

Introducción a los Conceptos Básicos de la Estadística

Leonardo Añez Vladimirovna
e-mail: toborochi98@outlook.com

26 de marzo de 2018

1. Reseña Histórica

La Estadística se remontan a Egipto, con los primeros registros de población y riqueza, atribuido a Ramses II en 1400 a.C., con el objetivo de hacer una nuevo reparto de tierras. Por otro lado, en Israel también se llevaron a cabo recuentos de la población hebrea. Los chinos y los griegos también han realizado censos desde hace cuarenta siglos, estos con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles).

Posteriormente, desde el siglo III a.C., en las civilización romana se llevan a cabo censos e inventarios de posesiones, que pueden considerarse precedentes institucionalizados de la recogida de datos demográficos y económicos de los Estados Modernos. Pero durante los mil años siguientes a la caída del imperio Romano se realizaron muy pocas operaciones Estadísticas.

2. Definición de Estadística

2.1. Concepto

Es la ciencia que proporciona un conjunto de métodos y técnicas que permiten recolectar, clasificar, organizar, analizar y presentar datos de forma adecuada para tomar decisiones frente a una incertidumbre o describir o afirmar algo acerca de la población.

Definición Etimológica

La palabra estadística es un término popularizado por Gottfried Aschenwall (economista alemán, 1719-1772). Este creó la palabra *Statistik* inspirándose en el latín *statisticus*, formada de *status* (estado) y el sufijo *-icus* (relativo a).

3. Clasificación de Estadística

3.1. Estadística Descriptiva

Es la parte de la estadística que proporciona el conjunto de métodos, con el fin de describir o analizar datos.

3.2. Estadística Inferencial

Es la cual proporciona la teoría necesaria para inferir (deducir) o estimar la generalización o toma de decisiones sobre la base de una información parcial.

4. Algunos Conceptos Clave

- (a) **Población:** Es un conjunto de sujetos o elementos que presentan características comunes. Sobre esta población se realiza el estudio estadístico.
- (b) **Tamaño de la Población:** Es el número de elementos que constituyen la población. Denotado por: N .
- (c) **Muestra:** Es un subconjunto de la población el cual debe ser representativo de la población.
- (d) **Tamaño de la Muestra:** Es la cantidad de elementos de la muestra. Denotado por: n .
- (e) **Parámetro:** Es una medida numérica que describe una característica de la población.
- (f) **Estadígrafo:** Es una medida numérica que describe una característica de la muestra.
- (g) **Dato:** Los datos estadísticos son los valores que se obtienen al llevar a cabo un estudio estadístico. Se trata del producto de observación de aquel fenómeno que se quiere analizar.
- (h) **Información:** Esta constituido por un conjunto de datos organizados.
- (i) **Observaciones:** Es el registro en forma sistemática de patrones, objetos y sucesos a fin de obtener información sobre el fenómeno de interés.
- (j) **Censo:** Es el recuento de individuos que conforman la población estadística. Básicamente consiste en obtener mediciones del número total de individuos mediante diferentes técnicas de recuento.
- (k) **Encuesta:** Es el procedimiento dentro de los diseños de una investigación descriptiva en el que se recopilan datos mediante un cuestionario previamente diseñado, sin modificar el entorno ni el fenómeno donde se recoge la información.

5. Variables Estadísticas

Se denomina variable estadística a la característica o combinación de características de una población.

5.1. Variables Cualitativas

Son aquellas que están asociadas con una característica cualitativa, es decir, son variables cuyos valores son cualidades, propiedades o atributos que presenta la población. Estas se clasifican de la siguiente manera:

5.1.1. Variables Nominales

Son aquellas cuyo valor se expresa en categorías sin ningún tipo de orden o clasificación.

Ejemplos: sexo, estado civil, nacionalidad, etc...

5.1.2. Variables Ordinales

Son aquellas cuyo valor se expresa en categorías pero su busca una clasificación de orden.

Ejemplos: Grado de Instrucción, Tamaño, grado de agresividad, etc...

5.2. Variables Cuantitativas

Son aquellas que están asociadas a una característica cuantitativa, es decir que se obtiene del resultado de mediciones o conteos. Estas se clasifican de la siguiente manera:

5.2.1. Variables Discretas

Son variables discretas cuando es resultado de la operación de contar.

Ejemplos: número de hijos, número de alumnos, dedos de las manos, etc..

5.2.2. Variables Continuas

Son aquellas que se obtienen por medición.

Ejemplos: edad, estatura, peso, PIB, etc...

