



ugr

Universidad  
de Granada

# Máster en Economía y Organización de empresas

Módulo III: Competencias para la preparación de trabajo  
fin de Máster

Dr. Eulogio Cordon Pozo



# ÍNDICE DE CONTENIDOS

1. Paquetes estadísticos disponibles
2. ¿Qué es la estadística? Estadística descriptiva vs. inferencial. Estadística paramétrica vs. no paramétrica.
3. Principales técnicas de análisis multivariante: objetivos y condiciones de uso.
4. Antes explorar... luego actuar
5. Evaluación de la validez y fiabilidad de una escala de medida:
  1. Análisis factorial exploratorio y confirmatorio.
  2. Coeficiente alfa de cronbach.
6. Bibliografía básica recomendada.



# SOFTWARE ESTADÍSTICO PARA ANÁLISIS DATOS

## Software propietario

BMDP (<http://www.statistical-solutions-software.com/products-page/bmdp-statistical-software/>)

EIEWS ver. 7 (<http://www.eviews.com/>)

SAS ver. 9.3 (<http://www.sas.com/software/sas9/#section=3>)

SPSS (IBM) ver. 20 (<http://www-01.ibm.com/software/es/analytics/spss/>)

STATA ver. 12 (<http://www.stata.com>)

STATISTICA ver. 10 (<http://www.statsoft.com/>)

STATGRAPHICS Centurion XVI (<http://www.statgraphics.net>)

## Software libre

ADaMSoft (<http://adamsoft.caspar.it/English/ADaMSoft4.html>)

Dataplot (<http://www.itl.nist.gov/div898/software/dataplot/homepage.htm>)

EPIINFO (<http://www.epidemiologia.vet.ulpgc.es/software.html>)

GRETl ([http://gretl.sourceforge.net/gretl\\_espanol.html](http://gretl.sourceforge.net/gretl_espanol.html))

R + RCommander (<http://www.r-project.org/>; <http://cran.r-project.org/web/packages/Rcmdr/index.html>)

Más software estadístico libre: <http://www.statisticalconsultants.co.nz/statssoftware.html>



# ¿QUÉ ES LA ESTADÍSTICA?

La Estadística es una Ciencia derivada de las matemáticas que estudia los métodos científicos para **recoger, organizar, resumir y analizar** datos, así como para **sacar conclusiones válidas y tomar decisiones** razonables basadas en dicho análisis.

# ESTADÍSTICA DESCRIPTIVA VS. INFERENCIAL

## ■ Estadística descriptiva o deductiva

Necesaria....  
pero insuficiente

## ■ Estadística inferencial o inductiva

- ✓ Permite detectar interrelaciones entre datos, las leyes que las rigen (eliminando las influencias del azar).
- ✓ Provee conclusiones o inferencias basándose en los datos simplificados y analizados. Permite realizar inferencias sobre la base de la muestra estudiada (sacar conclusiones en cuanto al universo o población, de donde se obtuvo dicha muestra)



# ESTADÍSTICA DESCRIPTIVA

Escala de medida	Medidas de posición central	Medidas de dispersión
Nominal	Moda	
Ordinal	Mediana	Rango intercuartílico
Intervalo	Media	Desviación típica
Razón	Media geométrica	Coeficiente de variación

# ESTADÍSTICA PARAMÉTRICA VS. NO PARAMÉTRICA

- La Estadística no paramétrica es una rama de la Estadística que estudia las pruebas y modelos estadísticos cuya distribución subyacente **no se ajusta a los llamados criterios paramétricos**.
- **La utilización de estos métodos se hace recomendable cuando** no se puede asumir que los datos se ajusten a una distribución normal, cuando el nivel de medida empleado no sea, como mínimo, de intervalo, o cuando el tamaño de la muestra es reducido.
- **Ejemplos de pruebas no paramétricas son:** prueba de la Chi-cuadrado, coeficiente de correlación de Spearman, prueba de Mann-Whitney, prueba de Wilcoxon, prueba de Kolmogorov-Smirnov, prueba de Kruskal-Wallis, etc.



# PRINCIPALES TEST NO PARAMÉTRICOS

Objetivos	Tests no paramétricos recomendados	SPSS
Comprobar la independencia de variables (escala nominal). Tablas de contingencia	Prueba de la Chi cuadrado de Pearson	Analizar / Estadísticos descriptivos / Tablas de Contingencia
Verificar si una distribución es normal	Kolmogórov-Smirnow Shapiro-Wilk	Analizar / Estadísticos descriptivos / Explorar (Gráficos)
Comprobar la existencia de diferencias entre grupos independientes (dos grupos)	U de Mann-Whitney	Analizar / Pruebas no paramétricas / 2 Muestras independientes
Comprobar la existencia de diferencias entre grupos independientes (más de dos grupos)	Kruskal-Wallis	Analizar / Pruebas no paramétricas / K Muestras independientes
Grado de correlación entre variables	Rho de Spearman Tau de Kendall	Analizar / Correlaciones

## EJEMPLO APLICADO

1º) Abrir el fichero noparametricas.sav (7 variables y 495 casos)

SALARIO-ESTUDIOS-SEXO: datos encuestado

GASTOID-INNOVA-EMPLEADOS-VENTAS: datos empresa

2º) Estudiar la asociación entre la variable ESTUDIOS y SEXO, SEXO y SALARIO

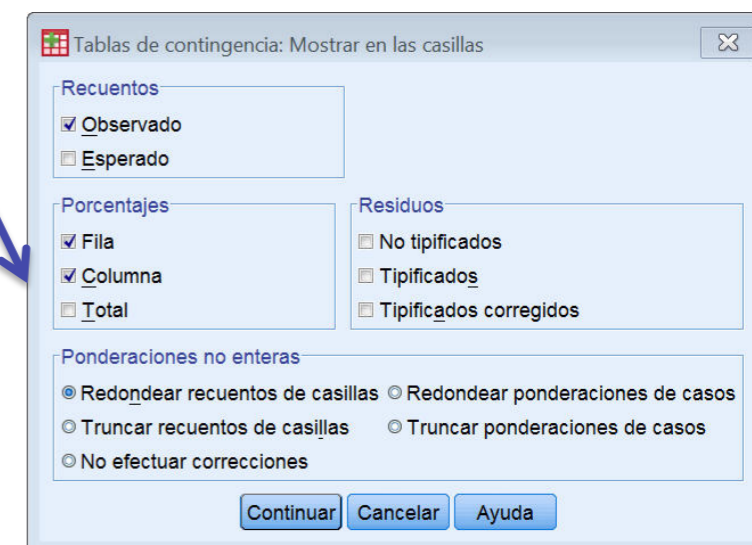
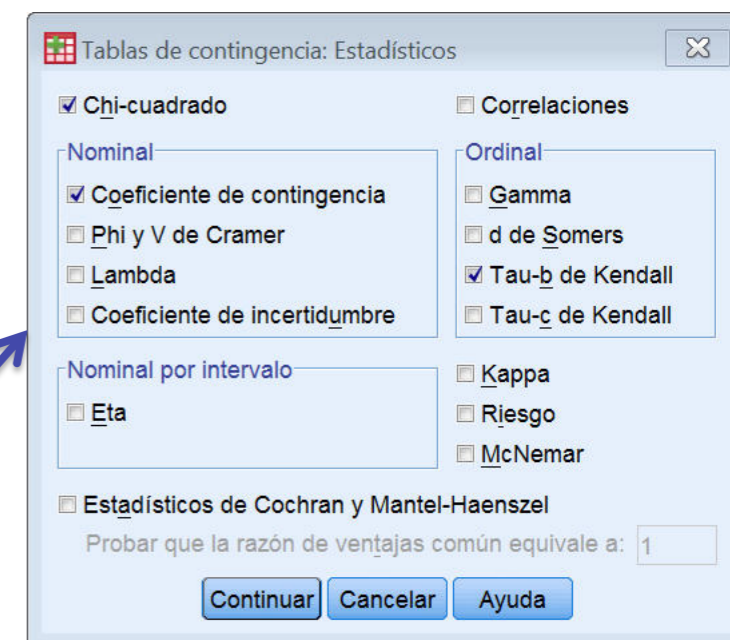
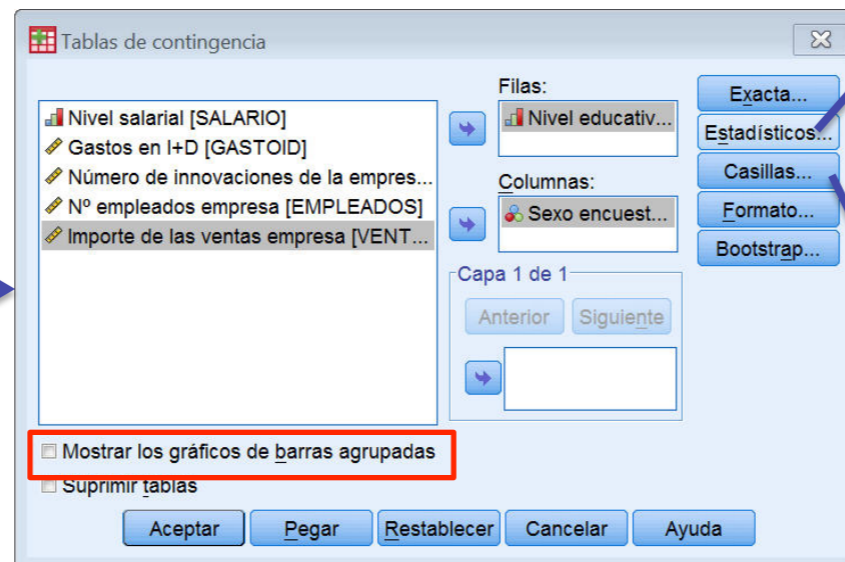
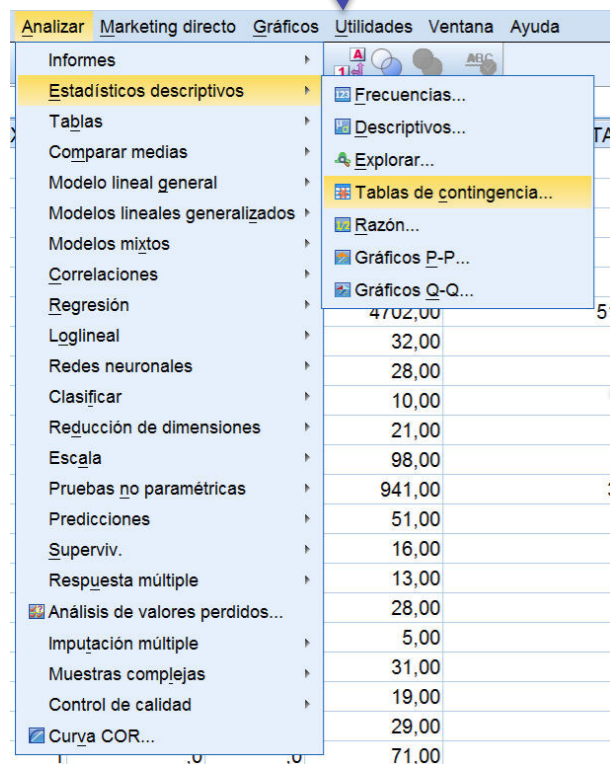
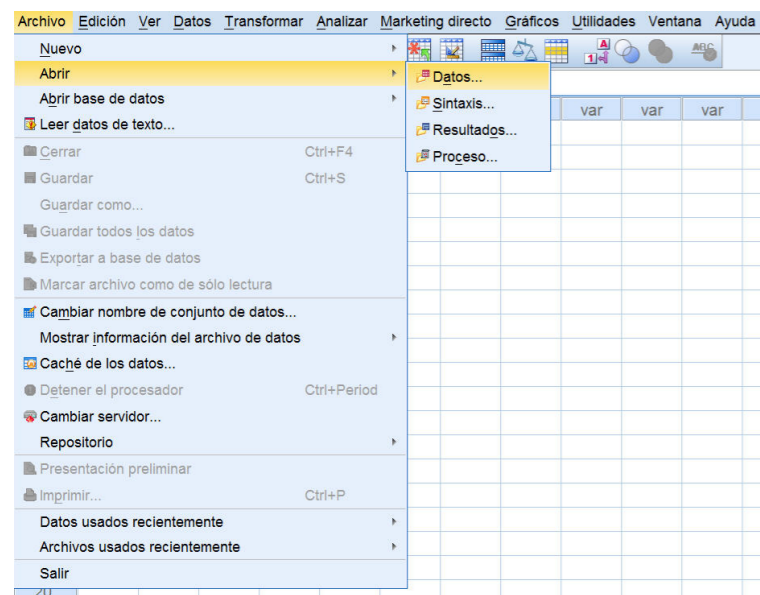
3º) Calcular el coeficiente de correlación entre las variables GASTOID, INNOVA,EMPLEADOS,VENTAS.

4º) Determinar qué relación existe entre SEXO y SALARIO

5º) Conclusiones de los análisis estadísticos realizados

## EJEMPLO APLICADO

### Asociación entre ESTUDIOS-SEXO (Tablas de contingencia)





## EJEMPLO APLICADO

### Asociación entre ESTUDIOS-SEXO (Tablas de contingencia)

#### → Tablas de contingencia

[Conjunto\_de\_datos1] Z:\Dropbox\Doctorado\Master Gerontologia\Metodologia investiga

Resumen del procesamiento de los casos

	Casos					
	Válidos		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Nivel educativo * Sexo encuestados	494	99,8%	1	,2%	495	100,0%

Tabla de contingencia Nivel educativo \* Sexo encuestados

			Sexo encuestados		Total
			Hombre	Mujer	
Nivel educativo	Primarios	Recuento	38	18	56
		% dentro de Nivel educativo	67,9%	32,1%	100,0%
		% dentro de Sexo encuestados	14,0%	8,1%	11,3%
	Medios	Recuento	153	135	288
		% dentro de Nivel educativo	53,1%	46,9%	100,0%
		% dentro de Sexo encuestados	56,5%	60,5%	58,3%
	Superiores	Recuento	80	70	150
		% dentro de Nivel educativo	53,3%	46,7%	100,0%
		% dentro de Sexo encuestados	29,5%	31,4%	30,4%
	Total	Recuento	271	223	494
		% dentro de Nivel educativo	54,9%	45,1%	100,0%
		% dentro de Sexo encuestados	100,0%	100,0%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	4,311 <sup>a</sup>	2	,116
Razón de verosimilitudes	4,425	2	,109
Asociación lineal por lineal	1,961	1	,161
N de casos válidos	494		

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5.  
La frecuencia mínima esperada es 25,28.

Medidas simétricas

		Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Nominal por nominal	Coefficiente de contingencia	,093			,116
Ordinal por ordinal	Tau-b de Kendall	,055	,043	1,276	,202
N de casos válidos		494			

a. Asumiendo la hipótesis alternativa.

b. Empleando el error típico asintótico basado en la hipótesis nula.

Ninguno de los test es  
significativo ( $p > 0.05$ )

## EJEMPLO APLICADO

### Asociación entre SEXO-SALARIO (Tablas de contingencia)

Tabla de contingencia Nivel salarial \* Sexo encuestados

			Sexo encuestados		Total
			Hombre	Mujer	
Nivel salarial	Inferior a la de la media de la categoría profesional	Recuento	25	32	57
		% dentro de Nivel salarial	43,9%	56,1%	100,0%
		% dentro de Sexo encuestados	9,2%	14,3%	11,5%
	En torno a la media de la categoría profesional	Recuento	156	141	297
		% dentro de Nivel salarial	52,5%	47,5%	100,0%
		% dentro de Sexo encuestados	57,6%	63,2%	60,1%
	Mayor a la media de la categoría profesional	Recuento	90	50	140
		% dentro de Nivel salarial	64,3%	35,7%	100,0%
		% dentro de Sexo encuestados	33,2%	22,4%	28,3%
	Total	Recuento	271	223	494
		% dentro de Nivel salarial	54,9%	45,1%	100,0%
		% dentro de Sexo encuestados	100,0%	100,0%	100,0%

Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)
Chi-cuadrado de Pearson	8,462 <sup>a</sup>	2	,015
Razón de verosimilitudes	8,538	2	,014
Asociación lineal por lineal	8,343	1	,004
N de casos válidos	494		

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5.  
La frecuencia mínima esperada es 25,73.

