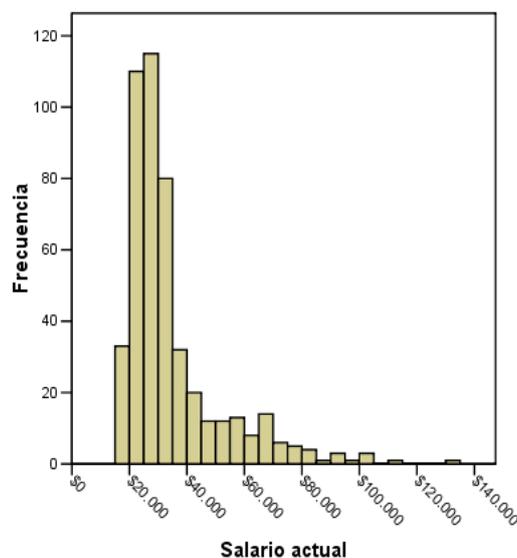


3.3 – Caracterización de grupos: Estadísticos de forma de la distribución

1. Simetría
2. Apuntamiento o curtosis
3. Descripción estadística de una variable: tabla resumen

- Ya ha sido abordado en temas precedentes el análisis de la forma de la distribución de frecuencias desde una aproximación gráfica. De hecho, se trata de la forma más directa e intuitiva de hacerse una idea acerca de la forma de la distribución de una variable.
- Como se vio en su momento, conocer la forma de una distribución resultaba importante, por ejemplo, a la hora de decidir qué estadísticos de posición y dispersión era oportuno aplicar con variables cuantitativas. En cualquier caso, el examen de la forma de la distribución de una variable va a aportar información relevante por sí misma a la hora de describir esa variable.

Ejemplo: ¿Qué nos dice la forma de la distribución de la variable “Salario actual” que se muestra en el siguiente histograma?



- En este tema se presentan diversos índices que permiten cuantificar esa forma de la distribución, en concreto, dos facetas de la misma: la simetría y el apuntamiento (o curtosis).

1. Simetría

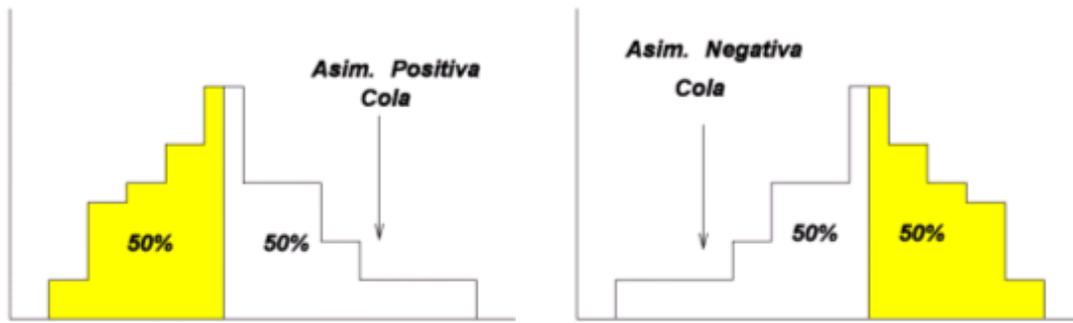
- La simetría de una distribución de frecuencias hace referencia al grado en que valores de la variable, equidistantes a un valor que se considere centro de la distribución, poseen frecuencias más o menos iguales. Cuanto más similares sean, más simétrica será la distribución; cuanto más distintas, más asimétrica.
- Es un concepto que resulta más intuitivo de comprender a nivel visual, especialmente, si se observa una representación gráfica de la distribución de frecuencias de la variable (gráfico de barras, histograma...). Ésta será simétrica si la mitad izquierda de la distribución es la imagen especular de la mitad derecha.

