

Tema 7. La variabilidad total frente a la variabilidad de la mayoría: la curtosis

Tema 7. La variabilidad total frente a la variabilidad de la mayoría: la curtosis 1

El concepto de curtosis y su aportación al análisis de variabilidad.....	1
Análisis gráfico de la curtosis.....	4
Cuantificación de la curtosis.....	5

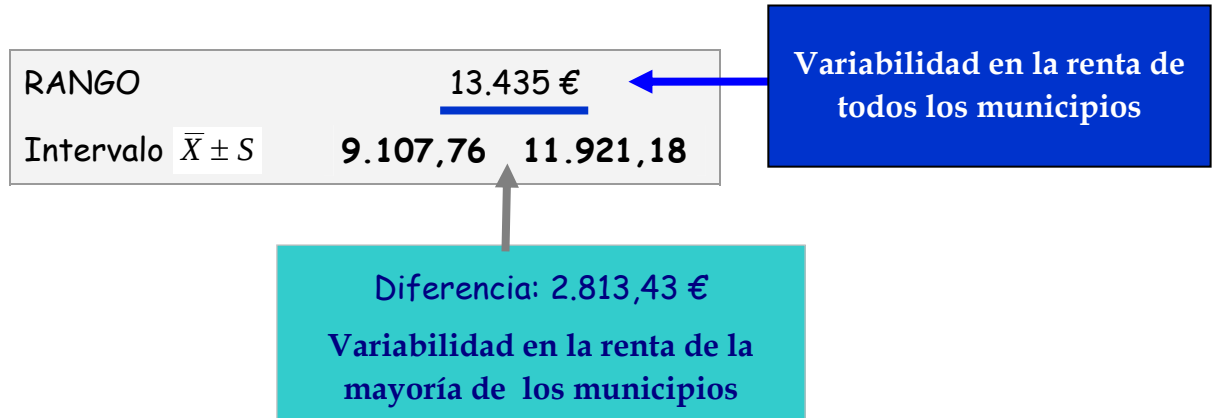
Tema 7:

La variabilidad total frente a la variabilidad de la mayoría: la curtosis

EL CONCEPTO DE CURTOSIS Y SU APORTACIÓN AL ANÁLISIS DE VARIABILIDAD

En el tema 6 hemos podido comprobar que mediante el intervalo que surge al sumar y restar a la media el valor de la desviación estandar logramos una buena descripción sobre la variabilidad del colectivo mayoritario de los valores de la variable.

Veamos de nuevo los resultados obtenidos para el intervalo $\bar{X} \pm S$ en el caso de la variable *Renta personal disponible* de los municipios de Esukadi en 2001:



Puesto que la distribución de los valores de la variable es de tipo campaniforme, sabemos que en un porcentaje de municipios superior al 60%¹ la

¹ En realidad son 200 municipios los que tienen valores comprendidos dentro del intervalo $\bar{X} \pm S$, lo que representa un 80% del total.

