

**Un caso de aplicación de las medidas de variabilidad en el análisis de distribuciones campaniformes ..... 1**

▣ Variabilidad comparada de las precipitaciones en dos observatorios meteorológicos.....1

▣ Observatorios de diferente nivel de humedad .....5

▣ Precipitaciones muy variables .....6

## Un caso de aplicación de las medidas de variabilidad en el análisis de distribuciones campaniformes

### Variabilidad comparada de las precipitaciones en dos observatorios meteorológicos

Para finalizar la explicación sobre la descripción estadística de distribuciones campaniformes, aplicaremos las medidas explicadas a un estudio comparativo de la precipitación de dos estaciones meteorológicas españolas. Se trata de las estaciones meteorológicas de Toledo y San Fernando (Cádiz).

Los datos que analizaremos se corresponden con los totales anuales de precipitación recogidos en cada uno de los observatorios desde 1910 a 1995. Nuestro objetivo será analizar la variabilidad en el tiempo de las precipitaciones en ambos observatorios.

Trataremos de responder a preguntas como las siguientes:

- ▣ ¿Ha sido regular el fenómeno de la lluvia en los observatorios de Toledo y San Fernando, en el período analizado?
- ▣ ¿Cuál de los dos observatorios ha tenido el comportamiento más regular?
- ▣ ¿Se podría hablar de un rango de precipitaciones habitual durante el período? ¿Cuál sería en cada caso?
- ▣ ¿Han existido momentos especialmente secos o especialmente húmedos?
- ▣ ¿Ha sido similar el comportamiento de la lluvia en los dos observatorios? Y en caso contrario ¿en que se diferencian?

Nuestro primer paso siempre será calcular el rango de la variable, que nos proporcionará la primera información sobre la variabilidad total de la lluvia en el período analizado:

	Toledo		San Fernando	
	Año	Precipitación	Año	Precipitación
Máximo	1955	573,3	1955	999
Mínimo	1994	188,1	1994	282,8
Rango		<b>385,2</b>		<b>716,2</b>

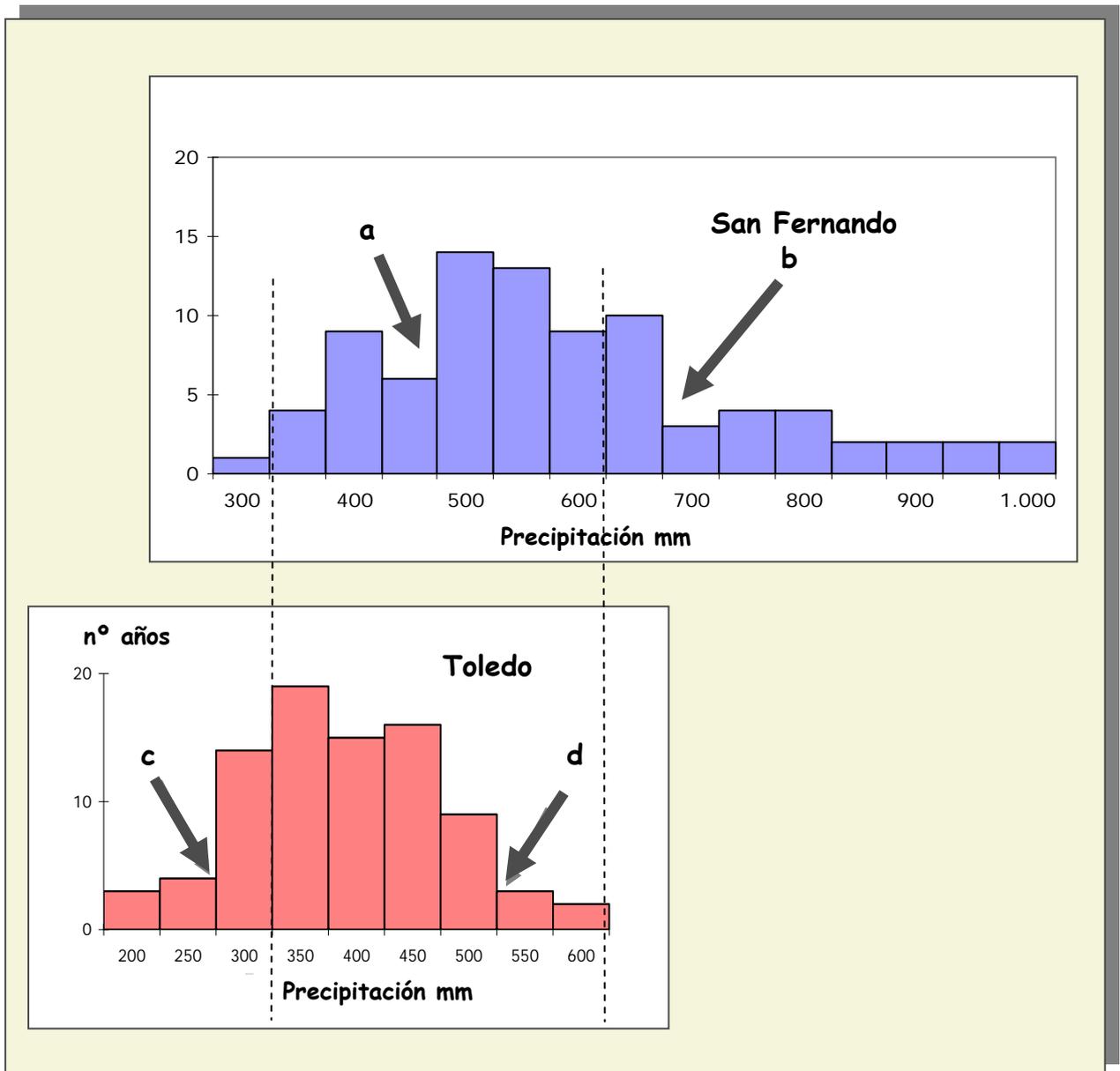
Vemos que el rango de una de las variables es muy superior al de la otra. La diferencia entre los valores del año más lluvioso y el más seco en la estación de San Fernando casi duplica al de la estación de Toledo. Vemos además que la diferencia no se produce tanto en el valor mínimo como en el máximo: en 1955, el año más lluvioso en ambas estaciones, en San Fernando se recogieron 425,7 mm más de lluvia que en Toledo.