Ejemplos de distribución simétrica:



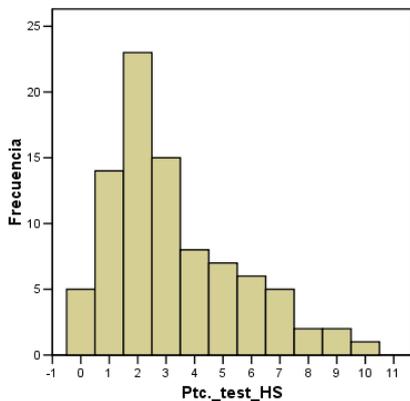
Media y mediana coinciden en las distribuciones simétricas. Si sólo hay una moda (distribución unimodal), el valor de ésta también será igual a las dos anteriores

- En distribuciones unimodales, el nivel de simetría se suele describir de acuerdo a tres grandes categorías: distribuciones simétricas, distribuciones asimétricas positivas (o asimetría a la derecha) y distribuciones asimétricas negativas (o asimetría a la izquierda). Tomando como eje de referencia a la moda, estas categorías de asimetría vienen definidas por el diferente grado de dispersión de los datos a ambos lados (colas) de ese eje virtual. La cola más dispersa en el lado de los valores altos de la variable caracteriza a la asimetría positiva; si en el lado de los más bajos, a la asimetría negativa; y si la dispersión es igual o muy similar a ambos lados, a una distribución de frecuencias simétrica.

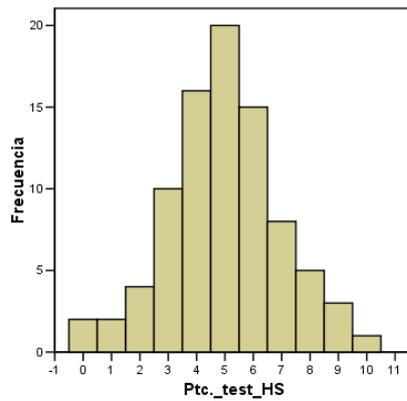


- En caso de asimetría, los valores de la media, de la mediana y de la moda difieren. En concreto, si la asimetría es positiva: $\bar{X} > Mdn \geq Mo$; si negativa: $\bar{X} < Mdn \leq Mo$.

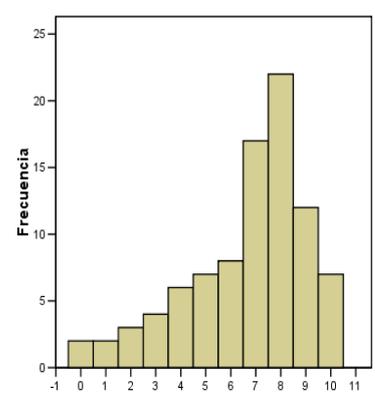
Ejemplo de las puntuaciones de un grupo de sujetos en un test de habilidades sociales antes, durante y después de recibir 6 sesiones de entrenamiento en habilidades sociales.



Antes ($\bar{X}=3,26$; $Mdn=3$; $Mo=2$)



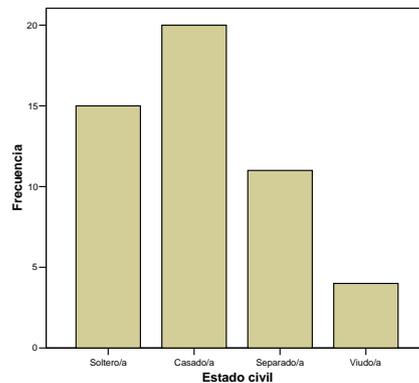
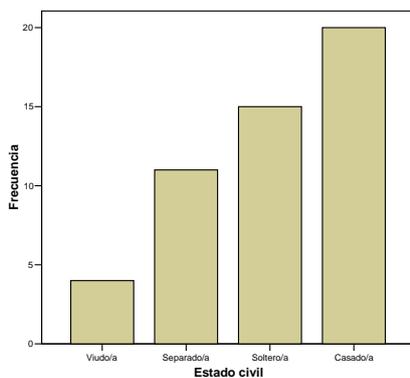
Durante ($\bar{X}=4,97$; $Mdn=5$; $Mo=5$)



Después ($\bar{X}=6,67$; $Mdn=7$; $Mo=8$)

- A continuación se presentan diferentes índices estadísticos que permiten cuantificar el grado de simetría de una variable. Destacar antes que para variables categóricas no tiene sentido el plantear este tipo de índices, dado que no existe un orden intrínseco a los valores de la variable.

