San Cayetano
Centro Técnico Educacional

Nivel: 1°Medio

Profesor: Maximiliano Bueno, Franco Orellana y Gonzalo Cofré.

Estadistica y Probabilidades

Objetivos

- ***Definir estadística y diferenciar entre estadística descriptiva e inferencial.***
- ***Conocer los tipos de variables con los cuales se trabaja en estadística, los niveles de medición y los datos que se generan.***
- ***Resumir la información obtenida atreves de la generación de tablas y gráficos estadísticos.***
- ***Calcular e interpretar las principales medidas de centralización, como la media aritmética, mediana, moda y media geométrica.***

¿Qué es la estadistica?

Es la ciencia que se encarga de la recolección, ordenamiento, representación, análisis e interpretación de datos generados en una investigación sobre hechos, individuos o grupos de los mismos, para deducir de ello conclusionesprecisasoestimacionesfuturas.

Conceptos usados en estdisticas.

Población

Es el colectivo que abarca a todos los elementos cuya característica o características queremos estudiar; dicho de otra manera, es el conjunto entero al que se desea describir o del que se necesita establecer conclusiones. Como ejemplos de poblaciones, podemos citar: todos los estudiantes de la Universidad Central del Ecuador, o los artículos producidos en una semana en una determinada fábrica.

Por su tamaño, las poblaciones pueden ser finitas o infinitas.

Muestra

Es un conjunto de elementos seleccionados de una población de acuerdo a un plan de acción previamente establecido (muestreo), para obtener conclusiones que pueden ser extensivas hacia toda la población.

Ejemplos constituyen las muestras que escogen las empresas encuestadoras en estudios de sondeos de opinión, o la selección de un grupo de artículos recibidos en una bodega para estimar las condiciones de todo un embarque.

Muestreo

Es la técnica que nos permite seleccionar muestras adecuadas de una población de estudio. El muestreo debe conducir a la obtención de una muestra representativa de la población de donde proviene, esta condición establece que cada elemento de la población tiene la misma probabilidad de ser incluida en la muestra.

El estudio de selección de muestras, en sí constituye todo un estudio pormenorizado, que no atañe al estudio en este texto.

Parámetro

Es cualquier medida descriptiva o representativa de una población. Generalmente se utilizan las letras griegas como símbolo. Ejemplos: media aritmética poblacional(μ)(mu), desviación estándar poblacional (σ) (sigma).

Principales tipos de parámetros son:

Parámetros de tendencia central o de resumen, siendo los más importantes:

- La media o promedio (μ)
- La mediana (Me)
- Lamoda (Mo)

Parametros de variabilidad, siendo los más importantes:

- La variancia o varianza ($\sigma 2$)
- La desviación estándar (σ)
- El coeficiente de variabilidad (C.V.)

Estadístico

Constituyen cualquiera de las medidas descriptivas de una muestra. Se las simboliza con letras minúsculas de nuestro alfabeto. Ejemplos: media aritmética (x), desviación estándar(s).

Tipos de Estadistica

- **Estadística descriptiva**: Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

- **Estadística inferencial:** Apoyándose en el cálculo de probabilidades y a partir de datos muestrales, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos. Su tarea fundamental es la de hacer inferencias acerca de la población a partir de una muestra extraída de la misma.

Variables o caracteres: característica observable que varía entre los diferentes individuos de una población. Las variables pueden dividirse en cualitativas y cuantitativas.

- **Variable cualitativa o categórica**: Los valores que toman no se pueden cuantificar. Cada uno de estos valores se denomina categoría, clase o modalidad. Pueden ser Ordinales o Nominales dependiendo de si se puede establecer un orden entre las diferentes categorías o no. Ejem- plos de este tipo de variables son el carácter rango militar y el sexo, respectivamente.

