



estadística

Julieth Tatiana Hernández cerquera

Que es la estadística

La Estadística es la parte de las Matemáticas que se encarga del estudio de una determinada característica en una población, recogiendo los datos, organizándolos en tablas, representándolos gráficamente y analizándolos para sacar conclusiones de dicha población

Población y muestra

Población se refiere al universo, conjunto o totalidad de elementos sobre los que se investiga o hacen estudios. Muestra es una parte o subconjunto de elementos que se seleccionan previamente de una población para realizar un estudio.

- Normalmente se selecciona la muestra de una población para su estudio, debido a que estudiar a todos los elementos de una población resultaría muy extenso y poco práctico.

Tipos de estadística

Estadística descriptiva

La estadística descriptiva o deductiva permite presentar de manera resumida y organizada los datos numéricos obtenidos tras un estudio o análisis en particular. Su objetivo, por lo tanto, es describir las características principales de los datos reunidos y evitar generalizaciones.

Estadística inferencial

La estadística inferencial o inductiva es el estudio que utiliza técnicas a partir de las cuales se obtienen generalizaciones o se toman decisiones en base a una información parcial o completa obtenida mediante técnicas descriptivas.

Su objetivo es extraer conclusiones de utilidad sobre el total de las observaciones posibles basándose en la información obtenida.

Estadística aplicada

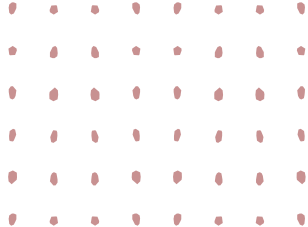
La estadística aplicada hace uso de los métodos expuestos anteriormente, y permite realizar inferencias a partir de una o varias muestras de una determinada población como objeto de estudio. De esta manera se pueden ofrecer resultados tanto específicos como generalizados.

La estadística aplicada se utiliza en diversas ciencias, como la historia, la economía, la educación o la sociología para realizar estudios y análisis estadísticos.

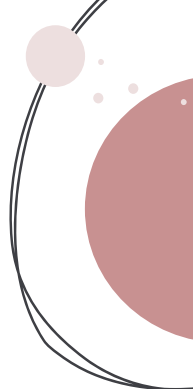
Estadística matemática

Se trata de la estadística que arroja datos aleatorios e inciertos, por ello hace uso de la teoría de la probabilidad, una rama de las matemáticas que estudia estos casos.

	Población	Muestra
Definición	Universo de elementos que se van a estudiar.	Selección de una parte de la población que se va a ser sujeto de estudio.
Características	- Se puede clasificar según la cantidad de individuos que la conforman. - Posee variables estadísticas.	- Forma parte de la población: debería comprender entre 5% y 10% para ser más efectiva. - Los elementos deben ser aleatorios. - Debe ser representativa de la población.
Objetivos	Analizar los datos recabados referentes a las características comunes que comparten los elementos con diversos propósitos.	Estudiar el comportamiento, características, gustos o propiedades de una parte representativa de la población.
Ejemplos	- Las personas que habitan un país. - La cantidad de carros en una ciudad. - Los estudiantes de un país.	Para el estudio del desempeño de los estudiantes de cinco universidades de una ciudad en una materia específica, se toma como muestra a 500 estudiantes aleatoriamente (100 de cada institución) que estén cursando el mismo nivel para que la muestra sea representativa.



Tipos de variables estadísticas



cuando hablamos de variable estadística estamos hablando de una cualidad que, generalmente adopta forma numérica. Por ejemplo, la altura de Juan es de 180 centímetros. La variable estadística es la altura y está medida en centímetros.





aunque hay decenas de tipos de variables estadísticas, por norma general podemos encontrarnos dos tipos de variables:

Variable cuantitativa: Son variables que se expresan numéricamente.

Variable continua: Toman un valor infinito de valores entre un intervalo de datos. El tiempo que tarda un corredor en completar los 100 metros lisos.

Variable discreta: Toman un valor finito de valores entre un intervalo de datos. Número de helados vendidos.

Variable cualitativa: Son variables que se expresan, por norma general, en palabras.

Variable ordinal: Expresa diferentes niveles y orden.

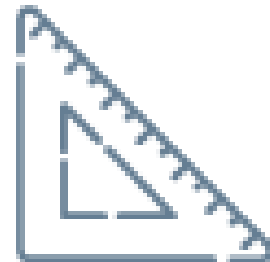
Variable nominal: Expresa un nombre claramente diferenciado.

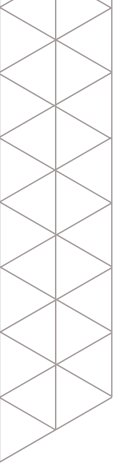
Por ejemplo el color de ojos puede ser azul, negro, castaño, verde, etc.

