**¿Qué es la estadística?**

Es una rama de la matemática la cual estudia las formas de extraer conclusiones, inferencias y relaciones de un conjunto de datos.

**Tablas de datos**

Las tablas de datos son necesarias para representar datos de manera ordenada.

**Algunas definiciones:**

**Tabla de datos:** una tabla de datos es un conjunto de datos representada en columnas y filas.

**Individuos:** son los elementos que pertenecen a una población o una muestra. Podemos representar a estos individuos en una tabla de datos en las filas.

**Variables:** en una tabla de datos las variablesson las características que los individuos tienen y se representan en columnas.

**Variables cuantitativas o numéricas:** aquellas que se representan con números.

**Variables cuantitativas continuas:** variables numéricas que pueden tomarse dentro de un rango de números, ese rango puede ser infinito y podemos decir que pertenecen al conjunto de los reales.

**Variables cuantitativas discretas:** variables numéricas las solo aceptan números finitos de un conjunto, podemos verlos como números enteros.

**Variables cualitativas:** aquellas que se expresan con palabras.

**Variables cualitativas ordinales:** aquellas variables **cualitativas** que se puede ordenar por ejemplo; muy mal, mal, normal, bien, muy bien.

**Variables cualitativas nominales:** aquellas variables **cualitativas** que NO se puede ordenar como las **Variables cualitativas ordinales.**

**Variable independiente:** Es la variable que el investigador manipula para observar su efecto sobre otras variables. Se considera la "causa" en la relación causal.

**Variable dependiente:** Es la variable que se mide para evaluar los efectos de la variable independiente. Se considera el "efecto" en la relación causal.

**Grupo experimental:** Es el grupo de participantes en el estudio que está expuesto a la variable independiente o tratamiento que se está probando.

**Grupo de control:** Es el grupo de participantes en el estudio que no está expuesto al tratamiento y se utiliza como referencia para comparar los resultados con el grupo experimental. Ayuda a controlar factores externos que podrían influir en los resultados.

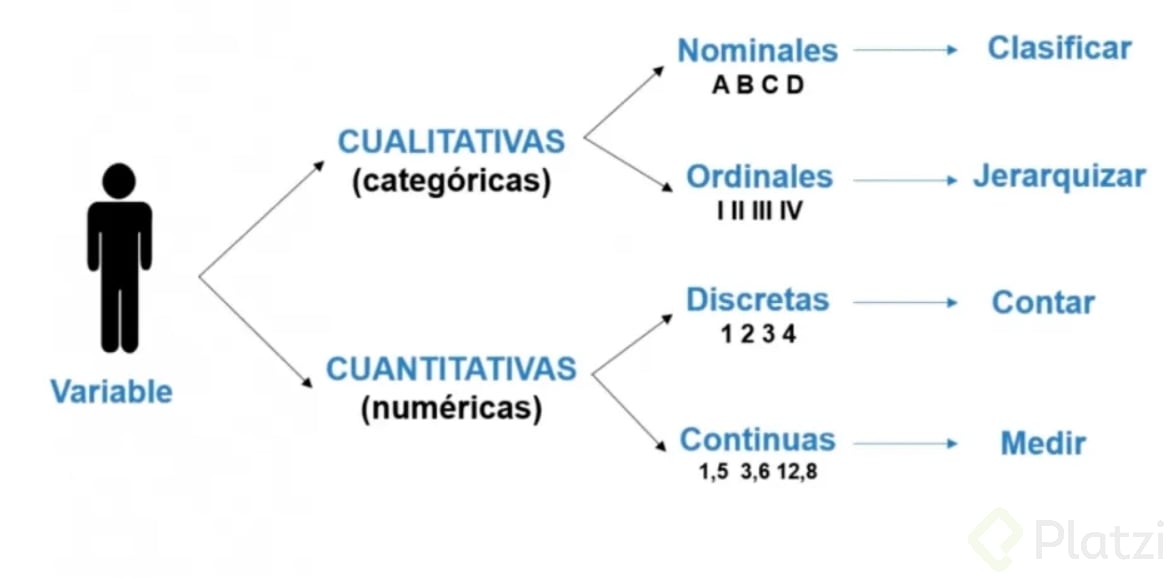
**Asignación aleatoria:** Para reducir el sesgo y garantizar la equivalencia entre los grupos, los participantes se asignan aleatoriamente al grupo experimental o al grupo de control.

**Cegamiento (o enmascaramiento):** Se refiere al ocultamiento de información sobre el tratamiento recibido para evitar que los participantes o los investigadores vean afectados los resultados.

**Diseño experimental:** Es la estructura y organización del estudio, incluidos el número de grupos, el tamaño de la muestra, la duración del estudio y otros aspectos relacionados con la manipulación de variables.