A juzgar por los valores del rango, la variabilidad en la estación de San Fernando es muy superior a la de Toledo; esto es innegable. Sin embargo, lo que no podemos saber por el momento es si esta variabilidad tan elevada es una característica de todo período analizado o se debe a la existencia puntual u ocasional de años especialmente húmedos, es decir de valores especialmente elevados. Los histogramas nos ayudarán a responder esta pregunta.

De todos modos, antes de elaborar los histogramas merece la pena recordar dos cuestiones que ya hemos planteado en el apartado destinado a la elaboración de histogramas:

- ▣ Que no se utilizan sólo en el momento de la presentación de los resultados de un análisis. Los histogramas son herramientas muy útiles también para explorar los datos, para descubrir rasgos de la distribución de valores de la variable. Utilizados con fines distintos, los criterios de elaboración también suelen ser distintos.
- ▣ Que la impresión visual que pueden producir los histogramas, en relación al fenómeno que describen, puede ser variable en función de las decisiones que haya tomado el autor para su realización.

En nuestro caso, dado que **haremos los histogramas con fines exploratorios, empezaremos con un gran número de clases. Intentaremos descubrir si existen descensos bruscos en las frecuencias de los valores, y en caso afirmativo, en qué puntos del rango se producen.**



Los histogramas muestran efectivamente que, en ambos casos, se producen disminuciones bruscas de la frecuencia de los valores en puntos concretos del rango.

- ▣ En el histograma de San Fernando, por debajo del intervalo 450-500, se produce un descenso de frecuencias (marcado en el gráfico con la flecha a) seguido de un nuevo aumento. Por el lado derecho el descenso es claro.
- ▣ En el histograma de Toledo hay un cambio brusco del intervalo de 200-250 al de 250-300 (flecha c del gráfico). Por la derecha el descenso de frecuencias es un poco más gradual.

Podemos concluir de nuevo que, dadas las dudas que se generan con respecto al intervalo en el que se concentran la mayoría de los valores, es necesario disponer de un método estándar para su cálculo.

