



PROBABILIDAD Y ESTADÍSTICA

UNIDAD 1: Estadística descriptiva

Profesora:

Claudia Villar

AÑO 2023

Unidad 1: Estadística descriptiva

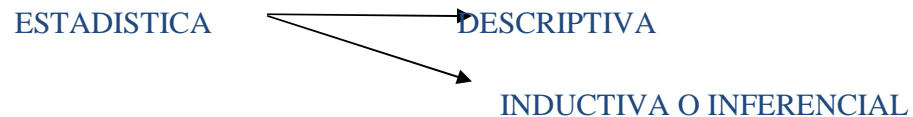
Como toda disciplina científica la Estadística tiene un lenguaje que le es propio. En el texto que sigue encontrarán los conceptos básicos de estadística que necesitan para afrontar el estudio de la materia.

Esperamos que al finalizar la lectura del siguiente texto puedan diferenciar los conceptos población de muestra, variable cualitativa de variable cuantitativa y censo de muestreo

¿QUÉ ES LA ESTADÍSTICA?

La Estadística es una ciencia que se ocupa de recolectar, organizar y analizar información sobre un grupo de datos de forma tal que podamos obtener conclusiones válidas sobre ellos.

Podemos clasificar la Estadística en dos ramas:



La **ESTADÍSTICA DESCRIPTIVA** trata sobre el procesamiento de los datos sin intenciones de formular inferencias sobre ellos. Se ocupa de recolectar, organizar y procesar los datos. Presenta datos por medio de gráficos y tablas que describen una determinada situación particular. Esta es la noción de estadística que tiene asociada la mayoría de la gente.

La **INFERENCIA ESTADISTICA** es la disciplina científica que se ocupa de sacar conclusiones sobre determinado grupo de estudio para poder realizar inferencias a partir de los datos obtenidos. En este rol de generar procedimientos válidos para el análisis de los datos, la estadística complace las necesidades de la investigación científica en sus diversas disciplinas.

La Teoría de la Probabilidad es la base del desarrollo y los fundamentos de la inferencia estadística

POBLACIÓN Y MUESTRA

Uno de los objetivos de la investigación estadística es explorar las características de una población utilizando sólo unas pocas observaciones. Por ejemplo: frente a una elección, si queremos pronosticar sus resultados, encuestamos a un grupo reducido de votantes (ya que encuestar a todos sería un proceso engorroso y muy caro). Es decir que, mediante el grupo que encuestamos estamos analizando el comportamiento de todos los votantes en las urnas.

Se denomina POBLACIÓN al conjunto de elementos de un grupo en estudio

Una población puede ser finita o infinita, Si podemos listar a los miembros de una población, entonces la población es finita. En caso contrario decimos que es infinita.

Una MUESTRA es el Conjunto de elementos que componen una parte representativa de la población de estudio.

Los términos población y muestra son relativos. Por ejemplo, todos los habitantes de Catamarca constituyen una población respecto de Catamarca. En cambio, son una muestra cuando la población en estudio son los habitantes de la Argentina.

Para obtener conclusiones sobre una población infinita o cuando el estudio de una población finita es costoso, es conveniente trabajar con muestras.

VARIABLES ESTADÍSTICAS

La característica que estudiamos de una población es lo que denominamos

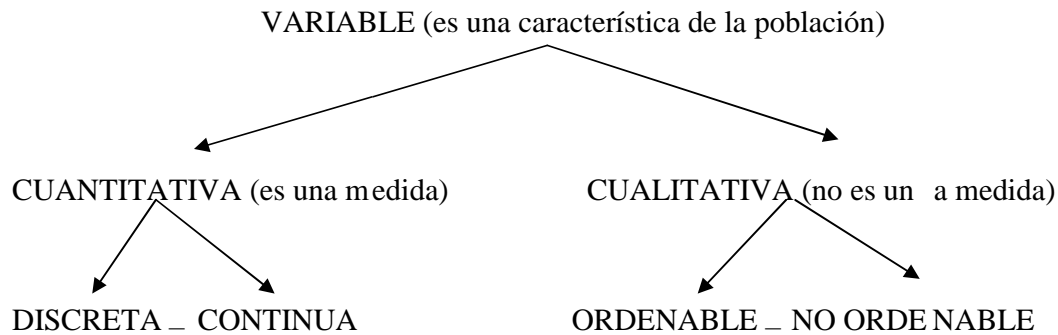
VARIABLE.