Las Tablas de frecuencias

- Las Tablas de frecuencias son herramientas de Estadística donde se colocan los datos en columnas representando los distintos valores recogidos en la muestra y las frecuencias (las veces) en que ocurren.
- Frecuencia absoluta
- La frecuencia absoluta (n_i) es el número de veces que aparece un determinado valor en un estudio estadístico. Número de veces que se repite el i -ésimo valor de la variable. La suma de las frecuencias absolutas es igual al número total de datos, que se representa por n
- Frecuencia absoluta acumulada
- La Frecuencia absoluta acumulada (N_i) es la suma de las frecuencias absolutas de todos los valores inferiores o iguales al valor considerado.
- $N_1 = n_1$
- $N_2 = n_1 + n_2 = N_1 + n_2$
- $N_3 = n_1 + n_2 + n_3 = N_2 + n_3$
- $N_k = n.$

- Frecuencia relativa
- La frecuencia relativa (f_i) es la proporción de veces que se repite un determinado dato.
- La frecuencia relativa es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos.
- $f_i = n_i/n$
- La suma de las frecuencias relativas es igual a 1.
- Frecuencia relativa acumulada
- La frecuencia relativa acumulada (F_i) es el número de observaciones menores o iguales al i -ésimo valor de la variable pero en forma relativa.
- $F_1 = f_1$
- $F_2 = f_1 + f_2 = F_1 + f_2$
- $F_3 = f_1 + f_2 + f_3 = F_2 + f_3$
- $F_k = 1$





medidas descriptivas

- Las medidas descriptivas son valores numéricos que resumen la información contenida en un conjunto de datos. Las medidas descriptivas pueden calcularse a partir de los datos de una muestra o los datos de una población.
 - El estudio de una variable estadística comienza con la obtención de datos, bien sondeando la población o tomando una muestra. El siguiente paso en el proceso es la ordenación de datos elaborando la tabla correspondiente.
- 

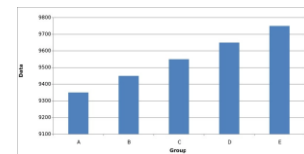
Tablas de frecuencias datos sin agrupar

- tablas de frecuencias con datos no agrupados.
Usamos este tipo de tablas cuando tenemos variables cualitativas, o variables cuantitativas con pocos valores. ... Es la suma de las frecuencias relativas, y se puede calcular también dividiendo la frecuencia acumulada entre el número de datos del estudio.

Tablas de frecuencias datos agrupados

- Por lo general una tabla de frecuencias con datos agrupados se realiza cuando la cantidad de datos es grande y/o la variable es continua. Básicamente consiste en agrupar los datos en intervalos de una misma amplitud, denominados clases. A cada clase se le asignan valores de cada tipo de frecuencias

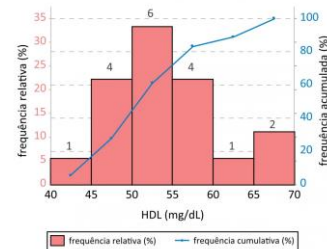
- Gráfico de barras
- El más conocido y utilizado de todos los tipos de gráficos es el gráfico o diagrama de barras. En éste, se presentan los datos en forma de barras contenidas en dos ejes cartesianos (coordenada y abscisa) que indican los diferentes valores. El aspecto visual que nos indica los datos es la longitud de dichas barras, no siendo importante su grosor.



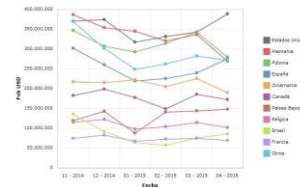
- Gráfico circular o por sectores
- El también muy habitual gráfico en forma de “quesito”, en este caso la representación de los datos se lleva a cabo mediante la división de un círculo en tantas partes como valores de la variable investigada y teniendo cada parte un tamaño proporcional a su frecuencia dentro del total de los datos. Cada sector va a representar un valor de la variable con la que se trabaja.



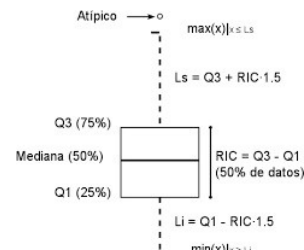
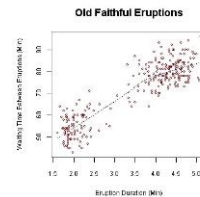
- Histograma
- Aunque a simple vista muy semejante al gráfico de barras, el histograma es uno de los tipos de gráfica que a nivel estadístico resulta más importante y fiable. En esta ocasión, también se utilizan barras para indicar a través de ejes cartesianos la frecuencia de determinados valores, pero en vez de limitarse a establecer la frecuencia de un valor concreto de la variable evaluada refleja todo un intervalo. Se observa pues un rango de valores, que además podrían llegar a reflejar intervalos de diferentes longitudes.



