



Análisis exploratorio de datos de calidad del aire



*"Statistics is
the grammar of science."
Karl Pearson*

¿Dónde se esconde la estadística?



1

Análisis exploratorio

(o el arte de hacer confesar a los datos)



Análisis exploratorio

Tiene como **objetivo** identificar:

- patrones subyacentes
- relaciones entre las variables
- el modelo teórico más adecuado para representar la población de la cual proceden los datos muestrales.



Qué veremos en este trabajo práctico?

Análisis exploratorio de los datos:

- Técnicas de descripción numérica
- Técnicas de descripción gráficas

Repaso

Variable

- **Cuantitativa**
 - Discreta
 - Continua
- **Cualitativa**
 - Nominal
 - Ordinal

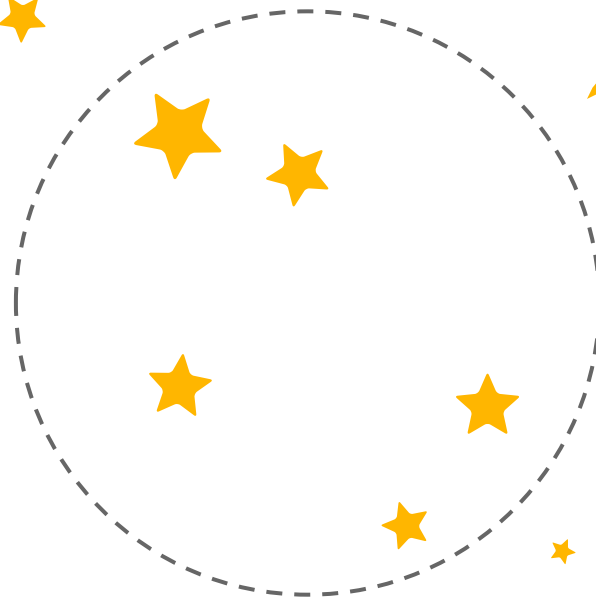


Repaso

Muestra

Subconjunto de una

Población



Repaso

Inferencia estadística

Proceso mediante el cual basándonos en una muestra sacamos conclusiones acerca de la población.

El objetivo final del análisis estadístico no es el estudio de la muestra, sino **el estudio de la población.**



