

Estadística General

Tema 2: Estadística descriptiva

Dispersión o Variabilidad, Coeficiente de Variación



Prof. José G. Páez

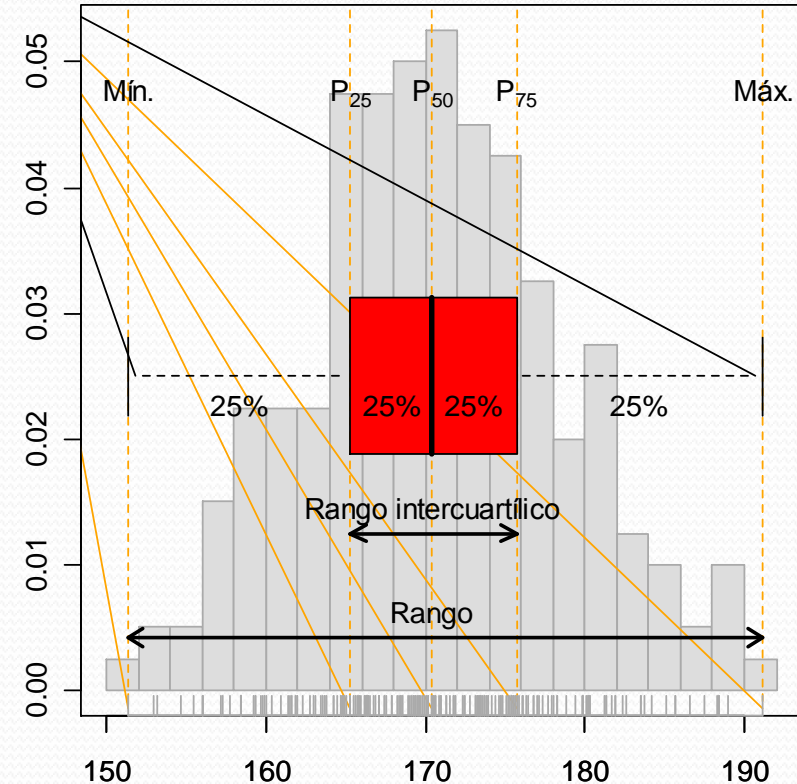
Variabilidad o dispersión

- Los estudiantes de Estadística reciben diferentes calificaciones en la asignatura (**variabilidad**). ¿A qué puede deberse?
 - **Diferencias individuales en el conocimiento** de la materia.
- ¿Podría haber otras razones (**fuentes de variabilidad**)?
- Por ejemplo supongamos que todos los alumnos poseen el mismo nivel de conocimiento. ¿Las notas serían las mismas en todos? Seguramente No.
 - Dormir poco el día del examen, el croissant estaba envenenado...
 - **Diferencias individuales en la habilidad** para hacer un examen.
 - El examen no es una medida perfecta del conocimiento.
 - **Variabilidad por error de medida.**
 - En alguna pregunta difícil, se duda entre varias opciones, y al azar se elige la mala
 - **Variabilidad por azar, aleatoriedad.**

Medidas de dispersión

Miden el grado de dispersión (variabilidad) de los datos, independientemente de su causa.

- **Amplitud o Rango** ('range'): Diferencia entre observaciones extremas.
 - 2,1,4,3,8,4. El rango es $8-1=7$
 - Es muy sensible a los valores extremos.
- **Rango intercuartílico** ('interquartile range'):
 - Es la distancia entre primer y tercer cuartil.
 - $\text{Rango intercuartílico} = P_{75} - P_{25}$
 - Parecida al rango, pero eliminando las observaciones más extremas inferiores y superiores.
 - No es tan sensible a valores extremos.