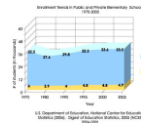
- Gráfico de líneas
- En este tipo de gráfico se emplean líneas para delimitar el valor de una variable dependiente respecto a otra independiente. También puede usarse para comparar los valores de una misma variable o de diferentes investigaciones utilizando el mismo gráfico (usando diferentes líneas). Es usual que se emplee para observar la evolución de una variable a través del tiempo.



- Gráfico de dispersión
- El gráfico de dispersión o gráfico xy es un tipo de gráfico en el cual mediante los ejes cartesianos se representa en forma de puntos todos los datos obtenidos mediante la observación. Los ejes x e y muestran cada uno los valores de una variable dependiente y otra independiente o dos variables de la que se esté observando si presentan algún tipo de relación.
- Gráfico de caja y bigotes
- Los gráficos de caja son uno de los tipos de gráficas que tienden a utilizarse de cara a observar la dispersión de los datos y cómo éstos agrupan sus valores. Se parte del cálculo de los cuartiles, los cuales son los valores que permiten dividir los datos en cuatro partes iguales. Así, podemos encontrar un total de tres cuartiles (el segundo de los cuales se corresponderían con la mediana de los datos) que van a configurar la “caja” en cuestión. Los llamados bigotes serían la representación gráfica de los valores extremos.



- Gráfico de áreas
- En este tipo de gráfico se observa, de manera semejante lo que ocurre con los gráficos de líneas, la relación entre variable dependiente e independiente. Inicialmente se hace una línea que une los puntos que marcan los diferentes valores de la variable medida, pero también se incluye todo lo situado por debajo: este tipo de gráfica nos permite ver la acumulación (un punto determinado incluye a los situados por debajo).



- Pictograma
- Se entiende por pictograma a un gráfico en el que, en vez de representar los datos a partir de elementos abstractos como barras o círculos, se emplean elementos propios del tema que se está investigando. De este modo se hace más visual. Sin embargo, su funcionamiento es semejante al del gráfico de barras, representando frecuencias de la misma manera

- Cartograma
- Este gráfico resulta de utilidad en el terreno de la epidemiología, indicando las zonas o áreas geográficas en las que aparece con mayor o menor frecuencia un determinado valor de una variable. Las frecuencias o rangos de frecuencias se indican mediante el uso del color (requiriéndose una leyenda para comprenderse) o el tamaño.

5. Estadística descriptiva (Medidas de tendencia central y de dispersión)

- Las medidas de tendencia central son medidas estadísticas que pretenden resumir en un solo valor a un conjunto de valores. ... Las medidas de tendencia central más utilizadas son: media, mediana y moda. Las medidas de dispersión en cambio miden el grado de dispersión de los valores de la variable
- Las medidas de centralización o de tendencia central expresan el valor en torno al cual se sitúan los datos de una muestra. La medida más habitual de este tipo es la media aritmética, que se obtiene sumando todos los datos y dividiendo el resultado entre el número total de datos sumados.
- Las medidas de dispersión tratan, a través del cálculo de diferentes fórmulas, de arrojar un valor numérico que ofrezca información sobre el grado de variabilidad de una variable. En otras palabras, las medidas de dispersión son números que indican si una variable se mueve mucho, poco, más o menos que otra.

Medidas de tendencia central datos no agrupados

- Las medidas de tendencia central son medidas estadísticas que pretenden resumir en un solo valor a un conjunto de valores. ... Si los datos se encuentran ordenados en una tabla estadística diremos que se encuentran “agrupados” y si los datos no están en una tabla hablaremos de datos “no agrupados”.

MEDIDAS DE TENDENCIA CENTRAL PARA DATOS AGRUPADOS

- Son medidas estadísticas que se usan para describir cómo se puede resumir la localización de los datos. Ubican e identifican el punto alrededor del cual se centran los datos. Las más utilizadas son: la media, la mediana y la moda.

Ejemplo:

En la siguiente tabla se registran el número de materias insuficiente del grado 9°C durante el segundo periodo en la Institución Educativa Santa Teresita.

Materias insuficientes	x_i	Número de estudiantes f_i	$x_i \cdot f_i$
0-2	1	13	
2-4	3	7	
4-6	5	15	
6-8	7	10	
8-10	9	3	
Total			