**Análisis estadístico:** Involucra el uso de técnicas estadísticas para analizar los datos recopilados y determinar si hay diferencias significativas entre los grupos experimental y de control.

**VARIABLES**



**NIVELES DE MEDIDA EN ESTADISTICA**

Son las formas en las que podemos clasificar las formas de medir nuestros datos. En general la medición para mí es la forma de darle una jerarquización a un conjunto de elementos.

* **Escala Nominal:** podríamos clasificar en esta categoría a variables que no tienen un orden, y no son variables con las cuales podamos hacer operaciones aritméticas. Además, podemos clasificarlas en dos las dicotómicas y las multicótomicas o policótomicas.
* **Escala ordinal:** las escalas ordinales las podemos interpretar como las variables las cuales podemos ordenar jerárquicamente según la característica que posean por ejemplo; muy frio, frio, tibio, caliente, muy caliente. **En este caso tampoco podemos hacer operaciones aritméticas, si con la variable se pueden hacer operaciones aritméticas como +,-.\*,/ no pertenecería a una escala ordinal.**
* **Escala de intervalos:** podemos clasificar a las escalas de intervalo como las variables consecutivas cuales establezcan intervalos iguales o constantes en la medición, por ejemplo, hay la misma distancia entre 1->2 que entre 6->7, que además también podremos concluir que la diferencia entre estos intervalos es la misma. Una característica de estas escalas es donde el cero no es nulo, o no es “nada”, es decir, el cero solo estaría representando un intervalo más.
* **Escala de Razón o Proporción:** por el contrario en las escalas de proporción el cero representa donde el cero es la “nada”, representa la ausencia de algo.

**TABLAS UNIDIMENCIONALES Y BIDIMENCIONALES**

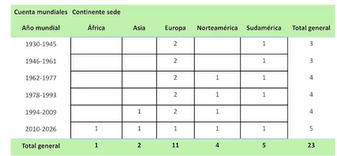
Las **tablas unidimensionales** son aquellas que con 1 pregunta podemos identificar el valor que se está buscando.

Las **tablas bidimensionales** son aquellas que con 2 preguntas podemos identificar el valor que se está buscando.

**FRECUENCIA ESTADISTICA O ABSOLUTA**

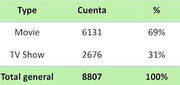
La frecuencia en estadística es el número de veces que un valor se repite dentro de un conjunto de datos.

La siguiente es una tabla de frecuencia bidimensional:



**FRECUENCIA RELATIVA**

La frecuencia relativa es la relación entre el número total de los eventos de un mismo tipo y el total de los eventos de todo el conjunto de datos, y esto se traduce en el porcentaje.

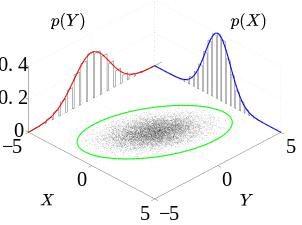
****

**DISTRIBUCIONES**

Las distribuciones es la forma de determinar la probabilidad de que un evento ocurra.

**DISTRIBUCIONES CONJUNTAS**

La distribución conjunta es la intersección entre dos variables aleatorias X y Y.



**DISTRIBUCION MARGINAL**

La distribución marginal es el conjunto de valores de un evento individual en que no dependen de otra variable.



**DISTRIBUCION CONDICIONAL**

La distribución condicional es la probabilidad de que ocurra un evento dado que un evento ya allá ocurrido, por ejemplo, si tenemos dos dados y ya sabemos que el primer dado es 6 cuál es la probabilidad de que el otro dado caiga en 2.



**MEDIDAS DE TENDENCIA CENTRAL**

Las medidas de tendencia central son fórmulas que se utilizan para calcular números que tiendan al centro de una distribución. Las principales:

* Media
* Mediana
* Moda

**MEDIDAS DE DISPESIÓN**

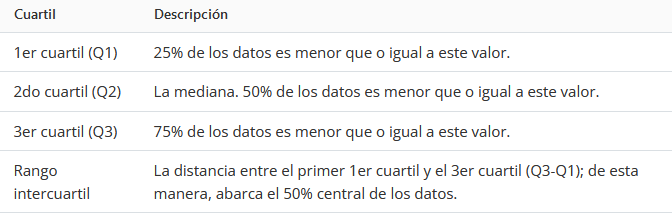
Las medidas de dispersión son formulas diseñadas para encontrar la separación (que tan lejos o que tan cerca) de las variablesrespecto al centro de una distribución de un conjunto de datos.

