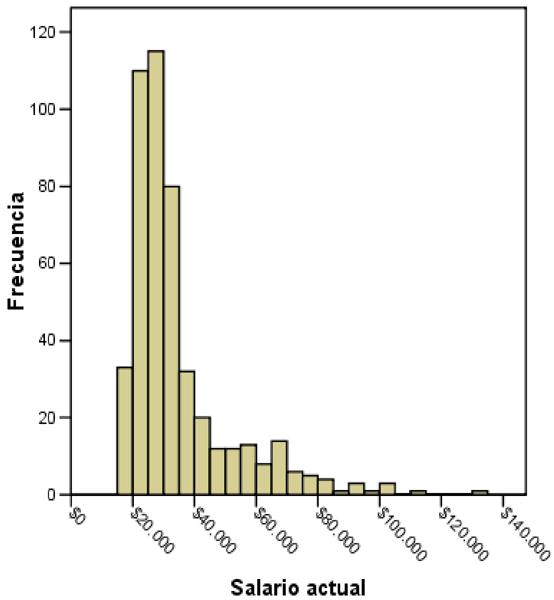
T. 5 – Estadísticos de forma de la distribución

1. **Asimetría**
2. **Apuntamiento o curtosis**

• 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 relevante a la hora de decidir qué estadísticos de posición y dispersión era oportuno utilizar con variables cuantitativas. En cualquier caso, su examen va a aportar información relevante por sí misma a la hora de describir a una variable.

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



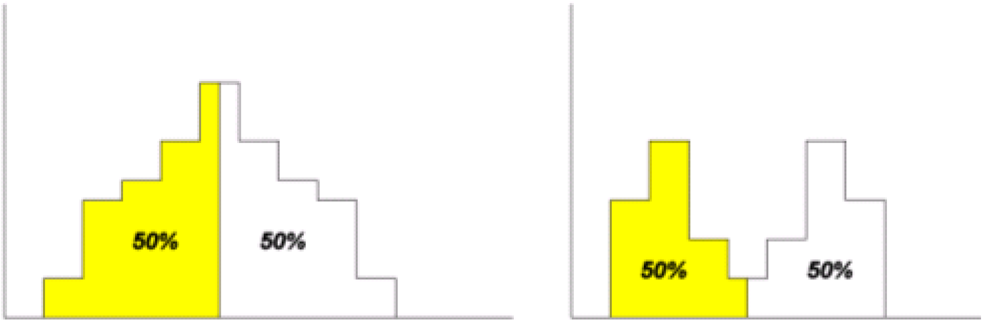
• En este tema se presentan diversos índices que permiten cuantificar esa forma de la distribución, en concreto, dos facetas de la misma: la asimetría y el apuntamiento (o curtosis).

1. Asimetría

• La simetría (también denominada sesgo) 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 similares.

• Es un concepto más intuitivo a nivel visual, especialmente, si se observa una representación gráfica (diagrama de barras, histograma…) de la distribución de frecuencias. Ésta será simétrica si la mitad izquierda de la distribución es la imagen especular de la mitad derecha.

Ejemplos de distribución simétrica:

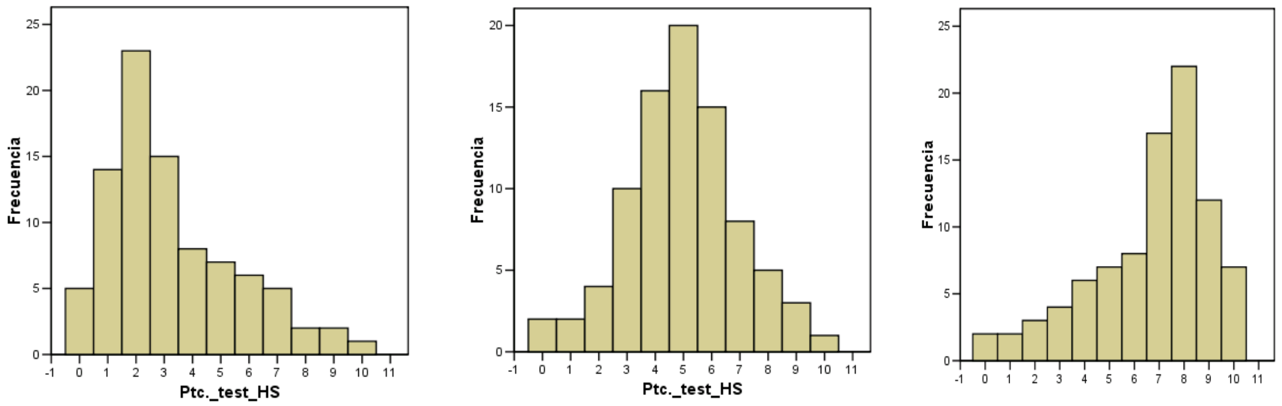


Media y mediana coinciden en las distribuciones simétricas. Si sólo hay una moda (distribución unimodal), el valor de ésta también será igual a las dos anteriores