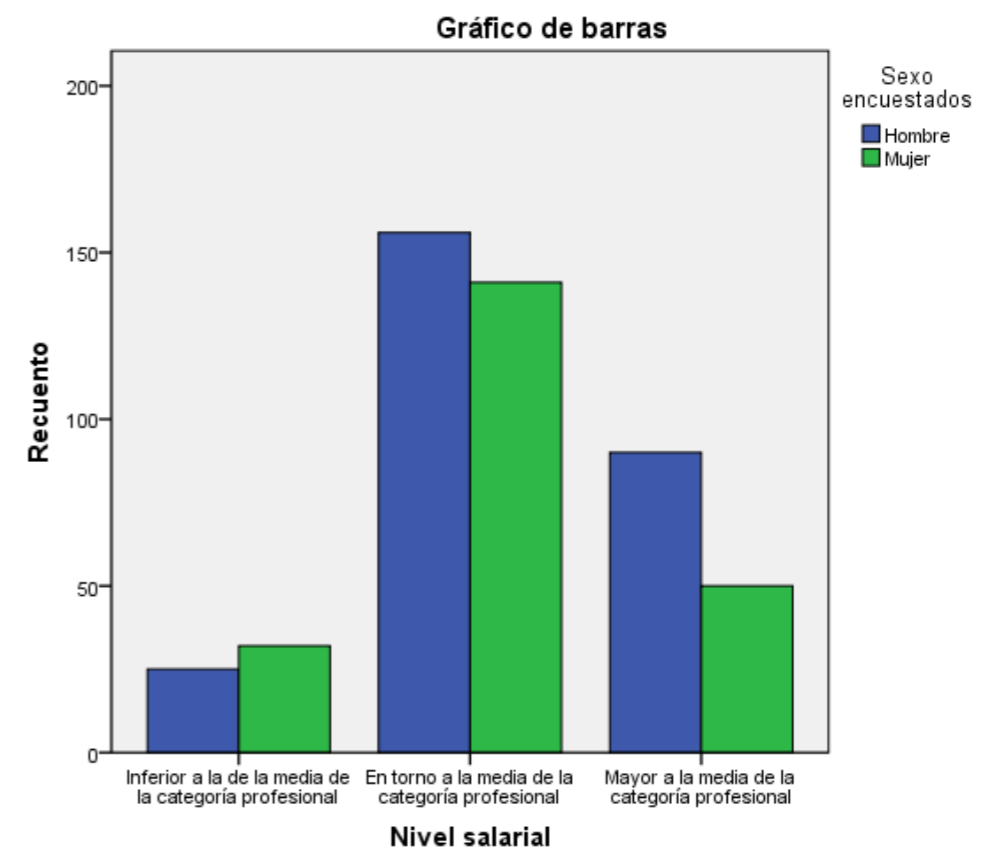
Los test son significativos  
( $p < 0.05$ )

Medidas simétricas

	Valor	Error típ. asint. <sup>a</sup>	T aproximada <sup>b</sup>	Sig. aproximada
Nominal por nominal Coeficiente de contingencia	,130			,015
Ordinal por ordinal Tau-b de Kendall	-,126	,042	-2,954	,003
N de casos válidos	494			

a. Asumiendo la hipótesis alternativa.

b. Empleando el error típico asintótico basado en la hipótesis nula.





## EJEMPLO APLICADO

### Correlaciones variables métricas (ratio o intervalo)

Analizar	Marketing directo	Gráficos	Utilidades	Ventana	Ayuda
Informes					
Estadísticos descriptivos					
Tablas					
Comparar medias					
Modelo lineal general					
Modelos lineales generalizados					
Modelos mixtos					
Correlaciones					
Regresión					
Loglineal					
Redes neuronales					
Clasificar					
Reducción de dimensiones					
Escala					
Pruebas no paramétricas					
Predicciones					
Superviv.					
Respuesta múltiple					
Análisis de valores perdidos...					
Imputación múltiple					
Muestras complejas					
Control de calidad					
Curva COR...					

Estadísticos descriptivos

	Media	Desviación típica	N
Gastos en I+D	7545,2727	43016,73617	495
Número de innovaciones de la empresa (producto)	1,0848	7,16639	495
Nº empleados empresa	279,7111	742,35215	495
Importe de las ventas empresa	9,1508E7	4,08994E8	495

Correlaciones

		Gastos en I+D	Número de innovaciones de la empresa (producto)	Nº empleados empresa	Importe de las ventas empresa
Gastos en I+D	Correlación de Pearson	1	,192**	,261**	,343**
	Sig. (bilateral)		,000	,000	,000
	N	495	495	495	495
Número de innovaciones de la empresa (producto)	Correlación de Pearson	,192**	1	,046	,058
	Sig. (bilateral)	,000		,306	,194
	N	495	495	495	495
Nº empleados empresa	Correlación de Pearson	,261**	,046	1	,884**
	Sig. (bilateral)	,000	,306		,000
	N	495	495	495	495
Importe de las ventas empresa	Correlación de Pearson	,343**	,058	,884**	1
	Sig. (bilateral)	,000	,194	,000	
	N	495	495	495	495

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

Correlaciones bivariadas

Variables:

Nivel salarial [SAL...]  
Nivel educativo [E...]  
Sexo encuestados ...

Gastos en I+D [GA...]  
Número de innovac...  
Nº empleados emp...  
Importe de las vent...

Opciones...  
Bootstrap...

Coefficientes de correlación

☒ Pearson ☐ Tau-b de Kendall ☐ Spearman

Prueba de significación

☒ Bilateral ☐ Unilateral

☒ Marcar las correlaciones significativas

Aceptar Pegar Restablecer Cancelar Ayuda

Correlaciones bivariadas: Opciones

Estadísticos

☒ Medias y desviaciones típicas

☐ Productos cruzados diferenciales y covarianzas

Valores perdidos

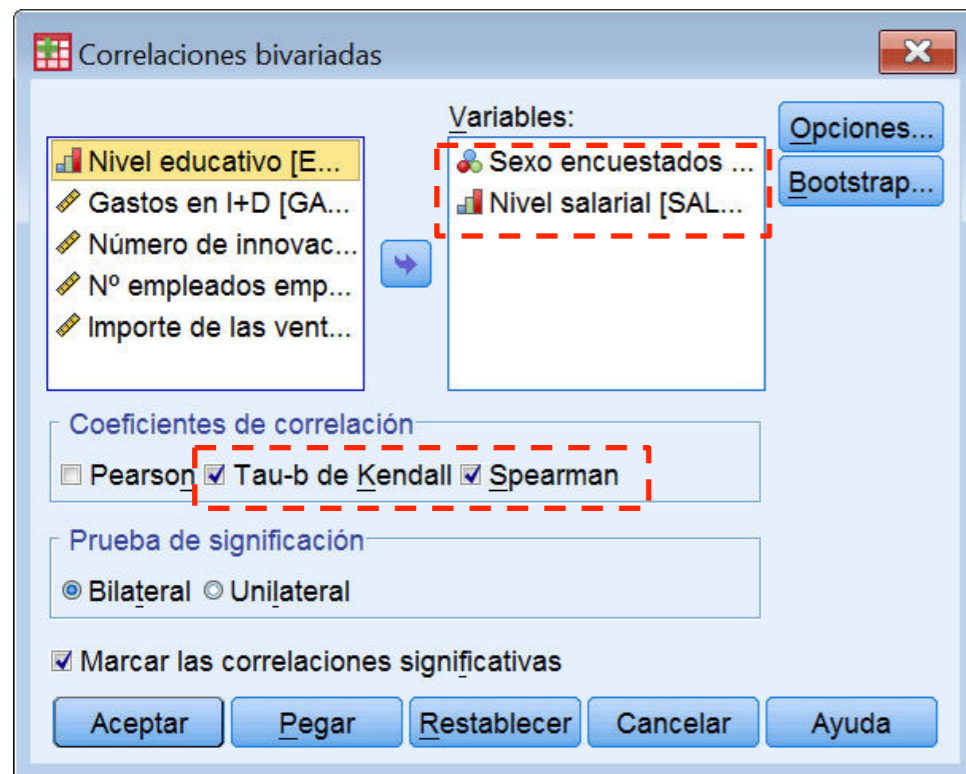
☒ Excluir casos según pareja

☐ Excluir casos según lista

Continuar Cancelar Ayuda

## EJEMPLO APLICADO: RELACIÓN SEXO-SALARIO

Correlaciones variables no métricas (nominal u ordinal)



Correlaciones			Sexo encuestados	Nivel salarial
Tau_b de Kendall	Sexo encuestados	Coeficiente de correlación	1,000	-,126**
		Sig. (bilateral)	.	,004
		N	494	494
	Nivel salarial	Coeficiente de correlación	-,126**	1,000
		Sig. (bilateral)	,004	.
		N	494	495
Rho de Spearman	Sexo encuestados	Coeficiente de correlación	1,000	-,131**
		Sig. (bilateral)	.	,004
		N	494	494
	Nivel salarial	Coeficiente de correlación	-,131**	1,000
		Sig. (bilateral)	,004	.
		N	494	495

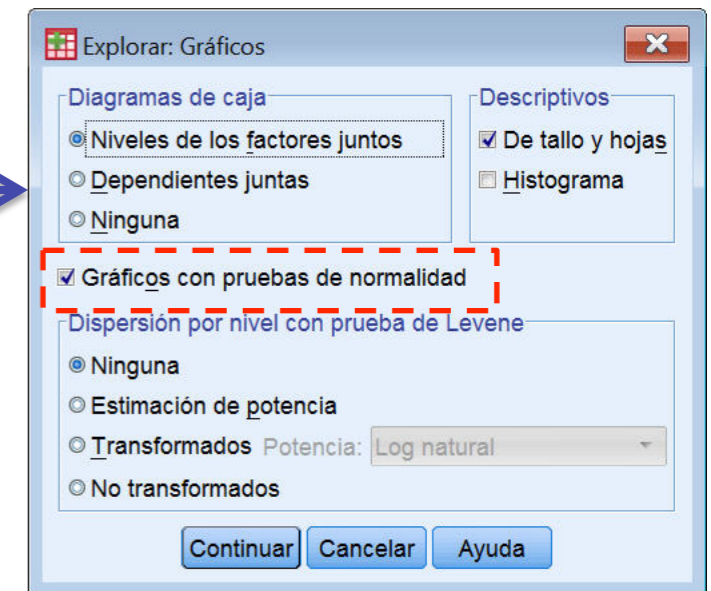
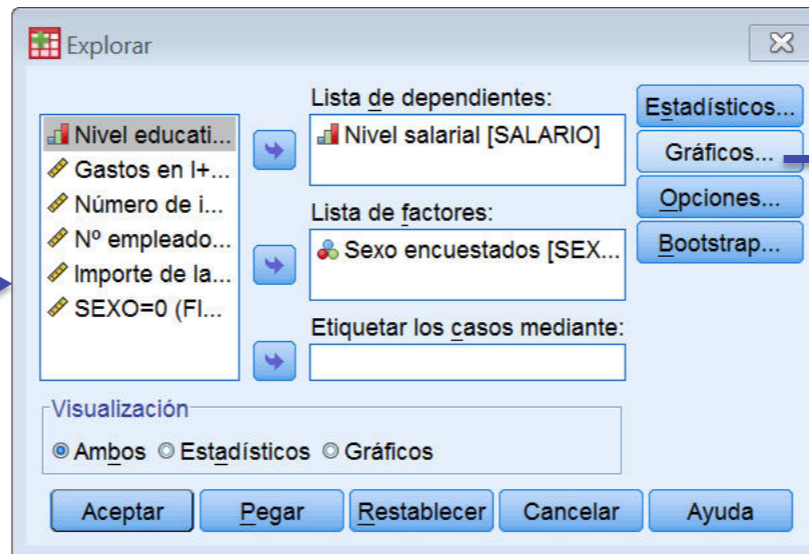
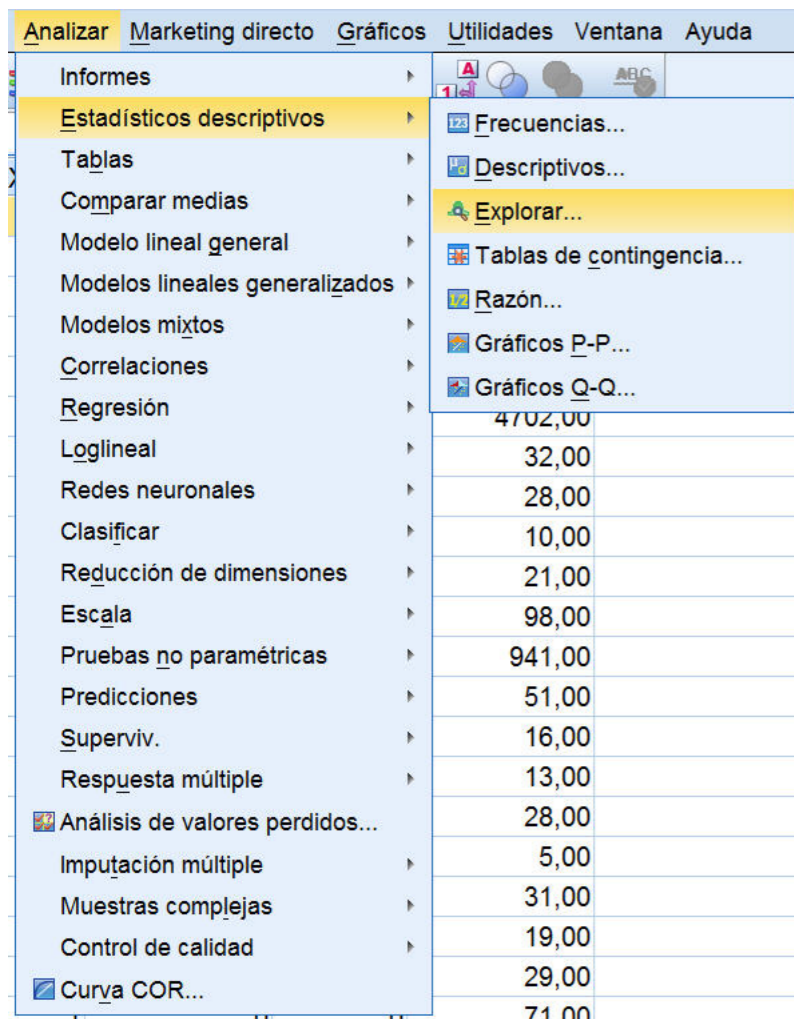
\*\* . La correlación es significativa al nivel 0,01 (bilateral).

La correlación es **NEGATIVA** y estadísticamente significativa ( $p < 0.01$ )  
¿Cómo se interpretaría este dato?

## EJEMPLO APLICADO: RELACIÓN SEXO-SALARIO

¿Hay diferencias estadísticamente significativas entre los niveles salariales de hombres y mujeres? (¿qué test es el apropiado?)

	Paramétricos	No paramétricos
2 grupos	t-test	U Mann-Whitney
Más de 2 grupos	ANOVA	Kruskal-Wallis





## EJEMPLO APLICADO: RELACIÓN SEXO-SALARIO

Descriptivos

Sexo encuestados			Estadístico	Error típ.
Nivel salarial	Hombre	Media	2,24	,037
		Intervalo de confianza para la media al 95%	Límite inferior	2,17
			Límite superior	2,31
		Media recortada al 5%	2,27	
		Mediana	2,00	
		Varianza	,368	
		Desv. típ.	,607	
		Mínimo	1	
		Máximo	3	
		Rango	2	
		Amplitud intercuartil	1	
		Asimetría	-,172	,148
		Curtosis	-,529	,295
	Mujer	Media	2,08	,040
		Intervalo de confianza para la media al 95%	Límite inferior	2,00
			Límite superior	2,16
		Media recortada al 5%	2,09	
		Mediana	2,00	
		Varianza	,363	
		Desv. típ.	,602	
		Mínimo	1	
		Máximo	3	
		Rango	2	
		Amplitud intercuartil	0	
		Asimetría	-,034	,163
		Curtosis	-,251	,324

Estadísticas  
descriptivas de interés

Pruebas de normalidad

Sexo encuestados		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Estadístico	gl	Sig.	Estadístico	gl	Sig.
Nivel salarial	Hombre	,322	271	,000	,761	271	,000
	Mujer	,329	223	,000	,765	223	,000

a. Corrección de la significación de Lilliefors

Pruebas de normalidad  
( $p < 0.01$ )



## EJEMPLO APLICADO: RELACIÓN SEXO-SALARIO

Analizar Marketing directo Gráficos Utilidades Ventana Ayuda

Informes

Estadísticos descriptivos

Tablas

Comparar medias

Modelo lineal general

Modelos lineales generalizados

Modelos mixtos

Correlaciones

Regresión

Loglineal

Redes neuronales

Clasificar

Reducción de dimensiones

Escala

Pruebas no paramétricas

Predicciones

Superviv.

Respuesta múltiple

Análisis de valores perdidos...

Imputación múltiple

Muestras complejas

Control de calidad

Curva COR...

EMPLEADOS VENTAS filter\_\$ va

387,00 58812522,00 0

61,00 10119732,00 0

150,00 12103968,00 0

221,00 32143823,00 0

31,00 2470925,00 0

4702,00 5134541594,00 0

32,00 2671179,00 0

28,00 1848016,00 0

10,00 509553,00 0

21,00 1475237,00 0

98,00 24136729,00 0

302168202,00 0

9559040,00 0

2332051,00 0

20,00

5,00

31,00

19,00

29,00

71,00

15,00

25,00

35,00

Chi-cuadrado...

Binomial...

Rachas...

K-S de 1 muestra...

2 muestras independientes...

K muestras independientes...

2 muestras relacionadas...

K muestras relacionadas...

Pruebas para dos muestras independientes

Nivel educativo [EST...]

Gastos en I+D [GAS...]

Número de innovacio...

Nº empleados empre...

Importe de las ventas...

SEXO=0 (FILTER) [fi...]

Nivel salarial [SALARIO]

SEXO(0 1)

Definir grupos...

Tipo de prueba

☒ U de Mann-Whitney

☐ Z de Kolmogorov-Smirnov

☐ Reacciones extremas de Moses

☐ Rachas de Wald-Wolfowitz

Aceptar Pegar Restablecer Cancelar Ayuda

Dos muestras independie...