En cualquier caso, la variabilidad de la lluvia a lo largo del período analizado es elevada en los dos observatorios: el grupo mayoritario de valores no se concentra en el centro del rango sino que aparece bastante disperso en un amplio sector del mismo. Esto es todavía más evidente en la distribución de los valores del observatorio de Toledo: el intervalo en el que se concentran los valores es casi tan amplio como el rango.

El histograma de San Fernando responde también claramente a la pregunta que formulábamos en relación a la variabilidad: en San Fernando los valores de precipitación elevados se han producido sólo ocasionalmente; los valores por encima de 800 mm. se han producido en muy pocas ocasiones. Los cuatro últimos intervalos (precipitaciones entre 800 y 1.000 mm) contienen en total sólo 8 valores.

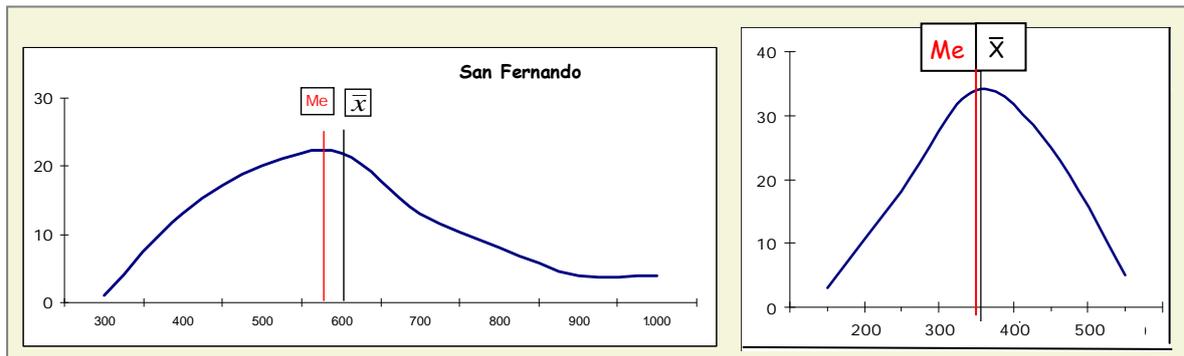
Los datos correspondientes a los años más húmedos forman la cola de la campana en la distribución de San Fernando. Como vemos, se trata de una cola que no tiene equivalente en el otro extremo de la distribución y que, por tanto, los valores de la estación de San Fernando conforman una distribución asimétrica.

Llegados a este punto podemos concluir que a través de la observación de los histogramas hemos obtenido mucha información sobre cómo se ha distribuido la lluvia en los observatorios de San Fernando y de Toledo a lo largo de los 85 años analizados. Nuestro siguiente paso será intentar realizar un análisis y descripción de la variabilidad de la variable precipitación mediante la utilización de las técnicas estadísticas estudiadas.

### Observatorios de diferente nivel de humedad

	Toledo		San Fernando	
Promedio	367,1		558,0	
Mediana	359,8		522,1	
Desviación estándar	83,2		160,0	
Promedio $\pm$ desviación estándar	283,9	450,3	397,93	717,98
Asimetría	0,12		0,81	
Coeficiente de variación	22,67		28,68	
Curtosis	-0,27		0,19	
Valor máximo	573,3		999	
Valor mínimo	188,1		282,8	
Rango	385,2		716,2	

Las precipitaciones que se recogieron en los observatorios de Toledo y San Fernando fueron, al menos en el período analizado, bastante diferentes. La zona en la que se ubica el observatorio de San Fernando es más lluviosa que la de Toledo. Según los resultados de la media, hay una diferencia de 189,6 milímetros entre ambos observatorios. Ahora bien, si nos fijamos en los datos de la mediana, la diferencia en el promedio anual de lluvia entre los dos observatorios es algo inferior: desciende hasta 162,3 milímetros. Esta diferencia entre la media y la mediana se produce debido a la existencia, en el caso de San Fernando, de algunos años cuyos valores de precipitación son sensiblemente superiores a los del resto del período. La explicación es sencilla: los valores de los años especialmente lluviosos en San Fernando *tiran* de la media hacia arriba (la *empujan* hacia la derecha de la distribución), es decir, incrementan su valor.



En el gráfico podemos ver que, mientras en el caso de Toledo media y mediana se encuentran muy próximas, en el centro de la distribución, en el caso de San Fernando, la media aparece a cierta distancia de la mediana, desplazada hacia la derecha de la distribución.

