



3 datos
tratamiento estadístico

ANÁLISIS DESCRIPTIVO DE LOS DATOS DE VARIABLES CUANTITATIVAS



Variables CUANTITATIVAS



Números con unidad de medida
(con un instrumento, o procedimiento, de medición formal)

Ej.: Tasa cardiaca; Presión arterial; Fiebre (temp. corp.); Índice Tobillo/Brazo; IMC; Número de plaquetas; Nivel de Fetuína; LDL-Colesterol; HDL-Colesterol; Troponina; Mioglobina; CPK; Glucemia; Edad Menopausia, Número de Hijos; Tiempo de Hospitalización, de Recidiba, de Supervivencia;; Glasgow; Barthel;...

ANÁLISIS DESCRIPTIVO

1º

EXPLORACIÓN de datos "outliers"

Su objetivo es comprobar si existen en la variable valores extraños (anómalos, atípicos) fuera del rango habitual en la población de estudio

- Gráfico: diagrama de caja y bigotes (box-plot)

2º

Estudio de la FORMA de la variable

Su objetivo es comprobar si la variable sigue, o no, el modelo de la distribución NORMAL de la Campana de Gauss

- Índices descriptivos de ASIMETRÍA y CURTOSIS
- Prueba (inferencial) de bondad de ajuste al modelo normal

3º

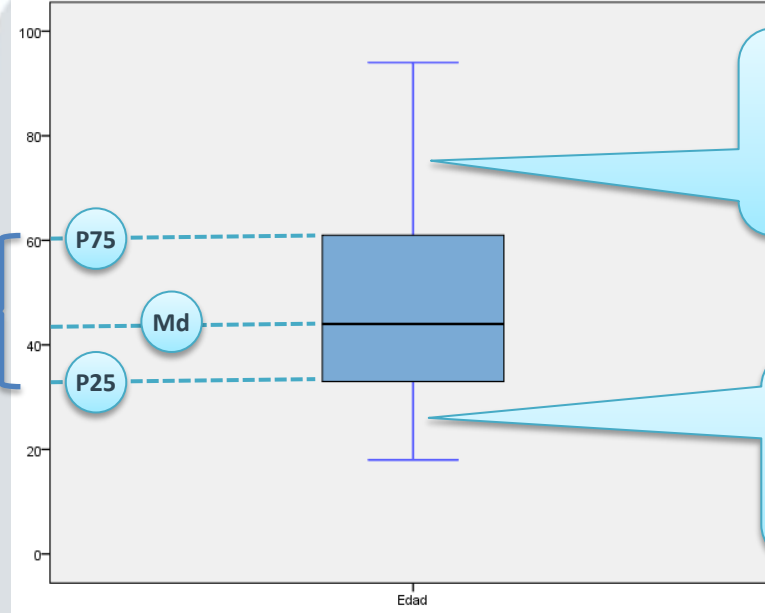
Descriptiva de la CENTRALIDAD y la VARIABILIDAD

Su objetivo es explicar las características (el "¿cómo son?") de los sujetos en la variable medida

- CENTRALIDAD: media; mediana
- VARIABILIDAD: desviación típica (estandar); amplitud semiintercuartil

Exploración de "outliers" (diagrama box-plot = caja y bigotes)

La CAJA contiene al 50% central de los casos. Por ello el límite superior es el percentil 75 y el inferior el 25. La línea central más oscura es la mediana.

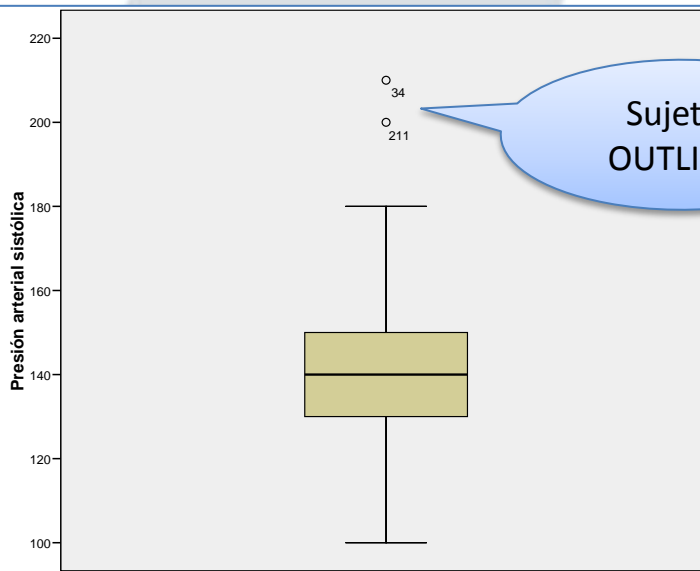


Bigote superior.
Se desplaza en función de la variabilidad hasta +3 d.t., por encima de la media. Llega hasta los percentiles 98-99, aprox.

