
TEORÍA SESIÓN 5: *CALIDAD DE DATOS*

PEDRO M. VALERO MORA

ESQUEMA

| | | |
|----|--|---|
| 1. | Introducción | 2 |
| 2. | Objetivos del control de datos | 3 |
| 3. | Métodos de control de calidad de datos | 4 |

1. Introducción

Aspectos generales de la calidad de datos

- Los datos pueden perder calidad en dos momentos a lo largo del proceso de convertirlos en ficheros para su análisis:
 - Al recogerlos
 - Al introducirlos
- Para tener datos de calidad hay que cuidar esos aspectos.
- No obstante, una vez los datos están introducidos...¿podemos controlar que los datos tienen suficiente calidad?

2. Objetivos del control de datos

¿Qué podemos mirar una vez tenemos los datos?

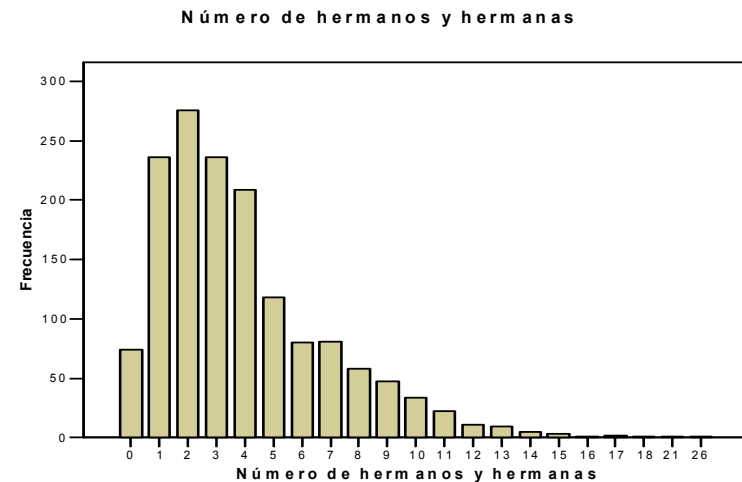
- Podemos controlar varias cosas en los datos:
 - Valores perdidos
 - Casos repetidos
 - Valores imposibles
 - *Sexo varón, dos embarazos*
 - Valores implausibles
 - *Altura 1.60 ms, peso 100 kilos*
- Los valores imposibles deben ser corregidos, los valores implausibles pueden ser valores legítimos y necesitan ser investigados con detalle

3. Métodos de control de calidad de datos

¿Cómo podemos mirar la calidad?

- Comprobaciones generales en la hoja de cálculo
 - Lineas repetidas (SPSS 12 tiene un comando para esto)
 - Casillas en blanco
 - Valores repetidos
- Utilizando técnicas estadísticas. Estas técnicas son diferentes para variables categóricas o numéricas

- Categóricas
 - Frecuencias para una variable
 - Cruces de Frecuencias para más de una variable



- Numéricas
 - Máximos y mínimos para variables numéricas. También histogramas

– Diagramas de dispersión para dos variables