### Precipitaciones muy variables

La variación de un año a otro en los totales de precipitación que recogió cada uno de los observatorios es verdaderamente elevada. Si nos fijamos en la variación total de la precipitación en todo el período estudiado (rango) veremos que dicha variación, en ambos observatorios, es superior incluso al valor del promedio:

	Toledo	San Fernando
Promedio	367,1	558,0
Rango	385,2	716,2

La diferencia más notoria entre los dos observatorios se produce con respecto a los años más lluviosos. En el observatorio de San Fernando, durante algunos años se registraron precipitaciones próximas a los 1.000 milímetros, es decir, valores muy alejados de su media. En el de Toledo, sin embargo, no se produjeron en ningún año valores de precipitación tan alejados de su media.

Para describir la diferencia que acabamos de explicar entre los dos observatorios, la medida estadística más apropiada es el coeficiente de asimetría:

	Toledo	San Fernando
Coeficiente de asimetría	0,12	0,81

Efectivamente, los resultados obtenidos para el coeficiente de asimetría reflejan bien la diferencia mencionada:

- ▣ El valor obtenido para San Fernando, 0,81, nos habla de una distribución asimétrica. La asimetría en este caso es positiva, consecuencia de la presencia de un alargamiento hacia la derecha de la distribución o, lo que es lo mismo, de la existencia de algunos valores bastante superiores al promedio.
- ▣ El valor obtenido para el observatorio de Toledo, 0,12, nos habla de una ligera asimetría y, por tanto, de una distribución que se puede considerar bastante simétrica.

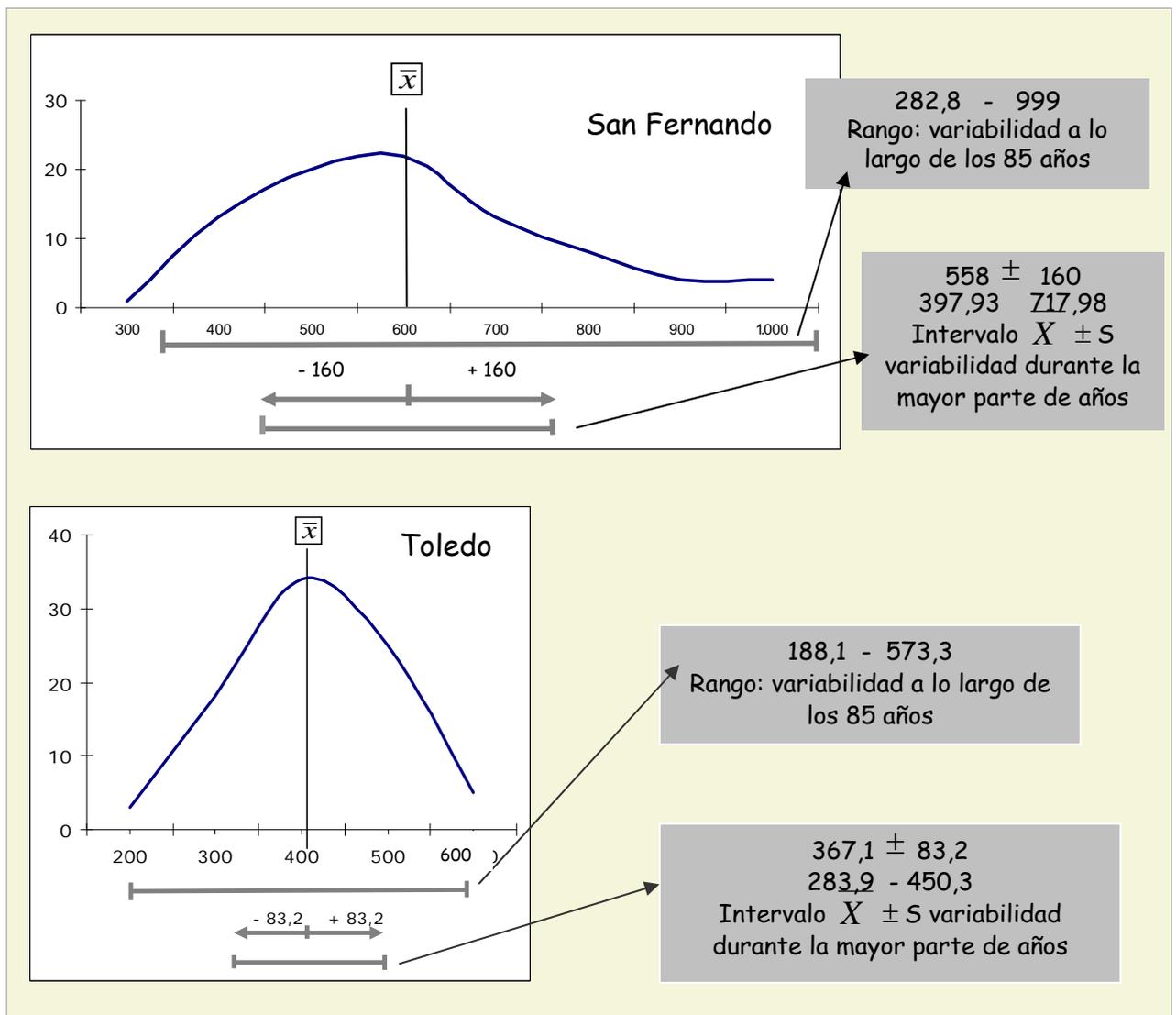
Lo dicho hasta ahora, sirve para caracterizar la variabilidad total del período estudiado. Lo que nos falta ahora es describir -dejar constancia- de la variabilidad habitual en las precipitaciones, es decir, de la variación real que se ha producido durante la mayoría de los años analizados. A este respecto, lo que hemos podido comprobar desde el momento en que hemos realizado los primeros histogramas, es que, en ambos observatorios, la variación que podemos considerar más habitual es también importante: dejando aparte los valores de los años especialmente húmedos o secos, durante el resto de años hay importantes diferencias en el total de precipitaciones.

Una de las herramientas más útiles para describir la variabilidad que afecta a la mayor parte de los elementos de la población con respecto a la variable de estudio, es el intervalo formado por la media +/- el valor de la desviación estándar. Veamos los resultados obtenidos en este caso:

	Toledo		San Fernando	
Promedio	367,1		558,0	
Desviación estándar	83,2		160,0	
Promedio ± desviación estándar	283,9	450,3	397,93	717,98

Como podemos comprobar, la amplitud del intervalo es destacable en ambos casos y, especialmente, en el del observatorio de San Fernando. Si tenemos la tentación de pensar que la variación habitual en Toledo -entre 283,9 y 450 mm- puede considerarse pequeña, deberemos fijarnos en el hecho de que esta

diferencia corresponde a un lugar en el que las precipitaciones son realmente escasas. Durante los años menos lluviosos del período que comprende la  $x \pm S$ , se recoge un 45% menos de lluvia que en los más lluviosos del mismo período;iii



Nos quedaría, por último, proporcionar una medida indicadora de la elevada variabilidad de la precipitación habitual, tanto en San Fernando como en Toledo. Recordemos que para ello disponemos de la curtosis, que mide la relación entre la variación total (el rango) y la variación más habitual (el intervalo  $X \pm S$ )

	Toledo	San Fernando
Curtosis	-0,27	0,19

Los resultados obtenidos para la curtosis de las dos distribuciones son realmente un buen reflejo de la elevada variabilidad habitual de las precipitaciones en San Fernando y Toledo. Ambos resultados son próximos a 0, lo que nos indica que se trata de distribuciones planas, características de variables cuyos valores se reparten ampliamente a lo largo del rango. Ya debemos tener claro que la ausencia de concentración de los valores en la parte central del rango se debe a la variabilidad de la característica que estemos analizando (en este caso, la lluvia)

En el caso de Toledo, y según el resultado de la curtosis, el valor negativo es claro indicador de una distribución platicúrtica. En el caso de San Fernando, aunque próximo a 0, el valor es positivo. Esto indicaría un mayor agrupamiento de los valores que en la distribución de Toledo. Y realmente, es cierto: aunque la variabilidad de San Fernando es, en general, más elevada que la de Toledo, el alargamiento de la distribución hacia la derecha implica que la concentración de los valores centrales con respecto al rango es superior.