

TIPOS:

1.- Descriptiva: La que describe las características de las muestras.

2.- Inferencial o analítica: La que analiza esos datos buscando inferencias. Una inferencia es, a partir de unos datos obtenidos de una pequeña muestra, predecir el comportamiento de la mayor parte de ella y fijar un margen de error en la predicción.

CONCEPTOS:

- Población: el total del grupo a estudio. Por su gran tamaño, normalmente no puede ser estudiado pues sería muy costoso en tiempo y dinero.
- Muestra: Como la población es difícil estudiarla, extraemos un subconjunto de ella para poder hacer el estudio, que debe ser representativa (reunir todas las características de la población).

VARIABLES:

Son aquellas características que estudiamos, y que como su nombre indica, son cambiantes (variables).

Las clasificamos:

- Según su relación:
 - o Independientes: no influidas por nada, es la causa que estudiamos.
 - o Dependientes: influidas de algún modo por un suceso (la VI). Es el resultado que observamos.
- Según su naturaleza:
 - o Cualitativa: Cuando las características de las variables se refieren a cualidades no medibles numéricamente de los sujetos. A su vez se subdividen en:
 - Cualitativas puras: no se pueden ordenar, es decir, cualquier aspecto de la variable tiene el mismo valor. Si una variable cualitativa solo puede admitir dos valores, se le llama DICOTÓMICA (si/no, hombre/mujer, verdadero/falso).
 - Cualitativas ordinales: Las distintas modalidades que puede adquirir la variable, se pueden ordenar de una forma lógica.
 - o Cuantitativas: Cuando las características de las variables se refieren a cualidades medibles numéricamente de los sujetos. A su vez se dividen en:
 - Cuantitativas discretas: Sus modalidades solo admiten números enteros (n° de hijos)
 - Cuantitativas continuas: sus modalidades admiten cualquier número (decimales, etc. Por ejemplo colesterol en sangre medido en mg/dl, T^a corporal).

ESCALAS DE MEDIDA:

Es la forma de ordenar o medir la variable, el tipo de medida que elijo. Tipos de escala de medida:

- Nominales: Corresponde a la medida de variables cualitativas puras, como no tiene orden tenemos que citar a todos los individuos (estado civil, sexo...)
- Ordinales: Hay un orden establecido previamente, generalmente del tipo mayor que, igual que.
- De intervalo: el cero es un valor que tiene significado en sí mismo. Admite, por tanto, valores negativos (T^a).
- De razón: El cero es absoluto: longitud, peso.

REPRESENTACIÓN GRÁFICA DE LAS VARIABLES:

- Para las cualitativas:

- Diagrama de barras (generalmente ordinales). En el eje de abcisas (horizontal) los valores de la variable, en el eje de ordenadas (vertical) las frecuencias.



- Diagramas de sectores (tartas).

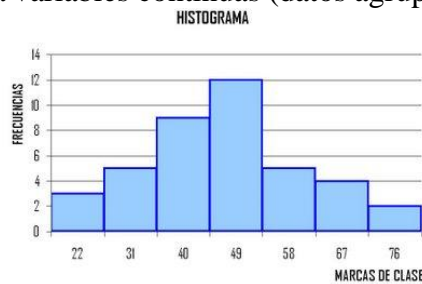


- Pictogramas: dibujo de la forma de la variable, y el tamaño del dibujo representa la frecuencia.

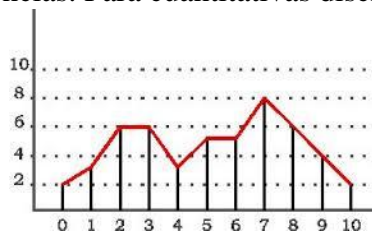


- Para las cuantitativas:

- Histograma. Para variables continuas (datos agrupados en intervalos)



- Polígono de frecuencias. Para cuantitativas discretas.



ERROR MUESTRAL.

Para que nuestro estudio tenga validez, la muestra dijimos que debía ser representativa de la población de la que es referencia, lo que origina problemas de selección muestral. Aplicaremos técnicas de muestreo, pero siempre pueden aparecer errores.

Tipos de errores:

- Aleatorio: aparece siempre en cualquier estudio, es inevitable. Es causado por el azar. La única forma de reducirlo es aumentar la muestra, pero no podemos aumentarlo demasiado porque no podemos acercarnos a la población, que lo convierte en costoso.
- Sistemático: son los errores de diseño y planificación. Son, por tanto, evitables.

ESTADÍSTICA DESCRIPTIVA:

La estadística descriptiva busca describir como son las muestras. Estudia variables y las detalla, las expone para que todos sepamos como son, sus características. Lo lleva a cabo de forma estadística, por lo que nos interesan determinados valores:

- Número de individuos y frecuencia de los sucesos.
- Tendencia central.
- Dispersión.

MEDIDAS DE TENDENCIA CENTRAL.

1.- **MEDIA:** suma de las observaciones partido por el número de las mismas.