Bigote inferior.
Se desplaza en función de la variabilidad hasta -3 d.t., por debajo de la media. Alcanza entre los percentiles 1-2, aprox.



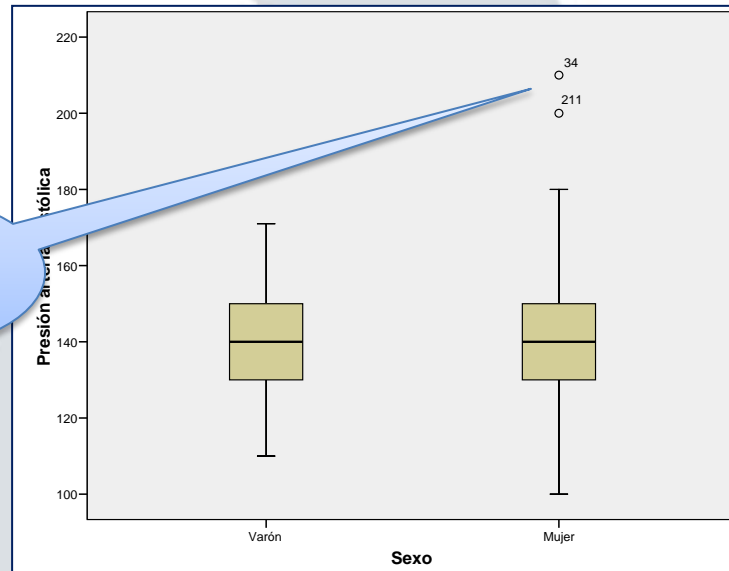
Ejemplo, resuelto con IBM-SPSS.19



Sujetos OUTLIERS

Casos OUTLIERS

Lo más adecuado es definirlos como valores perdidos



Estudio de la forma: ajuste al modelo normal



La distribución “normal” es un fenómeno natural que permite asociar probabilidades a cada valor de una variable cuantitativa



Es importante determinar si una variable cuantitativa sigue, o no, un modelo de **CURVA NORMAL**

La distribución normal de la variable es, frecuentemente, una condición necesaria para la utilización de bastante técnicas **paramétricas** especialmente de **contraste de hipótesis**

Análisis estadístico

Herramientas DESCRIPTIVAS

Índice de
ASIMETRÍA

Índice de
CURTOSIS

Generan, cada uno de ellos, un valor numérico que se interpreta bajo ciertas reglas (ver siguiente diapositiva)

