

Apuntes Stats

Oskar Denis Siodmok

19 de octubre de 2020

1 Introducción

1.1. Conceptos Básicos

- Individuos o elementos: Contienen la información a estudiar.
- Población: Conjunto de individuos o elementos que presentan la variable a estudiar.
- Muestra: Subconjunto representativo de una población.
- Variables: Propiedades de los elementos de la población que tendrán sus respectivos valores
- Clases: Conjunto de valores que cumplen una propiedad. Por definición un valor solo puede pertenecer a una clase (por ejemplo, intervalos en variables continuas).
- Parámetro: Será una función que operará con las diferentes variables y valores sobre la población con un propósito.
- Estadístico: Será una aproximación al parámetro a partir de una muestra.

1.1.1. Variables Estadísticas

Las variables estadísticas, ya definidas en el anterior apartado, se denotarán mediante una letra mayúscula (por lo general con X o Y). Podrán tomar cualquier valor de cualquier conjunto. El dominio de la variable será el conjunto de todos los posibles valores de dicha variable.

1.1.2. Tipos de variables

- Variables Cuantitativas: Se expresan en cantidades numéricas o cualquier otro sistema que se pueda ordenar. A su vez se clasifican en:
 - Variables Discretas: Toman valores concretos de conjuntos finitos o infinitos.
 - Variables Continuas: Toman valores de conjuntos infinitos y no concretos (el dominio son valores continuos, como todos los reales en un intervalo).

Para muchas variables resulta complicado distinguir el tipo. Por ejemplo, aunque la altura matemáticamente sea continua, en la vida real nadie va a determinar una altura por encima de 2 decimales.

- Variables Cualitativas: No se pueden medir, solo clasificar. Un tipo concreto de este tipo de variables son las *Variables Ordinales*, las cuales pese a no tener un valor numérico si que pueden tener relaciones de orden.

1.1.3. Representación de datos

Esta se puede realizar de varias formas:

- Tablas y Gráficos: representan información de forma rápida y visual.
- Medidas Descriptivas: describen la información de forma numérica.

1.2. Tablas de Frecuencias

Se pueden realizar sobre cualquier conjunto de datos y sobre cualquier variable. Por ejemplo, dada una población de n individuos que presenta la variable X se obtienen las clases $c = \{c_1, c_2, \dots, c_k\}$ posibles. En este caso, n_i hará referencia al número de observaciones para $i \in \{1, \dots, k\} \subseteq \mathbb{N}$. De esta forma, $n = \sum_i n_i$ será el número total de observaciones de nuestra variable, independientemente de la clase de cada observación. Por otro lado, la frecuencia relativa de una variable representará la frecuencia de una variable (n_i) sobre 1 y se calculará como $\frac{n_i}{n}$.

Clase c_i	c_1	c_2	\dots	c_k	
Freq. Absoluta n_i	n_1	n_2	\dots	n_k	$n = \sum_i n_i$
Freq. Relativa f_i	f_1	f_2	\dots	f_k	$f = \sum_i f_i$

Se podrían añadir las frecuencias acumuladas, las cuales se van sumando a los datos anteriores. Siendo N_i la frecuencia absoluta acumulada y F_i la frecuencia relativa acumulada:

Clase c_i	c_1	c_2	\dots	c_k	
Freq. Absoluta n_i	n_1	n_2	\dots	n_k	$n = \sum_i n_i$
Freq. Relativa f_i	f_1	f_2	\dots	f_k	$f = 1$
F. abs. acum. $N_i = \sum_{j=1}^i n_j$	N_1	N_2	\dots	N_k	$N = N_k = n$
F. rel. acum $F_i = \sum_{j=1}^i f_j$	F_1	F_2	\dots	F_k	$F = F_k = 1$

Si la variable es cualitativa entonces las clases serán nominales. Si la variable es discretas las clases serán valores numéricos dentro del rango y si es continua serán intervalos $(l_{i-1}, l_i] \forall i$. Si las clases son intervalos habrá varios parámetros y conceptos de interés:

- Amplitud del intervalo: $a_i = l_i - l_{i-1}$.
- Marca de clase: Será el valor representativo del intervalo. Por ejemplo, el punto medio: $c_i = \frac{l_i + l_{i-1}}{2}$.
- Número de intervalos: Se realizará mediante aproximación pues se pueden plantear los intervalos que se quiera. Una elección típica es:

$$k = \begin{cases} \sqrt{n} & , n \text{ no es muy grande} \\ 1 + \log_2(n) & \end{cases}$$