La estadística es la que recolecta y organiza los datos acerca de personas, sucesos o casos. Para esto tenemos la matriz de datos que es una disposición de números o símbolos donde cada fila horizontal representa a un individuo que posee la información de interés y cada columna es un aspecto del individuo que se ha seleccionado para estudiar. Este individuo no es necesariamente es una persona, pueden ser lugares, casos o entidades colectivas. Se las llama unidades de análisis a estos entes individuales que se analizan sus cualidades. Y la población es el conjunto de elementos, individuos o unidades de análisis que presentan una o más características o atributos que son objeto de un estudio particular. La estadística existe por la variabilidad, que genera el concepto de variable que es el aspecto de la columna de la matriz. Este aspecto puede representar distintas modalidades. La variable es una característica de los individuos o unidades de análisis de una población que puede presentar distintos valores o categorías.

Una variable puede ser directamente observable o no. Esto quiere decir que la variable no observable es una variable latente que es parte de un constructo, el cual es una construcción teórica, hipotética sobre un rasgo psicológico y variable compleja. Esto quiere decir que estos constructos no son directamente observables, pero pueden manifestarse a través de ciertas características observables que son susceptibles de medición. Para que este constructo sea abordado se necesita registrar sus manifestaciones observables a partir de su definición operacional. Esto quiere decir que se lo operacionaliza para tener una variable observable y de esta manera puede ser medido. Para esto indicamos los procedimientos para obtener el valor de la variable que corresponde a cada individuo.

La cualidad central de una variable es que puedan variar y que sus valores sean asignados a través de un proceso de medición. La medición consiste en asignar propiedades a los objetos medidos. Estas medidas tienen escalas, estas tienen ciertas propiedades de las que depende el significado y determinan nivel de medición. El nivel de medición de una variable depende de las propiedades de la escala y esta determinado por el significado que tengan los símbolos numéricos que se asignan a las categorías.

De acuerdo con los tipos de valores que pueden tomar las variables, se las puede clasificar en cuantitativas (discretas o continuas) o cualitativas. Las cualitativas son las variables cuyos valores expresan atributos o cualidades. Las variables cuantitativas son aquellas cuyos valores expresan cantidades numéricas. Las cuantitativas pueden ser discretas, esto quiere decir que los valores que puede tomar la variable son acotados, no puede tomar cualquier valor. No toman todos los valores y estos valores se pueden enumerar. En cambio, las continuas, son las que sus valores posibles constituyen un intervalo de números reales, puede ser cualquier numero real.

La medición es el proceso por el cual se asignan números a las características, conforme a ciertas reglas de modo que los números asignados reflejan las propiedades de los objetos medidos. La medición de las variables presenta niveles de medición que son jerarquías, reúnen propiedades del nivel anterior y agrega otra propiedad nueva. Estos niveles son:

* **Nivel nominal**: es el nivel más elemental de medición. Sus valores vislumbran atributos y solo cumplen con la condición de establecer igualdades y diferencias. (sexo, estado civil, nacionalidad)
* **Nivel ordinal**: es el nivel en el que sus valores cumplen con la condición de poder establecer igualdades y diferencias, y a parte esos valores pueden ser ordenados de forma jerárquica. (nivel socioeconómico, grado de satisfacción)
* **Nivel intervalar**: es el nivel en el que sus valores cumplen con la condición de poder establecer igualdades y diferencias, se pueden ordenar jerárquicamente y a su vez se pueden realizar sumas y restas. Su cero es arbitrario y no implica ausencia del atributo que se esta midiendo. (tiempo en C°, nota de un examen, año del calendario)
* **Nivel de razón**: es el nivel en el que sus valores cumplen con la condición de poder establecer igualdades y diferencias, se pueden ordenar jerárquicamente y a su vez se pueden realizar sumas y restas, multiplicaciones y divisiones. Su cero es absoluto porque lo que representa o implica ausencia de la característica. (edad, tiempo en minutos, respuestas correctas)

DISTRIBUCIÓN DE FRECUENCIAS

Una vez expuestos los datos, es necesario darle a la matriz un formato que permita hacer lectura de los resultados, para esto existe la **distribución de frecuencias**.

* **Frecuencia absoluta**(fi): numero de casos que asume ese valor, la cantidad de veces que ese valor se repite en el conjunto de datos.
* **Frecuencia relativa**(f’i): proporción de casos que asumen ese valor, indico la proporción de casos que se encuentran en una categoría. f:n= sumatoria=1 (frecuencia absoluta dividido la muestra y a suma de todos los f’i tienen que dar 1).
* **Frecuencia porcentual**(f%i): el porcentaje de veces que ese valor esta en el conjunto de datos. f’x100->sumatoria= 100 (frecuencias relativas por 100 y la suma de todos los f%i tiene que dar 100).
* **Frecuencia absoluta acumulada**(F): es la cantidad de casos que asumen ese valor y todos los valores menores a él. Se acumulan con las absolutas último número.
* **Frecuencia relativa acumulada**(F’): es la proporción de casos que asumen ese valor y todos los valores menores a él.