- **Varianza S^2** ('Variance'): Mide el promedio de las desviaciones (al cuadrado) de las observaciones con respecto a la media.

$$S^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2$$



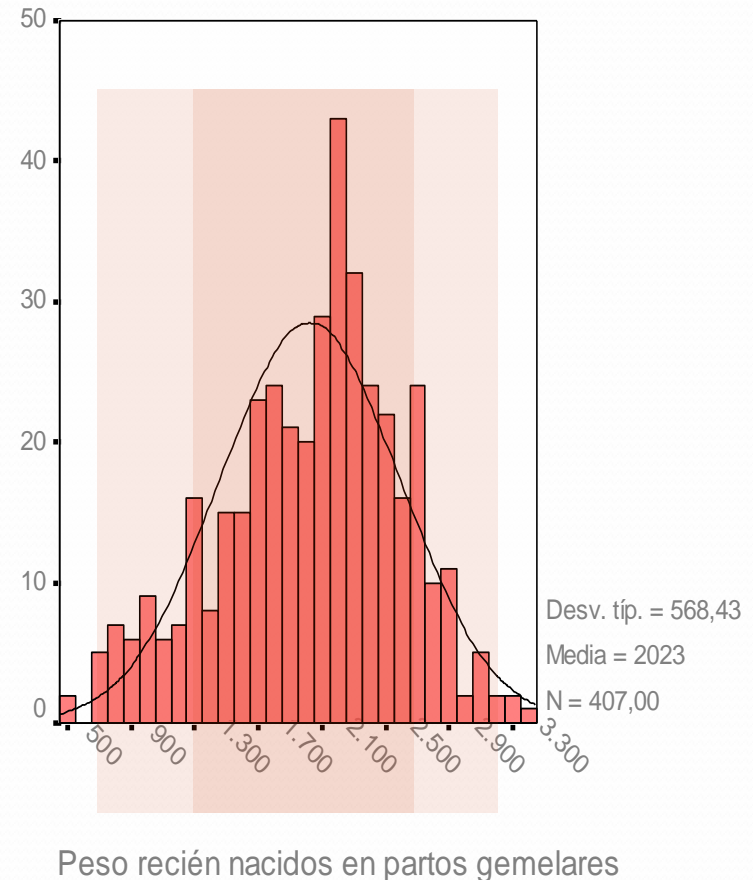
- Es sensible a valores extremos (alejados de la media).
- Sus unidades son el cuadrado de las de la variable. De interpretación difícil para un principiante.
- La expresión es fea, pero de gran belleza 'natural' (físicamente). Contiene la información geométrica relevante en muchas situaciones donde la energía interna de un sistema depende de la posición de sus partículas.
 - Energía de rotación (vía el coeficiente de inercia): patinadores con brazos extendidos (dispersos) o recogidos (poco dispersos)
 - Energía elástica: Muelles 'estirados' con respecto a su posición de equilibrio (dispersos) frente a muelles en posición cercana a su posición de equilibrio (poco dispersos)

Desviación típica ('standard deviation')

