

CM ESTADISTICA

Jennifer Lizeth
Rodríguez Forero

2021

MCS

ESTADISTICA

¿QUE ES?

La estadística es una ciencia que nos ayuda con el análisis de datos, los reúne, analiza y organiza. Generando conclusiones para tomar decisiones.

¿PARA QUE SIRVE?

Saber manejar la información respecto a la fluctuación del dinero, aprender a estudiarla nos permite tomar decisiones mas efectivas. Se suele usar en la vida cotidiana.

Clasificación

- Estadística descriptiva: La estadística descriptiva es la técnica que obtiene, organiza, presenta y describe un conjunto de datos con el propósito de facilitar el uso, generalmente con el apoyo de tablas, medidas numéricas o gráficas.
- Estadística inferencial: Abarca un conjunto de procedimientos estadísticos que se utilizan para hacer generalizaciones sobre una población a partir de los datos obtenidos de una muestra.

Poblacion

- una población es un conjunto de elementos o eventos similares que son de interés para alguna pregunta o experimento. Una población estadística puede ser un grupo de objetos existentes o una hipotética

Muestra

- se denomina así a un grupo de individuos seleccionados para representar a una población, especialmente cuando esta es demasiado amplia para ser estudiada en su totalidad.

Estadística Paso a Paso :

Tipos de Variables

Cualitativas

se dividen en nominales y ordinales se caracterizan por tener categorías. Las variables nominales tienen categorías que pueden estar en cualquier orden, las variables ordinales tienen categorías ordenadas y no pueden ser alterados.

Cuantitativas

se dividen en continuas y discretas se caracterizan por tener números, las variables continuas son las que pueden tomar valores decimales, las variables discretas no pueden tener decimales o es uno o dos.

Tipos de estadística

- Estadística descriptiva: es la técnica matemática que obtiene, organiza, presenta y describe un conjunto de datos con el propósito de facilitar el uso, generalmente con el apoyo de tablas, medidas numéricas o gráficas.

- Estadística inferencial: es una parte de la estadística que comprende los métodos y procedimientos que por medio de la inducción determina propiedades de una población estadística, a partir de una parte de esta.

organizar información

- Primero se halla la situación problemática y realizar un encuesta luego se debe mirar como organizar la información recolectada en un gráfico ya sea de barras, líneas, Pareto, sectores, pictograma, dispersión, cartograma o pirámide de población.

combinaciones diagrama de árbol.

- Colección no ordenada de elementos u objetos para realizar las combinaciones sencillas se debe usar grafico de árbol y la tabla de doble entrada.

Método del árbol:

es una representación gráfica de los posibles resultados del experimento, el cual consta de una serie de pasos, donde cada uno de estos tiene un número infinito de maneras de ser llevado a cabo. Se utiliza en los problemas de conteo y probabilidad.

- también llamadas tablas de contingencia, brindan información estadística de dos variables relacionadas entre sí, independientemente de si son cualitativas o cuantitativas. Son útiles en casos en los cuales un experimento es dependiente del otro.
-
-
-
-

Medidas para estadística Descriptiva.

- 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.

Las medidas que se utilizan dependen de la variable que quieran utilizar, para la cualitativa las 4 medidas son frecuencia absoluta, relativa, acumulada y porcentaje.

Las variables cuantitativas son 3 medidas de tendencia central (media, mediana y moda), dispersión (mínimo, máximo, desvió estándar y varianza) y medidas de posición no centrales (percentiles, cuartiles y deciles).

Tablas de frecuencias datos sin agrupar

- es una tabla que muestra la distribución de los datos mediante sus frecuencias. Se utiliza para variables cuantitativas o cualitativas ordinales.
- La tabla de frecuencias es una herramienta que permite ordenar los datos de manera que se presenten numéricamente las características de la distribución de un conjunto de datos o muestra.

x	f	Fr	%	F
10	2	0.066	6.6	2
11	5	0.166	16.6	7
12	12	0.4	40	19
13	5	0.166	16.6	24
14	2	0.066	6.6	26
15	4	0.133	13.3	30
TOTALES	30	0.997	99.7	

- La x son los datos que tenemos
- La f es la frecuencia absoluta y representa el número de veces que se repite un dato
- la fr es la frecuencia relativa entonces $fr = f \text{ sobre } n$ que es el número total de datos
- el signo de porcentaje % representa el porcentaje entonces $fr \times 100 = \%$
- la F mayúscula representa la frecuencia absoluta acumulada

Tablas de frecuencias datos agrupados

Datos Agrupados - Distribución de Frecuencias

Edades de 50 personas: 38 - 15 - 10 - 12 - 62 - 46 - 25 - 56 - 27 - 24 - 23 - 21 - 20 - 25 - 38 - 27 - 48 - 35 - 50 - 65 - 59 - 58 - 47 - 42 - 37 - 35 - 32 - 40 - 28 - 14 - 12 - 24 - 66 - 73 - 72 - 70 - 68 - 65 - 54 - 48 - 34 - 33 - 21 - 19 - 61 - 59 - 47 - 46 - 30 - 30



