

Modulo 3 ACT. Parte nº 8.
Tema 5:
Estadística descriptiva e inferencial
aplicada al entorno cotidiano

ÍNDICE

1) INTRODUCCIÓN.

2) CONCEPTOS.

3) ESTUDIO ESTADÍSTICO.

3.1. Recogida de datos.

3.2. Organización de los datos.

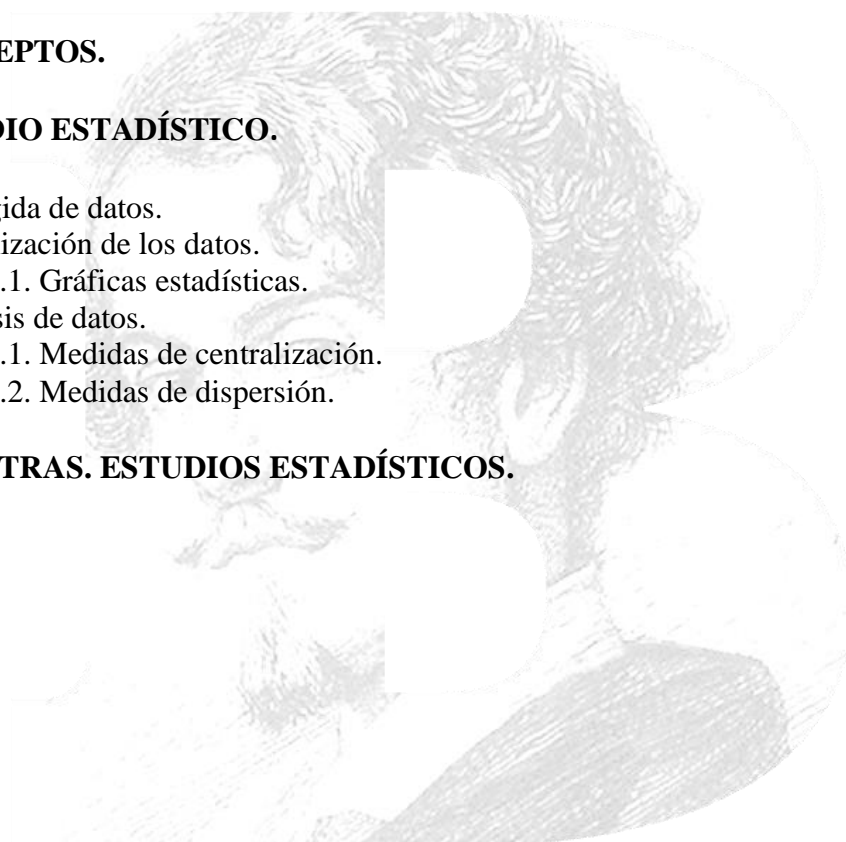
3.2.1. Gráficas estadísticas.

3.3. Análisis de datos.

3.3.1. Medidas de centralización.

3.3.2. Medidas de dispersión.

4) MUESTRAS. ESTUDIOS ESTADÍSTICOS.



1) LAS FRACCIONES.

La **estadística** es una ciencia matemática especializada en el **análisis de grandes volúmenes de información** para de ella extraer conclusiones. Tras analizar los datos se deducen determinadas características de dicha información.

También se podría decir que la **estadística** trata del recuento, la ordenación y clasificación de datos obtenidos por las observaciones, para poder hacer comparaciones y sacar conclusiones.

Un **estudio estadístico** consta de las siguientes fases:

- Recogida de datos.
- Organización y representación de datos.
- Análisis de datos.
- Obtención de conclusiones.

2) CONCEPTOS.

En un estudio estadístico distinguimos:

- ❖ **Población**: es el conjunto de todos los elementos sobre los cuales se va a estudiar una determinada característica.
Por ejemplo: si vamos a analizar la estatura media de los españoles, la población sería todos los ciudadanos españoles.
 - ❖ **Muestra**: del total de la población se selecciona un grupo representativo que es el que vamos a estudiar.
Por ejemplo: para analizar la estatura media de los españoles no podemos recoger esta información de todos los ciudadanos españoles, sino que tenemos que definir un grupo de estudio, por ejemplo, seleccionar a 2.000 personas. Este grupo tiene que ser representativo de la sociedad española, por lo que tiene que incluir a hombres y mujeres, gente de la ciudad y del campo, gente de diversos niveles de renta, de diversas edades, etc. Es decir, la muestra tiene que ser como una imagen “en miniatura” de la población.
 - ❖ **Variable estadística**: el aspecto que se va a estudiar. Si se puede medir se llama **variable cuantitativa** (por ejemplo: altura, peso...).
Si la variable estadística toma un número determinado de valores se llama **variable discreta**. Por ejemplo: números de años en el colegio (de 1 a 15).
Si la variable estadística puede tomar cualquier valor entre dos valores dados se llama **variable continua**. Aquí la variable puede tomar un número casi ilimitado de valores. Por ejemplo: estatura (... 161 cm, 162 cm, 163 cm...)
- Si no se pueden medir se llama **variable cualitativa** (por ejemplo: sexo, color de pelo...).
- ❖ **Valor**: es cada uno de los distintos resultados que se pueden obtener en un estudio estadístico.