- **Variable cuantitativa o medibles**: Los valores que toman se pueden cuantificar o medir. Pueden ser Discretas (los valores que pueden tomar son aislados) o Continuas (pueden tomar cualquier valor de la recta real o de un intervalo). Ejemplos de este tipo de variables son el carácter número de hermanos y el precio de unas acciones respectivamente.

Tabla de Frecuencias

- Exponen la información recogida en la muestra de manera inteligente:
  - Frecuencias absolutas: Contabilizan el número de individuos de cada modalidad.
  - Frecuencias relativas (porcentajes unitarios): Ídem, pero dividido por el total, normalizadas.
  - Frecuencias acumuladas absolutas y relativas: Acumulan las frecuencias absolutas y relativas. Son especialmente útiles

Tabla de Frecuencias cont.

Ordenamos los datos en forma creciente:

60	65	65	70	70	70	75	75	75	80
80	80	85	85	85	85	90	90	90	90
90	95	95	95	95	100	100	110	110	120

La **amplitud** total $A = 120 - 60$

Número de clases: $K = 30^{1/2} = 5.48$. Aprox. 6 clases

Extensión del intervalo: $H = A/K = 60/6 = 10$

Variable X_i	Frecuencia Absoluta n_i	Frecuencia Absoluta Acumulada N_i	Frecuencia Relativa h_i	Frecuencia Relativa Acumulada H_i
60 - 70	3	3	0.1	0.1
70 - 80	6	9	0.2	0.3
80 - 90	7	16	0.23	0.53
90 - 100	9	25	0.3	0.83
100 - 110	2	27	0.07	0.90
110 - 120	2	29	0.07	0.97
120 - 130	1	30	0.03	1.00
total	30		1.0	

[1] La frecuencia relativa corresponde al cociente (división) entre cada una de las cantidades de las clases por el total de esta, ejemplo $h_1 = n_1/\text{total}$

Moda, Mediana y Media Aritmética

La Media

La media o media aritmética, usualmente llamada promedio, se obtiene sumando todos los valores de los datos y divide el resultado entre la cantidad de datos. Si los datos proceden de una muestra la media se representa con una x testada (x) y si provienen de la población se representan con la letra griega miu (μ).

Media aritmetica para datos no agrupados muestrales

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Media aritmetica para datos no agrupados poblacionales

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Media aritmetica para datos agrupados

$$\bar{x} = \frac{\sum_{clases} f * xc}{\sum_{clases} fi}$$

Donde

X: promedio muestral (estadístico).

μ: promedio poblacional (parámetro).

∑: signo de sumatoria.

N = número de datos de la población.

n: número de datos de la muestra.

fi: frecuencia absoluta.

Xc: Marca de clase o punto medio

Ejemplo 1.-

De cómo se emplea la media o promedio con el siguiente ejemplo para datos no agrupados:

a) A continuación, se presenta una muestra de las puntuaciones en un examen de un curso de estadística:

70, 90, 95, 74, 58, 70, 98, 72, 75, 85, 95, 74, 80, 85, 90, 65, 90, 75, 90, 69

Podemos calcular el promedio de las puntuaciones para conocer cuántos estudiantes obtuvieron puntuaciones por encima y por debajo del promedio.

Primero, sumamos todos los valores de los datos y el resultado lo divide entre el total de datos o tamaño de la muestra. Al sumar todas las puntuaciones en el ejemplo anterior obtendrás un total de 1600, que dividido por 20(total de datos), es igual a 80. Si empleamos la fórmula obtenemos:

$$\bar{x} = \frac{\sum x}{n} \qquad \bar{x} = \frac{1600}{20} = 80$$

Ejemplo 2.-

La media para datos agrupados.