Estadísticos

☒ Descriptivos ☒ Cuartiles

Valores perdidos

☒ Excluir casos según prueba

☐ Excluir casos según lista

Continuar Cancelar Ayuda

### Prueba de Mann-Whitney

#### Rangos

Sexo encuestados		N	Rango promedio	Suma de rangos
Nivel salarial	Hombre	271	262,24	71066,00
	Mujer	223	229,59	51199,00
	Total	494		

#### Estadísticos de contraste<sup>a</sup>

	Nivel salarial
U de Mann-Whitney	26223,000
W de Wilcoxon	51199,000
Z	-2,904
Sig. asintót. (bilateral)	,004

a. Variable de agrupación: Sexo encuestados

Existen diferencias significativas  
( $p < 0.001$ )

## ANÁLISIS MULTIVARIANTE: CONCEPTO Y APLICACIÓN

- **Definición amplia:** “conjunto de métodos que analizan las relaciones entre varias medidas (variables), tomadas sobre cada objeto o unidad de análisis, en una o más muestras simultáneamente”
- **Característica principal:** tratan con relaciones simultáneas entre variables.
- **¿Cuáles son las potenciales áreas de aplicación?**
  - **Reducción de datos:** simplificar la estructura del fenómeno estudiado
  - **Clasificación y agrupación:** crear grupos de objetos, tipologías, etc
  - **Análisis de relaciones de dependencia entre variables,** ya sea con propósito de explicación o de predicción
  - **Construcción de modelos y pruebas de hipótesis**

# CLASIFICACIÓN MÉTODOS MULTIVARIANTES

## 1º) Según se distinga o no entre las variables que intervienen en el análisis

- a) Métodos de interdependencia: no hay distinción entre variables (todas son de naturaleza similar). Permiten sintetizar información (FACTORIAL, PRINCALS), mostrar la estructura de los datos o establecer clasificaciones (CLUSTER).
- b) Métodos de dependencia: distinguen entre variables explicativas (independientes o predictivas) y variables a explicar (o dependientes). Son métodos de carácter explicativo (REGRESIÓN LINEAL, REGRESIÓN LOGÍSTICA, ANOVA, ANÁLISIS DISCRIMINANTE).

## 2º) Según la naturaleza de las escalas de medida de las variables

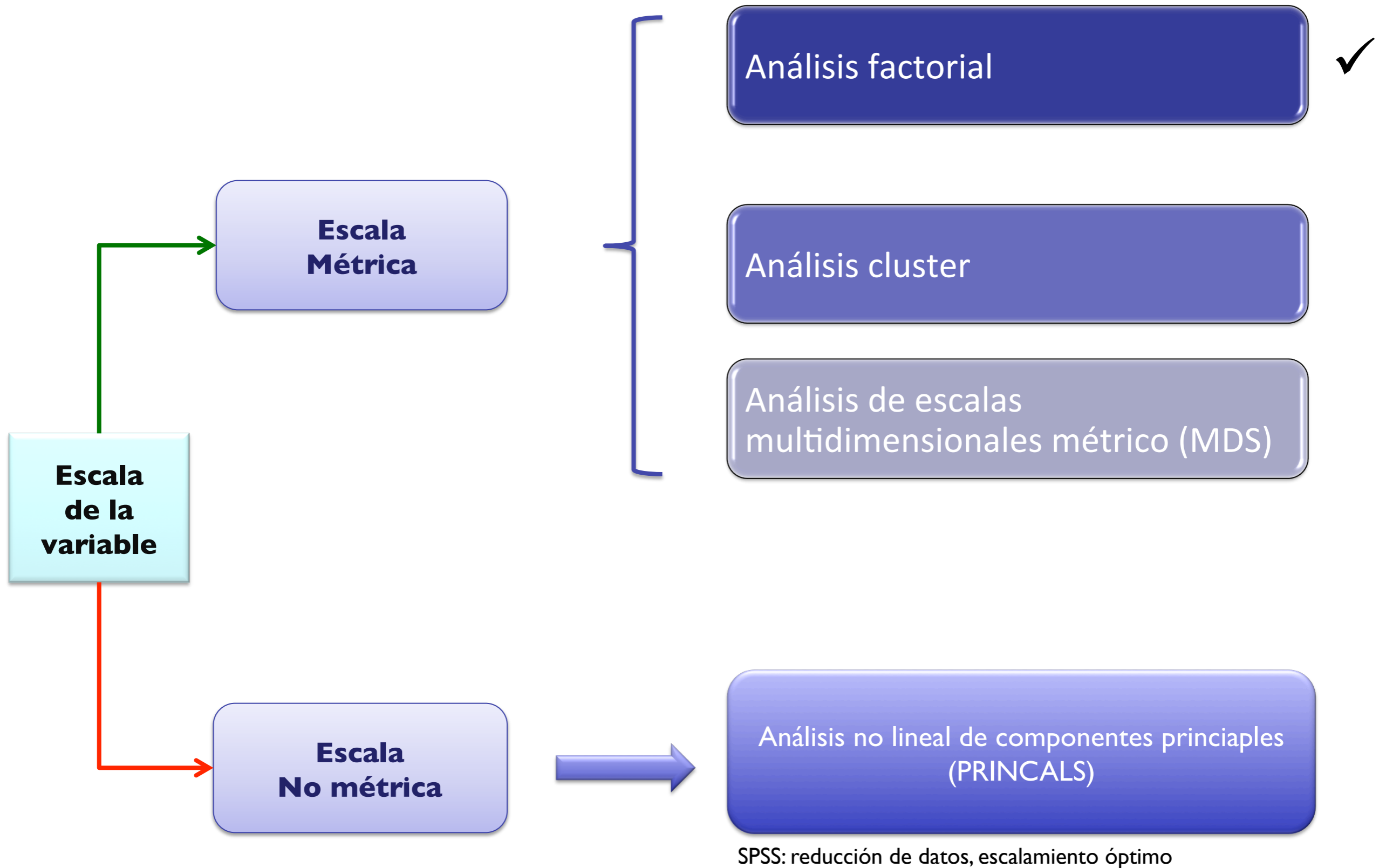
Hay métodos que requieren que todas las variables vengan medidas en una escala métrica, otras que las variables sean categóricas, y otras permiten combinar ambos tipos de escala de medida.

## 3º) Según el número de variables analizadas de manera simultánea

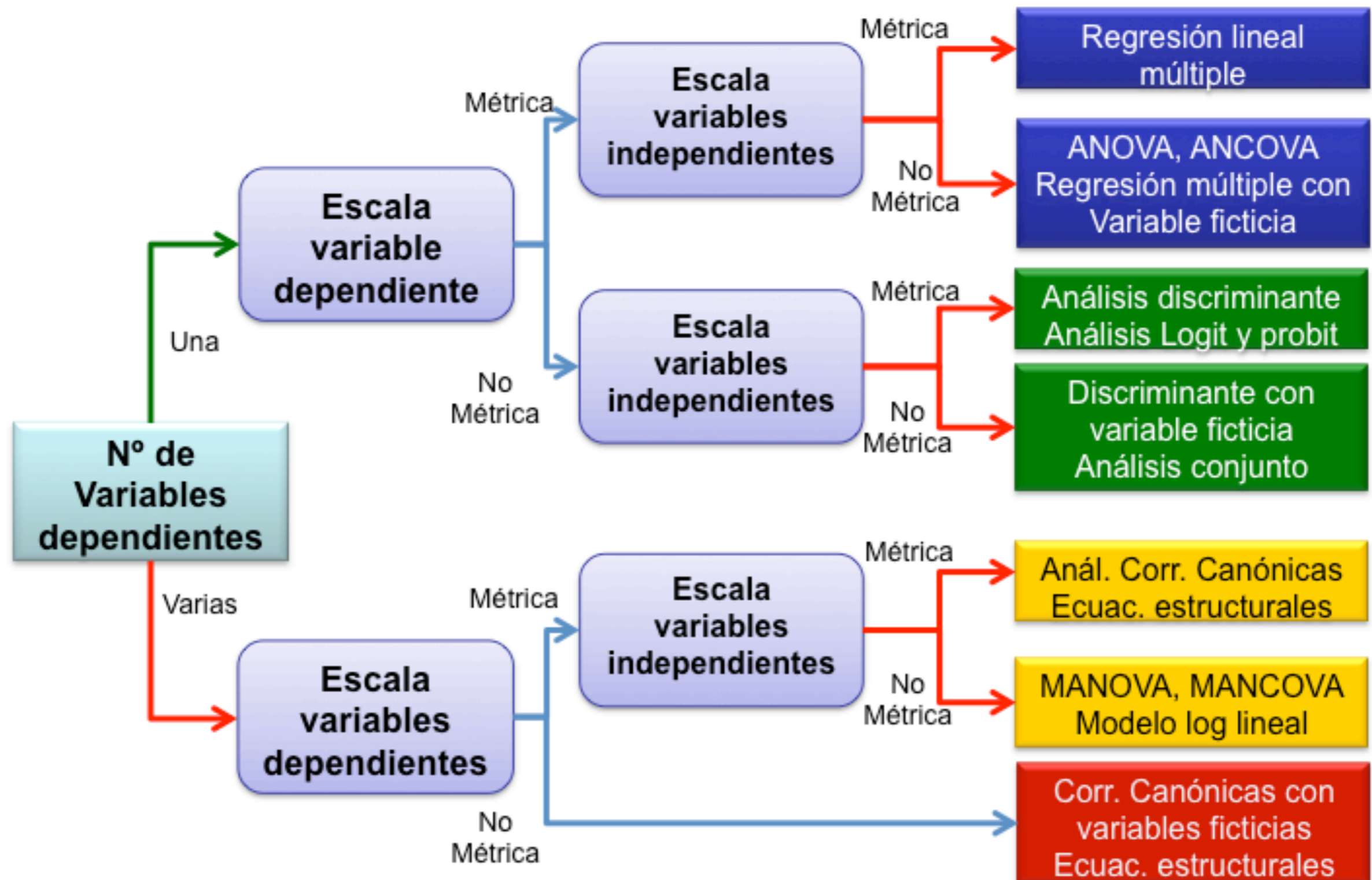
Podemos considerar métodos específicamente diseñados para casos en los que existe una/varias variables dependientes/independientes.



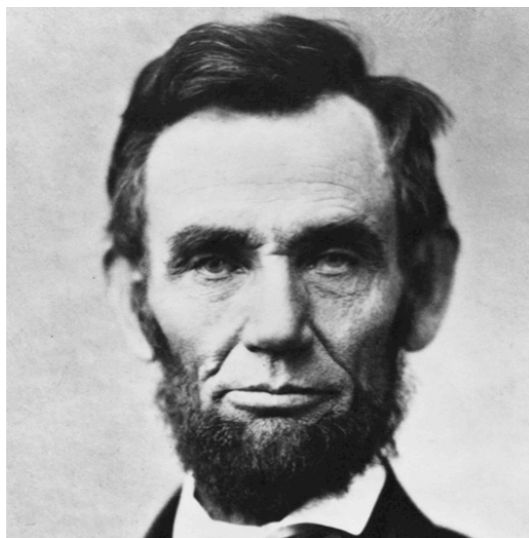
# MÉTODOS DE INTERDEPENDENCIA



# MÉTODOS DE DEPENDENCIA



Antes EXPLORAR.....  
.....Luego ACTUAR



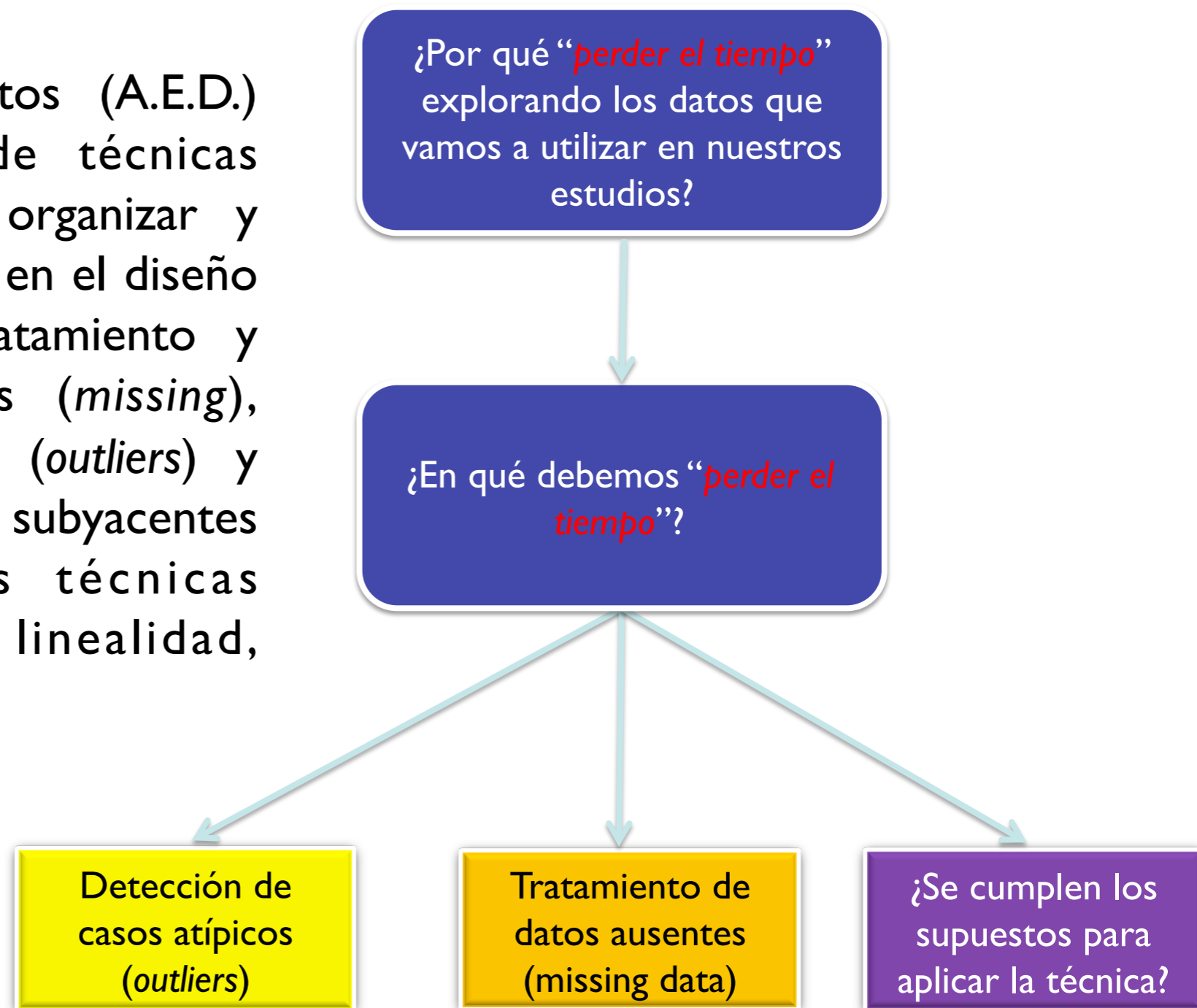
***Si dispusiera de ocho horas para cortar  
un árbol, dedicaría seis a afilar mi  
hacha***

## CUESTIONES PREVIAS A PLANTEARNOS



# EXPLORACIÓN INICIAL DE DATOS

El Análisis Exploratorio de Datos (A.E.D.) comprende un conjunto de técnicas estadísticas que nos permiten organizar y preparar los datos, detectar fallos en el diseño y recogida de los mismos, tratamiento y evaluación de datos ausentes (*missing*), identificación de casos atípicos (*outliers*) y comprobación de los supuestos subyacentes en la mayor parte de las técnicas multivariantes (normalidad, linealidad, homocedasticidad).



## CASOS ATÍPICOS Y OBSERVACIONES INFLUYENTES

- ▶ Los casos atípicos (outliers) son **observaciones con valores extremos** (notablemente diferentes de las restantes observaciones).
- ▶ Perspectiva univariante o multivariante
- ▶ A veces, pueden convertirse en observaciones influyentes que **distorsionan los resultados** (vg: relaciones no significativas, normalidad, etc.)
- ▶ Las razones por las que pueden surgir estos casos aislados son variadas, pero **debemos controlar posibles errores en la introducción de datos** (analizar frecuencias para detectar valores erróneos).

# CASOS ATÍPICOS Y OBSERVACIONES INFLUYENTES

Comprobando outliers  
Fichero: outliers.sav

- Dos variables. Escala ordinal (tipo Likert) con 7 puntos
- Analizar /Estadísticos descriptivos/Frecuencias

Variable número 1: escala 1 a 7

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	3	6,0	6,0	6,0
	2,00	6	12,0	12,0	18,0
	3,00	8	16,0	16,0	34,0
	4,00	11	22,0	22,0	56,0
	5,00	10	20,0	20,0	76,0
	6,00	4	8,0	8,0	84,0
	7,00	1	2,0	2,0	86,0
	8,00	3	6,0	6,0	92,0
	9,00	4	8,0	8,0	100,0
	Total	50	100,0	100,0	

¡ Claro error en  
la introducción  
de datos !

Variable número 2: escala 1 a 7

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	1,00	6	12,0	12,0	12,0
	2,00	11	22,0	22,0	34,0
	3,00	11	22,0	22,0	56,0
	4,00	9	18,0	18,0	74,0
	5,00	7	14,0	14,0	88,0
	6,00	4	8,0	8,0	96,0
	7,00	2	4,0	4,0	100,0
	Total	50	100,0	100,0	

# DETECCIÓN DE CASOS ATÍPICOS (UNIVARIANTE)

Utilizar el procedimiento explorar para detectar casos atípicos

The image illustrates the steps to use the 'Explorar' (Explore) procedure in SPSS for detecting univariate outliers. It shows the 'Analizar' (Analyze) menu, the 'Estadísticos descriptivos' (Descriptive Statistics) sub-menu, and the 'Explorar...' (Explore...) option. The 'Explorar' dialog box is shown with 'Dependientes:' (Dependent) set to 'Variable número 1: escala 1 a 7' and 'Variable número 2: escala 1 a 7'. The 'Mostrar' (Display) section has 'Gráficos' (Plots) selected. The 'Explorar: Gráficos' (Explore: Plots) sub-dialog box is also shown, with 'Diagramas de caja' (Boxplots) set to 'Niveles de los factores juntos' (All factors together), 'Descriptivos' (Descriptives) set to 'Tallo y hojas' (Stem-and-leaf) and 'Histograma' (Histogram), and 'Gráficos con pruebas de normalidad' (Normality tests) set to 'Ninguno' (None).