GRÁFICOS | NIVEL DE MEDICIÓN | TIPO DE VARIABLE

Gráfico de **sectores o diagrama circular/de teoría** 🡪 sirve para variables cualitativas y nivel nominal.

Gráfico de **barras o rectángulos** 🡪 sirve para variables cualitativas y nivel nominal y ordinal.

Gráfico de **bastones** 🡪 sirve para variables cuantitativas discretas.

Gráfico de **tallo y hoja** 🡪 sirve para variables cuantitativas discretas y continuas.

Gráfico de **caja y bigotes** 🡪 sirve para variables cuantitativas discretas y continuas.

Gráfico **escalonado** 🡪 sirve para variables cuantitativas discretas.

**Histograma, polígono de frecuencias y ojiva de Galton** 🡪 cuantitativa continua y nivel intervalar y razón.

+ comparación de variable bivariados

🡪adyacentes= 2 variables cualitativas

🡪de dispersión= 2 variables cuantitativas

MEDIDAS DE TENDENCIA CENTRAL: LA MODA

Estas representan a la totalidad de las observaciones. La **moda** es el valor de la variable observando con más frecuencia. Si hay dos valores se la llama bimodal.

En el caso en el que la variable sea cuantitativa y las dos modas sean adyacentes se toma como moda el promedio de esos dos valores adyacentes.

Advertencia: no confundir la moda con la mayor frecuencia. La moda es el nombre de la variable, no la cantidad.

La moda puede usarse en todos los niveles de medición.

MEDIDAS DE TENDENCIA CENTRAL: LA MEDIANA

Es el valor de la variable que deja por debajo a la mitad del total de las observaciones. n/2 y acumuladas, nivel ordinal, intervalar y de razón. También es el percentil P50 (q2 cuartil 2).

No mas de la mitad de las observaciones es menor a la mediada y no mas de la mitad de las observaciones es mayor a la mediana. Es el valor que ocupa la posición central.

Que pasa si la cantidad de observaciones es par, en las variables ordinales si coinciden las dos observaciones que “ocuparían la mediana” entonces ese valor es la mediana. Si no coinciden no hay mediana. Ahora, si son cuantitativas realizo un promedio entre los dos valores y esa es mi mediana.

Para poder calcular la mediana se debe calcular las frecuencias acumuladas.

MEDIDAS DE TENDENCIA CENTRAL: LA MEDIA

Es el valor de la variable obtenido sumando todas las observaciones multiplicadas por su frecuencia absoluta y dividiendo el resultado por la muestra. Si solo están los datos, sumarlos y dividir por la cantidad. f x variable= total/n (parecido a promedio).

Se utiliza en los niveles de razón e intervalar.

La medida no necesariamente coincide con alguno de los valores de la variable. La modificación de cualquier valor modifica la media.

Propiedades de la media aritmética:

1. Las observaciones pueden apartarse de la media, ya sea por defecto como por exceso; pero estos desvíos se compensan, suman 0. Esto permite interpretar a la media como el “centro de equilibrio de la distribución”. Ejemplo tabla con pesos. (distancia de los valores a la media)
2. La suma de los cuadrados de las desviaciones de las puntuaciones con respecto a su medida es menor que con respecto a cualquier otro valor.
* El calculo de la media solo tiene sentido para resumir datos que pertenecen a un nivel intervalar.

MEDIDAS DE POSICIÓN LOS PERCENTILES

Los percentiles son valores que se definen en función del porcentaje de observaciones a las que superan. El percentil r (Pr) de una distribución es el valor que deja a lo sumo el r porciento de los casos por debajo de él y como máximo el (1-r) por ciento de los casos por encima. En otras palabras, refiere a los valores de la variable que deja por debajo (acumuladas) distintos porcentajes de casos.

El subíndice r representa el porcentaje de observaciones superadas por P y se denomina rango percentilar.

MEDIDAS NO CENTRALES

Los percentiles mas usados son los cuartiles. El cuartil es el valor que dejo por debajo de un cuarto de casos o nombres el que dejo por debajo las tres cuartas partes de las observaciones.

1° cuartil🡪Q1: es el valor de la variable que supera a lo sumo el 25% de las observaciones. equivale al percentil 25 (P25)

2° cuartil🡪Q2🡪Mediana: es un valor de la variable que supera a lo sumo al 50% de las observaciones. equivale al percentil 50 (P50) también llamada mediana.

3° cuartil🡪Q3: es un valor de la variable que supera a lo sumo al 75% de las observaciones. equivale al percentil 75 (P75)

Como calcular los cuartiles:

1°cuartil: n\*(1/4)

2°cuartil: n\*(2/4)

3°cuartil: n\*(3/4)

Si el resultado da numero entero, entonces, es la media de los números en las posiciones n\*(1º2º3/4) y n\*(1º2º3/4)+1.

Si el resultado no es numero entero, entonces, se redondea hacia arriba y esa es la posición del cuartil.