3) ESTUDIO ESTADÍSTICO.

Una vez definidas las variables que vamos a estudiar y la muestra que vamos a analizar, hay que comenzar por obtener la información. Para realizarlo, debemos seguir los siguientes pasos:

3.1. Recogida de datos.

Planteado el test o encuesta oportuno, una vez elegido el tema al que se quiere hacer el estudio estadístico, y recogidos los datos que correspondan, el primer análisis que realizaremos es el del tipo de variable que pretendemos estudiar (**cuantitativa o cualitativa; discreta o continua**).

Esto condicionará en gran medida su posterior tratamiento.

3.2. Organización de los datos.

Determinado el modo de agrupamiento de las observaciones, procedemos a su **RECuento**. Los datos obtenidos en el punto anterior hay que ordenarlos y recogerlos en una tabla que se denomina **tabla estadística** o **tabla de frecuencias**.

Posteriormente podremos visualizar tales frecuencias de forma gráfica con el diagrama estadístico apropiado.

TABLA DE FRECUENCIAS:

Una tabla de frecuencias es una **ordenación** en forma de tabla de los **datos estadísticos**, asignando a cada dato su **frecuencia correspondiente**. Las columnas mínimas que debe tener una tabla de frecuencias son las siguientes:

- ❖ **Datos:** son los diferentes valores de la variable que se van a estudiar. Deben de estar ordenados.
- ❖ **Frecuencia absoluta:** es el número de veces que aparece un determinado valor en un estudio estadístico. Se representa por **f_i**.
La suma de las frecuencias absolutas es igual al número total de datos, que se representa por N. Es decir: $N = \sum f_i$.
- ❖ **Frecuencia absoluta acumulada:** es la suma de la frecuencia absoluta de todos los valores inferiores o iguales al valor considerado. Se representa por **F_i**.
El último valor de la frecuencia absoluta acumulada es el valor de la muestra (N).
- ❖ **Frecuencia relativa:** es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos. Se representa por **n_i** y se calcula como $n_i = \frac{f_i}{N}$.

La suma de las frecuencias relativas es igual a 1.

Si multiplicamos este valor obtenido por 100, lo podemos expresar en % (frecuencia relativa porcentual). La suma de las frecuencias relativas porcentuales es 100.

- ❖ **Frecuencia relativa acumulada:** es la suma de la frecuencia relativa de todos los valores inferiores o iguales al valor considerado. Se representa por N_i .
El último valor de la frecuencia relativa acumulada es 1.
El último valor de la frecuencia relativa porcentual acumulada es 100.

3.2.1. Gráficas estadísticas.

Las gráficas estadísticas permiten visualizar la información contenida en las tablas de manera rápida y sencilla.

Existen muchos tipos de gráficas estadísticas. Unas se emplean con variables cuantitativas y otras con variables cualitativas.

DIAGRAMA DE BARRAS:

Se utiliza para presentar **datos cualitativos** o **datos cuantitativos de tipo discreto**.

Se representa sobre unos ejes de coordenadas, en el eje de abscisas (eje x) se colocan los valores de la variable, y sobre el eje de ordenadas (eje y) las frecuencias absolutas, relativas, porcentuales o frecuencias acumuladas.

Los datos se representan mediante barras de una altura proporcional a la frecuencia.

POLÍGONO DE FRECUENCIAS:

Se realiza para **cualquier tipo de variable**. Es el polígono que se forma al unir los puntos medios de las barras, tanto en histogramas como en diagrama de barras.

DIAGRAMA DE SECTORES:

Cada sector es proporcional al porcentaje que representa. Los grados de cada sector son:
 $Grados = 360 \times n_i$.

PICTOGRAMA:

Es un gráfico con figuras.

PIRÁMIDE DE POBLACIÓN:

Consiste en dos histogramas, uno para hombres y otro para mujeres, correspondientes a habitantes de una misma comunidad más o menos extensa, repartidos por edades.

Es útil para estudiar su situación demográfica y buscar explicaciones a situaciones presentes, pasadas y futuras.

CLIMOGRAMA:

Son gráficas que representan la distribución de precipitaciones y temperaturas a lo largo de un año en un lugar determinado.

3.3. Análisis de datos.

Para este análisis se utilizan los llamados parámetros estadísticos. Son los siguientes:

- ❖ **Medidas de centralización:** media, mediana y moda.
- ❖ **Medidas de dispersión:** recorrido, desviación media, varianza, desviación típica y coeficiente de variación.

3.3.1. Medidas de centralización.

MEDIA ARITMÉTICA:

La media aritmética es el valor obtenido al sumar todos los datos y dividir el resultado entre el número total de datos.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{N} = \frac{\sum_{i=1}^n x_i}{N}$$

Si los datos vienen en una tabla de frecuencias, la expresión de la media es:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_n f_n}{N} = \frac{\sum_{i=1}^n x_i f_i}{N}$$

Evidentemente esta medida sólo se puede hallar para variables cuantitativas.