Ver, por **ejemplo**, los dos gráficos de barras de una misma distribución de frecuencias de la variable “Estado civil” en las que lo único que se ha cambiado es la posición de las barras:

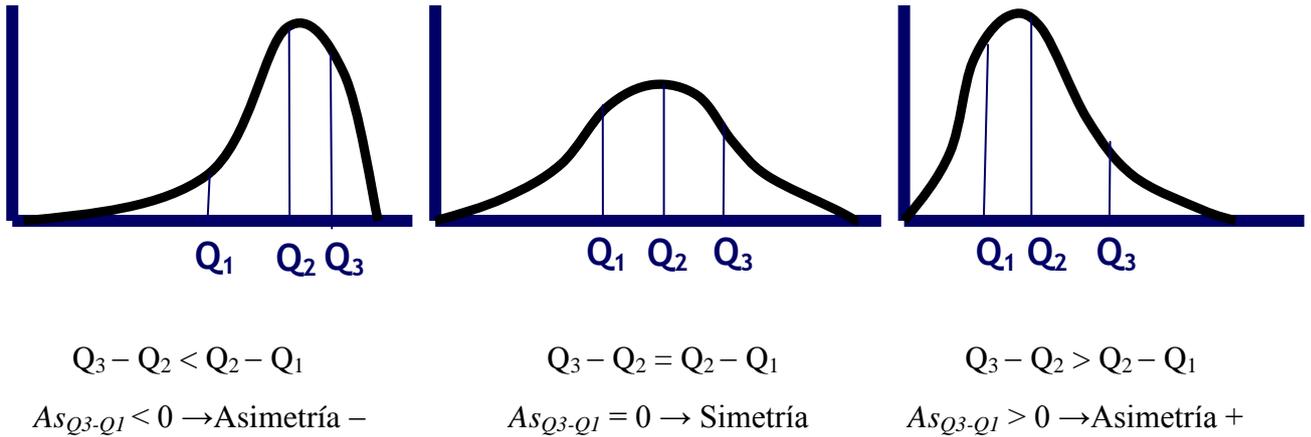


1.1. Variables ordinales: el índice de asimetría intercuartil

• El índice de asimetría intercuartil se basa en las distancias entre los cuartiles a fin de establecer un resumen de la asimetría de la distribución. La fórmula es la siguiente:

$$As_{Q_3-Q_1} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

• Interpretación: oscila entre -1 y 1, lo cual facilita su comunicación y comprensión.



Ejercicio 1: Obtener el $As_{Q_3-Q_1}$ para las distribuciones de frecuencias de 3 grupos de 100 casos cada uno (A, B y C) que, en el desarrollo de una investigación, cumplieron un test que constaba de 10 ítems, cada uno de los cuales era valorado con 1 punto si estaba ejecutado de forma totalmente correcta, un 0 en cualquier otro caso. La puntuación en el test para cada sujeto se obtenía como suma de las puntuaciones de los ítems, pudiendo por tanto oscilar entre 0 y 10 (aunque la variable podría considerarse como cuantitativa, asúmase para este ejercicio que es ordinal). Obtener también los respectivos gráficos de caja y bigotes.

	Grupo A	Grupo B	Grupo C
Nota	n_i	n_i	n_i
0	1	4	5
1	3	5	11
2	3	8	14
3	5	9	23
4	8	15	15
5	11	18	12
6	15	15	9
7	24	9	6
8	16	8	2
9	9	5	2
10	5	4	1
	100	100	100



1.2. Variables cuantitativas: el coeficiente de asimetría de Fisher.

• El coeficiente de asimetría de Fisher se basa en las desviaciones de los valores observados respecto a la media. La fórmula para su cálculo es la siguiente:

$$As_F = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n \cdot S_X^3} \quad (\text{versión para distribución de frecuencias: } As_F = \frac{\sum n_i (X_i - \bar{X})^3}{n \cdot S_X^3})$$

• Interpretación: los valores menores que 0 indican asimetría negativa; los mayores, asimetría positiva; y cuando sea 0, o muy próximo a 0, simetría. No está limitado a un rango de valores.

