hasta 40 letras) en los resultados (tablas, gráficas) de las variables. Sino no se definen las variables en este campo, los resultados aparecerán con el nombre que se identificó la variable. Permite establecer nombres más lógicos para las variables.
6. *Value* – se puede asignar un valor descriptivo (un nombre) a cada posible valor de la variable. Útil para variables “dummies” o variables con categorías. Ejemplo: se puede definir el valor 1 como femenino y el 0 como masculino. Se pudo haber definido en *Label* esta variable como género.



7. *Missing value* – se pueden designar valores como *missing values* del usuario. Se podría distinguir entre los datos que faltan porque el observado se negó a contestar o porque no le aplicaba la pregunta. Aquellos que se identifiquen como *missing value* se excluyen de los análisis estadísticos. (Para un resumen de lo que se considera *missing value* ver páginas 6 y 7).



8. *Align* – determina el lugar (izquierda, derecha o centro) en la celda de las columnas en donde aparecen los valores.

9. *Measure* – define la escala de las variables (ordinal, nominal o escalar).

MENÚ DE LOS DATOS

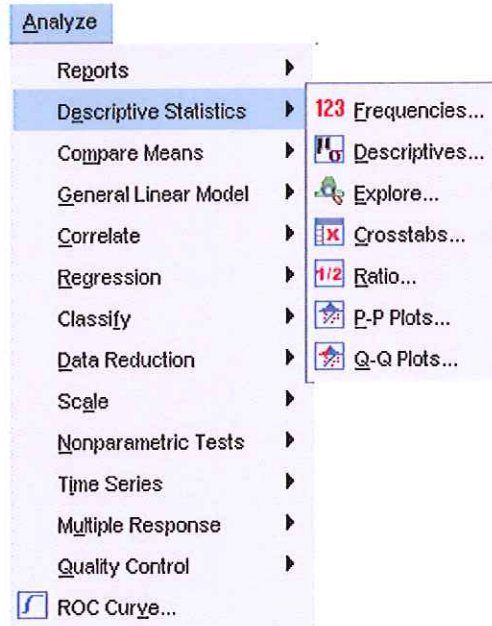
El menú de *Data* es muy útil. Contiene comandos para trabajar con los datos de una manera global. Los cambios que se hacen con los comandos de este menú son temporeros y no afectan el archivo original de los datos a menos que se guarden los cambios.

MENÚ DE TRANSFORMAR

El menú de transformar permite hacer cambios a variables seleccionadas y calcular nuevas variables con las variables ya existentes. Los cambios que se hacen con los comandos de este menú son temporeros y no afectan el archivo original de los datos a menos que se guarden los cambios.

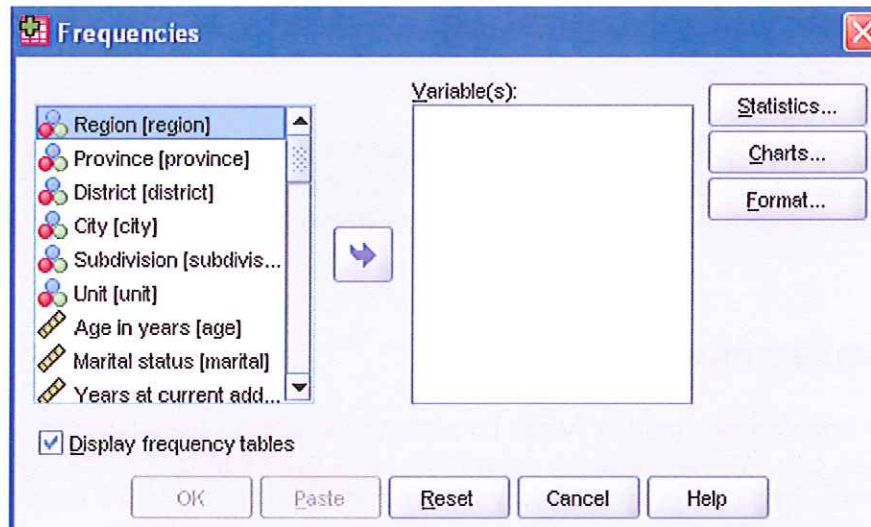
ESTADÍSTICAS DESCRIPTIVAS

SPSS ofrece un menú para calcular las estadísticas descriptivas bajo *Analyze*:

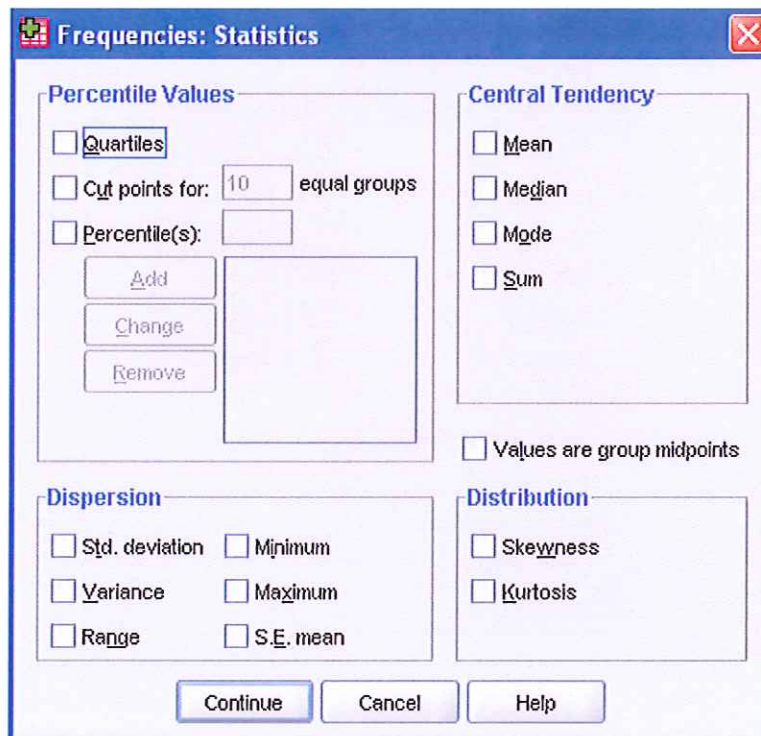


Frecuencias

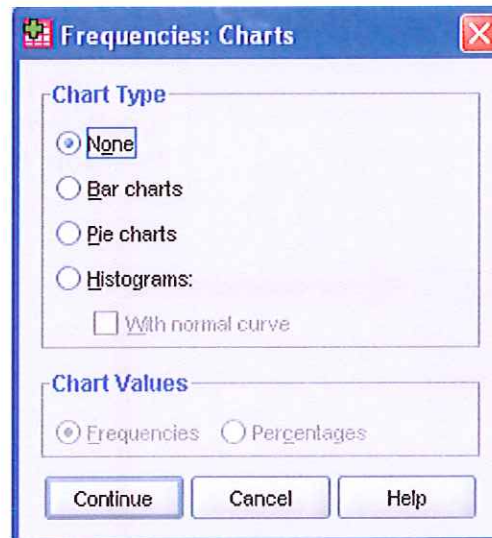
PASO #1 – SELECCIONAR LAS VARIABLES.



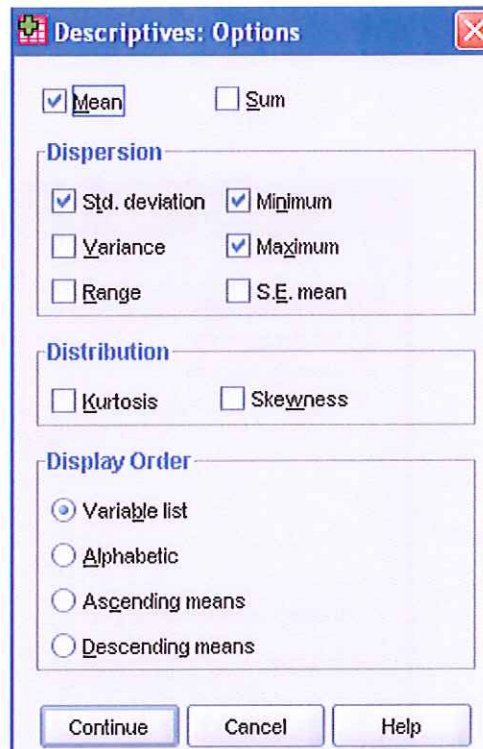
PASO #2 – SELECCIONAR EL TIPO DE ESTADISTICA.



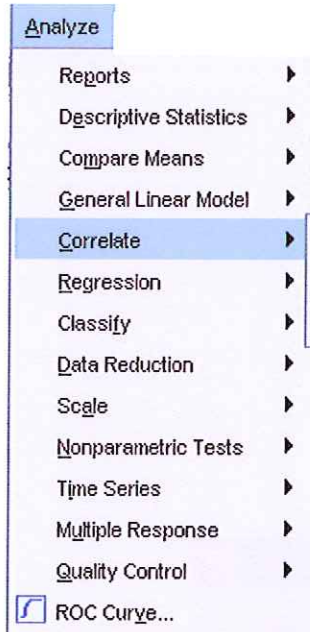
PASO #3 – SE PUEDEN SELECCIONAR GRAFICAS:



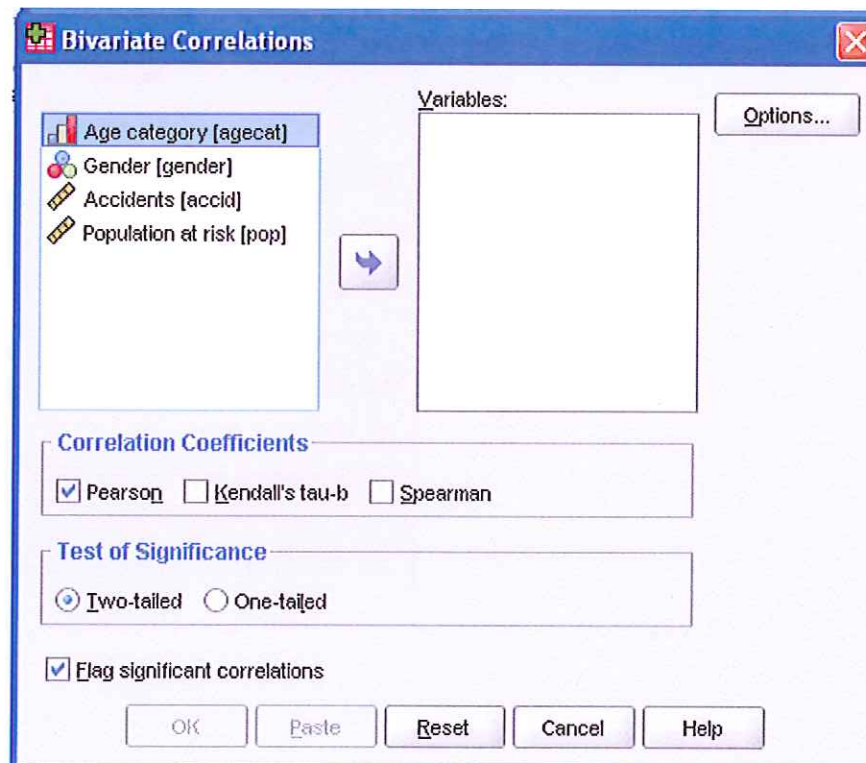
Descriptivas:



CORRELACIONES



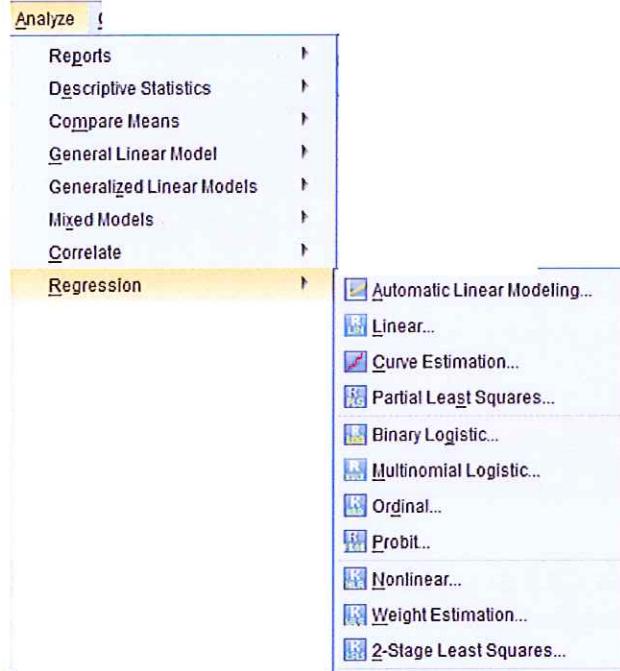
Permite calcular los coeficientes de correlación de Pearson, Kendall's y Spearman



