

VARIABLES CUALITATIVAS

(Tomado de:

http://www.liccom.edu.uy/bedelia/cursos/metodos/material/estadistica/var_cuali.html)

Datos nominales y ordinales:

Las variables pueden ser cualitativas o cuantitativas.

Variables cualitativas: aquellas que no aparecen en forma numérica, sino como categorías o atributos (sexo, profesión, color de ojos) y sólo pueden ser nominales u ordinales.

Variables nominales: lo único que puede hacerse es establecer frecuencias en cada atributo y la igualdad o desigualdad entre los diferentes casos, ver cuál es el grupo que tiene mayor frecuencia alcanzando el concepto de "moda" (y también obtener algunas medidas de asociación cuando se relacionan variables entre sí).

Variables ordinales: recogen la idea de orden pero no tiene sentido realizar operaciones aritméticas con ellas (acuerdo o desacuerdo con un proyecto de ley) ya que no puede medirse distancia entre una categoría y otra. Se puede establecer aquí igualdad y desigualdad, y relaciones como mayor que, y menor que. Puede establecerse orden, pero no medirse distancia dentro de ese orden. La medida estadística de tendencia central más apropiada para estas escalas es la "mediana".

Tabulación: frecuencia absoluta y relativa:

Una de las finalidades de la estadística descriptiva es resumir gran cantidad de información en pocos valores. La distribución de frecuencias consiste en un agrupamiento de datos en categorías (o clases) que muestren el número de observaciones registradas en cada categoría.

La frecuencia absoluta es el número de observaciones que la variable toma en cada

clase. Dicho de otra manera, la frecuencia absoluta de una clase c_i (con $1 \leq i \leq k$,

siendo k el número n_i de categorías de la variable) es el número de observaciones que presentan una modalidad perteneciente a esa variable.

Categorías de la variable	Frecuencia Absoluta
c_1	n_1
...	...
c_j	n_j

...	...
c_k	n_k
	N

Claramente puede verse que las frecuencias son números no negativos. Además, la suma de las frecuencias absolutas es el número total N de observaciones.

$$\sum_{i=1}^k n_i = N \quad \text{y} \quad 0 \leq n_i \leq N$$

La frecuencia relativa es la proporción del valor de la frecuencia absoluta de cada clase en relación al total de las observaciones de la variable. Dicho de otra manera, la

frecuencia relativa de la clase c_i (con $1 \leq i \leq k$, siendo k el número de categorías de la variable) es el cociente, entre las frecuencias absolutas de dicha clase y el número total de observaciones

$$f_i = \frac{n_i}{N}$$

Categorías de la variable	Frecuencia Absoluta	Frecuencia Relativa
c_1	n_1	$f_1 = \frac{n_1}{N}$
...
c_j	n_j	$f_j = \frac{n_j}{N}$
...
c_k	n_k	$f_k = \frac{n_k}{N}$
	N	1

$$\sum_{i=1}^k f_i = 1 \quad \text{y} \quad 0 \leq f_i \leq 1$$

Si se multiplica la frecuencia relativa por cien, obtenemos el porcentaje de cada categoría en relación al total de casos.

$$f_i \times 100 = x\%$$

Categorías de la variable	Frecuencia Absoluta	Frecuencia Relativa	Porcentajes
c_1	n_1	$f_1 = \frac{n_1}{N}$	$f_1 \times 100$
...
c_j	n_j	$f_j = \frac{n_j}{N}$	$f_j \times 100$
...
c_k	n_k	$f_k = \frac{n_k}{N}$	$f_k \times 100$
	N	1	1×100

La tabla que presenta las clases o categorías de la variable y sus respectivas frecuencias se llama distribución de frecuencias (indica cómo la frecuencia total se distribuye entre las clases).