6. Distribución de Frecuencias

Las Distribuciones de Frecuencia son tablas en las que los valores de las variables son colocadas en las filas, mientras que el número de frecuencias está en las columnas. Y con estas tablas tenemos los siguientes conceptos:

- (a) **Clase:** Conjunto de observaciones cualitativas o cuantitativamente iguales o muy próximas.
- (b) **Límite de Clase:** Son los valores de las observaciones, mínimo y máximo que pueden tener los datos incluidos en la clase.

$$y'_{i-1} , y'_i$$

- (c) **Intervalo de Clase:** Es la distancia entre los límites de clase. Denotados por:

$$y'_{i-1} - y'_i$$

- (d) **Marca de Clase:** Es el punto medio de cada uno de los intervalos de clase, este puede ser constante o variable, se expresa de la siguiente manera:

$$C_j = \frac{y'_{i-1} + y'_i}{2}$$

- (e) **Recorrido de la Variable:** Es el valor mínimo de la diferencia:

- Para datos originales(no agrupados): $R = y_k - y_i$
- Para datos agrupados: $R = y'_k - y'_i$

- (f) **Amplitud del Intervalo:** es la diferencia entre el límite superior e inferior del intervalo de clase.

$$C_j = y'_j - y'_{j-1}$$

- (g) **Número de Intervalos:** tenemos las siguientes maneras de determinarlos:

$$k = \frac{R}{C}, n \leq 25$$

$$k = \sqrt{n}, n > 25 \text{ (Regla del Ingeniero)}$$

Ademas de una forma general llamada *Regla de Sturges*:

$$k = 1 + 3,3 \log_{10}(n)$$

7. Frecuencia

- (a) **Frecuencia Absoluta:** Es el número de veces en que un evento se repite durante una muestra estadística. Representada por:

$$f_i$$

- (b) **Frecuencia Absoluta Acumulada:** la suma de las frecuencias absolutas de todos los valores iguales o inferiores al índice considerado.

$$F_i = \sum_{j=1}^i f_j$$

- (c) **Frecuencia Relativa:** el resultado de dividir la Frecuencia Absoluta de un determinado valor entre el número total de datos.

$$h_j = \frac{f_j}{n}$$

- (d) **Frecuencia Relativa Acumulada:** el resultado de dividir la frecuencia acumulada entre el número total de datos.

$$H_i = \frac{F_i}{n} = \sum_{j=1}^i h_j$$

7.1. Propiedades de las Frecuencias

$$(I) f_1 + f_2 + f_3 + \dots + f_k = \sum_{i=1}^k f_i = n$$

$$(II) h_1 + h_2 + h_3 + \dots + h_k = \sum_{i=1}^k h_i = 1$$

$$(III) \forall i \in I, \begin{cases} 0 \leq h_i \leq 1 \\ 0 \leq f_i \leq n \end{cases}$$

$$(IV) f_1 = F_1 \leq F_2 \leq F_3 \leq \dots \leq F_k = n$$

$$(V) h_1 = H_1 \leq H_2 \leq H_3 \leq \dots \leq H_k = 1$$

$$(VI) H_j = H_{j-1} + h_j$$

$$(VII) F_j = F_{j-1} + f_j$$

8. Tablas de distribución de frecuencias

8.1. Distribución de frecuencia para datos agrupados no en intervalos de clase

x_i	f_i	F_i	h_i	H_i
x_1	f_1	F_1	h_1	H_1
x_2	f_2	F_2	h_2	H_2
x_3	f_3	F_3	h_3	H_3
x_4	f_4	F_4	h_4	H_4
\vdots	\vdots	\vdots	\vdots	\vdots
x_k	f_k	$F_k = n$	h_k	$H_k = 1$
Totales	n		1	

8.2. Distribución de Frecuencias para datos agrupados en intervalos de clase

$y'_{i-1} - y_i$	y_i	f_i	F_i	h_i	H_i
$y'_0 - y_1$	y_1	f_1	F_1	h_1	H_1
$y'_1 - y_2$	y_2	f_2	F_2	h_2	H_2
$y'_2 - y_3$	y_3	f_3	F_3	h_3	H_3
$y'_3 - y_4$	y_4	f_4	F_4	h_4	H_4
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$y'_{k-1} - y_k$	y_k	f_k	$F_k = n$	h_k	$H_k = 1$
Totales		n		1	

8.3. Estructura de una Tabla Estadística

1. Título

2. Cuadro (Propiamente dicho):

- Encabezamiento
- Cuerpo
- Columna Matriz

3. Indicadores Complementarios

- Fuente
- Nota
- Comentario(s)

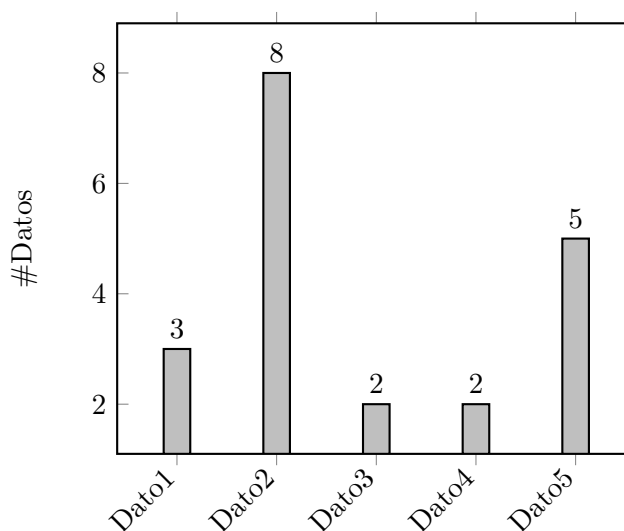
9. Representaciones Gráficas

Las gráficas son instrumentos estadísticos muy útiles para dar una idea global acerca de una situación estadística.