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

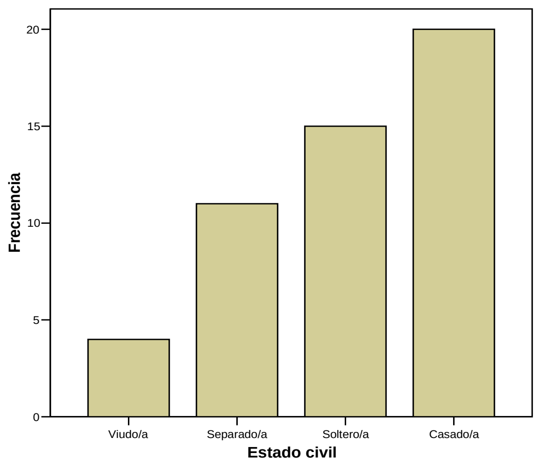
Ejemplo de las puntuaciones de un grupo de sujetos en un test de habilidades sociales antes, durante y después de recibir 6 sesiones de entrenamiento en habilidades sociales.

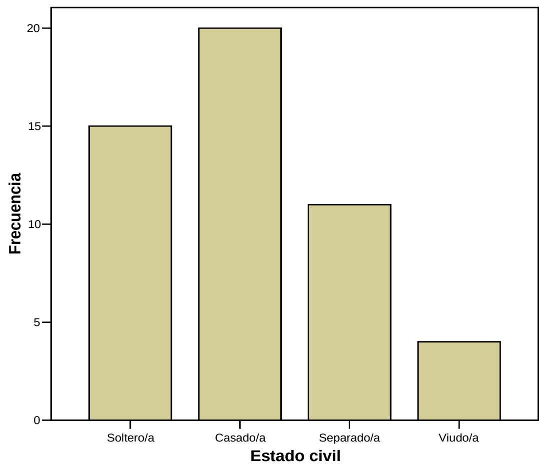


Antes ($\overline{X}$=3,26; *Mdn*=3; *Mo*=2) Durante ($\overline{X}$=4,97; *Mdn*=5; *Mo*=5) Después ($\overline{X}$=6,67; *Mdn*=7; *Mo*=8)

• A continuación se presentan diferentes índices estadísticos que permiten cuantificar el nivel de asimetría de una variable. Destacar antes que para <u>variables categóricas</u> 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 diagramas 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:



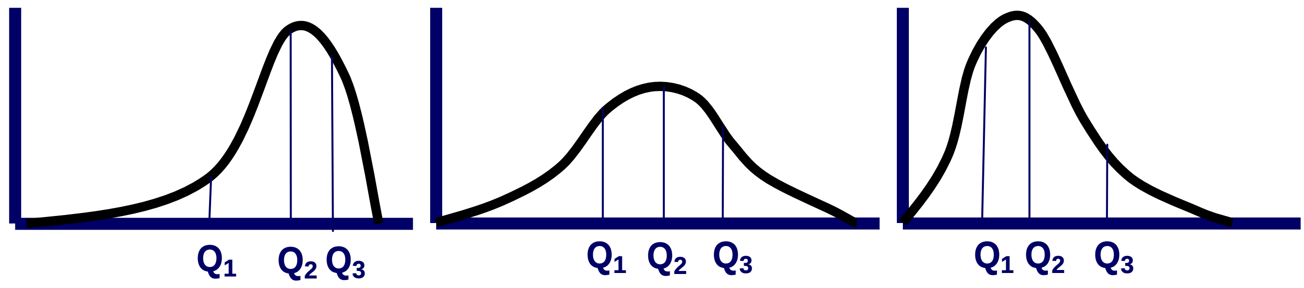


1.1. Variables ordinales: el índice de asimetría intercuartílico.

• El índice de asimetría intercuartílico 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 comprensión.