Ejemplo: Para los gastos diarios en periódicos del hotel agrupados en una tabla de frecuencia:

Intervalo de clase	Fi	Xc	Fi * Xc
5.2 - 6.0	3	5.6	16.8
6.1 – 6.9	5	6.5	32.5
7.0 – 7.8	9	7.4	66.6
7.9 – 8.7	7	8.3	58.1
8.8 – 9.6	5	9.2	46.0
9.7 – 10.5	3	10.1	30.3
Total	**32**		**250.4**

El promedio aritmético es:

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i X_i}{n} = \frac{250.4}{32} = 7.825$$

Se construye la tabla de distribución de frecuencias. 2do) se obtiene el total de la frecuencia absoluta de clase por el punto medio. 3ro) El resultado obtenido se divide entre el tamaño de la muestra.

Propiedades de la Media:

1ª) La suma de las desviaciones de los valores o datos de una variable X, respecto a su media aritmética es cero.

Ventajas e inconvenientes:

- La media aritmética viene expresada en las mismas unidades que la variable.
- En su cálculo intervienen todos los valores de la distribución.
- Es el centro de gravedad de toda la distribución, representando a todos los valores observados.
- Es única.
- Su principal inconveniente es que se ve afectada por los valores extremadamente grandes o pequeños de la distribución.

La Mediana

La segunda medida de tendencia central que analizaremos es la mediana, en ocasiones se le llama media posicional, porque queda exactamente en la mitad de un grupo de datos, luego de que los datos se han colocado de forma ordenada. En este caso la mitad (50%) de los datos estará por encima de la mediana y la otra mitad (50%) estará por debajo de ella. La mediana es el valor intermedio cuando los valores de los datos se han ordenado.

La Mediana (Me) para datos no agrupados:

1. Primero se ordenan los datos.
2. Luego se calcula la pocision de la mediana con la siguiente formula: (n+1)÷2

donde, n es el número de datos.

a) Por ejemplo, se tiene una muestra de tamaño 5 con los siguientes valores: 46, 54, 42, 48 y 32.

Primer paso, ordenar los datos: 32 42 46 48 54

Como la cantidad de datos es impar (5 datos), la mediana es el valor del dato que se encuentra ubicado en la posición (5+1)÷2=3, la mediana es: Me = 46.

b) Se ha obtenido una muestra con los valores de datos: 27, 25, 27, 30, 20 y 26. ¿cómo se determina la mediana en este caso?.

Primer paso, ordenar los datos de forma ascendente: 20 25 26 27 27 30

Como el número de datos es par (6), la mediana es el promedio de los datos que se

encuentran en las posiciones (6+1) ÷1 = 3.5. Por lo tanto la mediana es:

$$\boxed{Me = \frac{26+27}{2} = 26.5}$$

Para Datos Agrupados.

$$Me = L_i + \frac{\left(\frac{n}{2} - F_{i-1}\right)}{f_i} \quad \text{(i)}$$

Donde:

- Li: Limite inferior real de la clase que contiene la mediana.
- n: tamaño de la muestra.
- Fi-1 = AFA: Frecuencia acumulada anterior a la clase que contiene la mediana.
- Fi: frecuencia de clase absoluta de la clase mediana.

Para identificar la clase mediana se divide n/2 y la primera clase que contenga una frecuencia acumulada mayor que n/2.

n = 32, entonces n/2 = 32/2 = 16. Buscar la primera frecuencia acumulada mayor que 16, esa sera la clase mediana.

Intervalo de clase	fi	Xc	fi * Xc	fa	Limites reales
5.2 - 6.0	3	5.6	16.8	3	5.15 – 6.05
6.1 – 6.9	5	6.5	32.5	8	6.05 – 6.95
7.0 – 7.8	**9**	**7.4**	**66.6**	**17**	**6.95 – 7.85**
7.9 – 8.7	7	8.3	58.1	24	7.85 – 8.75
8.8 – 9.6	5	9.2	46.0	29	8.75 – 9.65
9.7 – 10.5	3	10.1	30.3	32	9.65 – 10.55
Total	**32**		**250.4**		

Ahora se aplica la formula:

1. *Me = (6.95+(((32/2 – 8)/9)*(0.9)) = 6.95 + (16 – 8) / 9)*(0.9)*
2. *Me = (6.95 +(8/9)*(0.9)) = 6.95 +(0.88*0.9)*
3. *Me = 6.95 + 0.79*
4. *Me = 7.75 ≈ 7.8*

Ventajas e inconvenientes:

- Es la medida más representativa en el caso de variables que solo admitan la escala ordinal.
- Es fácil de calcular.
- En la mediana solo influyen los valores centrales y es insensible a los valores
- extremos u “outliers ”.
- En su determinación no intervienen todos los valores de la variable.