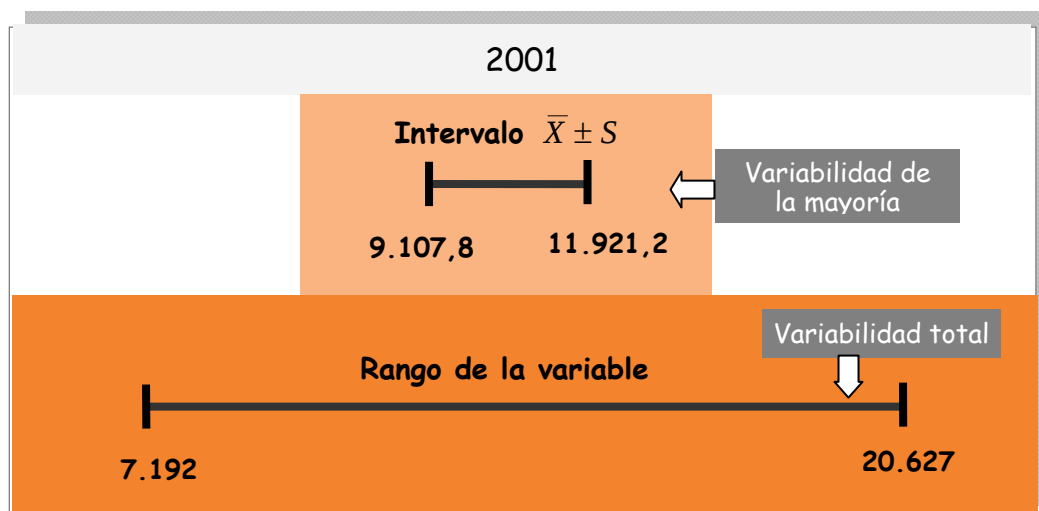
renta personal media de sus habitantes oscila entre 9.107,76 y 11.921,18 euros, lo que supone una variabilidad entre municipios de 2.813,43 euros. Esta variabilidad próxima a los 3.000 euros es pequeña, comparada con la variabilidad total, es decir, con el rango, que asciende a 13.435 €.

A la vista de semejante diferencia entre los dos intervalos (el de $\bar{X} \pm S$ y el del rango) podemos concluir que:

Aunque es cierto que existen algunos municipios cuyas rentas son sensiblemente más altas o más bajas que el promedio, una mayoría de los municipios tiene valores de renta bastante similares, que oscilan entre 9.107,76 y 11.921,18 €. La diferencia máxima entre sus rentas es de 2.813,43 €.

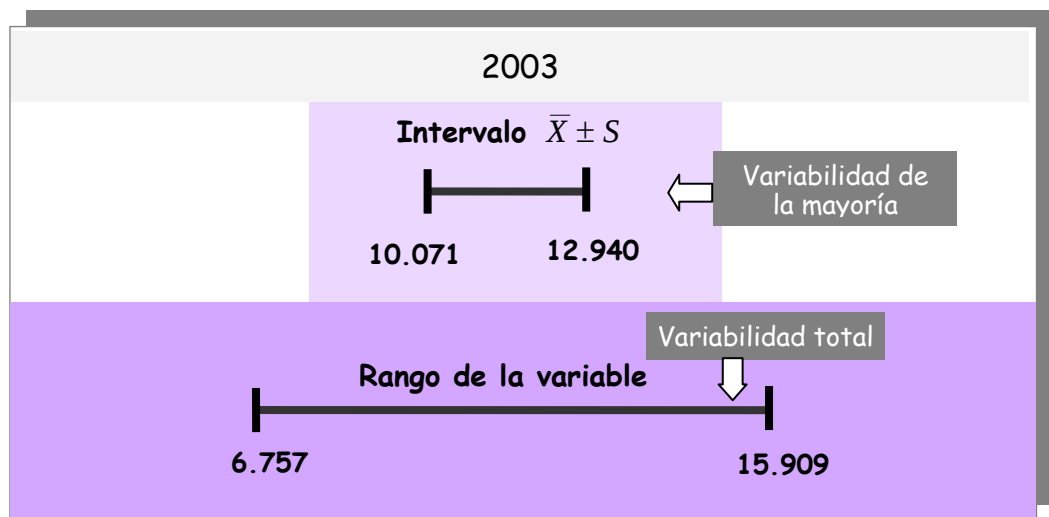
El grado de variabilidad de la mayoría no tiene nada que ver con el grado de variabilidad de toda la población.