Las variables pueden clasificarse en dos grandes grupos:

- **cuantitativa o numérica:** son aquellas variables que toman valores numéricos.
- **cualitativa o categórica:** son aquellas definidas por clases o categorías

Cuando los valores que toma una variable cuantitativa pertenecen a un conjunto finito o infinito numerable decimos que la variable es discreta. En caso contrario, decimos que es continua.

Cuando en una variable cualitativa podemos ordenar los datos, decimos que la variable es ordenable. En caso contrario decimos que es no ordenable.



GRÁFICOS

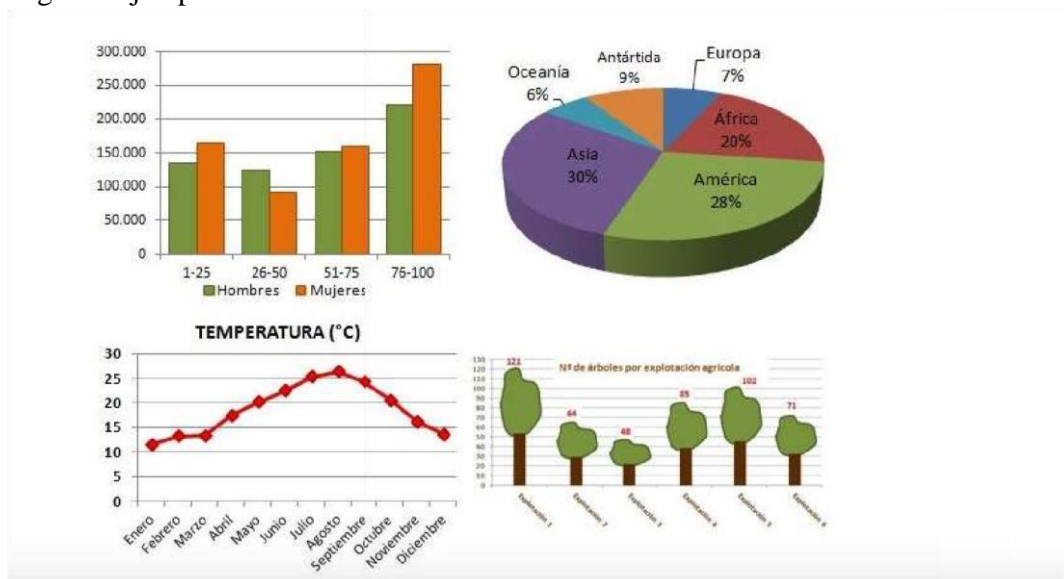
Se utilizan distintos tipos de gráficos, dependiendo de la variable a analizar.

Si es una variable cualitativa con pocas opciones, por ejemplo encuestas con opciones: SI – NO – PUEDE SER; se utiliza un gráfico circular o de tortas que nos permite ver rápidamente los valores obtenidos.

Si es una variable con cuantitativa, puedo volcar los datos en un gráfico de barras horizontales o verticales.

Si es una variable continua se utiliza el histograma, similar al gráfico de barras que termina uniendo los valores en un gráfico lineal.

Algunos ejemplos:



FRECUENCIAS ABSOLUTAS Y RELATIVAS

La frecuencia ABSOLUTA de un dato es la cantidad de veces que se presenta en la muestra analizada.

La frecuencia RELATIVA de un dato es la fracción del total de la muestra que le corresponde a ese dato, es decir, es el cociente entre la frecuencia absoluta y el total de elementos de la muestra.

La frecuencia ACUMULADA es el valor que acumula las frecuencias absolutas las variables menores o iguales a un valor dado.

Tablas de frecuencias.

Son tablas que muestran los datos ordenados y sus frecuencias, permiten leer la información con más claridad que cuando los datos están sueltos. Por ejemplo:

Un profesor tomó un examen en un curso de 25 estudiantes ($n=25$), obteniéndose los siguientes valores: 6, 3, 4, 4, 6, 6, 10, 10, 9, 2, 2, 2, 5, 7, 7, 4, 10, 10, 1, 1, 6, 1, 2, 8, 9, Los datos se ordenaron en la siguiente tabla:

Calificaciones	f	fr	f%	F	Fr	F%
1	3	0,12	12%	3	0,12	12%
2	4	0,16	16%	7	0,28	28%
3	1	0,04	4%	8	0,32	32%
4	3	0,12	12%	11	0,44	44%
5	1	0,04	4%	12	0,48	48%
6	4	0,16	16%	16	0,64	64%
7	2	0,08	8%	18	0,72	72%
8	1	0,04	4%	19	0,76	76%
9	2	0,08	8%	21	0,84	84%
10	4	0,16	16%	25	1	100%

f (frecuencia absoluta): cantidad de veces que se repite un valor.

fr (frecuencia relativa): es el cociente entre la frecuencia absoluta y el tamaño de la muestra

$fr = f\%$ (frecuencia porcentual): es el valor de la fr multiplicado por 100%

F (frecuencia acumulada): Es el valor que acumula las frecuencias absolutas menores o iguales al valor dado.