La Moda (Mo)

La moda es el dato que más se repite o el dato que ocurre con mayor frecuencia.. Un grupo de datos puede no tener moda, tener una moda (unimodal), dos modas (bimodal) o más de dos modas (multimodal).

Veamos los siguientes ejemplos:

a) Se tiene una muestra con valores 20, 23, 24, 25, 25, 26 y 30.

Mo = 25 es unimodal

b) Se tiene una muestra con valores 20, 20, 23, 24, 25, 25, 26 y 30. Mo= 20 y 25, se dice que es bimodal.

c) Se tiene una muestra con valores 20, 23, 24, 25, 25, 26, 30 y 30. Mo= 20, 25 y 30, se dice que es multimodal.

En los datos agrupados la Mo es la marca de clase de la clase que contenga la mayor frecuencia absoluta.

Intervalo de clase	fi	Xc	fi * Xc	fa	Limites reales
5.2 - 6.0	3	5.6	16.8	3	5.15 – 6.05
6.1 – 6.9	5	6.5	32.5	8	6.05 – 6.95
7.0 – 7.8	**9**	**7.4**	**66.6**	**17**	**6.95 – 7.85**
7.9 – 8.7	7	8.3	58.1	24	7.85 – 8.75
8.8 – 9.6	5	9.2	46.0	29	8.75 – 9.65
9.7 – 10.5	3	10.1	30.3	32	9.65 – 10.55
Total	**32**		**250.4**		

Mo = 7.4

Tambien se puede calcular a traves de la formula:

$$M_o = L_{ir} + \left[\frac{d_1}{d_1 + d_2} \right] i$$

donde

Lir: limite inferior verdadero de la clase modal.

$$\mathbf{d_1=(f_j - f_{j-1}) \qquad Y \qquad d_1=(f_j - f_{j+1})}$$

- fi es la frecuencia absoluta de la clase modal.
- fi-1 es la frecuencia de clase absoluta anterior a la clase modal
- fi+1 es la frecuencia de clase absoluta posterior a la de la clase modal.
- i es el intervalo de clase.

La clase modal es aquella que contiene la mayor frecuencia absoluta.

Intervalo de clase	fi	Xc	fi * Xc	fa	Limites reales
5.2 - 6.0	3	5.6	16.8	3	5.15 – 6.05
6.1 – 6.9	5	6.5	32.5	8	6.05 – 6.95
7.0 – 7.8	**9**	**7.4**	**66.6**	**17**	**6.95 – 7.85**
7.9 – 8.7	7	8.3	58.1	24	7.85 – 8.75
8.8 – 9.6	5	9.2	46.0	29	8.75 – 9.65
9.7 – 10.5	3	10.1	30.3	32	9.65 – 10.55
Total	**32**		**250.4**		

d1 = 9 – 4 = 4

d2 = 9 – 7 = 2

1. ***Mo=6.95+(4/4+2)*0.9 =6.95+(4/6)*0.9=6.95+0.66*0.9***
2. ***Mo = 6.95 + 0.59***
3. ***Mo = 7.55 ≈ 7.6***

- Es mejor utilizar la formula para el calculo de la moda.
- Ventajas e inconvenientes:
- Su cálculo es sencillo.
- Es de fácil interpretación.
- Es la única medida de posición central que puede obtenerse en las variables de tipo cualitativo.
- En su determinación no intervienen todos lo valores de la distribución.