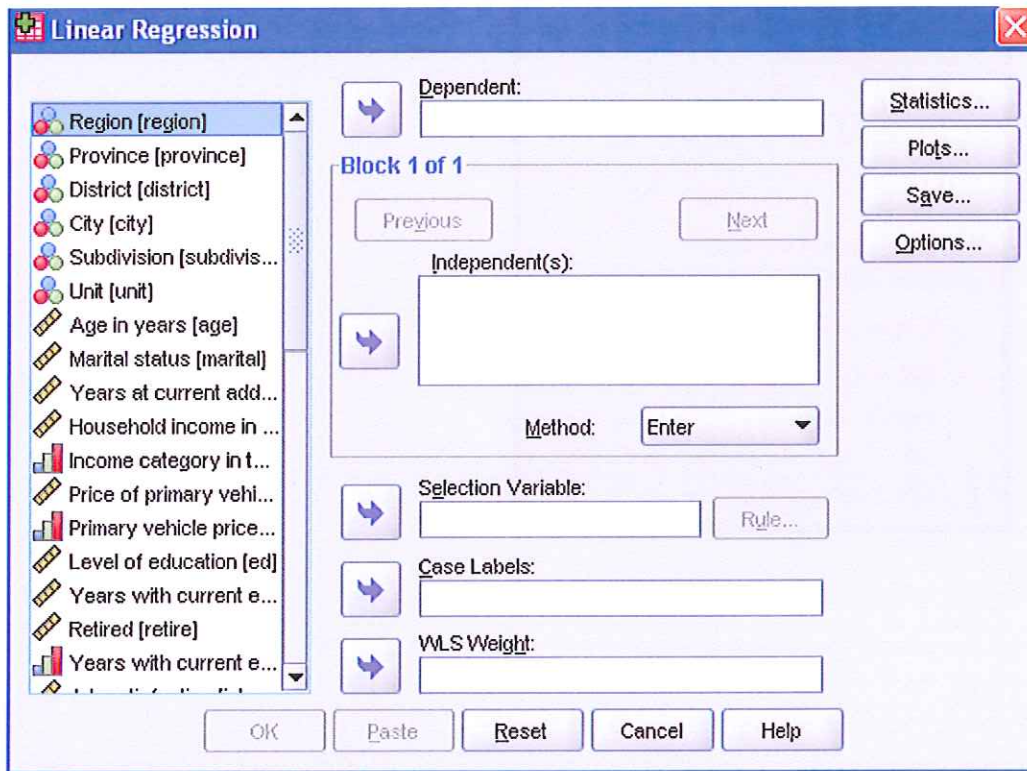
PRUEBAS ESTADISTICAS

REGRESIONES

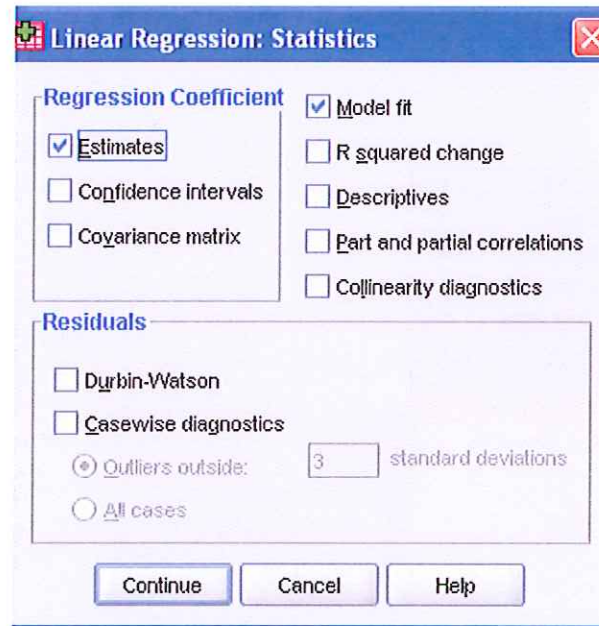
PASO #1



PASO #2

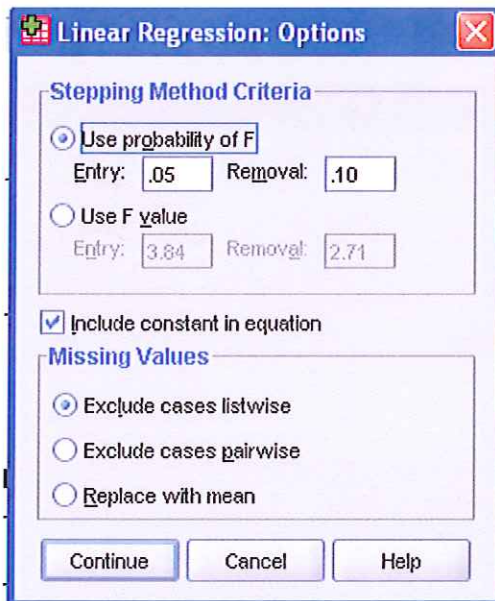


PASO #3



PASO #4 – SELECCIONAR *PLOTS* SI LO DESEA. SPSS tiene opciones de diferentes *scatter plots* aplicables a las regresiones.