La cuestión ahora es encontrar el modo de describir la diferencia de variabilidad entre el grupo mayoritario de municipios y el total de ellos, o lo que es lo mismo, de dejar constancia del grado de concentración de los valores del intervalo $\bar{X} \pm S$ con respecto al rango total de la variable. Volveremos a utilizar el gráfico que representa el rango para ilustrar la idea:



pequeña comparada con la segunda. Esto significa que la mayoría de los valores, agrupados dentro del intervalo $\bar{X} \pm S$ se encuentran verdaderamente concentrados en el sector central del rango. Esto equivale a decir que la variabilidad de esa mayoría es francamente inferior a la variabilidad del total de municipios. (Entre la mayoría de municipios no existe una diferencia de rentas superior a 2.813,42)

Veremos ahora una ilustración igual, pero esta vez con los datos de renta personal disponible de los municipios de Euskadi correspondientes al año 2003. (Ver tabla de datos en ANEXO)



Las diferencias entre las dos ilustraciones, y por tanto las diferencias en la variabilidad de las dos distribuciones, están a la vista: el rango de la variable correspondiente a 2003 es muy inferior al de 2001. También el agrupamiento de valores en el sector central del rango es ligeramente inferior en la distribución correspondiente a 2003.

Si quisiéramos hacer un análisis comparativo entre la renta personal disponible de los municipios en 2001 y 2003 sería necesario dejar constancia de las diferencias con respecto a la variabilidad que hay entre ambos años. Existe para ello una herramienta estadística denominada curtosis, que relaciona precisamente la amplitud del intervalo $\bar{X} \pm S$ y la amplitud total del rango.

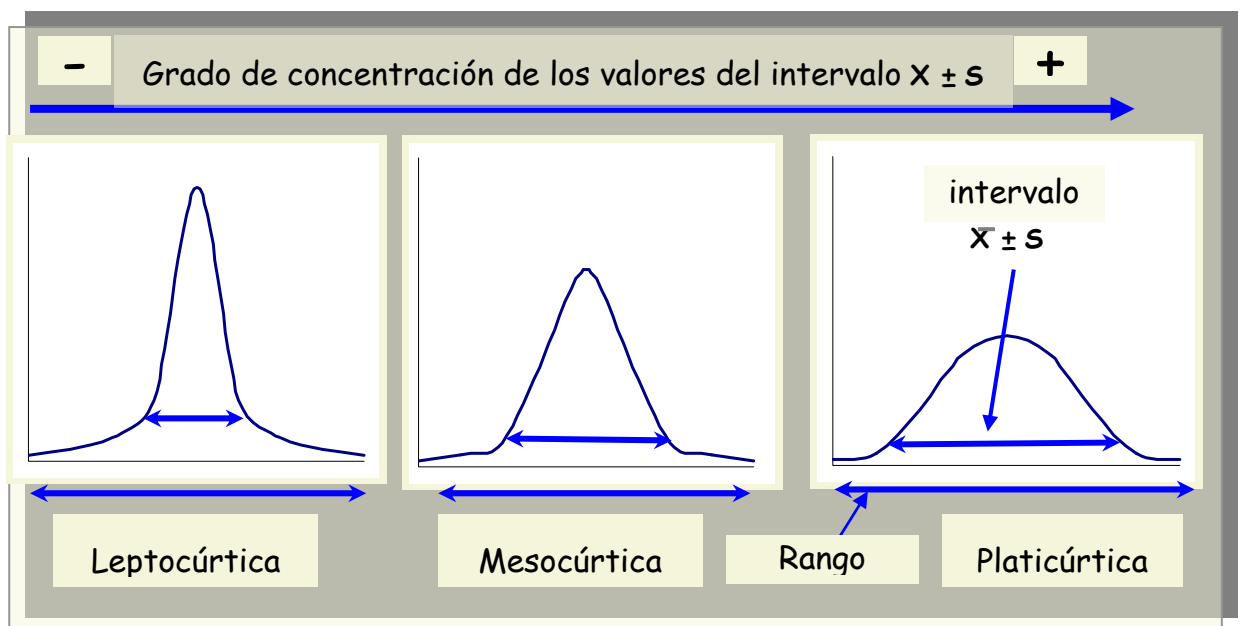
Al igual que ocurre con la simetría, la curtosis de la distribución de valores de una variable se puede analizar de dos modos:

- ▣ Mediante simple observación del histograma o del polígono de frecuencias
- ▣ Mediante la aplicación de un algoritmo que permite asignar un valor a la relación entre el rango de la variable y el del intervalo $\bar{X} \pm S$.