Valor máximo: 73 años

Valor mínimo: 10 años

Rango = 73 - 10 = 63 años

Intervalos $\left\{ \begin{array}{l} = \sqrt{n} \\ = 1 + 3.322 \text{ Log}(n) \end{array} \right.$
 $n = 50$

Intervalos = $\sqrt{50} = 7.07 \sim 7$

Amplitud = $R \div I = 63 \div 7 = 9$

Edad (x)	Marca de Clase (X _i)	Frecuencia absoluta (f _i)	Frecuencia absoluta acumulada (F _i)	Frecuencia relativa (f _i)		Frecuencia relativa acumulada (F _i)	
[10 - 19)	14.5	5	5	0.1	10%	0.1	10%
[19 - 28)	23.5	11	16	0.22	22%	0.32	32%
[28 - 37)	32.5	8	24	0.16	16%	0.48	48%
[37 - 46)	41.5	5	29	0.1	10%	0.58	58%
[46 - 55)	50.5	8	37	0.16	16%	0.74	74%
[55 - 64)	59.5	6	43	0.12	12%	0.86	86%
[64 - 73]	68.5	7	50	0.14	14%	1	100%
Total		50	Total	1	100%		

- Primero se identifica cual es el valor mínimo y cual es el valor máximo, luego calculamos el rango de manera que al valor máximo se le resta el valor mínimo, ahora calculamos la cantidad de intervalos por ultimo calculamos la amplitud de manera que rango dividido entre el numero de intervalos

Tipos de gráficos

- Histograma: es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia de los valores representados.
- Grafica de barras: es una forma de representar gráficamente un conjunto de datos o valores mediante barras rectangulares de longitud proporcional a los valores representados.
- Pictograma: es un signo icónico, es decir, que representa figurativamente un objeto real, o un significado. Los primeros sistemas de pictogramas fueron precursores o antecedentes de los sistemas de escritura propiamente dichos.
- Gráfica circular: es un recurso estadístico que se utiliza para representar porcentajes y proporciones.
- Grafica poligonal: es una gráfica lineal típicamente utilizada por la estadística para comparar datos y representar la magnitud o frecuencia de ciertas variables.
- Grafica de bigote: es un método estandarizado para representar gráficamente una serie de datos numéricos a través de sus cuartiles. De esta manera, se muestran a simple vista la mediana y los cuartiles de los datos, y también pueden representarse sus valores atípicos.
- Grafico pirámide poblacionales de barras dispuestas horizontalmente cuya longitud es proporcional a la cantidad de personas que representa la edad y sexo de la población en cada una de dichas barras y dicha información sirve para saber el porcentaje de la población.

Calcular Medidas de Tendencia Central en Excel

Se realiza una tabla con los datos obtenidos para luego calcular la media, mediana y moda. Para calcular la media se usa la formula promedio, entonces no posicionamos en l casilla donde queremos aparezca el resultado y la formula será =promedio o =avarage y seleccionamos los datos que estén en la variable que queremos tener resultado de tal manera que quede así =avarage(f2:f211) por ultimo se da enter y ya tenemos el resultado.

Para la mediana se utiliza la formula =mediana o =median.

Lo mismo para moda =moda o =mode y el proceso para mediana y moda es el mismo que para media.

Media

Empleados	Ventas
Empleado 1	10
Empleado 2	7
Empleado 3	5
Empleado 4	6
Empleado 5	8
Empleado 6	10
Empleado 7	10
Empleado 8	9

$$\bar{X} = \frac{10+7+4+6+8+10+10+9}{8}$$

$$\bar{X} = \frac{64}{8}$$

$$\bar{X} = 8$$

- ¿Qué es la media?
- La media, también conocida como promedio, es el valor que se obtiene al dividir la suma de un conglomerado de números entre la cantidad de ellos.
- Algunas características de la media son:
 - Considera todas las puntuaciones
 - El numerador de la fórmula es la cantidad de valores
 - Cuando hay puntuaciones extremas, no tiene una representación exacta de la muestra

mediana

Ejemplo de Mediana

- La cantidad de valores es impar

Si se tienen los valores: **9,5,4,2,7**, se ordenan: **2, 4, 5, 7, 9**. El elemento de en medio es el **5**, ya que se encuentra dos valores por encima y dos valores por debajo.

- La cantidad de valores es par

Si se tienen los valores **9,5,4,2**, se ordenan: **2,4,5,9**. En este caso se toman los dos valores centrales **5 y 4**, la mediana es el promedio de ambos: **9**

- ¿Qué es la mediana?
- La mediana es un conjunto es un valor que se encuentra a la mitad de los otros valores, es decir, que al ordenar los número de menor a mayor, éste se encuentra justamente en medio entre los que están por arriba.
- Algunas características de la media son:
- Las operaciones para calcular el valor son muy sencillas de realizar.
- La medida no depende de los valores de las variables, solamente de su orden.
- Generalmente, los valores son enteros.
- Se puede calcular aunque los números que se encuentren arriba y abajo no tengan límites.

moda

Moda	2,5,5,7,9,10	$\hat{M} = 5$
Bimodal	2,3,3,5,7,8,9,9	$\hat{M} = 3,9$
Multimodal	2,3,3,5,7,7,8,9,9	$\hat{M} = 3,7,9$
Amodal	2,4,5,7,9	$\hat{M} = \text{No}$
Adyacente	2, $\frac{3+3+5+5}{4}$, 7, 8	$\hat{M} = 4$

• ¿Qué es la moda?