* **RANGO**

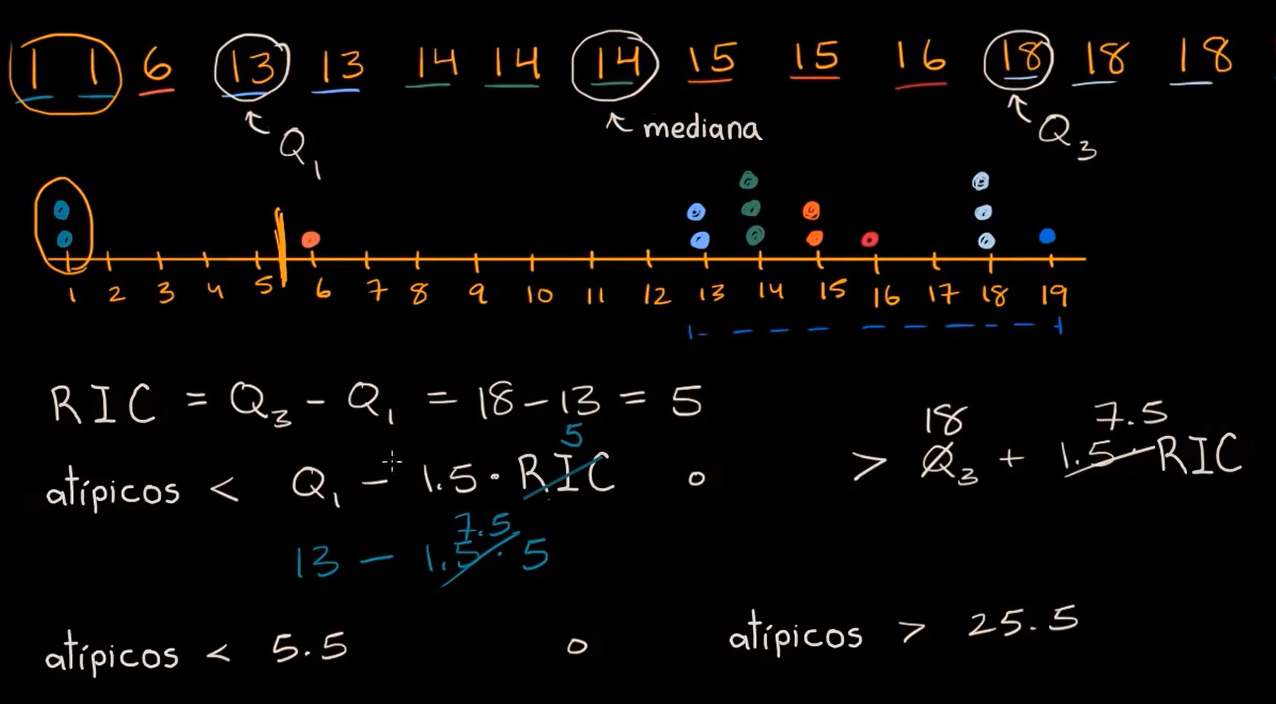
El rango es unamedida que ayuda a calcular la dispersión total de una muestra o una población.

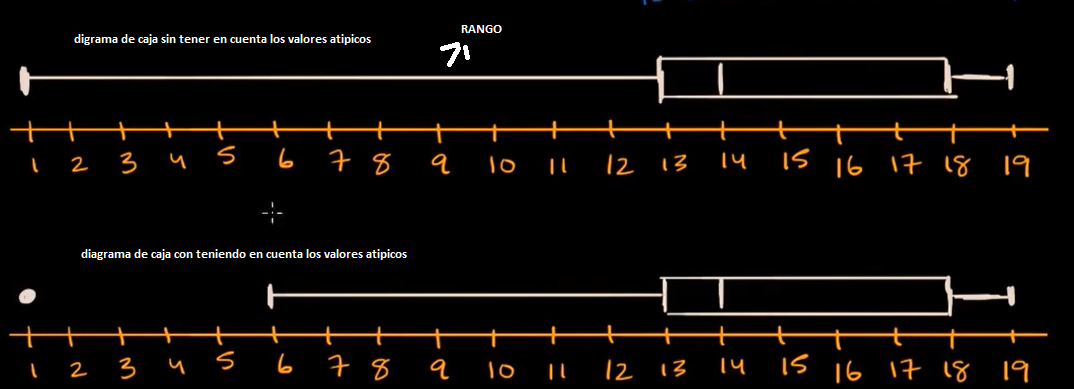
* **CUARTILES**

Los cuartiles son medidas que dividen un conjunto de datos en 4 partes iguales. Estas medidas nos ayudan a determinar la dispersión de los datos según su **mediana**.