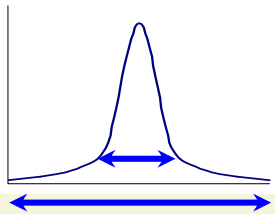
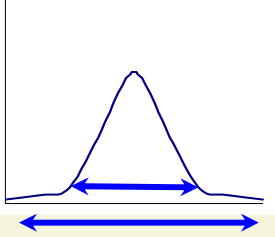
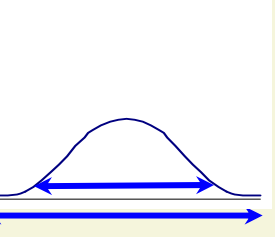
ANÁLISIS GRÁFICO DE LA CURTOSIS

La diferente relación entre el rango de la variable y el intervalo $\bar{X} \pm S$ produce distribuciones de formas muy variadas: cuando la concentración del intervalo $\bar{X} \pm S$ con respecto al rango total de la variable es muy elevada, obtenemos distribuciones con forma de campana apuntada. Por el contrario, cuanto mayor es la variabilidad del intervalo $\bar{X} \pm S$, la forma de la distribución es más achatada.

Habitualmente se clasifican las distribuciones, en función del grado de apuntamiento de la campana en tres tipos:



La curtosis es una medida que relaciona la amplitud del rango de la variable con la amplitud del intervalo $\bar{X} \pm S$. En la imagen, las flechas azules simbolizan la amplitud de los dos intervalos.

Se dice que la distribución es leptocúrtica cuando una mayoría de los valores se concentra en torno a los valores centrales de la distribución. Resultado de esta elevada concentración, este tipo de distribuciones se caracterizan por la diferente amplitud del rango de la variable y del intervalo $\bar{X} \pm S$	
Se dice que la distribución es mesocúrtica cuando presenta un grado de concentración medio alrededor de los valores centrales de la variable.	
Se dice que la distribución es platicúrtica cuando los valores presentan un reducido grado de concentración y se distribuyen prácticamente a lo largo de todo el rango. Resultado de esta baja concentración, este tipo de distribuciones se caracterizan por la similitud en la amplitud del rango de la variable y del intervalo $\bar{X} \pm S$.	

CUANTIFICACIÓN DE LA CURTOSIS

Existen además diferentes fórmulas estadísticas que permiten obtener una medida de la curtosis, es decir, cuantificar el grado de apuntamiento de una distribución. Se conocen con el nombre de *medidas del apuntamiento de una distribución*. Se denominan así debido al hecho de que cuanto mayor es el agrupamiento de los datos en torno al sector central de la distribución, más puntiaguda o apuntada es la forma de la *campana*. Aplicando alguna de las fórmulas existentes para medir la curtosis podemos cuantificar el grado de apuntamiento y definir así, con mayor precisión, el tipo de forma de la distribución.

De entre las fórmulas existentes presentaremos aquí una de las más conocidas: el coeficiente de Curtosis G2 de Fisher:

$$G_2 = \frac{\sum (x_i - \bar{x})^4}{N \cdot S^4} - 3$$

Los criterios para interpretar el resultado del coeficiente de curtosis son los siguientes:

- ▣ $G_2 > 0$. Cuando el resultado es positivo, es decir, cuando es superior a cero, significa que la distribución es leptocúrtica.
- ▣ $G_2 < 0$. Cuando el resultado es negativo, es decir, cuando es inferior a cero, significa que la distribución es platicúrtica.
- ▣ $G_2 \approx 0$. Cuando el resultado es cero, o muy próximo a cero, significa que la distribución es mesocúrtica.