Analizar Gráficos Utilidades Ventana

Informes

Estadísticos descriptivos

Tablas

Comparar medias

Modelo lineal general

Modelos lineales generalizados

Modelos mixtos

Correlaciones

Regresión

Loglineal

Clasificar

Reducción de datos

Escalas

Pruebas no paramétricas

Serie temporales

Supervivencia

Respuesta múltiple

Análisis de valores perdidos...

Muestras complejas

Control de calidad

Curva COR...

Frecuencias...

Descriptivos...

Explorar...

Tablas de contingencia...

Razón...

Gráficos P-P...

Gráficos Q-Q...

Explorar

Dependientes:

Variable número 1: escala 1 a 7

Variable número 2: escala 1 a 7

Factores:

Etiquetar los casos mediante:

Mostrar

Ambos Estadísticos Gráficos

Estadísticos...

Gráficos...

Opciones...

Aceptar

Pegar

Restablecer

Cancelar

Ayuda

Explorar: Gráficos

Diagramas de caja

Niveles de los factores juntos

Dependientes juntas

Ninguno

Descriptivos

Tallo y hojas

Histograma

Gráficos con pruebas de normalidad

Dispersión por nivel con prueba de Levene

Ninguno

Estimación de potencia

Transformados

No transformados

Potencia: Log natural

Continuar

Cancelar

Ayuda

## DETECCIÓN DE CASOS ATÍPICOS (UNIVARIANTE)

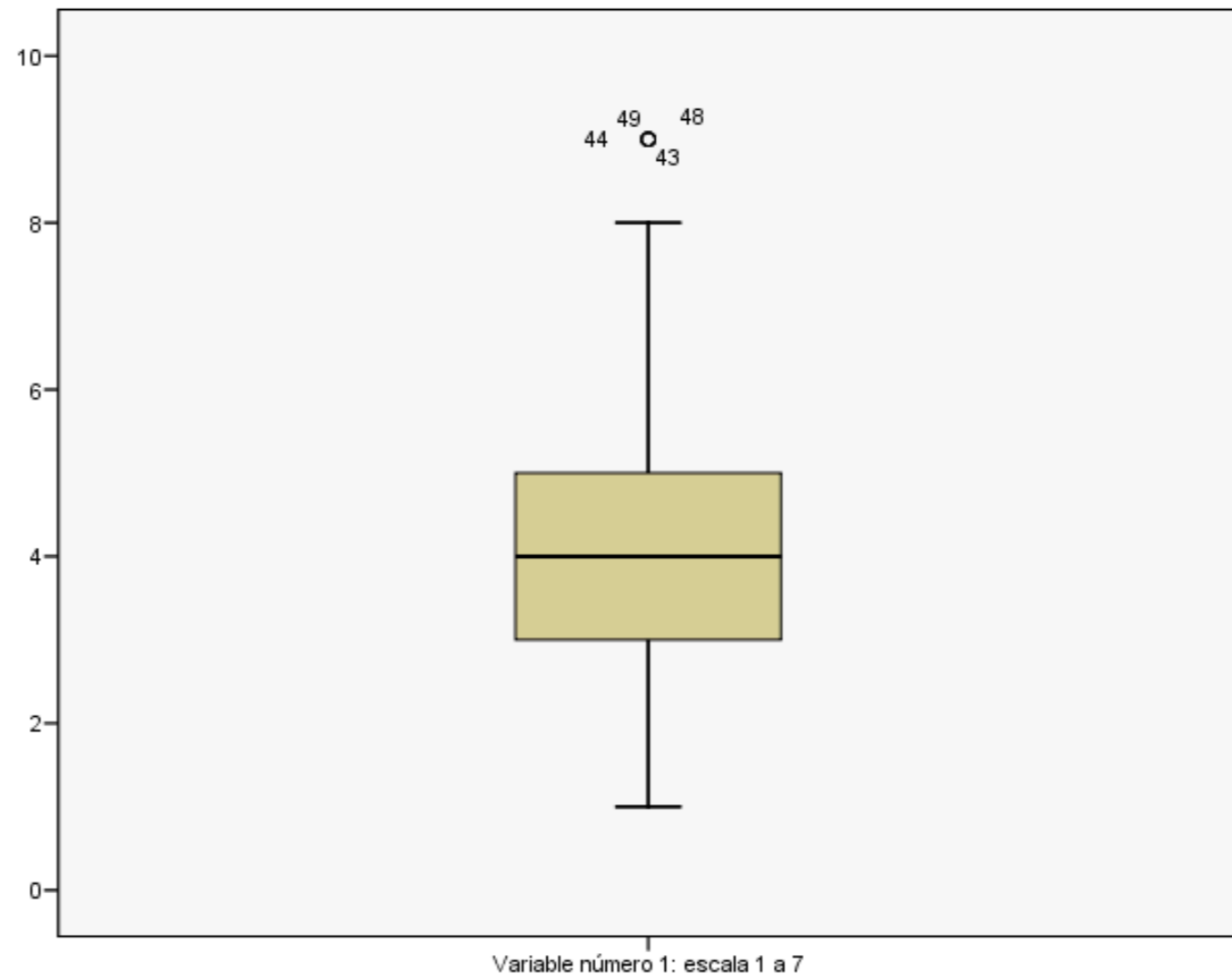
La variable I  
presenta varios  
casos atípicos  
(43, 44, 48 y 49)

**¿Eliminar?**



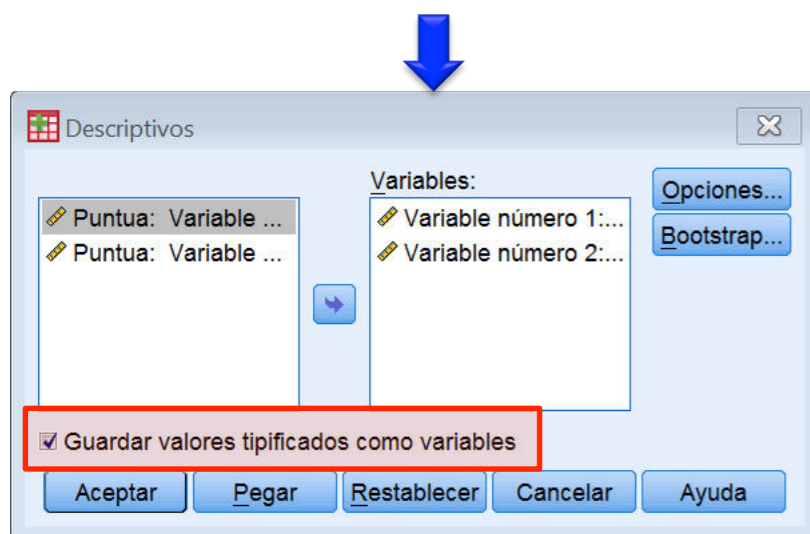
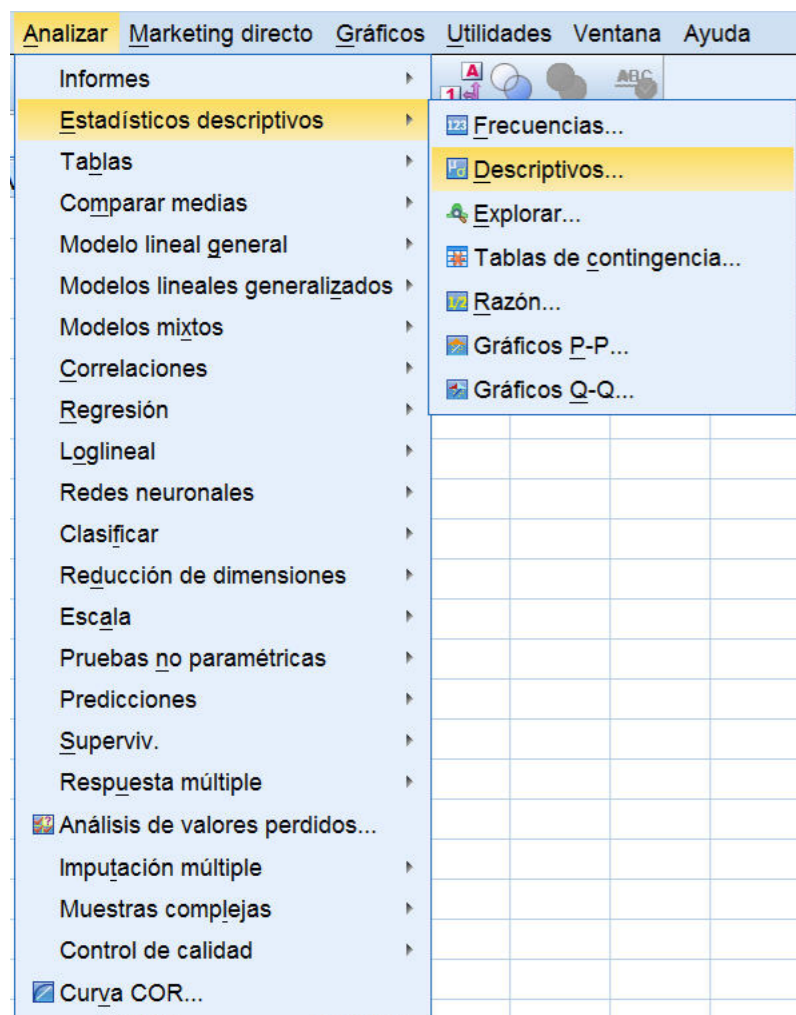
Si es un error en la  
introducción de  
datos rectificar, en  
otro caso .....

**¿Influencia?**





## DETECCIÓN DE CASOS ATÍPICOS (UNIVARIANTE)



	variable1	variable2	Zvariable1	Zvariable2
1	2,00	2,00	-1,14349	-,85298
2	3,00	2,00	-,68241	-,85298
3	4,00	3,00	-,22132	-,24371
4	3,00	2,00	-,68241	-,85298
5	4,00	3,00	-,22132	-,24371
6	5,00	4,00	,23976	,36556
7	5,00	4,00	,23976	,36556
8	6,00	4,00	,70085	,36556
9	5,00	5,00	,23976	,97483
10	4,00	4,00	-,22132	,36556
11	3,00	4,00	-,68241	,36556
12	2,00	2,00	-1,14349	-,85298
13	4,00	3,00	-,22132	-,24371
14	5,00	4,00	,23976	,36556
15	6,00	7,00	,70085	2,19338
16	1,00	2,00	-1,60458	-,85298
17	2,00	1,00	-1,14349	-1,46225
18	3,00	2,00	-,68241	-,85298
19	4,00	3,00	-,22132	-,24371
20	5,00	5,00	,23976	,97483
21	4,00	5,00	-,22132	,97483
22	5,00	5,00	,23976	,97483
23	6,00	6,00	,70085	1,58411
24	5,00	6,00	,23976	1,58411
25	4,00	3,00	-,22132	-,24371
26	3,00	3,00	-,68241	-,24371
27	2,00	1,00	-1,14349	-1,46225
28	1,00	1,00	-1,60458	-1,46225
29	2,00	2,00	-1,14349	-,85298
30	3,00	4,00	-,68241	,36556
31	4,00	4,00	-,22132	,36556
32	5,00	5,00	,23976	,97483
33	6,00	6,00	,70085	1,58411
34	7,00	6,00	1,16194	1,58411
35	5,00	7,00	,23976	2,19338

Tamaño Muestra	Referencia valor tipificado
Menos de 80	2,5 o más
Más de 80	3 o más

(en valores absolutos)

## CASOS ATÍPICOS INFLUYENTES

Correlaciones con todos  
los casos

Correlaciones			
		Variable número 1: escala 1 a 7	Variable número 2: escala 1 a 7
Variable número 1: escala 1 a 7	Correlación de Pearson	1	,157
	Sig. (bilateral)		,276
	N	50	50
Variable número 2: escala 1 a 7	Correlación de Pearson	,157	1
	Sig. (bilateral)	,276	
	N	50	50

Correlaciones eliminando  
casos aislados  
(43, 44, 48 y 49)  
(sinoutlier.sav)

Correlaciones			
		Variable número 1: escala 1 a 7	Variable número 2: escala 1 a 7
Variable número 1: escala 1 a 7	Correlación de Pearson		,460**
	Sig. (bilateral)		,001
	N		46
Variable número 2: escala 1 a 7	Correlación de Pearson	,460**	1
	Sig. (bilateral)	,001	
	N	46	46

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

# DETECCIÓN DE CASOS ATÍPICOS (MULTIVARIANTE)

datos\_importar.sav [Conjunto\_de\_datos2] - PASW Statistics Editor de datos

1: item1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
item1	item2	item3	item4	item5	item6	item7	item8	item9	item10	item11	item12	item13	item14	item15	item16	item17	item18	item19	item20	item21	item22	item23
1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1
2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
6	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
7	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
8	2	1	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
9	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
10	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
13	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
16	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
17	1	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
18	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
19	3	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
22	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
23	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

[datos\_importar.sav]

Como dependiente podemos utilizar cualquier variable (item 9). La detección de observaciones atípicas se basará en las variables independientes

Regresión lineal

Dependientes: Variable latente 3: indicador 4 [item9]

Bloque 1 de 1

Independientes: Variable latente 1: indicador 1 [item1], Variable latente 1: indicador 2 [item2], Variable latente 2: indicador 1 [item3], Variable latente 2: indicador 2 [item4], Variable latente 2: indicador 3 [item5], Variable latente 3: indicador 1 [item6], Variable latente 3: indicador 2 [item7], Variable latente 3: indicador 3 [item8]

Método: Introducir

Variable de selección: Regla...

Etiquetas de caso: Regla...

Ponderación MCP: Regla...

3

Aceptar Pegar Restablecer Cancelar Ayuda

2

1

Regresión lineal: Guardar

Valores pronosticados: No tipificados, Tipificados, Corregidos, E.T. del pronóstico promedio

Residuos: No tipificados, Tipificados, Método de Student, Eliminados, Eliminados estudentizados

Distancias: ☒ Mahalanobis, De Cook, Valores de influencia

Intervalos de pronóstico: Media, Individuos, Intervalo de confianza: 95 %

Estadísticos de los coeficientes: Crear estadísticos de los coeficientes, Crear un nuevo conjunto de datos, Escribir un nuevo archivo de datos

Exportar información del modelo a un archivo XML: Examinar...

☒ Incluir la matriz de covarianzas

Continuar Cancelar Ayuda

1

Regresión lineal: Estadísticos

Coeficientes de regresión: Estimaciones, Intervalos de confianza, Nivel(%): 95, Matriz de covarianzas

Ajuste del modelo, Cambio en R cuadrado, Descriptivos, Correlaciones parciales y semiparciales, Diagnósticos de colinealidad

Residuos: Durbin-Watson, Diagnósticos por caso, Valores atípicos fuera: 3 desviaciones típicas, Todos los casos

Continuar Cancelar Ayuda

2

# DETECCIÓN DE CASOS ATÍPICOS (MULTIVARIANTE)

\*datos\_importar.sav [Conjunto\_de\_datos2] - PASW Statistics Editor de datos

1 : item1	2	item1	item2	item3	item4	item5	item6	item7	item8	item9	MAH_1
1	2	1	2	2	2	2	1	1	1	2	3,45043
2	1	2	2	1	2	2	2	2	3	3	12,86785
3	2	2	2	3	2	3	2	2	2	2	8,31569
4	2	2	2	2	2	2	1	1	2	2	4,36613
5	2	2	2	3	2	2	2	2	2	2	2,76660
6	2	2	2	2	2	2	1	2	3	3	8,71028
7	3	2	2	2	2	2	1	1	1	2	6,66081
8	2	1	3	2	3	3	3	2	2	2	12,41832
9	3	3	3	3	3	3	3	3	3	3	7,01414
10	2	3	2	2	2	2	2	2	2	2	6,29083
11	1	1	1	1	1	1	1	1	1	1	3,88081
12	3	2	2	3	2	2	2	2	3	2	7,51611
13	2	1	1	1	1	1	1	1	1	3	4,47265
14	1	1	1	2	2	2	1	1	1	2	7,82586
15	2	2	3	2	3	3	3	2	3	3	9,18353
16	2	2	2	2	3	3	1	2	2	2	11,34041
17	1	2	3	3	3	3	3	2	2	3	13,45694
18	2	2	2	2	2	2	1	1	1	2	3,47938
19	3	1	2	2	2	2	2	3	3	3	15,39055
20	1	1	1	1	1	1	1	1	1	1	3,88081
21	2	3	3	2	3	3	2	2	2	3	11,44211
22	2	1	2	1	2	2	2	1	2	2	8,61118
23	3	3	2	2	1	2	2	2	3	3	13,44990

Valor de la distancia de Mahalanobis  $D^2$

Transformar /Calcular Variable

\*datos\_importar.sav [Conjunto\_de\_datos2] - PASW Statistics Editor de datos

1 : item1	2	item1	item2	item9	MAH_1
1	2	1	2	2	3,45043
2	1	2	2	3	12,86785
3	2	2	2	2	8,31569
4	2	2	2	2	4,36613
5	2	2	2	2	2,76660
6	2	2	2	3	8,71028
7	3	2	2	1	6,66081
8	2	1	3	2	12,41832
9	3	3	3	3	7,01414
10	2	3	2	2	6,29083
11	1	1	1	1	3,88081
12	3	2	2	3	7,51611
13	2	1	1	1	4,47265
14	1	1	1	2	7,82586

Calcular variable

Variable de destino: probabilidad

Expresión numérica: CDF.CHISQ(MAH\_1,8)

Tipo y etiqueta...