Repaso



Distribución muestral



Funciones de distribución poblacional



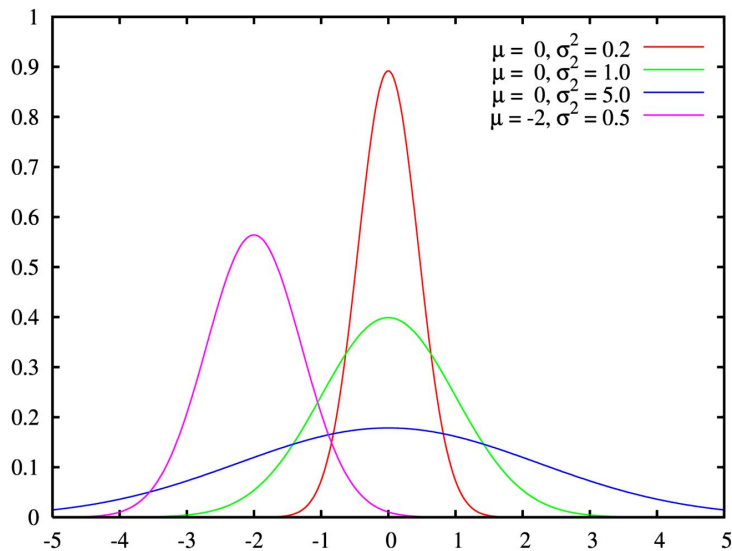
Variable continúa

- Normal
- Lognormal
- Gamma

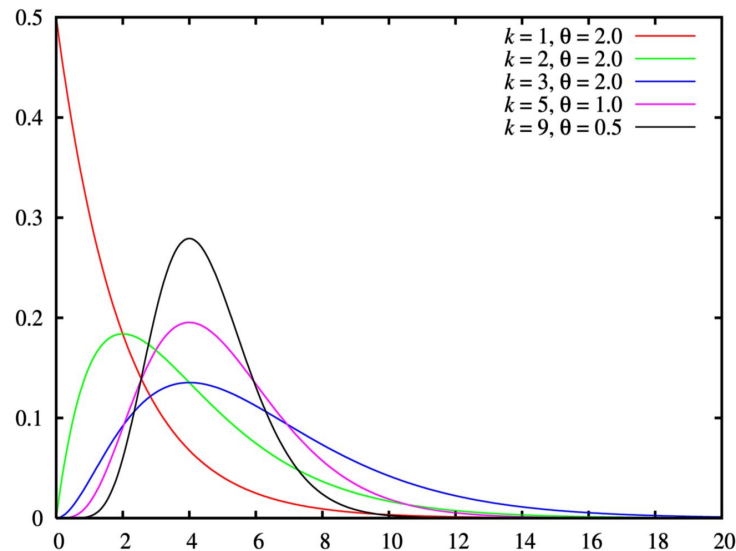
Variable discreta

- Poisson
- Binomial
- Negativa

Funciones de distribución en calidad del aire



Normal



Gamma



Parámetros estadísticos descriptivos

1. Posición

En torno a qué valor central toma valores la variable.

2. Dispersión

El grado de concentración de los valores que toma la variable alrededor de su posición central.

3. Forma

Qué forma toma la distribución de los valores.



1. Parámetros de posición

A. **De posición relativa:** Su construcción hace referencia al total de los datos muestrales.

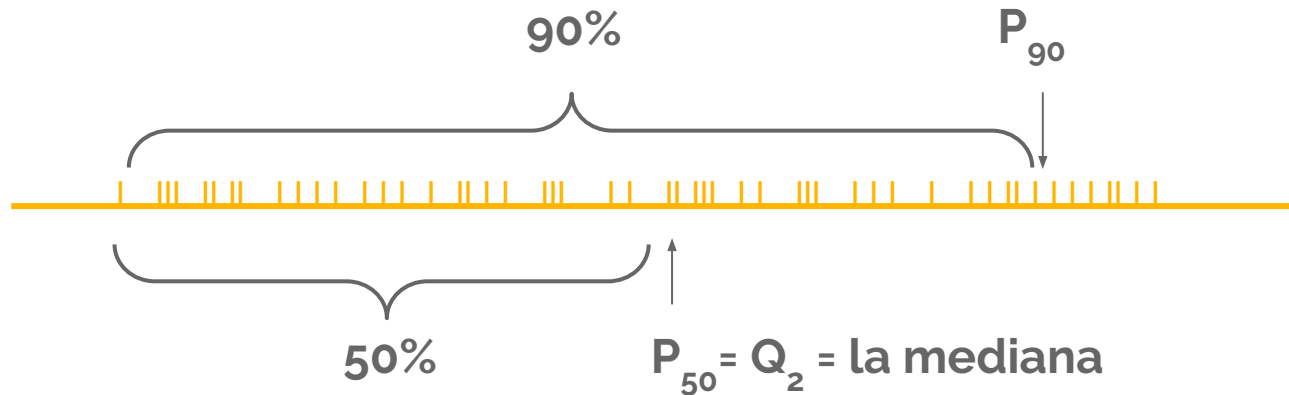
Percentiles: ordenados los datos de menor a mayor, el valor de la variable por debajo del cual se encuentra un porcentaje dado de observaciones en un grupo de observaciones. Cuartil.



1. Parámetros de **posición**

Percentiles: ordenados los datos de menor a mayor, el percentil p representa el valor de la variable por debajo del cual se encuentra el $p\%$ de las observaciones.