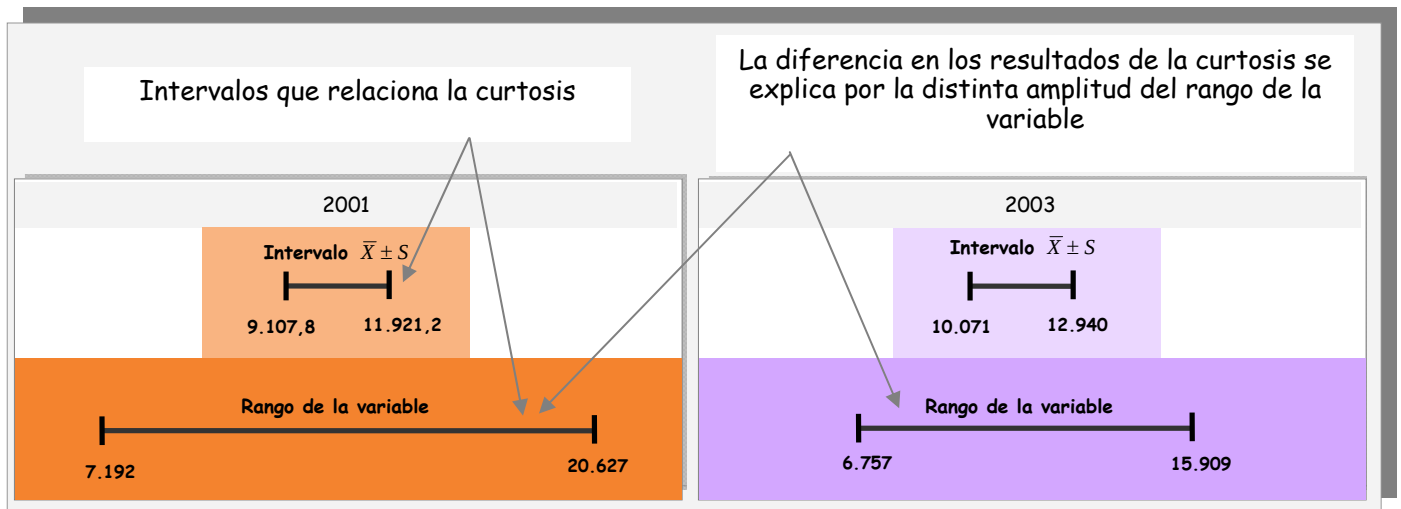
A continuación calcularemos los valores de curtosis de la variable Renta personal disponible de los municipios de Euskadi en los años 2001 y 2003.

	2001	2003
Suma de las diferencias elevadas a la 4ª potencia $\sum (x_i - \bar{x})^4$	13.315.132. 10^9	4.352.128. 10^9
Desviación estándar	1.403,91	1.461,65
Desviación estándar elevada a la 4ª potencia	3.884.685. 10^6	4.200.998. 10^6

2001
$G_2 = \frac{\sum (x_i - x)^4}{N \cdot S^4} - 3 = \frac{13.315.132 \cdot 10^9}{251 \cdot 3.884.685 \cdot 10^6} - 3 = \frac{13.315.132 \cdot 10^9}{975.056 \cdot 10^9} - 3 = 10,65$
2003
$G_2 = \frac{\sum (x_i - x)^4}{N \cdot S^4} - 3 = \frac{4.352.128 \cdot 10^9}{251 \cdot 4.200.998 \cdot 10^6} - 3 = \frac{4.352.128 \cdot 10^9}{1.054.450 \cdot 10^9} - 3 = 1,27$

Los resultados obtenidos ponen en evidencia las diferencias entre las dos distribuciones. En 2001 el rango de la variable es más amplio que en 2003, de ahí que el agrupamiento de los valores centrales de la distribución resulte también mayor en 2001 que en 2003 y que, consecuentemente, para el primero de los años se obtenga una curtosis más elevada. De todas formas, aunque el valor de la curtosis de las dos distribuciones es diferente,

las dos distribuciones pertenecen al tipo de distribuciones leptocúrticas ya que, en ambas, los valores centrales de la distribución se encuentran bastante concentrados.. Expresado en términos de renta personal disponible, el hecho de que la distribución sea leptocúrtica significa que en una mayoría de los municipios las rentas consideradas fueron bastantes similares, es decir, se situaron dentro de un rango pequeño de variación.

