## METODOS DE IMPUTACION:

## **1. Imputación por Media/Mediana/Moda**

Es un model que substitueix les dades per la mitjana, moda o mediana, és útil quan falten pocs valors i les dades **estan distribuïdes de manera simètrica.**

## **2. Imputación por Vecinos más Cercanos (KNN)**

**Descripció**: Utilitza la informació dels casos més semblants (veïns) per omplir els valors que falten.

**Quan utilitzar-lo**: Quan hi ha correlació entre les variables i es pot fer servir informació d'altres registres per imputar.

**Desavantatges**: Pot ser costós computacionalment amb grans volums de dades.

## **3. \*\*Imputación por Regresión\*\***

**Descripció**: Fa servir models de regressió per predir els valors que falten basant-se en altres variables.

**Quan utilitzar-lo**: Quan hi ha una relació clara entre les variables observades i les que tenen valors que falten.

**Desavantatges**: Pot introduir biaix si el model no captura correctament la relació entre variables.

## **4. \*\*Imputación por Múltiple Imputación por Cadena de Markov (MICE)\*\*:**

Funcionament:

**Imputacions inicials**: MICE comença amb una imputació inicial per a les cel·les buides. Aquesta imputació pot ser, per exemple, la mitjana, un valor aleatori o basat en models simples.

**Model regressió:** utilitza una variable o varies per predir les celes buides,

**Repetició:** es repeteix el process pero anar combinant diferentes variables.

**Resultat:** el resultat l’obte un cop les prediccions cada cop son mes semblants , és a dir no hi ha molta diferencia entre elles.

**Quan utilitzar-lo**: Quan es necessiten resultats robustos amb variabilitat reflectida de manera adequada.

**Desavantatges**: És més complex i requereix més recursos computacionals.

## **6. \*\*Imputación por Valor Aleatorio (Random Imputation)\*\*:**

Substitueix els valors que falten amb valors aleatoris generats a partir de la distribució de la variable.

**Quan utilitzar-lo**: Quan es vol mantenir la variabilitat de les dades.

**Desavantatges**: Pot afegir soroll i reduir la precisió del model si es fa de manera excessiva.

## **7. \*\*Imputación por Aproximaciones de Matriz (Matrix Completion)\*\*:**

**Descripció**: Utilitza tècniques de factorizació de matrius per completar matrius amb valors faltants, trobant relacions entre les files i columnes.

**Quan utilitzar-lo**: Quan es treballa amb dades matricials (com recomanadors o dades de sensors).

**Desavantatges**: Pot ser complex de configurar i necessita recursos computacionals elevats.

## **8. \*\*Imputación por Interpolación\*\***

**Descripció**: Utilitza valors coneguts abans i després del valor que falta per calcular una estimació.

**Quan utilitzar-lo**: Quan es treballa amb dades seqüencials o temporals.

**Avantatges**: Pot ser molt precís per a dades que tenen un patró temporal clar.

**Desavantatges**: No és aplicable a dades que no tenen una seqüència temporal.

## 8.1 TIPUS

### **7. Imputació per Interpolació**

* **Quan utilitzar-lo:**
  + Per dades seqüencials o temporals amb patrons clars.
* **Quan evitar-lo:**
  + Si les dades no tenen un ordre seqüencial.
* **Exemple:** Omplir la temperatura en un dia basant-se en els dies abans i després.

**Tipus d'interpolació:**

* **Lineal:** Per patrons simples i lineals.
* **Polinòmica:** Per patrons no lineals més complexos.
* **Spline:** Per dades amb canvis graduals.