Variable latente ...

Variable latente ...

Variable latente ...

Variable latente ...

Variable latente ...

Variable latente ...

Variable latente ...

Variable latente ...

Mahalanobis Dis...

Grupo de funciones:

Todo

Aritméticas

FDA y FDA no centrada

Conversión

Fecha/hora actual

Cálculo de fechas

Funciones y variables especiales:

Abs

Any

Arsin

Artan

Cdf.Bernoulli

Cdf.Beta

Cdf.Binom

Cdf.Bvnr

Cdf.Cauchy

Cdf.Chisq

Cdf.Exp

CDF.CHISQ(c, gl). Numérico. Devuelve la probabilidad acumulada de que un valor de la distribución de chi-cuadrado, con los grados de libertad dados, sea menor que c.

Si la opción... (condición de selección de casos opcional)

Aceptar Pegar Restablecer Cancelar Ayuda

$CDF.CHISQ(c,gl)$

c: variable con  $D^2$

gl: número de variables independientes Utilizadas en el análisis (8)



# DETECCIÓN DE CASOS ATÍPICOS (MULTIVARIANTE)

\*datos\_importar.sav [Conjunto\_de\_datos2] - PASW Statistics Editor de datos

Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda

1 : probabilidad ,09699747331031

	item1	item2	item3	item4	item5	item6	item7	item8	item9	MAH_1	probabilidad
1	2	1	2	2	2	1	1	1	2	3,45043	
2	1	2	2	1	2	2	2	3	3	12,86785	
3	2	2	2	3	2	3	2	2	2	8,31569	
4	2	2	2	2	2	1	1	2	2	4,36613	
5	2	2	2	3	2	2	2	2	2	2,76660	
6	2	2	2	2	2	1	2	3	3	8,71028	
7	3	2	2	2	2	1	1	1	2	6,66081	
8	2	1	3	2	3	3	2	2	2	12,41832	
9	3	3	3	3	3	3	3	3	3	7,01414	
10	2	3	2	2	2	2	2	2	2	6,29083	,39
11	1	1	1	1	1	1	1	1	1	3,88081	,13
12	3	2	2	3	2	2	2	3	2	7,51611	,52
13	2	1	1	1	1	1	1	1	3	4,47265	,19
14	1	1	1	2	2	1	1	1	2	7,82586	,55
15	2	2	3	2	3	3	3	2	3	9,18353	,67
16	2	2	2	2	3	1	2	2	2	11,34041	,82
17	1	2	3	3	3	3	2	2	3	13,45694	,90
18	2	2	2	2	2	1	1	1	2	3,47938	,10
19	3	1	2	2	2	2	3	3	3	15,39055	,95
20	1	1	1	1	1	1	1	1	1	3,88081	,13
21	2	3	3	2	3	3	2	2	3	11,44211	,82
22	2	1	2	1	2	2	2	1	2	8,61118	,62
23	3	3	2	2	1	2	2	3	3	13,44990	,90

\*datos\_importar.sav [Conjunto\_de\_datos2] - PASW Statistics Editor de datos

Archivo Edición Ver Datos Transformar Analizar Marketing directo Gráficos Utilidades Ventana Ayuda

1 : probabilidad ,00191518774626

	item1	item2	item3	item4	item5	item6	item7	item8	item9	MAH_1	probabilidad
1	2	2	2	2	2	2	2	2	2	1,02509	,00
2	2	2	2	2	2	2	2	2	2	1,02509	,00
3	2	2	2	2	2	2	2	2	1	1,02509	,00
4	2	2	2	2	2	2	2	2	2	1,02509	,00
5	2	2	2	2	2	2	2	2	2	1,02509	,00
6	2	2	2	2	2	2	2	2	2	1,02509	,00
7	2	2	2	2	2	2	2	2	2	1,02509	,00
8	2	2	2	3	2	2	2	2	2	2,76660	,05
9	2	2	2	3	2	2	2	2	2	2,76660	,05
10	2	2	2	3	2	2	2	2	3	2,76660	,05
11	2	2	2	3	2	2	2	2	2	2,76660	,05
12	2	2	2	3	2	2	2	2	2	2,76660	,05
13	3	2	2	2	2	2	2	2	3	3,15102	,08

Los valores inferiores  
a 0.001 pueden ser  
indicativos de  
casos «atípicos»

## DATOS AUSENTES (MISSING VALUES)

- **Completamente al azar (MCAR-Missing Completely At Random)**: cuando la presencia o ausencia de valores en una variable  $X$  es independiente tanto de los valores de otras variables como de sus propios valores.
- **Al azar (MAR –Missing At Random)**: si la presencia o ausencia de valores en  $X$  está relacionada con otra u otras variables del modelo, pero es independiente de los propios valores de  $X$
- **No al azar (MNAR-Missing Not At Random)**: si la presencia o ausencia de valores de la variable  $X$  depende de los propios valores de  $X$ .



# TRATAMIENTO DE DATOS AUSENTES CON SPSS

**mecanismos de perdida.sav - SPSS Editor de datos**

Archivo Edición Ver Datos Transformar **Analizar** Gráficos Utilidades Ventana ?

1 : Enero 169

	Enero	Febrero	MC
1	169,00	148,00	1
2	126,00	123,00	
3	132,00	149,00	
4	160,00	169,00	
5	105,00	138,00	
6	116,00	102,00	
7	125,00	88,00	
8	112,00	100,00	
9	133,00	150,00	
10	94,00	113,00	
11	109,00	96,00	
12	109,00	78,00	
13	106,00	148,00	
14	176,00	137,00	

- Informes ▶
- Estadísticos descriptivos ▶
- Tablas ▶
- Comparar medias ▶
- Modelo lineal general ▶
- Modelos mixtos ▶
- Correlaciones ▶
- Regresión ▶
- Loglineal ▶
- Clasificar ▶
- Reducción de datos ▶
- Escalas ▶
- Pruebas no paramétricas ▶
- Series temporales ▶
- Supervivencia ▶
- Respuesta múltiple ▶
- Análisis de valores perdidos...**
- Muestras complejas ▶

indic

<http://faculty.chass.ncsu.edu/garson/PA765/missing.htm>



# TRATAMIENTO DE DATOS AUSENTES

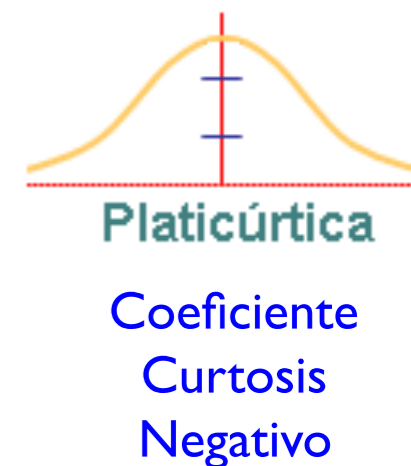
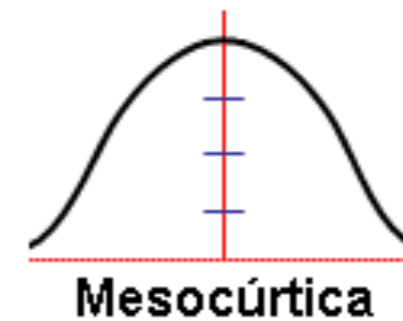
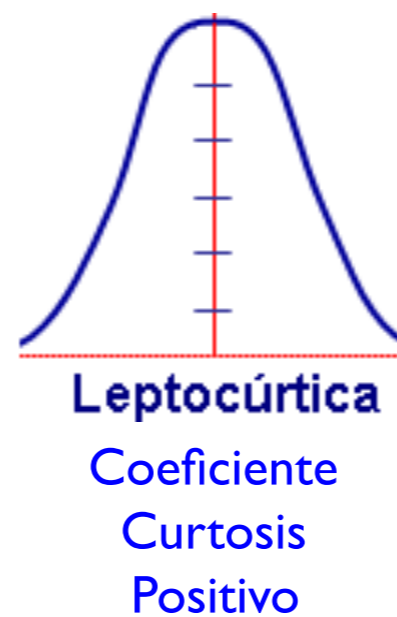
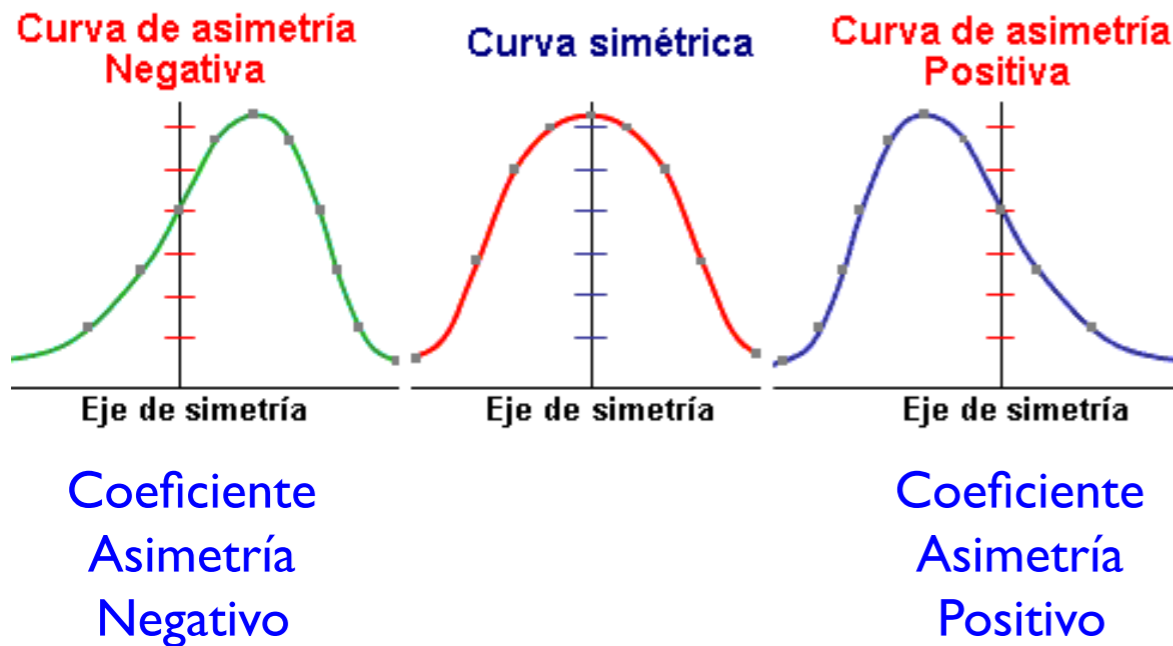
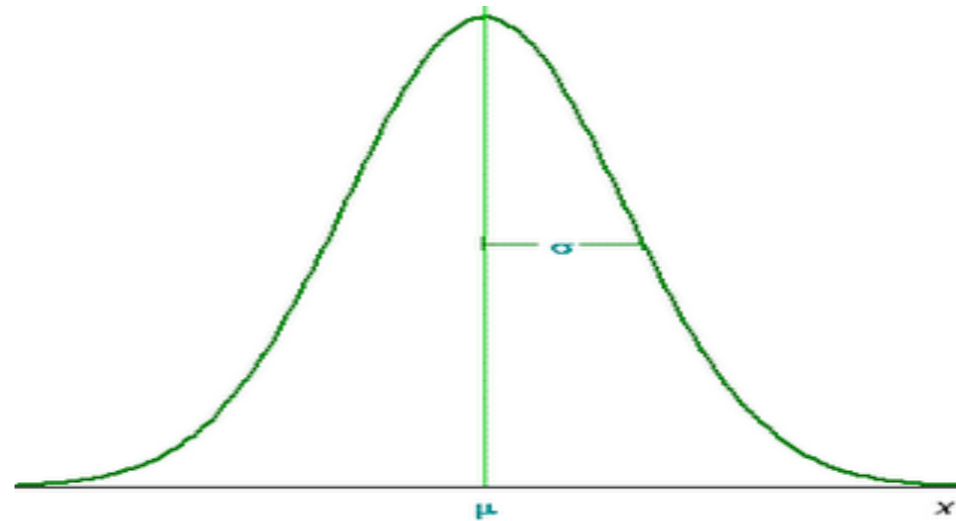
TRADICIONALES	CONSISTE EN.....	HIPÓTESIS BÁSICA	INCONVENIENTES	RECOMENDACIÓN
<b>Eliminación completa del caso (LISTWISE)</b>	Eliminar los individuos que no respondieron alguna de las preguntas	MCAR	Puede reducirse mucho el tamaño de la muestra (contrastes poco potentes y errores estándar elevado)	No utilizar, salvo que se asegure MCAR y no se eliminen muchos casos
<b>Eliminación por parejas (PAIRWISE)</b>	Usar individuos que si respondieron a las dos variables en cuestión	MCAR	No existe un tamaño de muestra único (los errores estándares no serán adecuados. Es frecuente que las matrices no sean definidas positivas)	No utilizar
<b>Sustitución por la media</b>	Reemplazar datos ausentes por la media de los individuos que si contestaron	MCAR	Se sesgan las varianzas y covarianzas	No utilizar
<b>Métodos de imputación</b>				
<i>Regresión lineal</i>	Para cada dato ausente se ajusta un modelo de regresión con las restantes variables como predictoras		Es un método no demasiado eficiente, pudiendo complicarse con patrones arbitrarios de valores perdidos	No utilizar
<i>Donación (Hot Deck)</i>	El valor ausente se sustituye por el de otro individuo que muestre valores muy similares en las variables en las que se dispone información		Varios estudios demuestran que este tipo de imputación en el contexto SEM puede implicar estimaciones sesgadas incluso bajo la hipótesis MCAR	PRELIS implementa un método mejorado (Similar Response Pattern Imputation-SRPI) que funciona correctamente bajo MCAR, pero no suponiendo MAR
MODERNOS	CONSISTE EN.....	HIPÓTESIS BÁSICA	INCONVENIENTES	RECOMENDACIÓN
Método de esperanza-maximización (Imputación Múltiple)	Calcular la matriz de varianzas-covarianzas por máxima verosimilitud en dos etapas	MAR ; Normalidad de las variables	Los errores estándares son desconocidos. No es fácil determinar cuál es el tamaño muestral correcto	Comprobar que se cumplen hipótesis de partida
Método de máxima verosimilitud directa	Calcular la matriz de varianzas-covarianzas por máxima verosimilitud en una etapa	MAR ; Normalidad de las variables	Consistente, eficiente y conduce a inferencias correctas sólo si se cumplen las dos hipótesis básicas	Es el más recomendable cuando existen datos ausentes (FIML)

## COMPROBANDO HIPÓTESIS DE PARTIDA

- Para que una técnica estadística esté “bien aplicada”, de manera que podamos extraer conclusiones válidas, **es importante comprobar que se cumplen las hipótesis básicas de aplicación** (¡ muchos dan por hecho que se cumplen sin verificar nada !)
- Aunque podemos encontrar diferentes situaciones, **varias comprobaciones son especialmente importantes en muchas técnicas estadísticas**:
  - 1) Normalidad de los datos y homocedasticidad (igualdad varianzas)
  - 2) Linealidad - Multicolinealidad
  - 3) Número mínimo de casos (vg: regresión, factorial, SEM, etc)
- Si no se cumplen las hipótesis requeridas, **NO DESESPERAR**. Siempre, o casi siempre, podemos encontrar métodos alternativos. Por ejemplo, para contrastar la diferencia entre medias podemos utilizar tanto una alternativa paramétrica (t-test) como no paramétrica (test de Mann-Whitney). Si no disponemos de datos suficientes, una alternativa a LISREL es PLS.

