




# VALORES PERDIDOS



# ¿QUE SON?

- 
- Son observaciones perdidas.
  - Se producen por un fallo en la recogida del dato o en el proceso de almacenamiento.
  - Genera problemas de completitud en los datos.
  - Deben ser tratados.




# TRATAMIENTOS



- Existen dos opciones principales:
  - ❖ Eliminación de las variables.
  - ❖ Imputación de los valores perdidos:
    - ✓ Valores estadísticos (media, mediana o moda)
    - ✓ Imputación basada en modelos.





# ELIMINACIÓN DE OBSERVACIONES

- Eliminar todas las observaciones que tengan algún valor perdido.
  - Importante pérdida de datos.
  - No se analiza si los valores perdidos se producen por condiciones especiales.
  - Por un valor perdido se pierde toda la observación.
- 



# IMPUTACIÓN POR VALORES ESTADÍSTICOS

- Sustitución de los valores perdidos por la media o mediana en las variables cuantitativas y por la moda en las cualitativas.
  - Falta de representatividad de estos valores en la mayoría de los casos.
  - No se produce una pérdida de información.
  - Se da una mayor importancia a los valores centrales lo que puede condicionar las relaciones.
- 



# IMPUTACIÓN BASADA EN MODELOS

- Método de imputación múltiple:
  - Imputación de los valores perdidos basándose en la relación con el resto de las variables.
  - Efectiva si existe relación real.
  - Riesgo de multicolinealidad.
  - Desconocimiento de la metodología exacta implementada.
  - Solución sencilla.



# IMPUTACIÓN BASADA EN MODELOS

- Creación de modelos específicos para cada variable:
  - Imputación de los valores perdidos basándose en la relación con el resto de las variables.
  - Efectiva si existe relación real.
  - Riesgo de multicolinealidad.
  - Conocimiento exacto de la metodología aplicada.
  - Solución mas laboriosa.