$Q_3 - Q_2 < Q_2 - Q_1$	$Q_3 - Q_2 = Q_2 - Q_1$	$Q_3 - Q_2 > Q_2 - Q_1$
$As_{Q3\text{-}Q1} < 0$ →Asimetría –	$As_{Q3\text{-}Q1} = 0$ → Simetría	$As_{Q3\text{-}Q1} > 0$ →Asimetría +

Ejercicio 1: Obtener el $As_{Q3\text{-}Q1}$ para las distribuciones de frecuencias de 3 grupos de 10 casos cada uno (A, B y C) que, en el desarrollo de una investigación, cumplimentaron 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 diagramas 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 primer coeficiente de Pearson y el coeficiente de asimetría de Fisher.

• El primer coeficiente de Pearson se basa en la relación existente entre la media y la moda en distribuciones unimodales asimétricas (ver más arriba). Su fórmula es la siguiente:

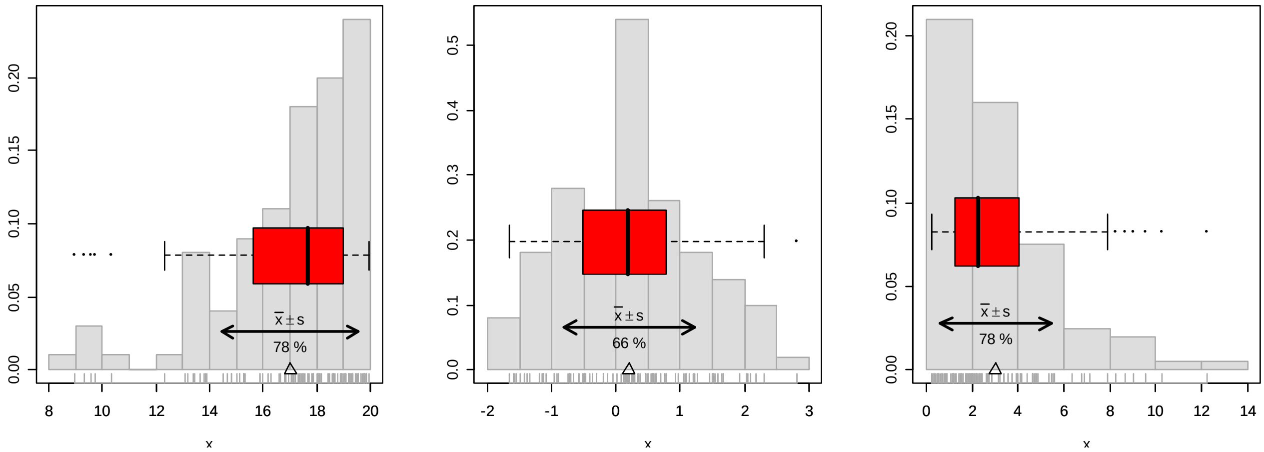
$$As_1 = \frac{\overline{X} - Mo}{S_X}$$

• 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 se basa en las desviaciones de los valores observados respecto a la media. La interpretación de los resultados proporcionados por este coeficiente es igual a la del primer coeficiente de Pearson. La fórmula para su cálculo es la siguiente:

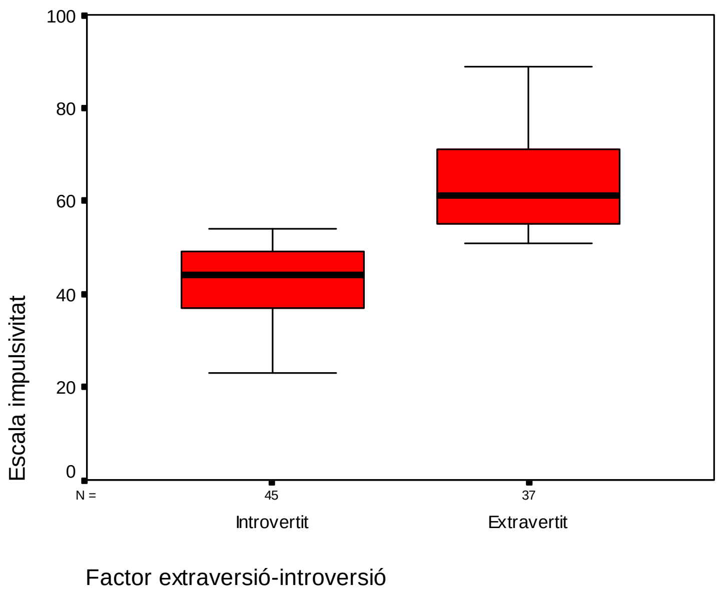
$$As_F = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{n \cdot S_X^3}$$ (versión para distr. de frecuencias: $$As_F = \frac{\sum n_i (X_i - \overline{X})^3}{n \cdot S_X^3}$$)