# NORMALIDAD (UNIVARIANTE Y MULTIVARIANTE)

## Distribución normal



Comprobación de la normalidad {

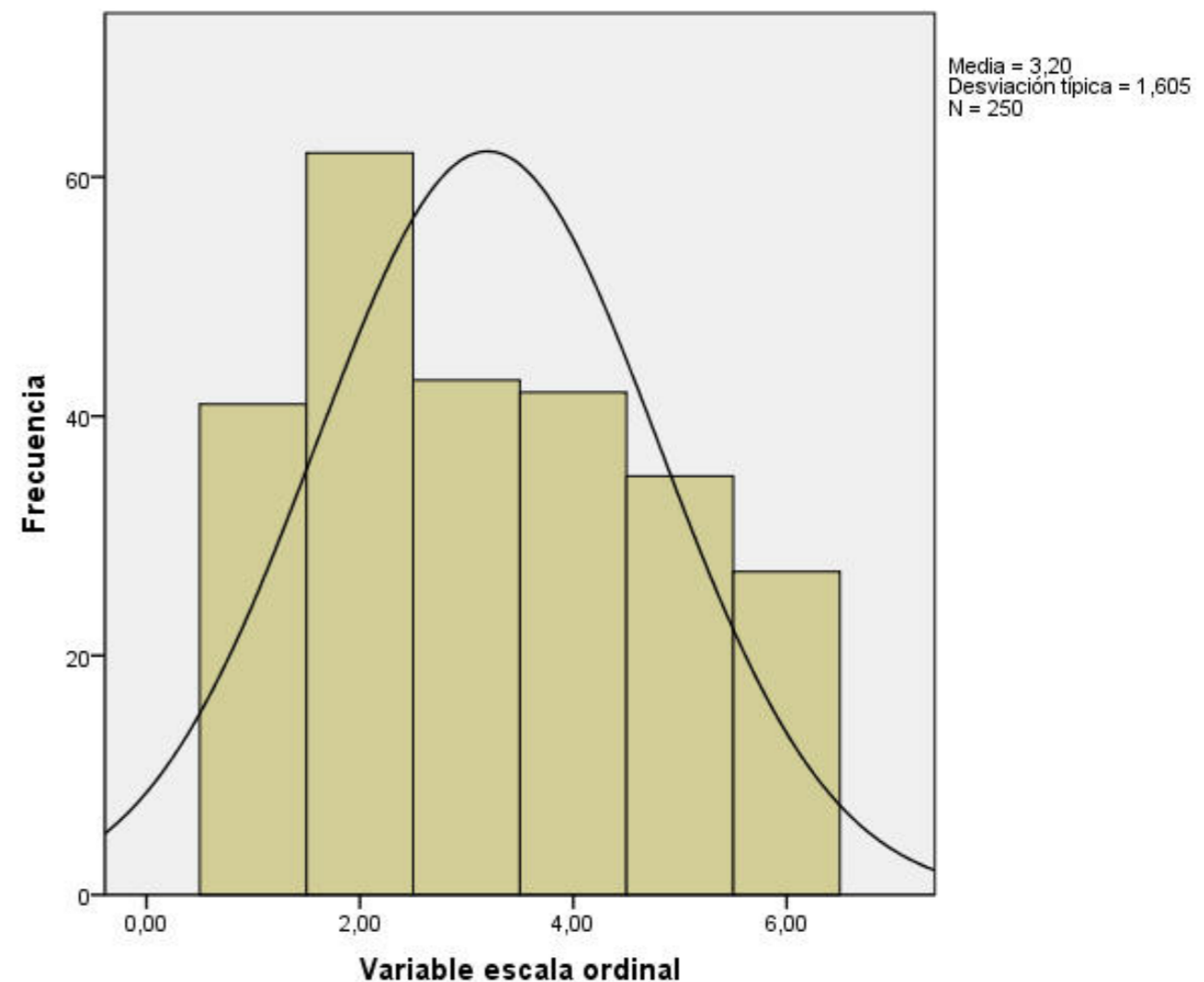
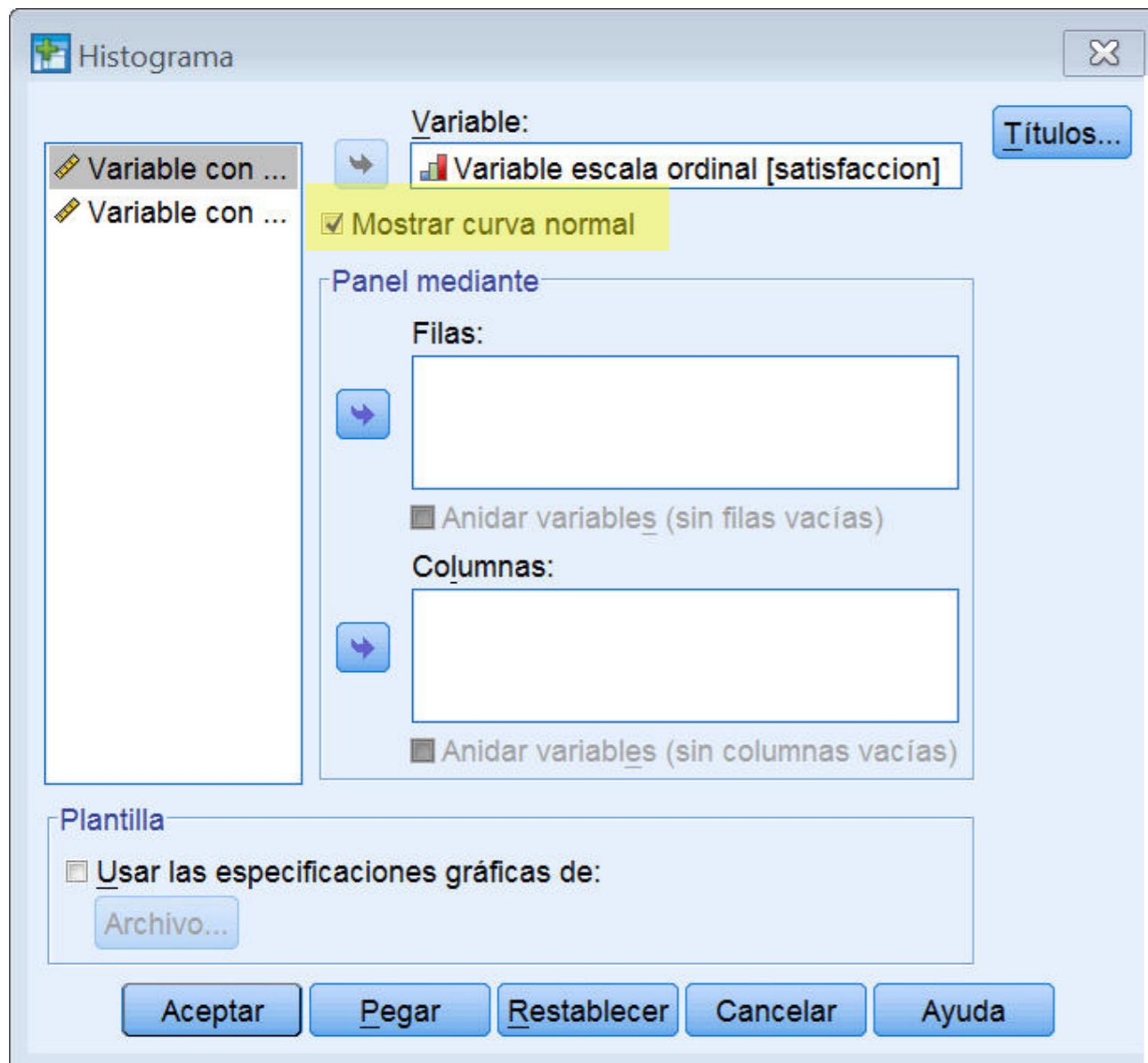
- Métodos gráficos
- Contraste de hipótesis

# MÉTODOS GRÁFICOS: HISTOGRAMAS

normalidad.sav (3 variables – 250 casos)

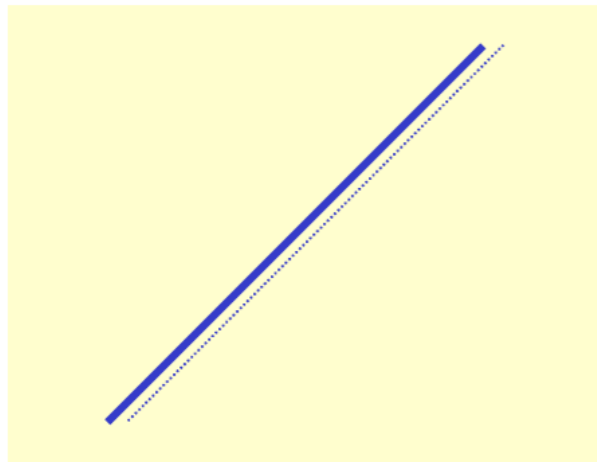
Variable SATISFACCION (ordinal-no métrica)

Gráficos/ Histograma (mostrar curva normal)

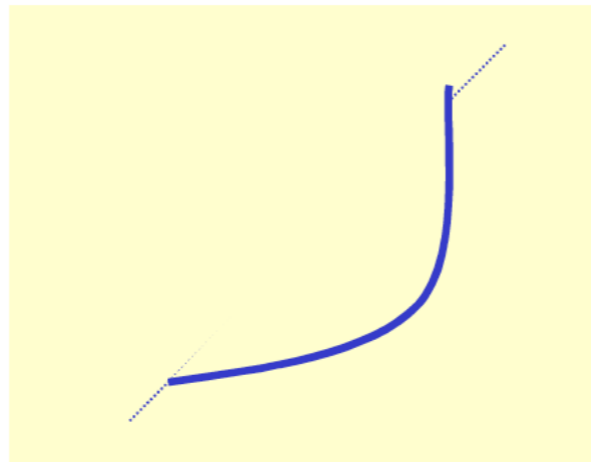


## MÉTODOS GRÁFICOS: Q-Q NORMAL

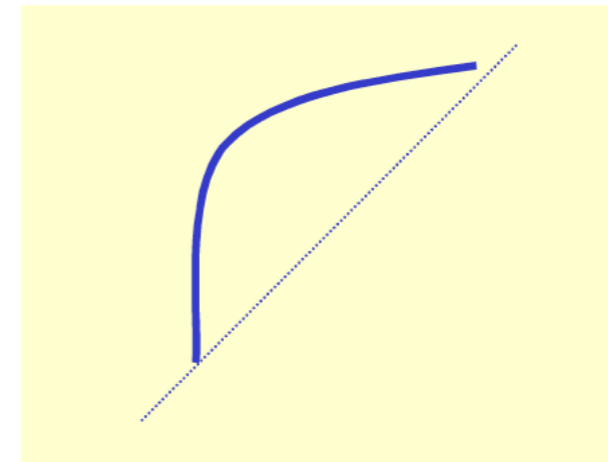
Estos gráficos muestran un diagrama de cuantiles en el que en el eje X se representan los cuantiles muestrales y en el eje Y los cuantiles esperados bajo una distribución normal



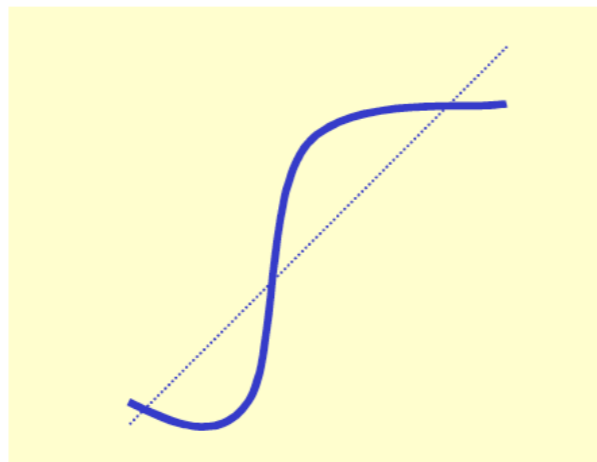
**Distribución Normal**



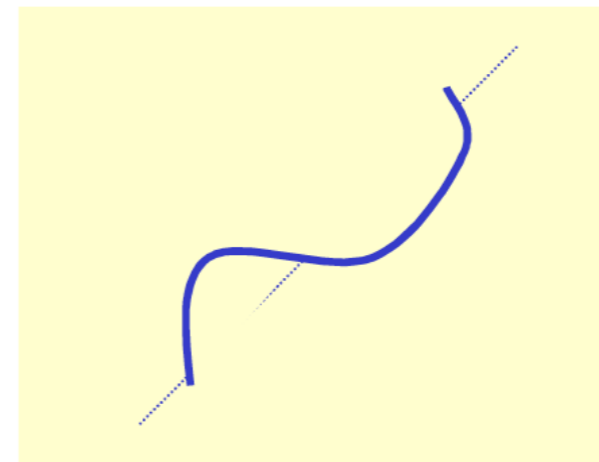
**Distribución Asimétrica  
a Izquierda**



**Distribución Asimétrica  
a Derecha**



**Distribución Leptocúrtica**



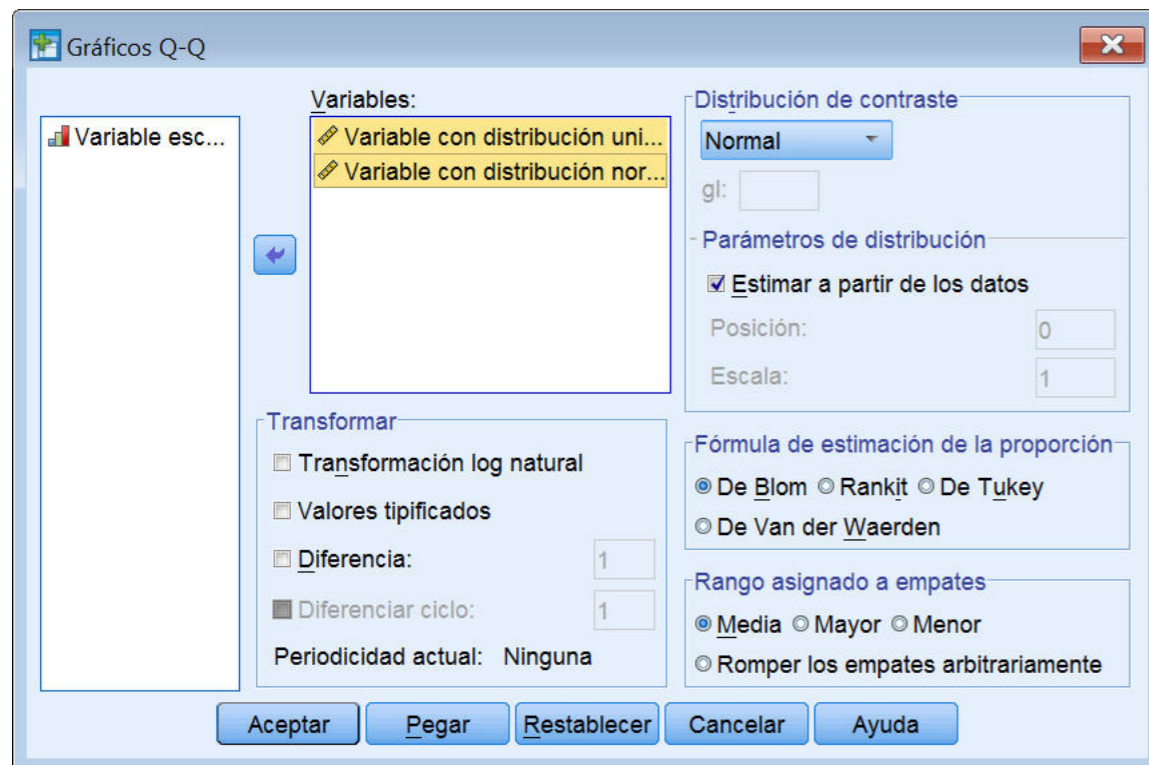
**Distribución Platicúrtica**

# MÉTODOS GRÁFICOS: Q-Q NORMAL

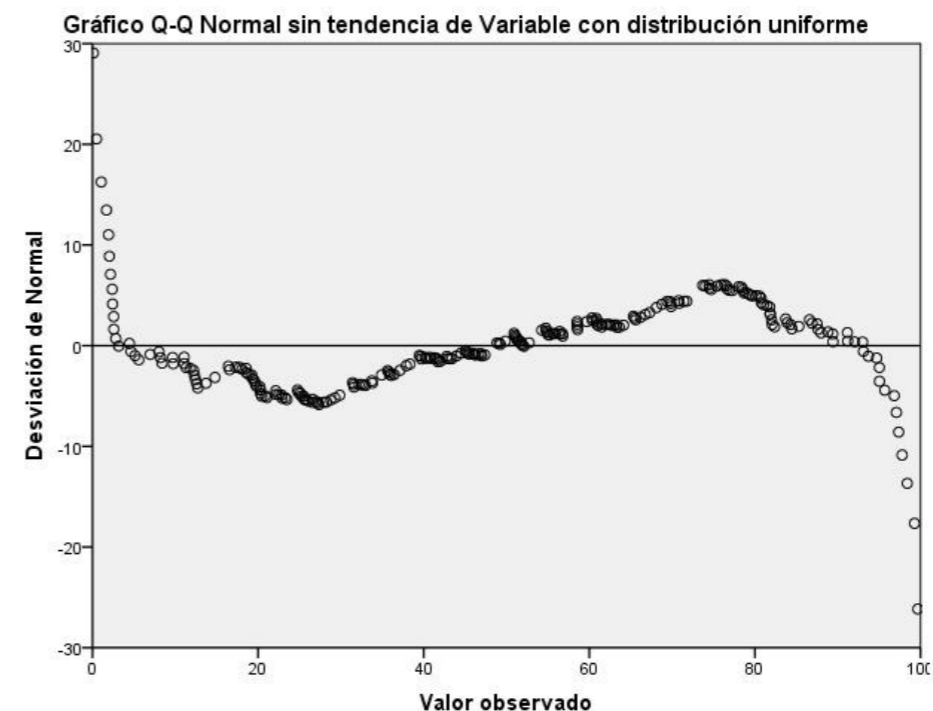
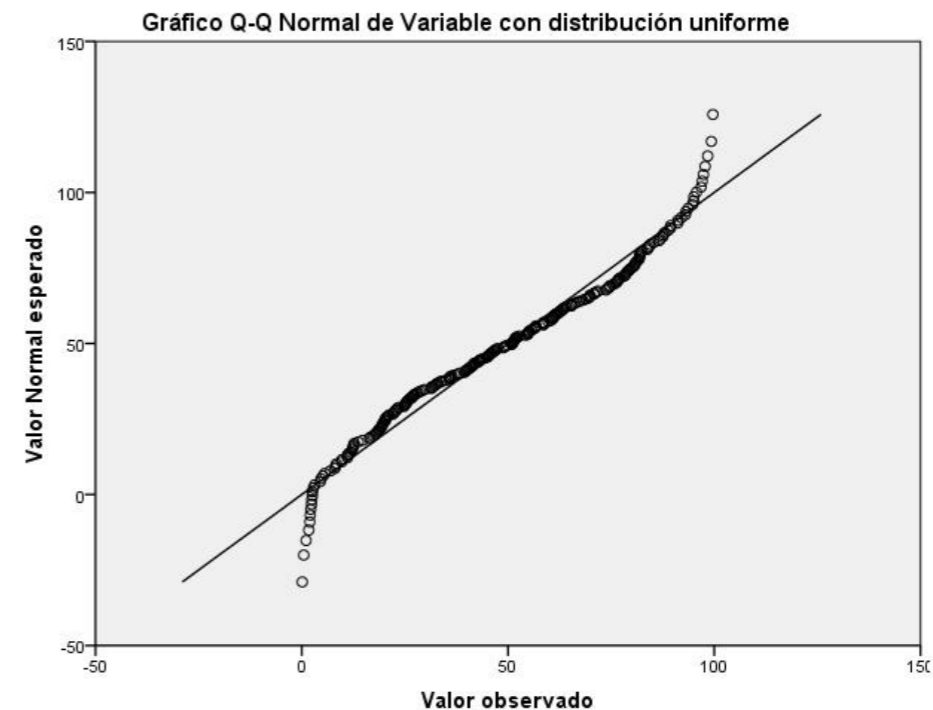
normalidad.sav (3 variables – 250 casos)