9.1. Caso de Variables Cualitativas

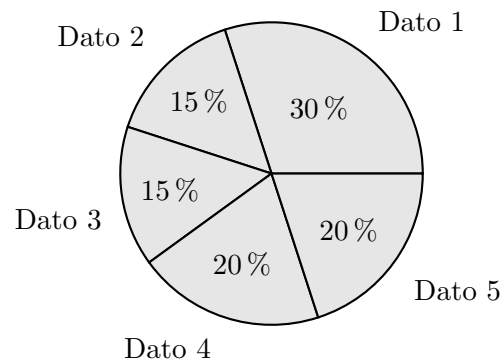
9.1.1. Diagrama de Barras

Un Diagrama de Barras es un gráfico que presenta datos categóricos con barras rectangulares, con alturas o longitudes proporcionales a los valores que representan. Se pueden representar de manera horizontal o vertical.



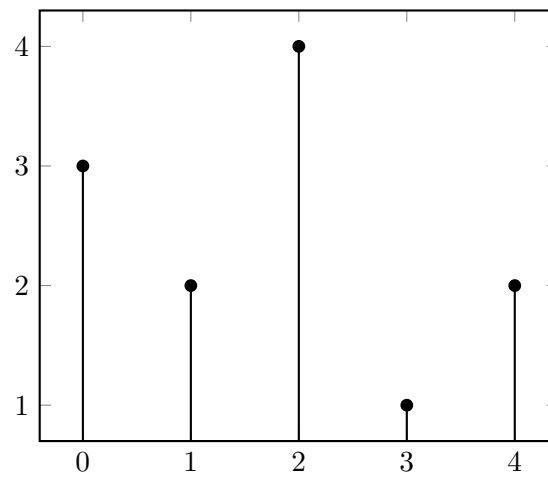
9.1.2. Diagrama de Sectores

También llamado *Diagrama Circular o de Pastel*, se utiliza para representar distribuciones porcentuales:



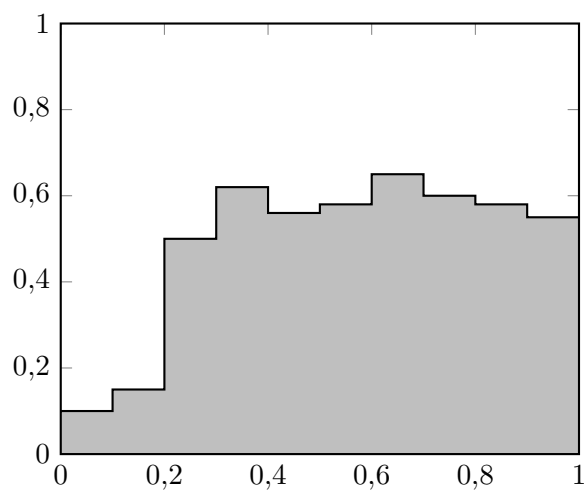
9.2. Caso de Variables Cuantitativas

9.2.1. Diagrama de Bastones



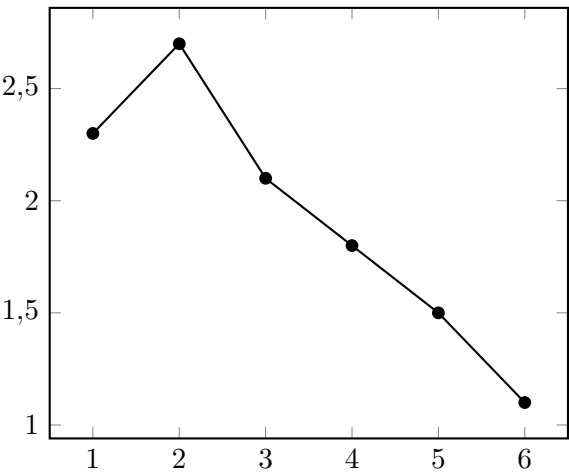
9.2.2. Histograma

Es una gráfica de la distribución de un conjunto de datos. Donde no hay espacios entre las barras.



9.3. Polígono de Frecuencia

Se construye uniendo los extremos de un diagrama de barras o puntos de los rectángulos de un histograma.



Referencias

- [1] Instituto de Ciencias y Humanidades - Perú, *Aritmética, análisis del número y sus aplicaciones*. Lumbrera Editores, 2014.
- [2] I. Espejo Miranda, F. Fernández Palacín, M. A. López Sánchez, M. Muñoz Márquez, A. M. Rodríguez Chía, A. Sánchez Navas, C. Valero Franco, *Estadística Descriptiva y Probabilidad*. Universidad de Cádiz, 2006.
- [3] Artículo: http://calculo.cc/temas/temas_e.s.o/estadistica/teoria/poligono-frec.html
- [4] Fuente para dibujar gráficas: pgfplots.sourceforge.net/gallery.html