**NOTA: Los valores atípicos en estos cuartiles se calcula generalmente Q1-1.5\*RI ó Q3+1.5\*RI donde RI es rango intercuartil.**

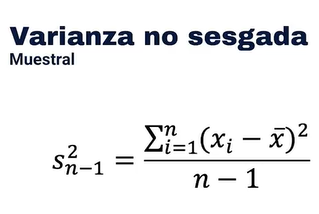


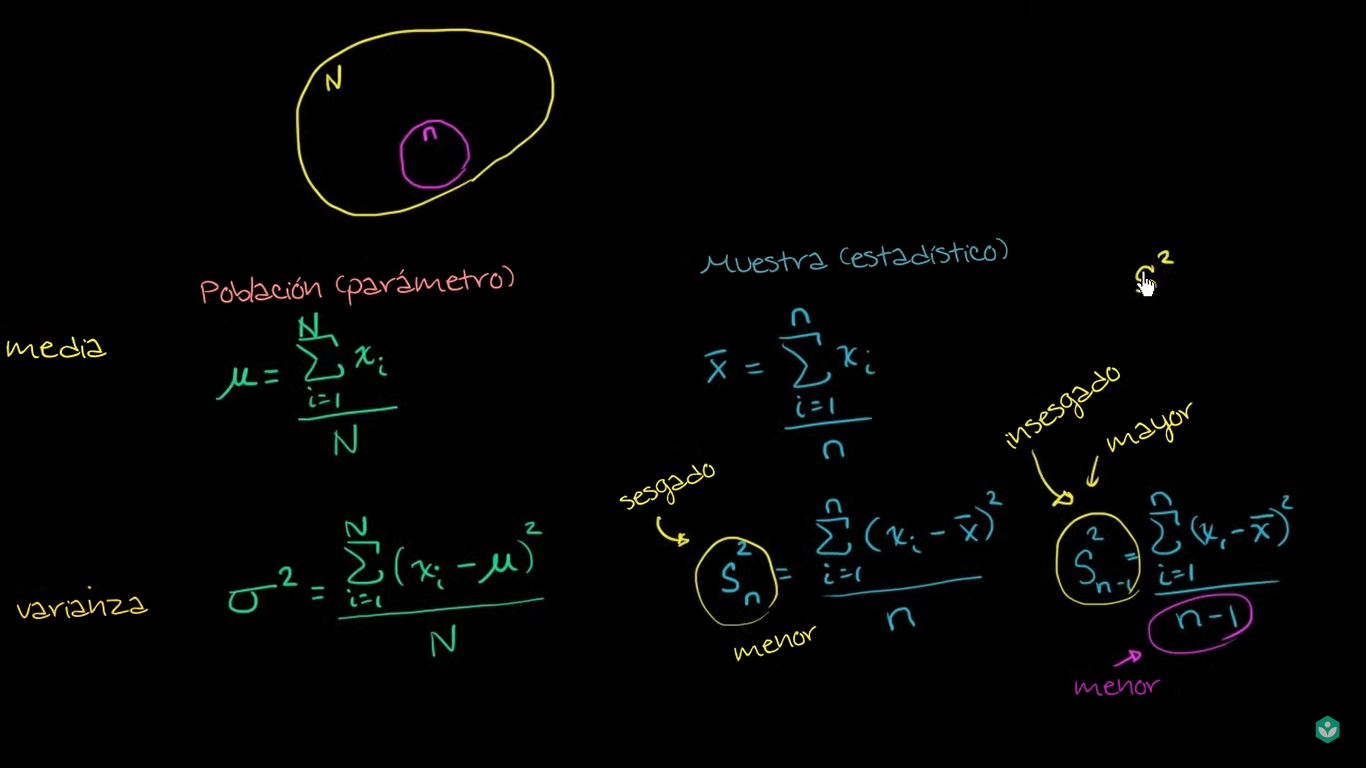


* **VARIANZA**

La varianza es una mediada de dispersión la cual puede ser utilizada para calcular que tan dispersos están los datos respecto a la media.

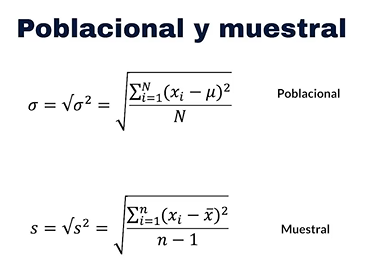
Nota: la varianza no sesgada se toma cuando es para una muestra en donde se remplaza n por n-1 ya que se aproxima mejor a la varianza poblacional.





* **DESVIACIÓN ESTANDAR**

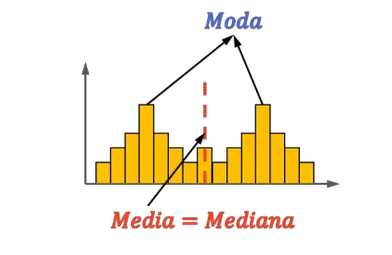
La desviación estándar es la raíz cuadrada de la varianza. Como la varianza eleva sus valores al cuadrado entonces tendríamos el cuadrado de las medidas originales por lo tanto con la desviación estándar al sacarle sus raíz tenemos un valor más acercado a su escala original.