Variable UNIFORME / NORMAL (razón-métricas)

Analizar/ Estadísticos descriptivos / Gráficos Q-Q

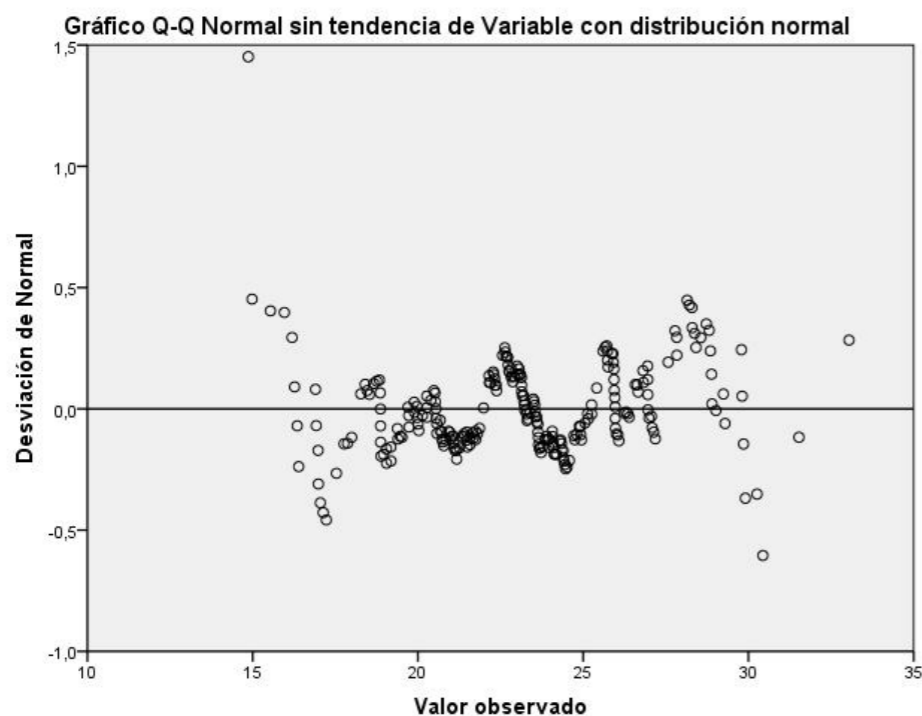
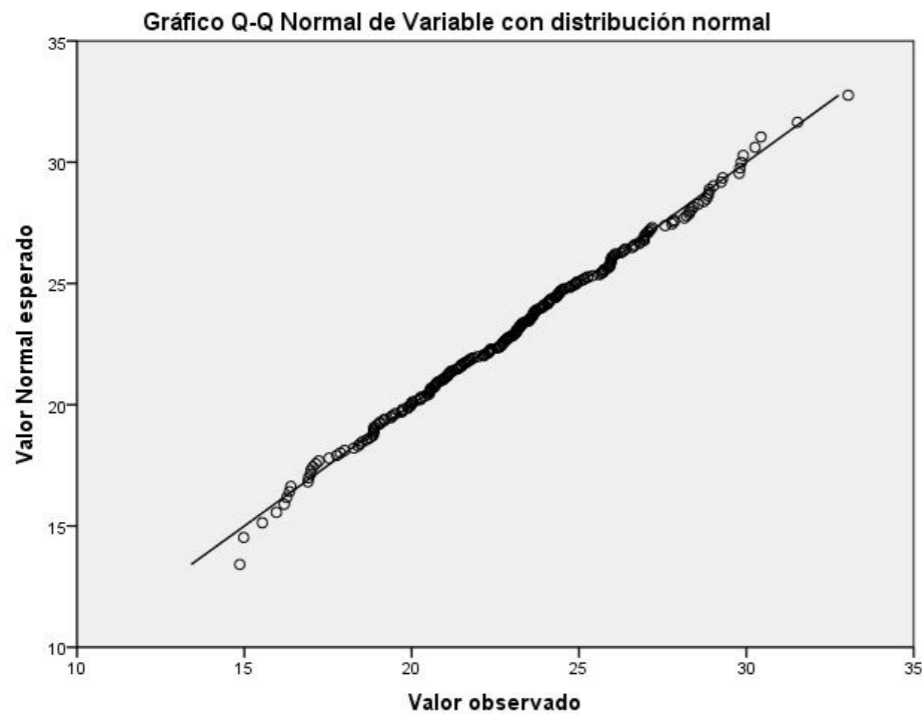


Variable con distribución uniforme



# MÉTODOS GRÁFICOS: Q-Q NORMAL

Variable con distribución normal



Variable con distribución normal

## Analizar / Explorar (Gráficos)

Explorar

Variable esc...

Lista de dependientes:

- Variable con distribuci...
- Variable con distribuci...

Lista de factores:

Etiquetar los casos mediante:

Visualización

☒ Ambos ☐ Estadísticos ☐ Gráficos

Aceptar Pegar Restablecer Cancelar Ayuda

Explorar: Gráficos

Diagramas de caja

- ☒ Niveles de los factores juntos
- ☐ Dependientes juntas
- ☐ Ninguna

Descriptivos

- ☐ De tallo y hojas
- ☒ Histograma

☒ Gráficos con pruebas de normalidad

Dispersión por nivel con prueba de Levene

- ☐ Ninguna
- ☐ Estimación de potencia
- ☐ Transformados Potencia: Log natural
- ☐ No transformados

Continuar Cancelar Ayuda

# TEST DE NORMALIDAD (EXPLORAR)

Descriptivos

			Estadístico	Error típ.
Variable con distribución uniforme	Media		48,4344	1,74362
	Intervalo de confianza para la media al 95%	Límite inferior	45,0003	
		Límite superior	51,8685	
	Media recortada al 5%		48,3455	
	Mediana		48,1094	
	Varianza		760,053	
	Desv. típ.		27,56906	
	Mínimo		,11	
	Máximo		99,67	
	Rango		99,56	
	Amplitud intercuartil		46,44	
	Asimetría		,033	,154
	Curtosis		-1,088	,307
Variable con distribución normal	Media		23,0867	,21793
	Intervalo de confianza para la media al 95%	Límite inferior	22,6575	
		Límite superior	23,5159	
	Media recortada al 5%		23,0814	
	Mediana		23,1490	
	Varianza		11,873	
	Desv. típ.		3,44575	
	Mínimo		14,86	
	Máximo		33,04	
	Rango		18,18	
	Amplitud intercuartil		5,00	
	Asimetría		,051	,154
	Curtosis		-,267	,307

ASIMETRÍA PRÓXIMA A CERO  
(dentro del intervalo [-0,5; 0,5])

CURTOSIS NEGATIVA (platicúrtica)  
(fuera del intervalo [-0,5;0,5])

normalidad.xls

Pruebas de normalidad

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Variable con distribución uniforme	,070	250	,005	,965	250	,000
Variable con distribución normal	,031	250	,200*	,996	250	,744

a. Corrección de la significación de Lilliefors

\*. Este es un límite inferior de la significación verdadera.

## TRANSFORMACIONES PARA LOGRAR NORMAL

Causa de la falta de normalidad	Transformación aconsejable
Asimetría positiva	$\text{Log } (X + C)$
Asimetría negativa	$\text{Log } (C - X)$
Leptocurtosis	$1/X$
Platicurtosis	$X^2$

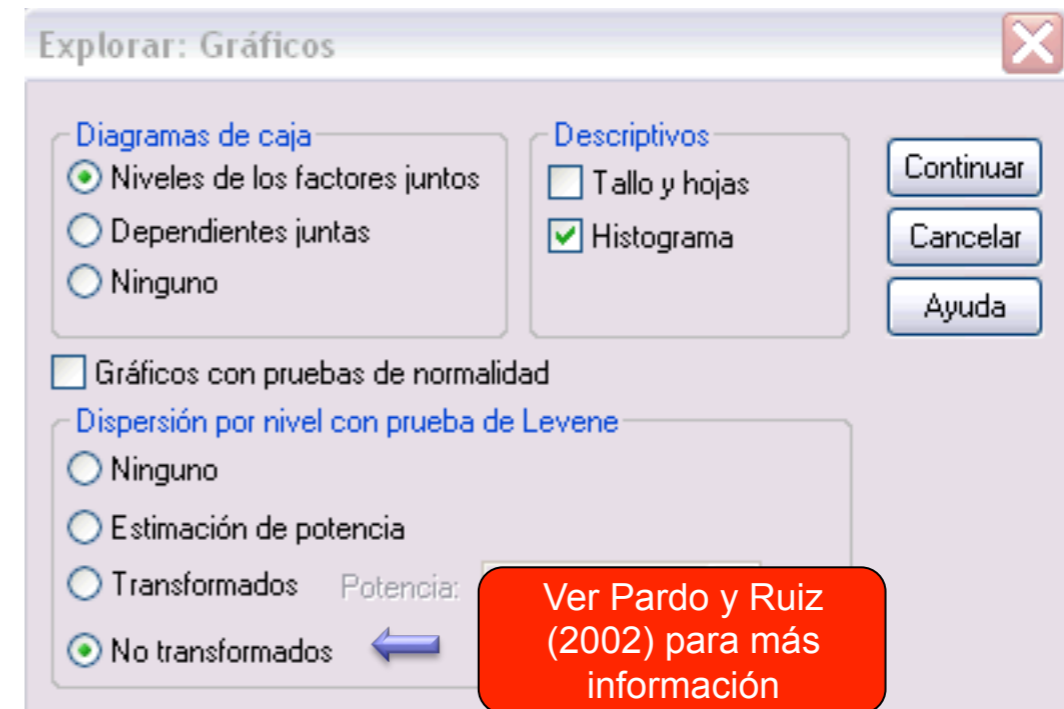
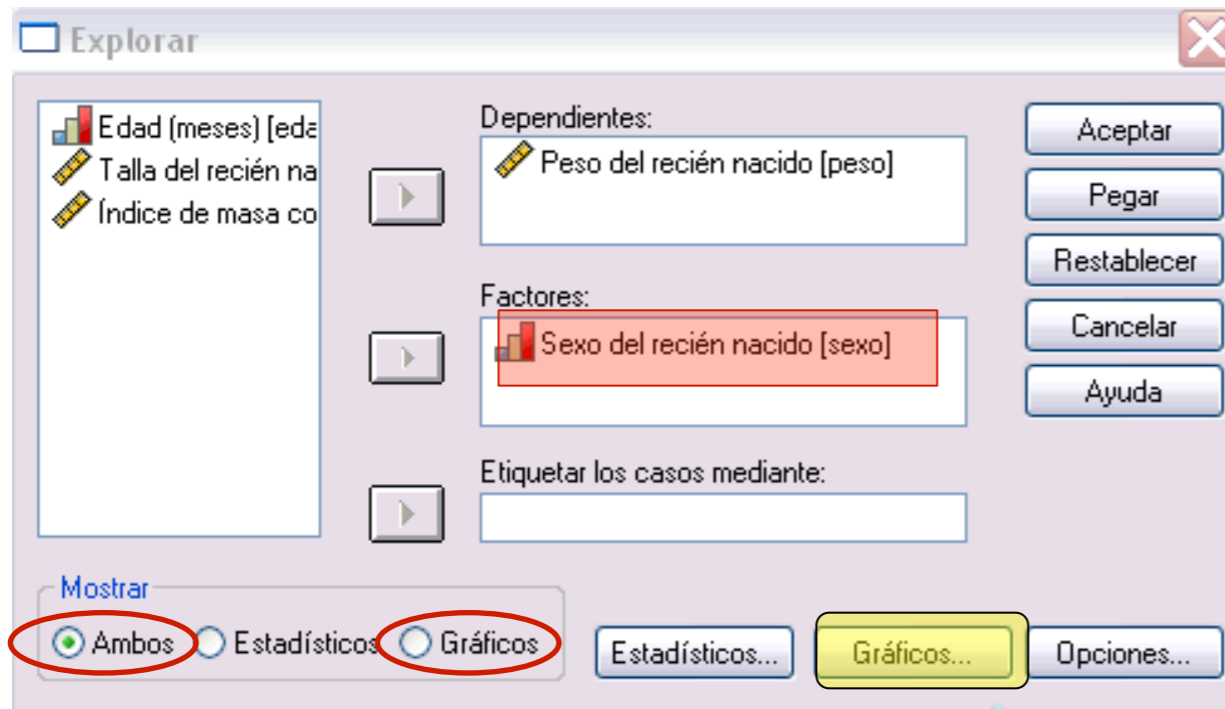
NOTA: C es una constante

¿Transformación aconsejable para la variable con  
distribución uniforme?

# HOMOCEDASTICIDAD

Comprobando homocedasticidad con SPSS  
Fichero: homocedasticidad.sav

- El supuesto de igualdad de varianzas (homocedasticidad) es también bastante común en algunos procedimientos estadísticos (vg: ANOVA, Regresión lineal).
- Para obtener el test, es necesario incorporar un factor en el diálogo inicial de el procedimiento explorar, y marcar la casilla Ambos o Gráficos en mostrar.



# HOMOCEDEASTICIDAD

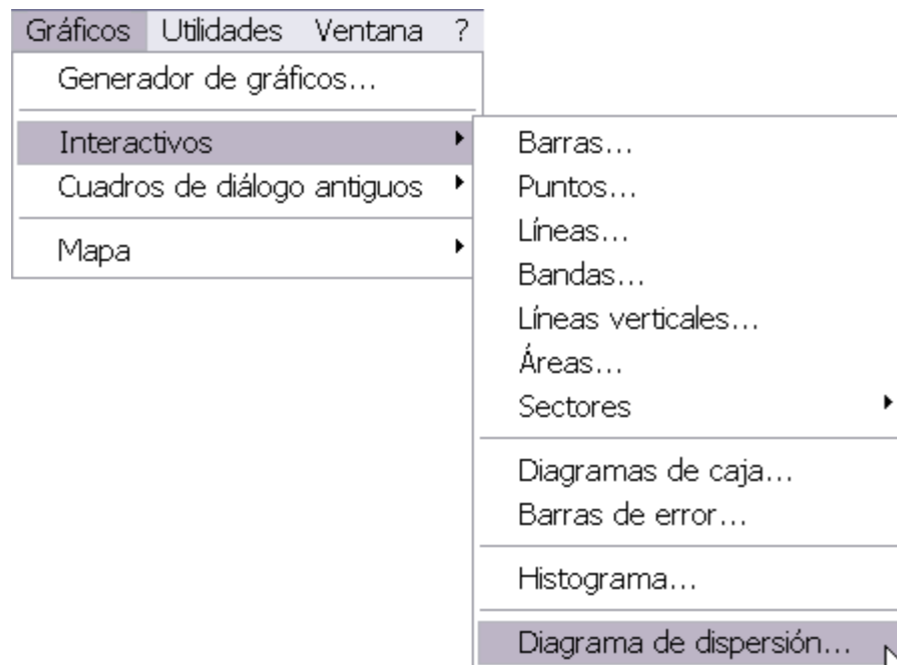
Prueba de homogeneidad de la varianza

		Estadístico de Levene	gl1	gl2	Sig.
Peso del recién nacido	Basándose en la media	11,066	1	3542	,001
	Basándose en la mediana.	9,590	1	3542	,002
	Basándose en la mediana y con gl corregido	9,590	1	3512,108	,002
	Basándose en la media recortada	10,416	1	3542	,001