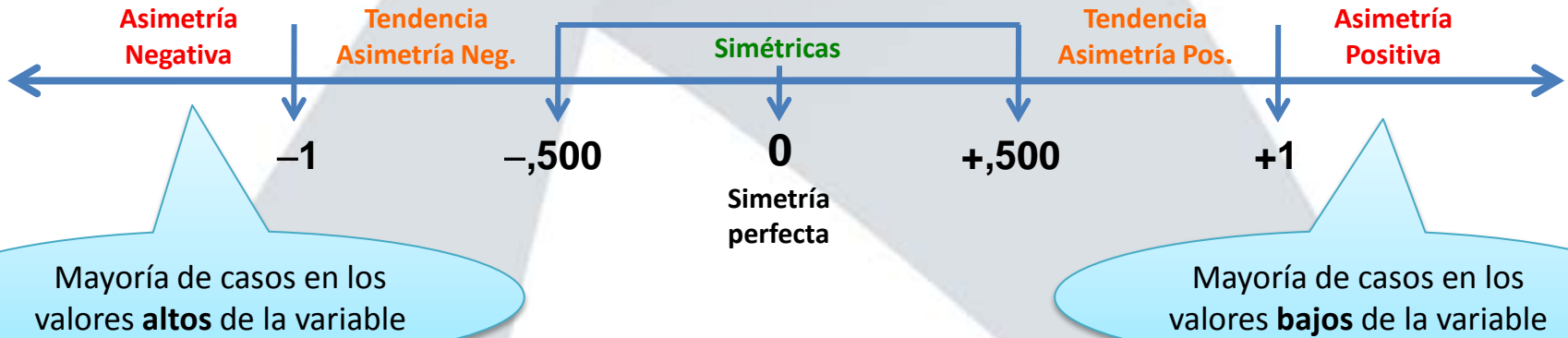
Procedimiento INFERENCIAL

Test de K-S (Kolmogorov-Smirnov)
de bondad de ajuste

Genera, un valor de p de significación en función del cual, se toma la decisión de aceptar o rechazar que la variable no difiere significativamente del modelo normal

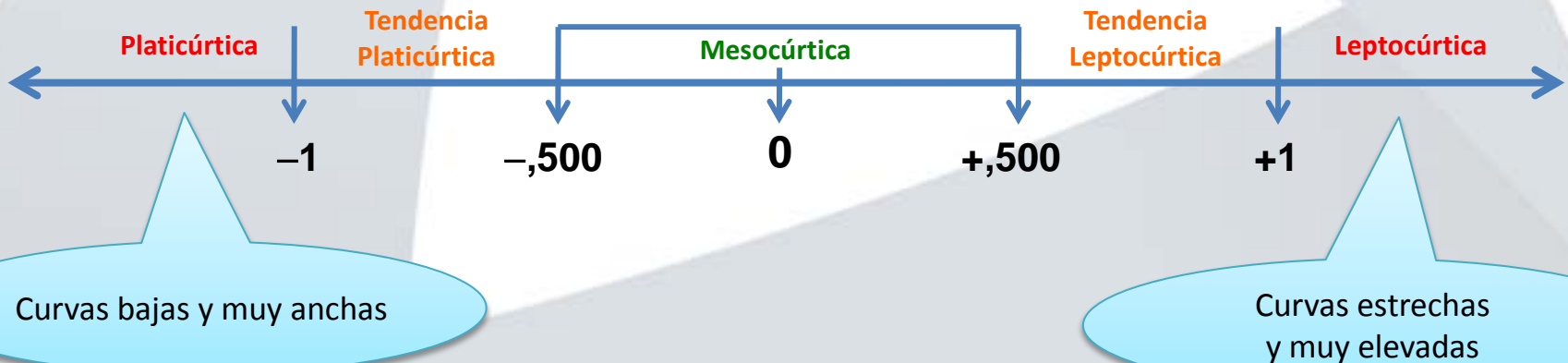
Índice de ASIMETRÍA

Grado de desviación de la "igualdad de distancias" entre puntos con respecto al valor central (la mediana)



Índice de CURTOSIS

Altura de la curva hasta el eje de abscisas, medida en el valor central (la mediana).
NO confundir con la altura máxima.



Test de K-S (Kolmogorov-Smirnov) de bondad de ajuste

Técnica inferencial de tipo no-paramétrico, cuyo objetivo es contrastar la hipótesis de que una variable cuantitativa se acomoda el modelo de una distribución normal

Hipótesis Nula: La variable se distribuye según el modelo de la normal

Hipótesis Alternativa: La variable NO se distribuye normalmente

Recuérdese que las hipótesis estadísticas aunque no se citan en las publicaciones científicas, son necesarias para entender la decisión estadística que se toma

Se resuelve mediante

Estadístico de contraste:
Chi-cuadrado
(Transformado en Z equivalente)

Si $p \geq ,010 \Rightarrow A Ho \Rightarrow$ Conclusión: La variable se distribuye según el modelo de la normal

Si $p < ,010 \Rightarrow R Ho \Rightarrow$ Conclusión : La variable NO se distribuye normalmente, porque la distribución empírica difiere significativamente del modelo

Advertencia: Si el valor de la p de significación está comprendido entre ,050 y ,010 (el 5% y el 1%) por lo que habitualmente se R Ho, el desvío de la normalidad NO es grave. El incumplimiento de la condición de normalidad puede tener efectos de gravedad sólo si $p < ,010$ (<1%)