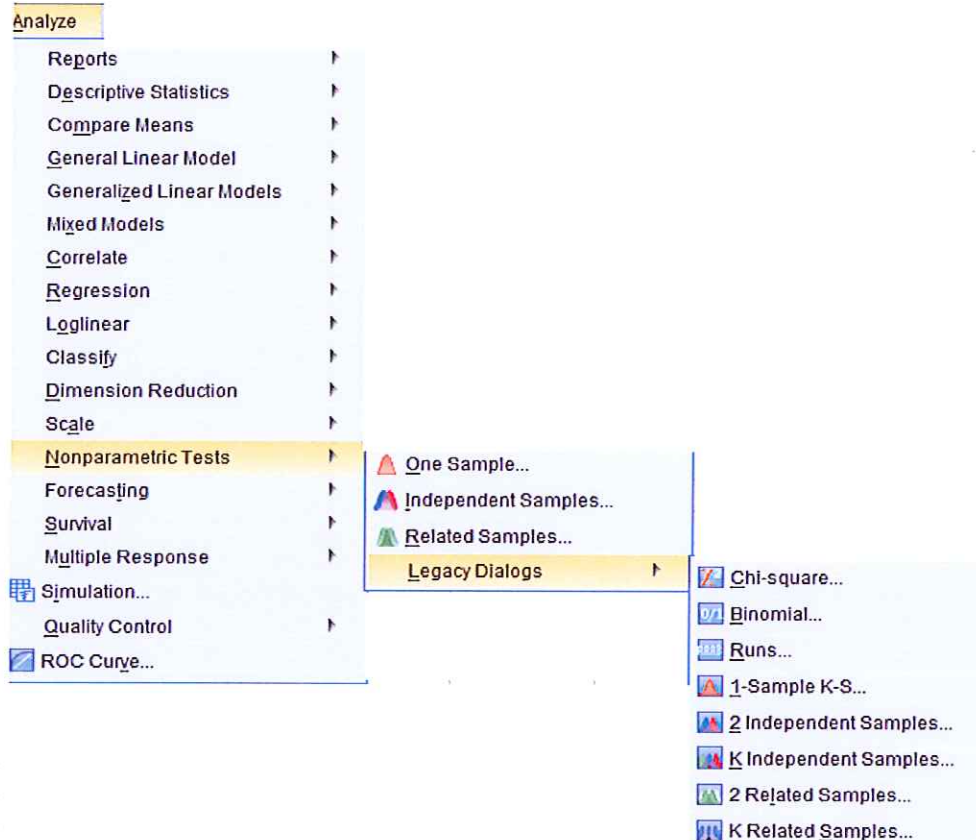
Menú de opciones:



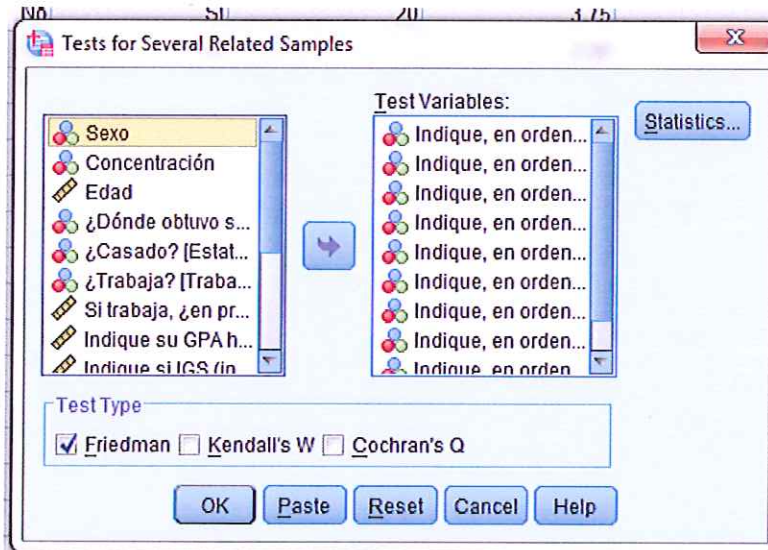
PRUEBA NO PARAMETRICA – “RANKING”

En el caso de los datos que se ordenen por preferencia se analizan con la Prueba de Friedman. La hipótesis nula (H_0) es que no hay diferencia estadística en el orden de las respuestas. La hipótesis alternativa (H_1) es que existe diferencia estadística en el orden de las respuestas. Para cada caso (cuestionario), las variables k se ordenan del 1 a k . La prueba estadística se basa en este orden.

PASO #1



PASO #2



Referencias:

Hair , J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (1998). *Multivariate Data Analysis (5th ed.)*. New Jersey: Prentice Hall.

Peterson, R. A. (1994). A Meta-analysis of Cronbach's Coefficient Alpha, *Journal of Consumer Research*, Vol. 21, 381 – 391.

SPSS 22.0 for Windows (Computer Software). (2013). SPSS, Inc.