Ventajas: apropiada para manipulaciones matemáticas. Normalmente es la más representativa, la mejor, por lo que es la que se utiliza en la mayoría de las mediciones.

Inconvenientes: muy afectada por valores extremos.

2.- **MEDIANA:** valor que deja el 50% de las observaciones por encima y el otro 50% por debajo.

Ventaja: no influida por valores extremos.

Inconveniente: es menos representativa que la media, y si el valor es constante, no se puede calcular.

Si hay un número par de datos, se saca la media de los dos centrales.

3.- **MODA:** El valor que más se repite.

Ventaja: muy fácil de calcular.

Inconveniente: a veces no hay (amodal) o hay más de una (bimodal, trimodal...)

Si es bimodal, y las dos modas son adyacentes, la moda es la media de los dos.

4.- **PERCENTILES, DECILES, CUARTILES...**: Indican la posición de un sujeto en el grupo.

Es, pues, la proporción de observaciones que caen dentro de unos valores específicos.

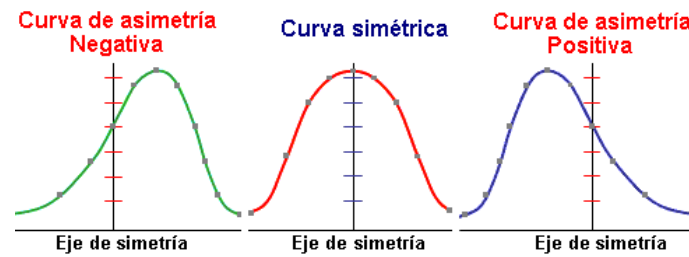
Centiles o percentiles: divide la muestra en 100 partes, cada una de ellas es el 1% de la muestra.

Cuartiles: divide la muestra en 4 partes, cada una de ellas con el 25% de las observaciones.

OJO: el percentil 50, el cuartil 2, y el decil 5, coinciden con la median cuando la distribución es normal.

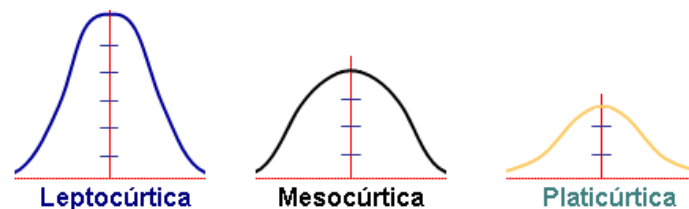
ASIMETRÍA O SESGO: Cuando no hay una curva normal (los datos no se concentran en un valor medio).

- Curva normal
- Curva con asimetría positiva: más sujetos puntúan alto.
- Curva con asimetría negativa: más sujetos puntúan bajo.



CURTOSIS: es el grado de apuntamiento de la distribución de frecuencias.

- Curva normal o mesocúrtica.
- Curva leptocúrtica: con mucho apuntamiento, tendencia a la verticalidad.
- Curva platicúrtica: con poco apuntamiento, tendencia a ser plana.



COMPARACIÓN DE LAS MEDIDAS DE TENDENCIA CENTRAL:

- Se preferirá siempre la media a cualquiera del resto.
- Cuando se elegirá la mediana antes que la media:
 - o Si la variable es ordinal.
 - o Si hay valores extremos.
 - o Cuando haya intervalos abiertos (el intervalo superior o inferior no tiene límites).
- Cuando se elegirá la moda:
 - o Si la variable es nominal.
 - o Cuando haya intervalos y la mediana esté incluida en uno de ellos).

MEDIDAS DE DISPERSIÓN:

Sirven para medir la variabilidad, es distinto de las medidas de tendencia central. Nos dice: ¿Cuánto se alejan las puntuaciones respecto a las medidas centrales).

Ejemplo:

País rico: renta per cápita 3000€/media/mes.

País pobre: renta per cápita 3000€/media/mes.

Pero todos sabemos que no es lo mismo. En el país pobre la media es la misma, pero solo hay un 5% de población rica. Las medidas de tendencia central no dicen, pues, toda la verdad.

Necesitamos complementarla con las **medidas de dispersión**, para ver cual es el reparto real.

1.- **RANGO:** Amplitud entre el valor menor y el último (el mayor). Es la resta entre el 1er valor y el último. Muy afectado por valores extremos.

2.- **VARIANZA (S):** sumatorio de la diferencia de nuestra observación respecto a la media, elevado al cuadrado (al elevar al cuadrado conseguimos eliminar media=0 o valores negativos) y dividido por n.

3.- DESVIACIÓN TÍPICA (DT): Raíz cuadrada de la varianza. Mejor descriptor de la variabilidad que la varianza. Se utiliza más porque la varianza tiene el problema de que al utilizar cuadrados, trabaja con cifras elevadas, que no “parecen” ser representativas; no se asocian bien. Al aplicar la raíz cuadrada de S^2 , se retorna a las cifras originales.

4.- DESVIACIÓN MEDIA (DM): en vez de utilizar el cuadrado para eliminar la media=0, en el caso de la varianza, utilizamos valores absolutos.