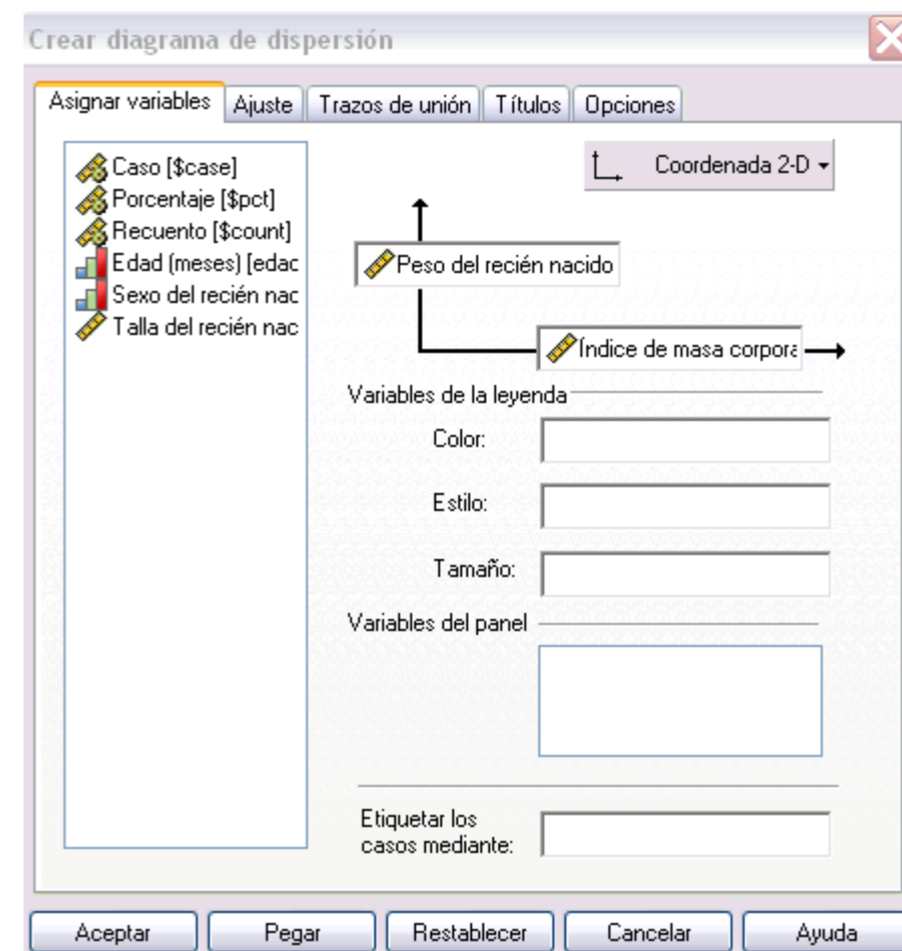
- 1.El estadístico de Levene contrasta la hipótesis de homogeneidad de varianzas.
- 2.La hipótesis nula es que las varianzas son homogéneas. Se rechaza la hipótesis nula cuando el nivel de significación es  $< 0,05$ .
- 3.Si no se cumple el supuesto, en algunos casos (vg: ANOVA) sería necesario realizar una transformación de los datos originales (ver pág. 220-222 del manual de Pardo y Ruiz, 2002)

## LINEALIDAD

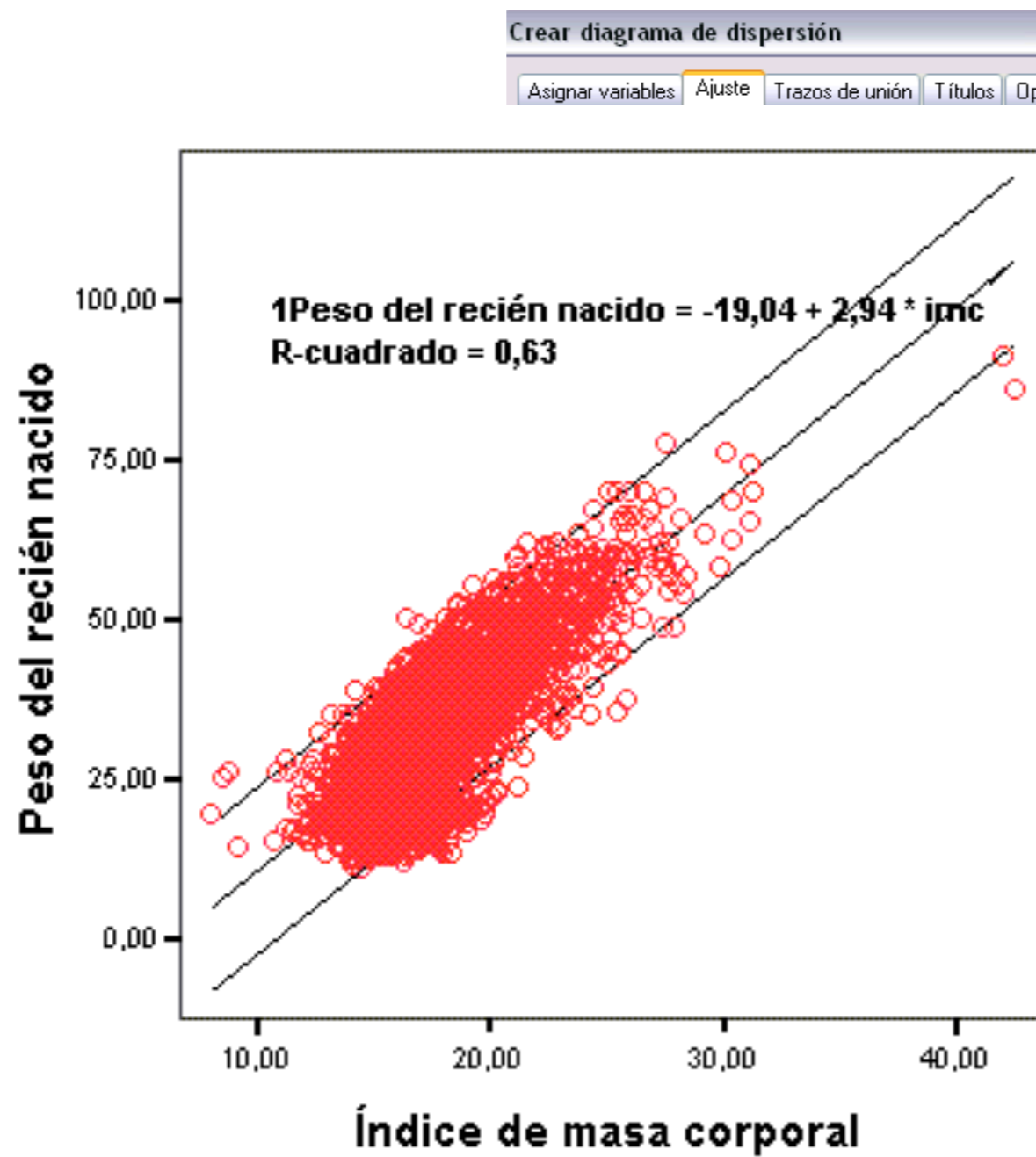
- Esta hipótesis es importante para todas las técnicas basadas en MLG.
- En algunas técnicas, la falta de linealidad puede detectarse analizando los diagramas de residuos.
- En SPSS, podemos utilizar la opción Gráficos – Diagramas de dispersión



Homogeneidad.sav



## LINEALIDAD



Regresión lineal con  
Intervalo de predicción individual al 95,00%

## MULTICOLINEALIDAD

■ Los datos deben estar libres de multicolinealidad. La multicolinealidad tiene lugar cuando variables diferentes miden de hecho el mismo constructo, lo que implica una correlación muy elevada entre ellas.

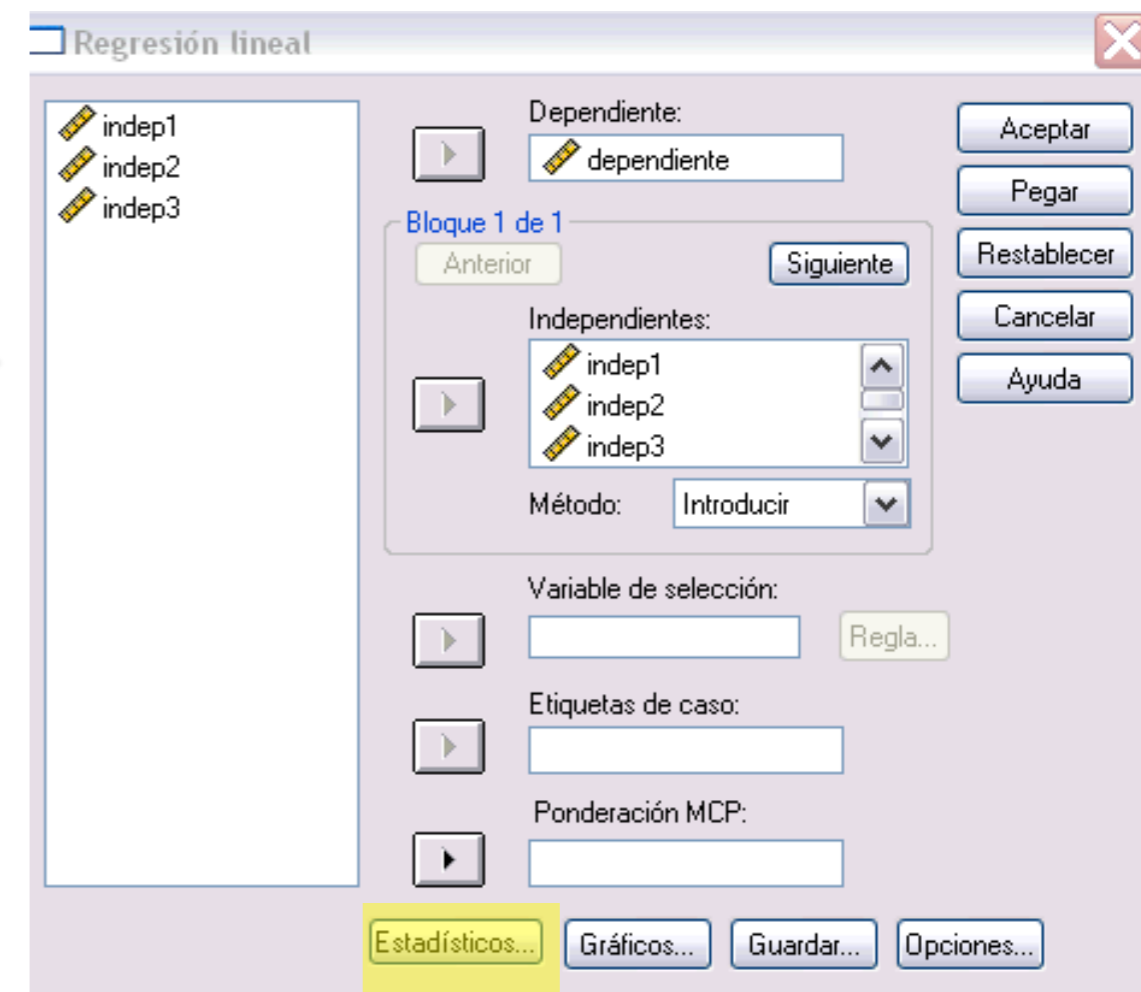
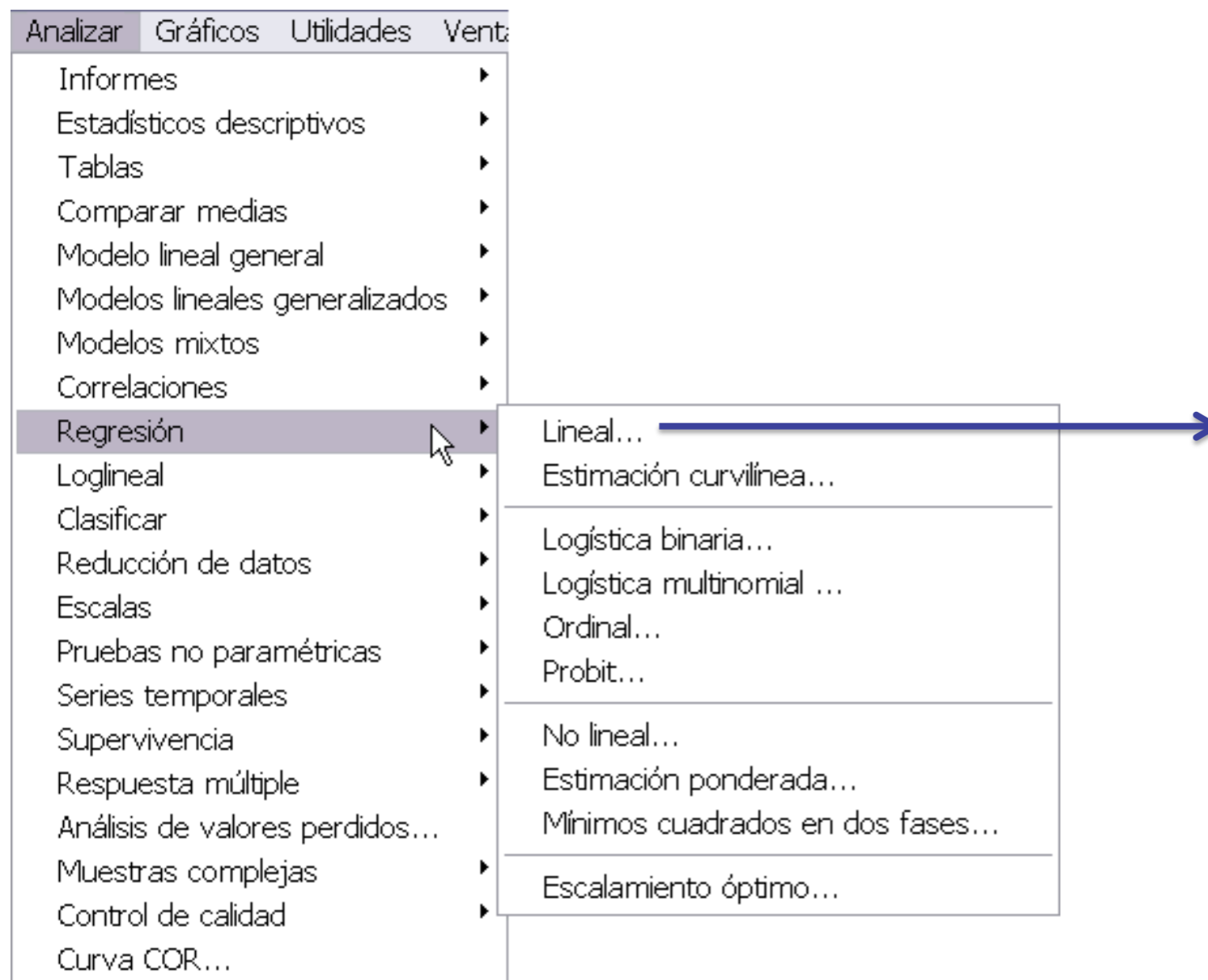
■ Existen dos modos principales de detectar la multicolinealidad:

1. Examinar la matriz de correlaciones (0,90 o superiores).
2. Calcular la correlación múltiple al cuadrado entre cada variable y todas las demás ( $R^2_{smc}$  - valores superiores a 0.90 sugieren multicolinealidad). En SPSS, deberemos revisar los indicadores **Tolerancia** ( $1 - R^2_{smc}$  ;  $>0,1$ ) y **FIV** ( $< 10$ ) disponibles en el procedimiento de regresión lineal.



# MULTICOLINEALIDAD

Comprobando multicolinealidad con SPSS  
Fichero: multicolinealidad.sav





# MULTICOLINEALIDAD

Comprobando multicolinealidad con SPSS  
Fichero: multicolinealidad.sav

Regresión lineal: Estadísticos

**Coeficientes de regresión**

- ☒ Estimaciones
- ☐ Intervalos de confianza
- ☐ Matriz de covarianzas

**Residuos**

- ☐ Durbin-Watson
- ☐ Diagnósticos por caso
  - ☒ Valores atípicos a más de:  desviaciones típicas
  - ☐ Todos los casos

☒ Ajuste del modelo

☐ Cambio en R cuadrado

☒ Descriptivos

☐ Correlaciones parcial y semiparcial

☒ Diagnósticos de colinealidad

Continuar

Cancelar

Ayuda



## MULTICOLINEALIDAD

### Correlaciones

		dependiente	indep1	indep2	indep3
Correlación de Pearson	dependiente	1,000	,492	,482	,433
	indep1	,492	1,000	,444	,449
	indep2	,482	,444	1,000	,971
	indep3	,433	,449	,971	1,000
Sig. (unilateral)	dependiente	.	,001	,001	,003
	indep1	,001	.	,002	,002
	indep2	,001	,002	.	,000
	indep3	,003	,002	,000	.
N	dependiente	40	40	40	40
	indep1	40	40	40	40
	indep2	40	40	40	40
	indep3	40	40	40	40

# MULTICOLINEALIDAD

Coeficientes<sup>a</sup>

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados	t	Sig.	Estadísticos de colinealidad	
		B	Error típ.	Beta			Tolerancia	FIV
1	(Constante)	37,130	7,668		4,842	,000		
	indep1	4,233	1,742	,363	2,431	,020	,797	1,254
	indep2	4,086	2,190	1,046	1,866	,070	,057	17,690
	indep3	-2,147	1,617	-,747	-1,327	,193	,056	17,796

a. Variable dependiente: dependiente

$$R_i^2 = R_{X_i | X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k}^2 \quad i = 1, \dots, k$$

$$T_i = \frac{1}{FIV_i} = 1 - R_i^2$$

$$FIV_i = \frac{1}{1 - R_i^2}$$

Tolerancia (debe estar próximo a la unidad. Si < 0,1 problemático)

FIV = Factor de Inflación de la Varianza (mayor de 10 problemático)

# MULTICOLINEALIDAD

Diagnósticos de colinealidad<sup>a</sup>

Modelo	Dimensión	Autovalor	Indice de condición	Proporciones de la varianza			
				(Constante)	indep1	indep2	indep3
1	1	3,723	1,000	,01	,02	,00	,00
	2	,220	4,118	,05	,93	,00	,00
	3	,055	8,255	,92	,06	,01	,02
	4	,002	39,072	,02	,00	,98	,98

a. Variable dependiente: dependiente

## Regla empírica (Belsley):

- Índices de condición entre 5 y 10 colinealidad débil.
- Entre 10 y 30 (no problemático)
- Entre 30 y 100 colinealidad moderada o fuerte.



ugr

Universidad  
de Granada

# Máster en Economía y Organización de empresas

Módulo III: Competencias para la preparación de trabajo  
fin de Máster

Dr. Eulogio Cordon Pozo