**DISTRIBUCIONES SIMETRICAS**

**CARACTERISTICAS:**

* Las distribuciones simétricas se caracterizan porque al dividir por su eje de simetría los dos lados divididos van a ser idénticos o van a tener el 50% de los valores cada uno.
* La media debe ser igual a la mediana.
* Puede ser bimodal (tener dos modas)

****

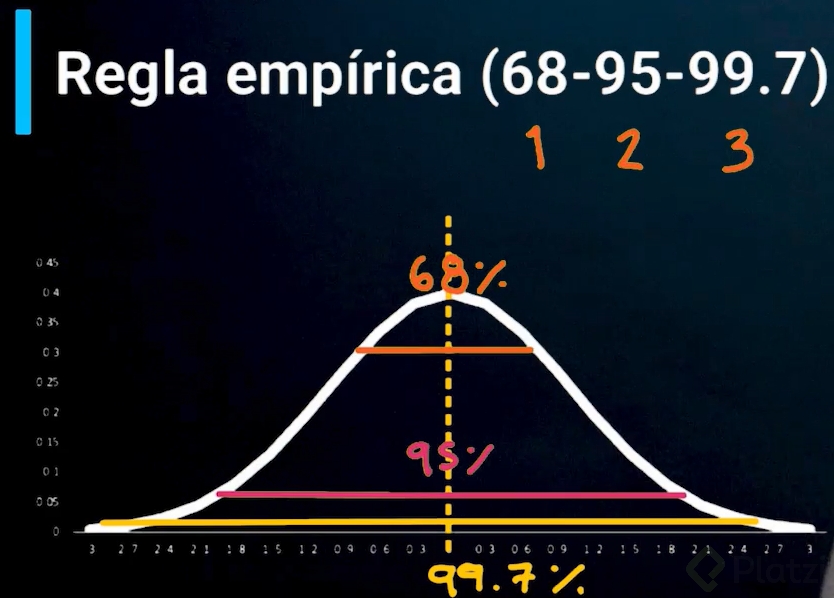
* Si la distribución es unimodal la moda, mediana y media van a ser iguales.

**DISTRIBUCIÓN NORMAL/CAMPANA DE GAUSS/GAUSSIANA**

Una de las distribuciones simétricas más populares es la distribución normal o campana de Gauss.

**CARACTERÍSTICAS:**

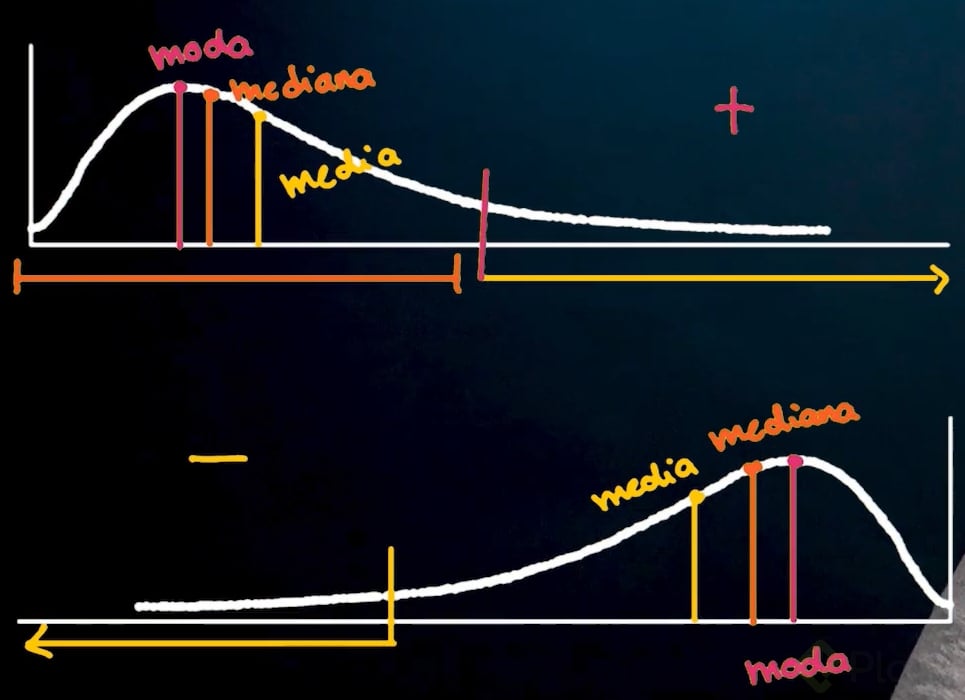
* Media, mediana y moda son iguales.
* Su “50%” central está dentro de 2/3 de desviación estándar hacia la izquierda y derecha de la media, o también llamada llamado el rango intercuartil.
* Se cumple la regla empírica, la cual nos explica que el con 1 deviación estándar a la izquierda y derecha de su eje de simetría agrupa el 68% de los datos, con 2 el 95% y con 3 el 99,7.