MODA:

Es el valor que tiene mayor frecuencia absoluta. Se representa por Mo.

Se puede hallar para cualquier tipo de variable, aunque para variables cuantitativas es poco útil.

La moda de la distribución 2, 3, 3, 4, 4, 4, 5, 5 es Mo = 4.

Si en un grupo hay dos o varias puntuaciones con la misma frecuencia, y esa frecuencia es la máxima, la distribución es bimodal o multimodal, es decir, tiene varias modas.

1, 1, 1, 4, 4, 5, 5, 5, 7, 8, 9, 9, 9; Mo = 1, 5, 9.

MEDIANA:

Es el valor que ocupa el lugar central de todos los datos cuando éstos están ordenados de menor a mayor. La mediana se representa por Me y se puede hallar sólo para variables cuantitativas.

Cálculo de la mediana con pocos datos:

- I. Ordenamos los datos de menor a mayor.
- II. Si la serie tiene un número impar de medidas, la mediana es la puntuación central de la misma.
2, 3, 4, 4, 5, 5, 5, 6, 6; Me = 5
- III. Si la serie tiene un número par de medidas, la mediana es la media entre las dos puntuaciones centrales.
7, 8, 9, 10, 11, 12; Me = 9.5

3.3.2. Medidas de dispersión.

Las medidas de dispersión sirven para comparar dos o más distribuciones y decidir cuál de ellas es más o menos dispersa.

RECORRIDO O RANGO:

Es la diferencia entre los valores extremos, es decir, entre el mayor valor y el menor.

Recorrido = Valor mayor – Valor menor

DESVIACIÓN MEDIA:

Es un parámetro asociado a la media; y es el promedio (o media) de las distancias de los valores de todos los individuos a la media.

$$DM = \frac{\sum |x_i - \bar{x}| \cdot f_i}{N}$$

VARIANZA:

Sirve para identificar si los datos están cercanos a la media o no. Se calcula sumando los valores que se obtienen al elevar al cuadrado la diferencia de cada dato con la media, y dividiendo este valor entre el número de datos.

$$\sigma^2 = \sum \frac{x_i^2 \cdot f_i}{N} - \bar{x}^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{N}$$

DESVIACIÓN TÍPICA:

Da un valor de las diferencias de los valores con respecto a la media, que se obtiene haciendo la raíz cuadrada de la varianza.

MÓDULO 3 ACT

Parte nº 8: Funciones como modelos de situaciones cotidianas, registro e inferencia sobre las mismas.

Tema 5: Estadística descriptiva e inferencial aplicada al entorno cotidiano.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum \frac{x_i^2 \cdot f_i}{N} - x^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{N}}$$

COEFICIENTE DE VARIACIÓN:

Se usa para comparar las dispersiones de dos distribuciones heterogéneas.

$$CV = \frac{\sigma}{\bar{x}}$$

Al calcular el coeficiente de variación estamos relativizando la dispersión. El resultado se da, a veces, en tanto por ciento.

4) MUESTRAS. ESTUDIOS ESTADÍSTICOS.

Si queremos hacer un estudio estadístico tenemos que:

- a) Recoger los datos.

Para recoger los datos y determinar los valores de la variable se puede utilizar a toda la población, todo el universo sobre el que se realiza el estudio, o hacer una muestra. En muchas ocasiones no es conveniente recoger valores de toda la población, porque es complicado o demasiado costoso, o incluso porque es imposible como en el caso de un control de calidad en que se destruya el objeto a analizar.

La parte de la estadística que se ocupa de cómo seleccionar adecuadamente las muestras se denomina Teoría de muestras, cuyos conceptos básicos son:

- Población o universo es todo el conjunto de individuos sobre el que se realiza el estudio.
- Una muestra es un subconjunto representativo de esa población.
- Cada uno de los elementos de la población es un individuo.

- b) Describir esos datos con tablas y gráficas, cálculo de parámetros estadísticos...

Las características de la población que se estudian se denominan variables estadísticas, que se clasifican en cuantitativas y cualitativas según que los valores que tomen sean o no numéricos.

- Las variables cuantitativas que toman valores aislados se denominan variables discretas. Un ejemplo es el número de calzado.
- Las variables que pueden tomar cualquier valor de un intervalo de la recta real, se denominan variables continuas. Por ejemplo la altura de una persona.

La parte de la estadística que ordena, analiza y representa un conjunto de datos para describir sus características se denomina Estadística descriptiva.

MÓDULO 3 ACT

Parte nº 8: Funciones como modelos de situaciones cotidianas, registro e inferencia sobre las mismas.

Tema 5: Estadística descriptiva e inferencial aplicada al entorno cotidiano.

c) Extraer conclusiones.

Para extraer conclusiones se utilizan las probabilidades, y la parte de la estadística que se ocupa de ello es la Inferencia estadística.