Cuartiles: Son los percentiles más importantes





1. Parámetros de posición

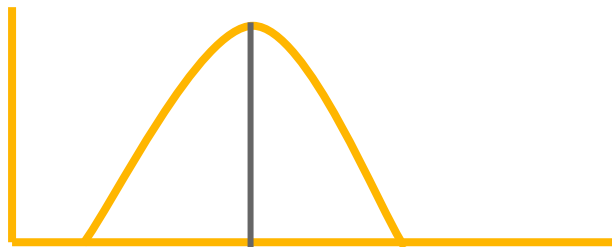
- B. **De tendencia central:** Nos dan una idea en torno a qué valor se encuentran concentrados nuestros datos. Ej: Media aritmética o promedio, mediana, moda.

Parámetros de **posición**

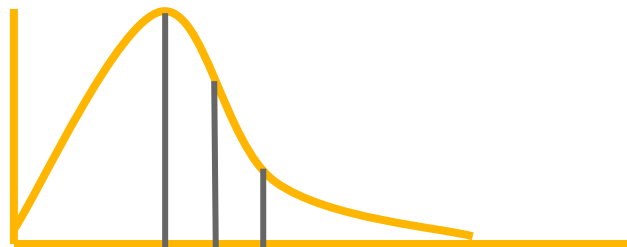
¿Cuál elegir?



¿Distribución simétrica o sesgada?



Media
Mediana
Moda

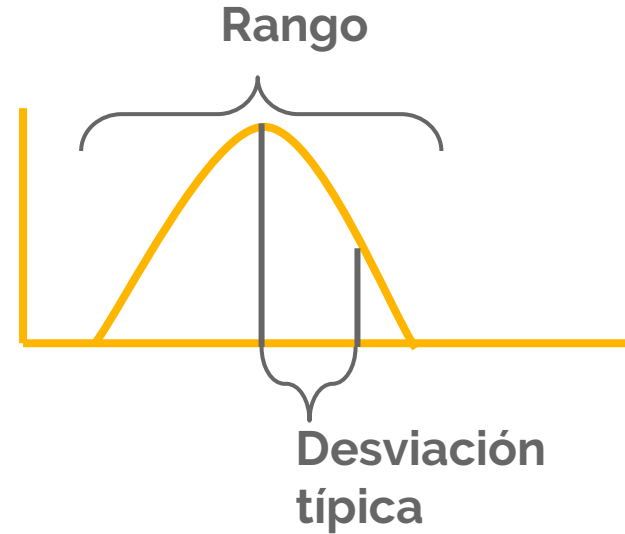


Media
Mediana
Moda



2. Parámetros de dispersión

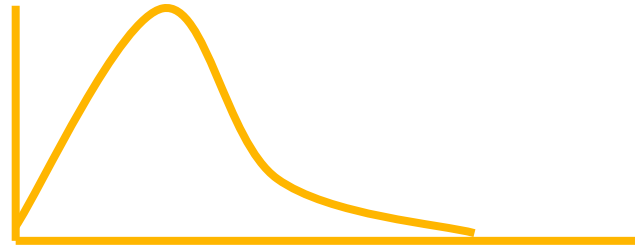
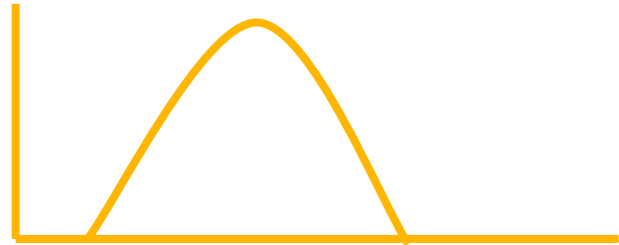
- Rango (Máximo - Mínimo)
- Varianza
- Desviación estándar o típica

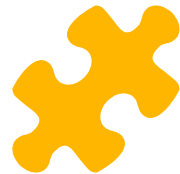




3. Parámetros de forma

- Coeficientes de asimetría
- Apuntamiento o curtosis





Histograma

- Rango de variación de los datos
- Simetría o Asimetría





Box-Plot

- Posición
- Dispersión
- Asimetría
- Valores extremos