5.- AMPLITUD SEMIINTERCUARTIL: $Q=$

ERROR ESTÁNDAR (ES) de la media: a partir de la desviación típica, podemos hacer una inferencia sobre cual será la desviación de la población. $ES=S/\text{raíz cuadrada de } n$.

$A > n^\circ$ de datos, $< ES$, porque el nivel de confianza es $> a > n$.

PUNTUACIÓN TÍPICA: N° de desviaciones típicas (DT) que una observación se separa de la media del grupo de observaciones. ¿Para qué sirve? Nos sirve para comparar grupos distintos.

LA CURVA NORMAL.

Conforme aumenta el tamaño muestral, observamos que al representarla gráficamente se distribuye de una forma característica (curva normal o gaussiana).

Características:

- Media, mediana y moda coinciden.
- Es simétrica con respecto a la media, y asintótica (la curva se acerca a los ejes, pero nunca llega a tocarlos (en \pm infinito).
- La media:
 - o ± 1 DT: comprende el 65% de los individuos.
 - o ± 2 DT: comprende el 95% de los individuos.
 - o ± 3 DT: comprende el 99% de los individuos.

NIVELES DE CONFIANZA.

Estadística inferencial. Utilizamos los niveles de confianza cuando queremos inferir parámetros de la muestra a la población.

El número mínimo de muestra (n) debe ser ≥ 30 (teorema del límite central), para que cualquier muestra sea representativa. Si es menor, aplicaremos medidas paramétricas especiales (t de student).

Niveles de confianza aceptable:

95%: $Z=1,96$

99%: $Z=2,58$

X (MEDIA) $\pm Z * ES$ (ERROR STANDARD)

Para calcular inferencias, necesito predeterminar el error (Standard). Lo que buscamos es el error respecto a la media (ver fórmula).

Por tanto, cuanto mayor es la muestra, mayor es el dividendo y menor es el error Standard. Cuando n tiende a infinito, el error estándar tiende a cer

Cuando trabajo con un nivel de confianza del 95%, significa que busco niveles de significación (diferencia estadísticamente significativa).

El nivel mínimo de confianza se ha prefijado por consenso en el 95% ($p=0,05$), y a ello se le llama nivel de significación alfa.

Cuando en las conclusiones de un estudio, determino que he encontrado una diferencia significativa, estoy asegurando que los resultados obtenidos solo se darían por azar con una probabilidad de $p=0,05$ o lo que es lo mismo, el azar solo modificaría el 5% de los casos.

Por tanto, para estimar el valor de un parámetro a nivel poblacional, debemos:

- 1- Fijar el nivel de confianza.
- 2- Calcular la media y la desviación típica de nuestra muestra.
- 3- Calcular el error estándar.
- 4- Construir el intervalo de confianza.

CONTRASTE DE HIPÓTESIS.

Empleamos el contraste de hipótesis para conocer hasta qué punto nuestros resultados pueden ser debidos al azar, o bien son fruto de una verdadera diferencia.

Una hipótesis estadística es un enunciado que se piensa acerca de parámetros poblacionales.

La hipótesis nula la simbolizamos como H_0 . La hipótesis alternativa se simboliza H_1 .

Hipótesis nula: no hay diferencia entre la media de la población y el valor teórico supuesto. Nula significa que no hay diferencia. Es decir, afirma que en la población no existe asociación entre las variables predictoras y el desenlace.

Ej: no hay diferencia en el hábito de consumir café de pacientes con cáncer de páncreas y sus controles.

Siempre partimos de que no hay asociación; por ello, para aceptar una hipótesis alternativa, debo ser capaz de desmentir estadísticamente la H_0 .

Hipótesis alternativa: sí hay diferencia entre la media de la población y el valor supuesto. Es decir, proponemos que existe una asociación entre variable predictoras y desenlace.

Ej: los pacientes con cáncer de páncreas tienen diferentes hábitos de tomar café que los controles.

ERROR TIPO ALFA O ERROR TIPO I.

Cuando en un estudio concluyo que hay diferencia estadísticamente significativa, cuando en realidad no la hay. Es decir, elijo equivocadamente la H_1 (alternativa) cuando la correcta es la H_0 (nula).

ERROR TIPO BETA O TIPO II.

Cuando concluyo que no hay diferencias significativas, cuando en realidad sí las hay.

CORRELACIONES.

Son cuantificaciones estadísticas de las relaciones entre 2 o más variables. Estudiamos relaciones lineales.

- Coeficiente de correlación de Pearson: interrelación entre dos variables cuantitativas dependientes. Puede ser relación directa o inversa. El resultado es un resultado entre +1 y -1 (relación perfecta). Si el resultado es 0, no hay ninguna asociación (excepciones, ej: gráfica del círculo).
- Chi cuadrado: para variables cualitativas (X^2).
- T de student: comparar 2 medias.
- Análisis multivariante (p. ej. ANOVA): interrelación entre múltiples variables.