Y un recordatorio, para cuando usamos SPSS



5%	1%
0,05	0,01
,05	,01
,050	,010

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desv. típ.	Asimetría		Curtosis	
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Error típico	Estadístico	Error típico
Presión arterial sistólica	226	100	210	139,11	16,070	,556	,162	1,815	,322
Presión arterial diastólica	226	50	110	77,56	9,596	-,113	,162	,193	,322
N válido (según lista)	226								

Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desv. típ.	Asimetría		Curtosis	
	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Estadístico	Error típico	Estadístico	Error típico
Presión arterial sistólica	224	100	180	138,52	14,868	,027	,163	-,036	,324
Presión arterial diastólica	224	55	100	77,54	9,208	-,199	,163	-,272	,324
N válido (según lista)	223								

Prueba de Kolmogorov-Smirnov para una muestra

	Presión arterial sistólica	Presión arterial diastólica
N	224	224
Parámetros normales ^{a,b}		
Media	138,52	77,54
Desviación típica	14,868	9,208
Diferencias más extremas		
Absoluta	,124	,208
Positiva	,110	,131
Negativa	-,124	-,208
Z de Kolmogorov-Smirnov	1,862	3,113
Sig. asintót. (bilateral)	,002	,000

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Descriptiva de: Centralidad y Variabilidad

CENTRALIDAD

Indicador de **nivel colectivo** del grupo en la variable medida

- **Media**
- **Mediana**

VARIABILIDAD

Indicador del **grado de semejanza/desemejanza** entre los casos del grupo

- **Desviación típica (Varianza)**
- **Amplitud Semi-intercuartil**

En función de la forma

Si la variable se distribuye normalmente

Media y Desv. Típica

Si la variable **NO** se distribuye normalmente

Mediana y Amp.Semi.

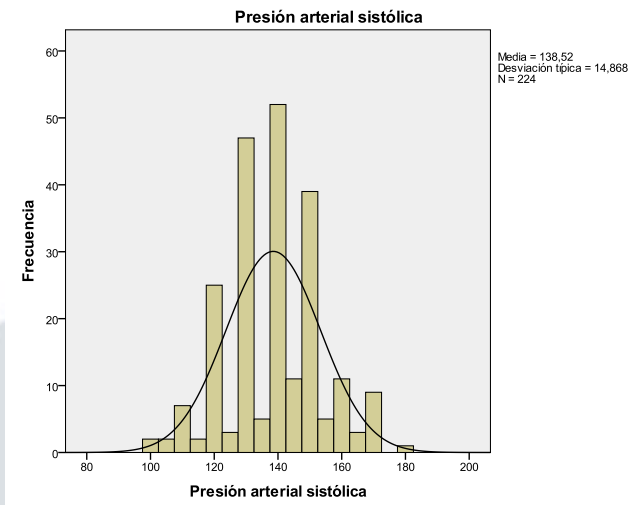


Ejemplo, resuelto con IBM-SPSS.19

Descriptivos			Estadístico	Error típ.
Presión arterial sistólica	Media		138,67	,986
	Intervalo de confianza para la media al 95%	Límite inferior	136,73	
		Límite superior	140,62	
	Media recortada al 5%		138,61	
	Mediana		140,00	
	Varianza		216,960	
	Desv. típ.		14,730	
	Mínimo		100	
	Máximo		180	
	Rango		80	
	Amplitud intercuartil		20	
	Asimetría		,051	,163
	Curtosis		-,034	,324
	Presión arterial diastólica	Media		77,57
Intervalo de confianza para la media al 95%		Límite inferior	76,36	
		Límite superior	78,79	
Media recortada al 5%			77,60	
Mediana			80,00	
Varianza			84,912	
Desv. típ.			9,215	
Mínimo			55	
Máximo			100	
Rango			45	
Amplitud intercuartil			12	
Asimetría			-,208	,163
Curtosis			-,267	,324

Estadísticos

		Presión arterial sistólica	Presión arterial diastólica
N	Válidos	224	224
	Perdidos	68	68
Media		138,52	77,54
Mediana		140,00	80,00
Desv. típ.		14,868	9,208
Varianza		221,049	84,788
Asimetría		,027	-,199
Error típ. de asimetría		,163	,163
Curtosis		-,036	-,272
Error típ. de curtosis		,324	,324
Mínimo		100	55
Máximo		180	100
Percentiles	25	130,00	70,00
	50	140,00	80,00
	75	150,00	82,00



Muchas gracias por su atención



3 datos
tratamiento estadístico

www.3datos.es