• Acorde al tipo de variable que nos ocupa, el histograma representa la mejor opción en la visualización de la asimetría de una variable, si bien, el diagrama 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 asimetría:



• Tal como ya se destacó en el capítulo previo, una ventaja importante de los diagramas 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 extravertidos:



Ejercicio 2: La distribución de frecuencias que se muestra a continuación corresponde a las puntuaciones en un test de habilidades sociales aplicado a una muestra de 86 sujetos tras la tercera de seis sesiones que recibieron a fin de mejorar este tipo habilidades. Valorar la asimetría de esta distribución, con el primer coeficiente de Pearson. Puede verse el histograma de esta distribución de frecuencias más arriba en la introducción de esta sección ("Durante").

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
0	2	2,3	2,3	2,3
1	2	2,3	2,3	4,7
2	4	4,7	4,7	9,3
3	10	11,6	11,6	20,9
4	16	18,6	18,6	39,5
5	20	23,3	23,3	62,8
6	15	17,4	17,4	80,2
7	8	9,3	9,3	89,5
8	5	5,8	5,8	95,3
9	3	3,5	3,5	98,8
10	1	1,2	1,2	100,0
Total	86	100,0	100,0	

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.

2.1. Variables ordinales: el índice K_2.

• El índice K_2 se basa en la comparación de la dispersión existente en el 80% central de la distribución con la existente en el 50% central. Su fórmula es la siguiente:

$$K_2 = \frac{P_{90} - P_{10}}{1,9 \cdot (P_{75} - P_{25})}$$

• Interpretación: valores igual o muy próximos a 1 corresponden a una distribución mesocúrtica (apuntamiento como la distribución normal); valores mayores que 1 ponen de manifiesto que la distribución es leptocúrtica (más 'puntiaguda' que la normal); mientras que si son menores que 1 indican que la distribución es platicúrtica (más 'aplastada' que la normal). Este coeficiente no está limitado a un rango de valores.

Ejercicio 3: Obtener el índice de apuntamiento K_2 para la distribución de frecuencias presentada en el ejercicio 1 para las notas en el grupo B.

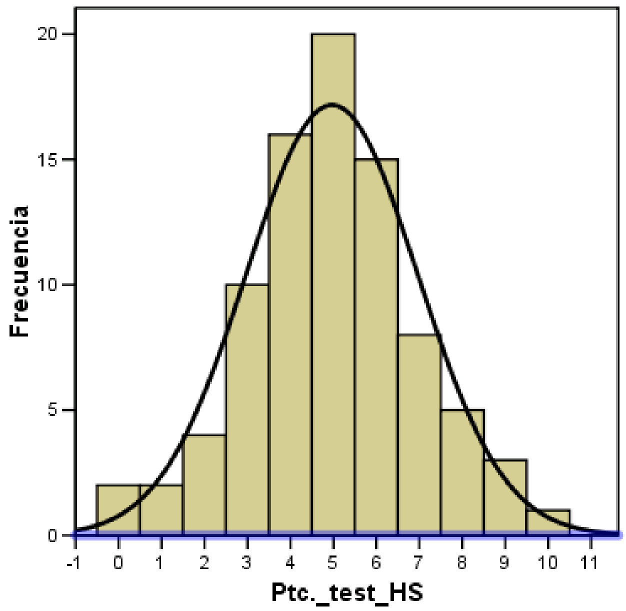
2.2. Variables cuantitativas: el coeficiente de apuntamiento de Fisher

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

$$Ap_F = \frac{\sum (X_i - \bar{X})^4}{N \cdot S_X^4} - 3$$ (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 4: 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.



Referencias:

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

Rius, F., Barón-López, F. J., Sánchez, E. y Parras, L. (2006). *Bioestadística: métodos y aplicaciones.* Disponible en www.bioestadistica.uma.es/libro/

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