Fr (frecuencia relativa acumulada): Es el valor que acumula las frecuencias relativas.

$F\%$ (frecuencia porcentual acumulada): es el valor que se obtiene al sumar las frecuencias porcentuales menores o iguales a un valor dado.

Las f en minúsculas representan frecuencias y las F en mayúsculas representan frecuencias acumuladas

En el ejemplo para calcular $F(4)$, hay que sumar las frecuencias absolutas de los valores menores o iguales a 4, es decir sumar $(3 + 4 + 1 + 3)$ entonces $F(4)=11$.

De la misma manera podríamos calcular $Fr(6)$, sumando las fr de todos los valores menores o iguales a 6, sumar $(0,12 + 0,16 + 0,04 + 0,12 + 0,04 + 0,16)$ entonces $Fr(6)=0,64$

Estos datos también pueden representarse en un gráfico de barras

DATOS AGRUPADOS

Cuando una muestra está constituida por un conjunto numeroso de datos numéricos, es conveniente agruparlos en clases.

Observá el ejemplo:

. Los datos de la lista corresponden al nivel de glucosa en la sangre de 50 individuos (en gramos/1 000 cm³) Construir una tabla de frecuencias con los datos agrupados en 8 clases. Los valores son:

0,85; 0,89; 1,10; 0,85; 1,21; 0,87; 0,91; 0,85; 1,10; 0,89; 1,20; 0,87; 1,20; 0,89, 0,85;

1,20; 0,87; 1,22; 0,87; 1,12; 0,85; 0,89; 1,19; 0,85; 0,81; 1,20; 0,91; 0,92; 1,15; 0,87; 1,21; 1,10; 0,87; 1,12; 1,12; 1,12; 0,83; 1, 10, 085, 1,12; 0,87; 0,90; 0,90; 1,10; 0,90; 0,81; 0,81; 1,14; 1,2; 1,10.

Es habitual tomar los intervalos de igual longitud. En ese caso, la amplitud de cada uno (d) es la diferencia entre el mayor valor y el menor valor de la muestra dividida por el número de clases a formar

En el ejemplo $d = \frac{1,22 - 0,81}{8} = 0,05125$ Conocida d se forman los intervalos de las clases, como se indica al costado. Si un dato coincide con el límite superior de un intervalo se lo incluye en el siguiente.

La tabla de frecuencias de los datos agrupados es:

Clase	Intervalo	f	f _r
1. ^a	[0,81; 0,86125)	11	0,22
2. ^a	[0,86125; 0,9125)	16	0,32
3. ^a	[0,9125; 0,96375)	1	0,02
4. ^a	[0,96375; 1,015)	0	0
5. ^a	[1,015; 1,06625)	0	0
6. ^a	[1,06625; 1,1175)	6	0,12
7. ^a	[1,1175; 1,16875)	8	0,16
8. ^a	[1,16875; 1,22]	8	0,16
Total		50	1

La frecuencia de la 1° clase es 11, porque hay 11 datos que pertenecen al intervalo (0,81; 0,86125). Las frecuencias de la 4° y la 5° clase son nulas, porque no hay datos en los intervalos correspondientes.

ACTIVIDAD:

- 1) Definir muestra y población, ejemplificar.
- 2) Dar ejemplos de encuestas dónde utilicen variables cualitativas y cuantitativas,

- 3) Se midió el colesterol total de 50 pacientes. Armá una tabla de frecuencias con 8 clases de igual amplitud, y hacé el Histograma correspondiente. Los datos son: 171, 178, 175, 173, 182, 189, 184, 188, 181. 187, 186,
- 188, 198, 191, 191. 195, 197. 196, 199, 194, 199. 197, 194, 193, 205, 209, 207, 204, 207, 206, 208, 201,203, 202, 202, 201, 207. 205. 208, 207, 215, 211, 213, 218, 212, 222, 227, 228, 234, 243.