• La moda es el valor que aparece más dentro de un conglomerado. En un grupo puede haber dos modas y se conoce como bimodal, y más de dos modas o multimodal cuando se repiten más de dos valores; se llama amodal cuando en un conglomerado no se repiten los valores.

• Por último, se conoce como moda adyacente cuando dos valores continuos tienen la misma cantidad de repeticiones. En este caso se saca el promedio de ambos.

• Las principales características de la moda son:

• Es una muestra muy clara

• Las operaciones para determinar el resultado son muy fáciles de elaborar

• Los valores que se presentan pueden ser cualitativos y cuantitativos

Medidas de tendencia central datos no agrupados

- Las medidas de tendencia central para datos no agrupados son un conjunto de indicadores estadísticos que van a mostrar hacia qué valores se agrupan los datos numéricos, es decir, son medidas estadísticas que buscan resumir en un solo valor un conjunto de valores.

Media Aritmética

\bar{X}

Promedio

Moda

M_o

Dato que mas se repite.

Mediana

M_e

Valor central cuando los datos están acomodados de menor a mayor.

<u>Nombre</u>	<u>Símbolo</u>	<u>Definición</u>	<u>Ventajas</u>	<u>Desventajas</u>
Media	\bar{x}	$\bar{x} = \frac{\sum_{i=1}^n x_1 + x_2 \dots x_i}{n}$	<ol style="list-style-type: none"> 1) Refleja cada valor 2) <u>Propiedades Algebraicas</u> 3) Es la más usada en el análisis estadístico 	<ol style="list-style-type: none"> 1) Puede ser excesivamente influenciada por valores extremos
<u>Mediana</u>	Md	50% de los valores son mayores y 50% menores que ella	<ol style="list-style-type: none"> 1) Menos sensible a valores extremos que la media 	<ol style="list-style-type: none"> 1) Difícil de calcular si hay muchos datos 2) No tiene propiedades algebraicas
<u>Moda</u>	M	Valor con la frecuencia más alta	<ol style="list-style-type: none"> 1) Fácil de calcular 2) Valor "típico" 3) Más valores reunidos en este punto que en cualquier otro 	<ol style="list-style-type: none"> 1) No se emplea en el análisis estadístico 2) Puede haber más de una moda o ninguna

Medidas de dispersión

- 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.

El rango es la diferencia entre el valor máximo y mínimo de un conjunto de datos

$$\text{Rango} = (\text{Max}) - (\text{Min})$$

La desviación media (DM) es el promedio de las distancias de cada punto a su media.

$$D_{\bar{x}} = \frac{\sum_{i=1}^N |X_i - \bar{x}|}{N}$$

MEDIDAS DE DISPERSIÓN

VARIANZA	DESVIACIÓN ESTÁNDAR
$\sigma^2 = \frac{\sum_1^N (x_i - \bar{X})^2}{N}$	$\sigma = \sqrt{\frac{\sum_1^N (x_i - \bar{X})^2}{N}}$

- **X** → Variable sobre la que se pretenden calcular la varianza.
- **x_i** → Observación número i de la variable X. i puede tomará valores entre 1 y n.
- **N** → Número de observaciones.
- **\bar{x}** → Es la media de la variable X.

RANGO ESTADÍSTICO	COEFICIENTE DE VARIACIÓN
$R = Máx_x - Mín_x$	$CV = \frac{\sigma_x}{ \bar{X} }$
<ul style="list-style-type: none">• R → Es el rango.• Máx → Es el valor máximo de la muestra o población.• Mín → Es el valor mínimo de la muestra o población estadística.• x → Es la variable sobre la que se pretende calcular esta medida.	<ul style="list-style-type: none">• X → Variable sobre la que se pretenden calcular la varianza.• σ_x → Desviación típica de la variable X.• \bar{x} → Es la media de la variable X en valor absoluto con $\bar{x} \neq 0$.

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.

Media geométrica para datos agrupados

$$G = \text{Antilog} \frac{\sum F(\log X)}{n}$$

G = Media Geometrica

$$\sum F(\log X) = \text{Sumatoria de la frecuencia por el logaritmo de la marca de clase}$$

n = No. de datos

MEDIA, MEDIANA Y MODA

DATOS AGRUPADOS

EJEMPLO 3

HORAS	x	f	F	xf
55-60	57,5	5	5	287
60-65	62,5	18	23	1125
65-70	67,5	20	43	1350
70-75	72,5	50	93	3625
75-80	77,5	17	110	1317

$$\bar{x} = \frac{\sum xf}{n}$$

$$\bar{x} = \frac{630}{20} = 31,5$$