• El coeficiente de asimetría de Fisher es precisamente el que obtiene el paquete estadístico SPSS cuando se le solicita el cálculo de la asimetría de una variable. Obtendremos también con este programa, por defecto, el valor del error típico asociado a este estadístico, un concepto que se tratará más adelante (ver figura a continuación). Sin embargo, conviene señalar ya ahora que ambos valores nos van a permitir obtener un criterio más objetivo en la valoración de la asimetría de la distribución de una variable: si se divide el valor del coeficiente de asimetría de Fisher entre el correspondiente error típico, se obtendrá un valor que: si es inferior a -2, se considera como indicativo de distribución asimétrica negativa; si entre -2 y 2, de distribución simétrica, y si superior a 2, de distribución asimétrica positiva.

Estadísticos

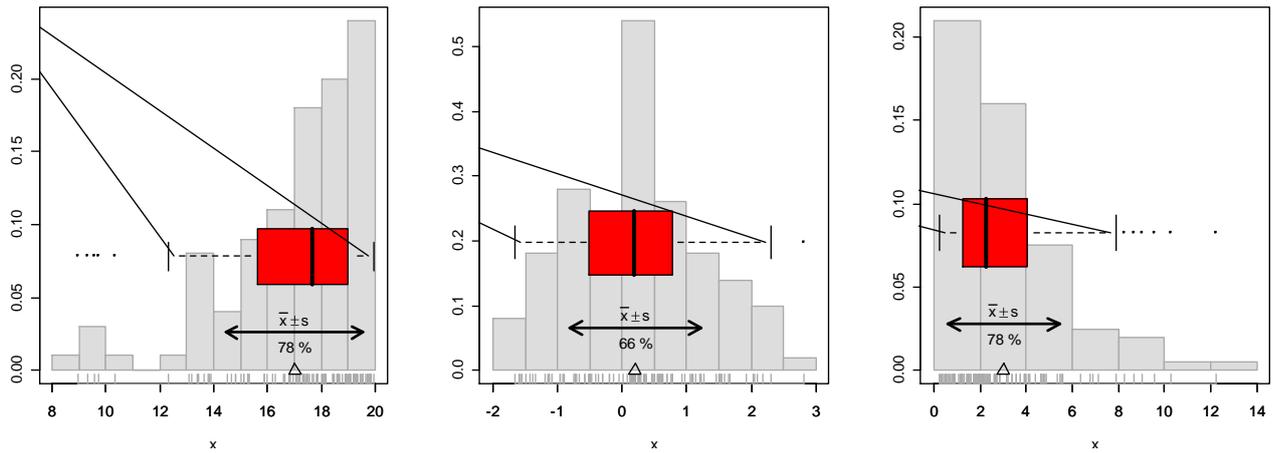
Nota media de acceso

N	Válidos	169
	Perdidos	5
Media		6,3885
Mediana		6,3000
Moda		6,60
Desv. típ.		,55429
Asimetría		1,139
Error típ. de asimetría		,187
Curtosis		2,733
Error típ. de curtosis		,371
Rango		3,99
Mínimo		5,06
Máximo		9,05

Ejercicio 2: Valora la simetría de la variable “Nota media de acceso” a partir de la salida de SPSS obtenida para esta variable.