**DISTRIBUCIONES ASIMETRICAS**

**CARACTERISTICAS:**

* Las distribuciones asimétricas cuentan con gran cantidad de valores atípicos.
* Cuentan con colas: las colas son concentraciones de valores atípicos.
* Hay distribuciones asimétricas positivas con colas hacia la derecha y la negativa con cola hacia la izquierda. Las medidas de tendencia central suelen ser en estas distribuciones asimétricas en orden 1 moda, 2 mediana y 3 media en distribuciones positivas y 1 media, 2 mediana y 3 moda en negativas.



**ESTADISTICOS**

Los estadísticos son funciones reales de una muestra, es decir, que ingreso un valor de X (datos de la muestra) que al operarlos va dar valores Y (estadístico) condicionados por X.

Los estadísticos son funciones reales que se aplican a conjuntos de datos o muestras. En otras palabras, son medidas o características numéricas derivadas de los datos de una muestra. Cuando se ingresa un conjunto de valores (datos de la muestra) en una función estadística, esta función genera un valor específico (estadístico) que resume alguna propiedad o característica de la muestra.

**PARAMETROS**

Los parámetros medidas numéricas de una población, sin embargo, como la población muchas veces no la podemos calcular totalmente llegamos a aproximaciones de estos mediante los estadísticos.

**METODOS DE RECOPILACIÓN DE DATOS**

Hay dos tipos de estudios:

**OBSERVACIONALES**

Son aquellos en donde no hay intervención en los datos analizados por parte del investigador, y se dedican a identificar las variables que definen el estudio.

**EXPERIMENTALES**

Los estudios experimentales, son estudios los cuales modifican una variable independiente para comprobar el comportamiento de una variable dependiente, con ellos podemos evaluar la causalidad (relación entre causa y efecto).

**CARACTERÍSTICAS**

* **Aleatorización**: La muestra tiene que ser aleatorio
* **Se usan 2 grupos**: grupo de control (control ciego o doblemente ciego) y grupo de estudio.
* **Replicación:** el estudio se tiene que poder replicar.
* **Bloqueo (matched pairs)** Que el grupo sea equitativo respecto a género, edad etc.

**MUESTREO**

Poder obtener una muestra de una población aleatoria y significativa, que se pueda utilizar para inferir valores de una población.

**SESGOS**

Al tomar las muestras, puede haber cierto tipo de sesgos que nos alejen de los valores a inferir de la población y no podamos aproximarnos correctamente a una estimación. Por ejemplo:

* **SESGOS DE RESPUESTA:** Respuestas de los participantes que no reflejan con precisión la realidad.
  + Medición incorrecta
  + Deseabilidad social
  + Preguntas inductivas o capciosas
* S**ESGOS DE INFRACOBERTUTA:** La muestra no es lo suficientemente representativa.
  + Selección
  + Respuesta voluntaria
  + Muestreo a conveniencia
* **SESGO DE NO RESPUESTA:** se produce cuando las personas se limitan a no participar.

**PROBABILIDAD**

Es la medida que calcula la posibilidad de que un evento ocurra.

La probabilidad simple calcula el evento entre el espacio muestral. Donde el evento es el evento que cumple los requisitos específicos que estoy buscando y el espacio muestra es el conjunto de eventos que cumplen con ciertos requisitos. Además va de 0 a 1 o de 0% a 100%.

Hay varios tipos de probabilidad:

**Probabilidad Clásica:**

Se basa en la idea de que todos los eventos posibles son igualmente probables. Por ejemplo, al lanzar un dado justo de seis caras, cada cara tiene una probabilidad de 1/6 de aparecer.

**Probabilidad Empírica o Frecuencial:**

Se calcula observando la frecuencia relativa de un evento en un conjunto de datos. Por ejemplo, la probabilidad de que un equipo de fútbol gane un partido se puede estimar observando su historial de victorias y derrotas.

**Probabilidad Subjetiva o Personal:**

Se basa en la opinión o juicio subjetivo de una persona sobre la probabilidad de que ocurra un evento. Por ejemplo, un experto en meteorología podría asignar una probabilidad subjetiva a la ocurrencia de lluvia.

**Probabilidad Condicional:**

Es la probabilidad de que ocurra un evento dado que otro evento ya ha ocurrido. Se denota como P(A|B), la probabilidad de A dado B.

**Probabilidad Conjunta:**

Es la probabilidad de que dos o más eventos ocurran simultáneamente. Se denota como P(A y B).

**Probabilidad Marginal:**

Se refiere a la probabilidad de un evento individual en un conjunto de eventos. Por ejemplo, la probabilidad marginal de que llueva, independientemente de otras condiciones meteorológicas.