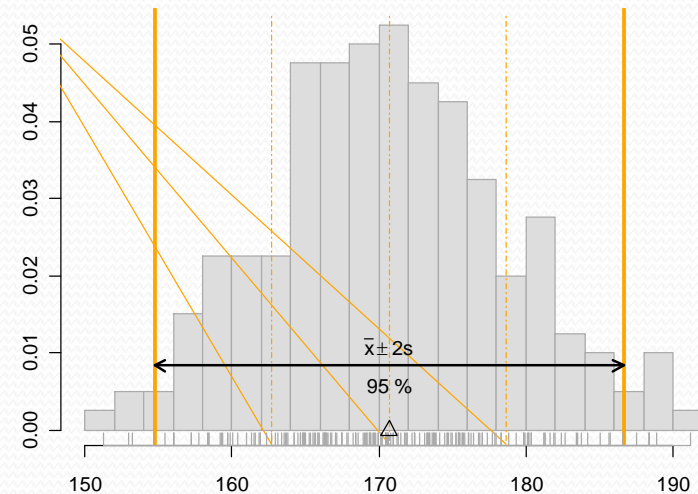
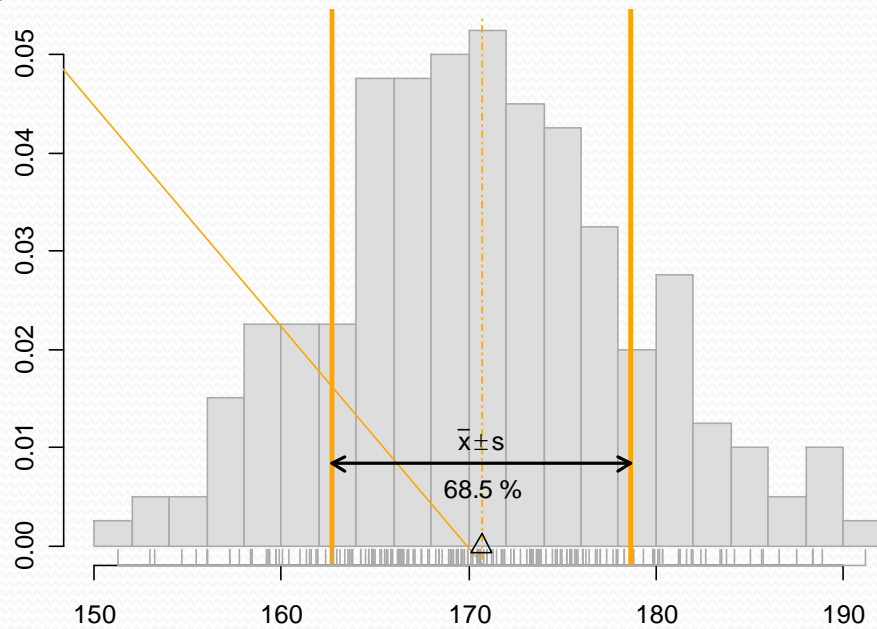
Es la raíz cuadrada de la varianza

$$S = \sqrt{S^2}$$

- Tiene la misma dimensionalidad (unidades) que la variable. Versión 'estética' de la varianza.
- Cierta distribución que veremos más adelante (**normal o gaussiana**) quedará completamente determinada por la media y la desviación típica.
 - A una distancia de una desviación típica de la media hay más de la 'mitad'.
 - A una distancia de dos desviaciones típicas de la media las tendremos casi todas.



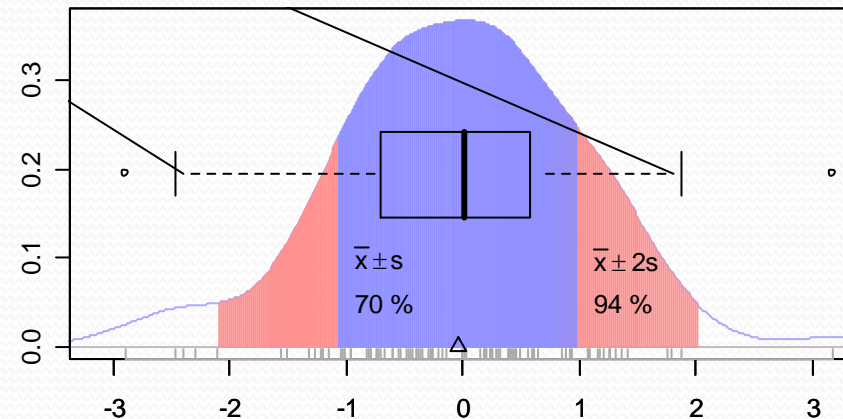
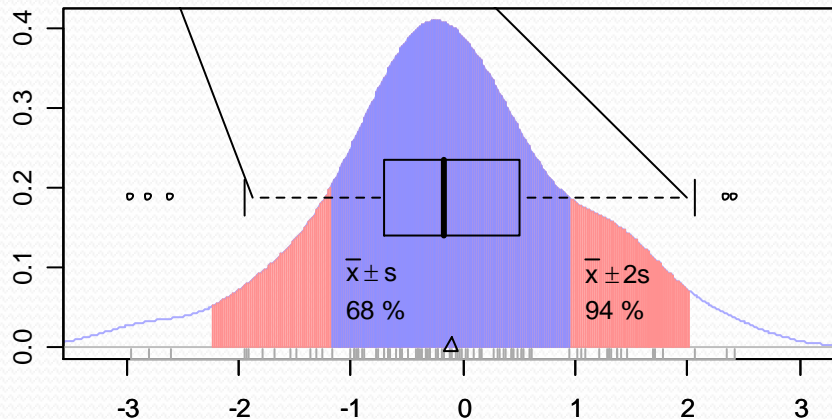
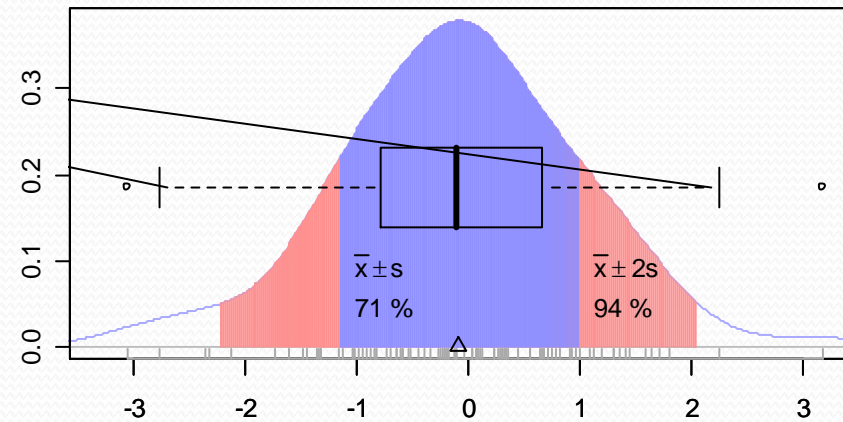
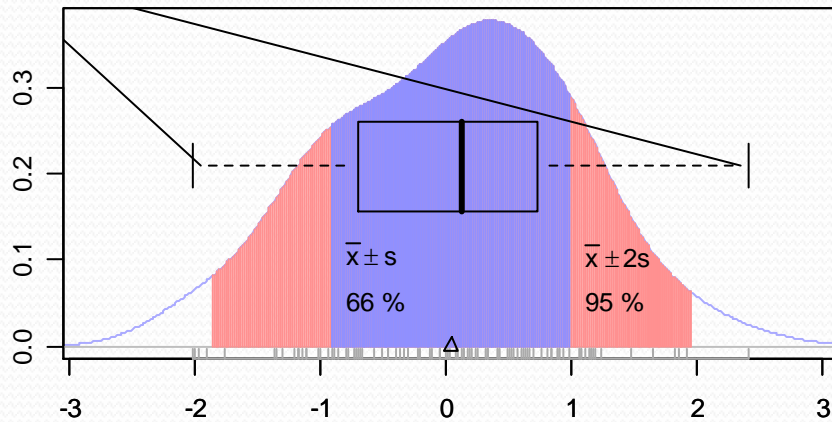
Dispersión en distribuciones 'normales'



- Centrado en la media y a una desv. típica de distancia hay aproximadamente el 68% de las observaciones.
- A dos desviaciones típicas tenemos el 95% (aprox.)

• Datos 'casi normales'. Eje 'x' medido en desviaciones típicas...

- ¿Encuentras relación entre rango intercuartílico y desviación típica?
- ¿Podrías caracterizar las observaciones anómalas?



Coeficiente de variación

$$CV = \frac{S}{\bar{x}}$$

Es la razón entre la desviación típica y la media.

- Mide la desviación típica en forma de “qué tamaño tiene con respecto a la media”
- También se la denomina **variabilidad relativa**.
- Es frecuente mostrarla en porcentajes
 - Si la media es 80 y la desviación típica 20 entonces $CV=20/80=0,25=25\%$ (variabilidad relativa)
- Es una cantidad **adimensional**. Interesante para comparar la variabilidad de diferentes variables.
 - Si el peso tiene $CV=30\%$ y la altura tiene $CV=10\%$, los individuos presentan más dispersión en peso que en altura.
- No debe usarse cuando la variable presenta valores negativos o donde el valor 0 sea una cantidad fijada arbitrariamente
 - Por ejemplo $0^{\circ}\text{C} \neq 0^{\circ}\text{F}$
- Los ingenieros electrónicos hablan de la razón ‘señal/ruido’ (su inverso).