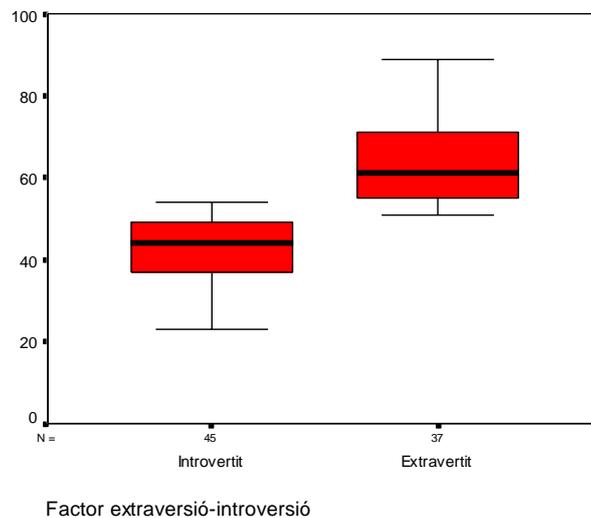


• Acorde al tipo de variable que nos ocupa, el histograma representa la mejor opción en la visualización de la simetría de una variable cuantitativa, si bien, el gráfico de caja y bigotes también constituye una opción válida para tal fin. A continuación se presenta un ejemplo con ambos tipos de gráficos superpuestos (Barón-López, 2005) en que se muestran 3 variables que ilustran distribuciones con diferente nivel de simetría. Señalar que el gráfico de caja y bigotes aparece tumbado a la derecha, algo que no es habitual en la presentación de este tipo de gráfico, aunque se ha considerado así en esta ilustración con fines didácticos.



• Tal como ya se destacó en el capítulo previo, una ventaja importante de los gráficos de caja y bigotes es la facilidad para representar varios de ellos conjuntamente y, en consecuencia, para realizar comparaciones entre diferentes distribuciones.

Ejemplo con las puntuaciones en un test de impulsividad para un grupo de sujetos introvertidos y para otro grupo de extrvertidos en que se pone de manifiesto como para los primeros la variable es algo asimétrica negativa, mientras que para los segundos es asimétrica positiva:



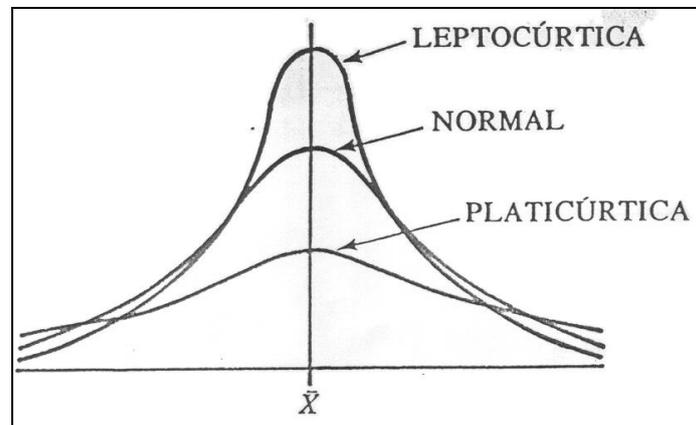
2. Apuntamiento (curtosis)

• El apuntamiento o curtosis de una distribución de frecuencias no tiene un referente natural como en el caso de la simetría, sino que se sustenta en la comparación respecto a una distribución de referencia, en concreto, la distribución normal o campana de Gauss. En consecuencia, su obtención sólo tendrá sentido en variables cuya distribución de frecuencias sea similar a la de la curva normal –en la práctica ello se reduce, básicamente, a que sea unimodal y más o menos simétrica.

• El apuntamiento expresa el grado en que una distribución acumula casos en sus colas en comparación con los casos acumulados en las colas de una distribución normal cuya dispersión sea equivalente (Pardo y Ruiz, 2002). Así, de forma análoga a la asimetría, se diferencian 3 grandes categorías de apuntamiento:

- Distribución platicúrtica (apuntamiento negativo): indica que en las colas hay más casos acumulados que en las colas de una distribución normal.
- Distribución leptocúrtica (apuntamiento positivo): justo lo contrario.
- Distribución mesocúrtica (apuntamiento normal): como en la distribución normal.

Ejemplos gráficos de estas tres formas de apuntamiento en la distribución:

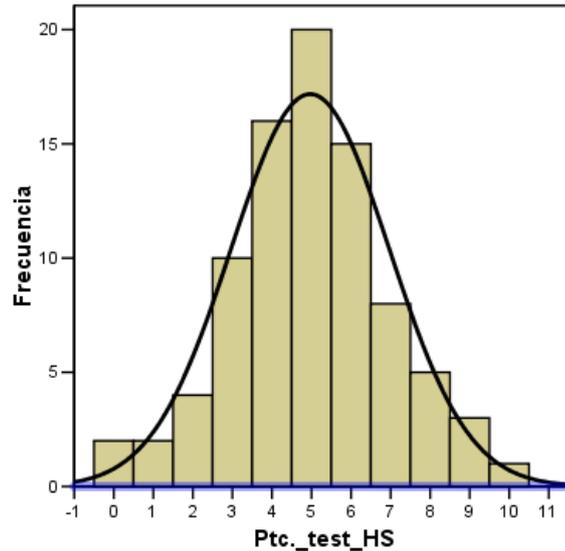


• El coeficiente de apuntamiento de Fisher permite describir el apuntamiento o curtosis de una distribución de datos y se basa en las desviaciones de los valores observados respecto a la media. La fórmula para su cálculo es la siguiente:

$$Ap_F = \frac{\sum (X_i - \bar{X})^4}{N \cdot S_x^4} - 3 \quad (\text{versión para distr. de frecuencias: } Ap_F = \frac{\sum n_i (X_i - \bar{X})^4}{N \cdot S_x^4} - 3)$$

- **Interpretación:** el valor de este coeficiente para la distribución normal será igual a 0, o sea que cualquier distribución para la que se obtenga un valor de Ap_F igual o próximo a 0 significará que su nivel de apuntamiento es como el de la distribución normal (mesocúrtica). Valores mayores que 0, expresan que la distribución es leptocúrtica, mientras que si son menores que 0 ponen de manifiesto que la distribución es platicúrtica. No está limitado a un rango de valores.

Ejercicio 3: Valorar la curtosis a partir del histograma de la distribución de la variable con las puntuaciones en el test de habilidades sociales. En el mismo aparece superpuesta la curva suavizada de esta variable en el caso en que se distribuyese según la curva normal.



- El coeficiente de apuntamiento de Fisher es el que obtiene el paquete estadístico SPSS cuando se le solicita el cálculo de la curtosis de una variable. Obtendremos también con este programa, por defecto (ver figura a continuación), el valor del error típico asociado a este estadístico. Ambos valores nos van a permitir obtener un criterio más objetivo en la valoración de la curtosis de la distribución de una variable: si se divide el valor del coeficiente de apuntamiento de Fisher entre el correspondiente error típico, se obtendrá un valor que: si es inferior a -2 , se considera como indicativo de distribución platicúrtica; si entre -2 y 2 , de distribución mesocúrtica, y si superior a 2 , de distribución leptocúrtica.

Estadísticos

Nota media de acceso

N	Válidos	169
	Perdidos	5
Media		6,3885
Mediana		6,3000
Moda		6,60
Desv. típ.		,55429
Asimetría		1,139
Error típ. de asimetría		,187
Curtosis		2,733
Error típ. de curtosis		,371
Rango		3,99
Mínimo		5,06
Máximo		9,05

Ejercicio 4: Valora el apuntamiento de la variable “Nota media de acceso” a partir de la salida de SPSS obtenida para esta variable.

3. Descripción estadística de una variable: tabla resumen

• En este tema y los precedentes (temas del 2 al 5) se han presentado una serie de procedimientos estadísticos, tanto numéricos como gráficos, adecuados para describir y/o resumir los valores obtenidos al medir una variable en un conjunto de casos. En la tabla siguiente se resume esta información clasificada en función de la escala de medida de la variable que se desea describir.

	Categorica	Ordinal	Cuantitativa simétrica	Cuantitativa asimétrica
Gráficos	Gráfico de sectores Gráfico de barras Pictograma	Polígono de frecuencias Gráfico de caja y bigotes Histograma (sólo para variables continua)		
Tendencia central	Moda	Mediana	Media aritmética	Mediana Media recortada
Variabilidad	Índice de variación cualitativa	Rango (Amplitud) Rango intercuartil	Varianza Desviación estándar Coeficiente de variación	Rango intercuartil
Simetría		Índice de asimetría intercuartil	Coeficiente de asimetría de Fisher	
Curtosis			Coeficiente de apuntamiento de Fisher	



Referencias:

Barón-López, J. (2005). *Bioestadística: métodos y aplicaciones*. Apuntes y material disponible en <http://www.bioestadistica.uma.es/baron/apuntes/>

Pardo, A. y Ruiz, M. A. (2002). *SPSS: Guía para el análisis de datos*. Madrid: McGraw-Hill.