**Probabilidad Discreta:**

Se aplica a variables aleatorias discretas, como el número de caras al lanzar un dado. Se representa mediante funciones de masa de probabilidad.

**Probabilidad Continua:**

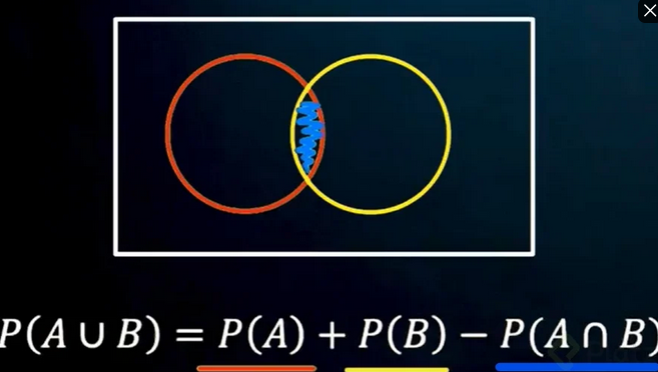
Se aplica a variables aleatorias continuas, como la altura de una persona. Se representa mediante funciones de densidad de probabilidad.

**Probabilidad Marginal y Condicional:**

Las probabilidades marginales son las probabilidades individuales de eventos, mientras que las probabilidades condicionales se calculan cuando se tiene información adicional sobre otros eventos.

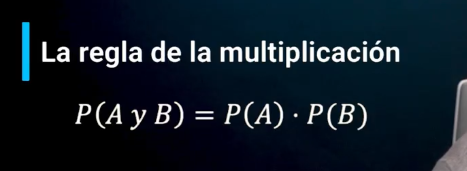
**REGLA DE LA SUMA**

La probabilidad de tener un evento u otro va a ser igual a la suma de la probabilidad de mi primer evento más la probabilidad de mi segundo evento menos la probabilidad de la intersección de ambos eventos.



**REGLA DE LA MÚLTIPLICACIÓN**

La probabilidad de tener un evento A y B se calcula multiplicando la probabilidad A por la probabilidad de B.

****

**EVENTOS INDEPENDIENTES**

Los eventos independientes son eventos cuales no son influéncianos por otros eventos. “Se dice que dos eventos A y B son independientes si y solo si la probabilidad del evento B no está influida por el suceso del evento A o viceversa.”

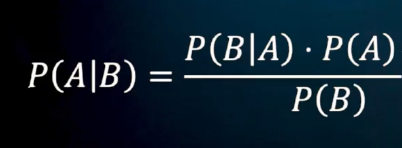
**EVENTOS DEPENDIENTES**

Los eventos dependientes son eventos si el cambio de A afecta a B o viceversa. Formalmente, dos eventos A y B son dependientes si la probabilidad de que ocurra el evento A (o su complemento, A') está influenciada por si el evento B ha ocurrido o no. Matemáticamente, la dependencia de dos eventos se expresa mediante la fórmula:

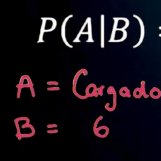
P(A∣B)=P(A)

**TEOREMA DE BAYES**

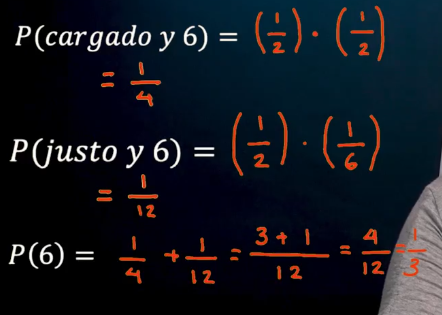
Calcula la probabilidad condicional de A tal que B es cierto, y se calcula multiplicando la probabilidad de que B ocurra si A es cierto entre la probabilidad de A y todo lo anterior dividido entre la probabilidad de B.



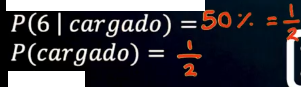
Ejercicio imaginemos que tenemos dos dados, uno de ellos está cargado y el 50% de las veces cae en 6, ahora queremos saber la probabilidad de que el dado este cargado tal que cae 6.