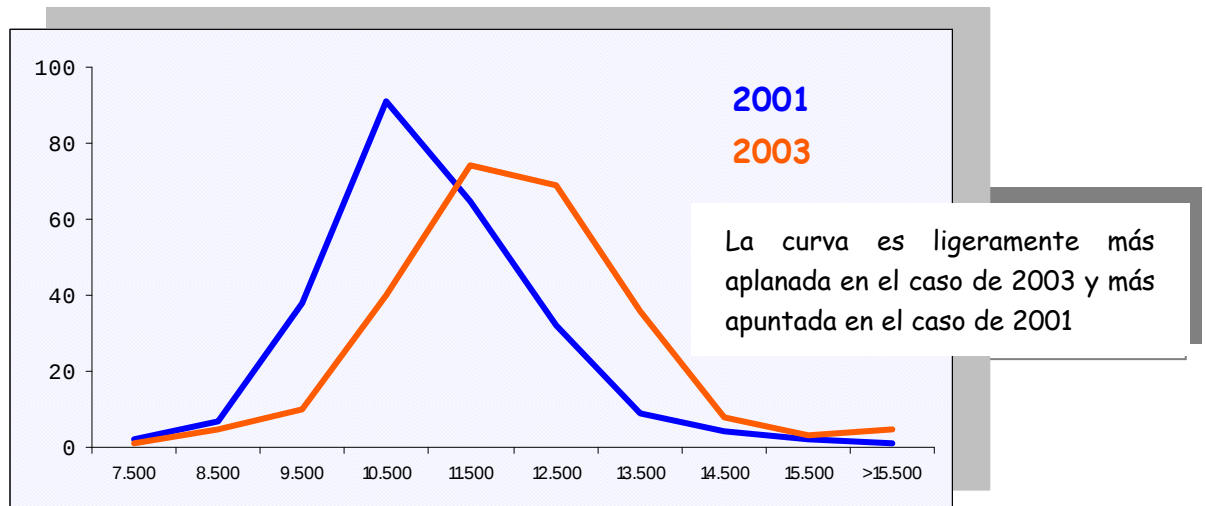


Puesto que la curtosis mide la relación entre los dos intervalos, aunque las dos distribuciones presentan un grado de concentración similar en torno a los valores centrales, la diferencia en el rango hace que el resultado de la curtosis sea distinto en las dos.

Por otra parte, ya hemos visto anteriormente que, en el caso de los datos de 2001, la amplitud del rango de la variable se debe a la presencia de un valor extremo (Laukiz con 20.627 €). Este valor eleva el rango de 15.346€ a 20.627 € y es responsable también del elevado resultado de la curtosis. Para comprobarlo, realizaremos de nuevo los cálculos de la curtosis pero eliminando esta vez el valor extremo:

$$G_2 = \frac{\sum (x_i - \bar{x})^4}{N \cdot S^4} - 3 = \frac{2.857.357.374 \cdot 10^6}{250 \cdot 2.459.110 \cdot 10^6} - 3 = \frac{2.857.357.374 \cdot 10^6}{614.776.103 \cdot 10^6} - 3 = 1,65$$

Tal como esperábamos, el resultado de la curtosis ha descendido notablemente, acercándose mucho al valor obtenido para el año 2003. De todos modos, aun eliminado el valor extremo, la curtosis de los datos de 2001 sigue siendo algo superior que la correspondiente a 2003. La propia gráfica nos permite comprobar que la concentración en torno a los valores centrales es algo mayor en los datos de 2001.



En la gráfica podemos ver la pequeña diferencia que existe en la curtosis entre la distribución correspondiente al año 2001 y la de 2003: la primera, marcada en azul (2001), muestra un apuntamiento ligeramente superior a la segunda (2003), marcada en rojo. (Atención! la gráfica no permite ver la diferencia en el rango de las dos distribuciones, ya que el eje de las abcisas sólo llega hasta los 15.500 €)