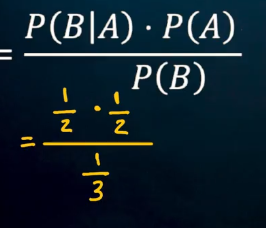
Sabemos que la probabilidad que salga 6 es:



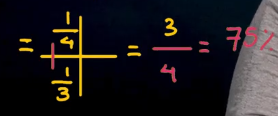
Además:



Por lo tanto tenemos que:



Entoncess:



**PERMUTACIONES**

Es la cantidad de maneras distintas en las cuales puedo ordenar un conjunto de datos. La principal diferenciación es que en las permutaciones importa el orden en las posiciones de los elementos. Los tipos de permutaciones son:

**PERMUTACIONES CON REPETICIÓN**

[**https://matemovil.com/permutacion-con-repeticion-ejercicios-resueltos/**](https://matemovil.com/permutacion-con-repeticion-ejercicios-resueltos/)

[**https://www.studysmarter.es/resumenes/matematicas/estadistica-y-probabilidad/permutacion-y-combinacion/**](https://www.studysmarter.es/resumenes/matematicas/estadistica-y-probabilidad/permutacion-y-combinacion/)

Son permutaciones en las cuales un elemento puede reutilizarse en todas las posiciones del ordenamiento, por ejemplo si queremos formar grupos de dos del conjunto A, B y C podríamos repetir elementos como AA, BB, CC.

En este caso la formula sería:

Donde n es el total de los elementos que se pueden seleccionar y r el total de elementos seleccionados.

En este caso

Sin embargo otro tipo de permutaciones con repetición es a diferencia de la anterior, es cuando tenemos elementos fijos repetidos, es decir, hay un conjunto de elementos en el cual tenemos elementos repetidos como RED, BLUE, RED, GREEN. Para estos casos utilizamos la formula.

En donde n es el número total de elementos y a, b, c son los elementos teniendo en cuenta cuantas veces se repiten en el conjunto n.

**Ejemplo**

En una urna, hay 5 bolas del mismo tamaño y peso, de los cuales, 3 son rojas y 2 son azules. ¿De cuántas maneras se pueden extraer una a una las bolas de la urna?

Solución:

Coloquemos algunas formas de extraer las bolas:

Roja – Roja – Azul – Roja – Azul.

Azul – Roja – Roja – Azul – Roja.

Roja – Azul – Roja – Azul – Roja.

En cada forma de extraer las bolas, importa el orden, hay elementos repetidos y participan todos los elementos (bolas), por ello, usaremos la fórmula de permutación con elementos repetidos.

Número de bolas rojas: 3.

Número de bolas azules: 2.

Número total de elementos: n = 3+2 ➜ n=5

permutación con elementos repetidos

**PERMUTACION SIMPLE O SIN REPETICION**

Este tipo de permutaciones no se repite ningún elemento y el ordenamiento es lineal. Formula:

**Ejemplo.**

Que tantas combinaciones puedo obtener al ordenar 3 libros diferentes.

**PERMUTACION SIMPLE O SIN REPETICION DE UN SUBCONJUNTO**

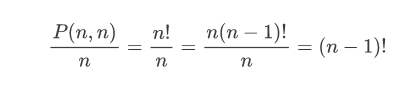
La diferencia con este tipo de permutaciones es que se toman de un conjunto “n” los elementos a combinar que en este caso llamaremos “r”, su fórmula seria.

Ejemplo:

Tenemos que ver de cuantas formas se pueden ordenar dos canicas de un grupo de cuatro canicas diferentes entre sí.

**PERMUTACION CIRCULAR**

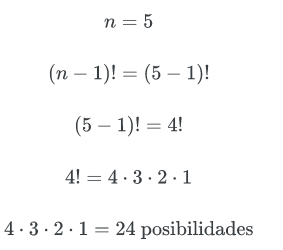
En este caso el ordenamiento no es en línea recta, sino que es circular, por lo tanto **un círculo comienza en un punto y termina en el mismo punto.**



Ejemplo

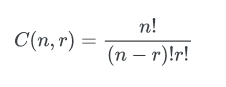
A Juan le han pedido que coloque a 5 alumnos en una mesa redonda de comedor. ¿De cuántas formas puede conseguirlo?

Solución:



**COMBINACIONES**

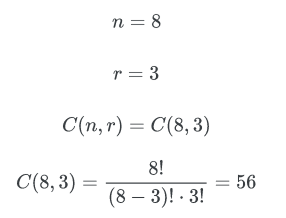
La diferencia entre permutación y combinación es que la combinación no tiene en cuenta el orden de los elementos. Para la combinaciones se utiliza la fórmula:



Donde N es el número total de elementos y r es el número de elementos elegidos.

POR EJEMPLO

De una clase de 8 alumnos, hay que elegir tres para visitar el museo. ¿De cuántas maneras se puede tomar la decisión?



**COMBINACIÓN DE SUCESOS MÚLTIPLES**

En estas combinaciones se más de una clase de elección y se multiplican los resultados.

POR EJEMPLO

Si en una clase de 14 personas hay 8 chicas y 6 chicos, y hay que elegir 4 chicos y 5 chicas, habrá que considerar la elección por género.

Por lo tanto, se convierte en elegir 4 chicos de 6 y 5 chicas de 8. Esto significa que:

