Fundamental si queres aprobar:

* Dar la clase al profesor cómo si le estuvieras explicando los temas a un amigo tuyo.
* Saber para qué sirve cada tema.

**Estadística:**

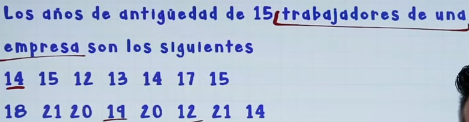
+Vamos a utilizar la estadística para recopilar, organizar y analizar datos.

Población: Son todos los elementos.

Individuo: es un elemento de la población.

Muestra: Si tomamos una muestra debe ser representativa de la población. Es un subconjunto de la población (una parte, esta debe ser seleccionada con datos aleatorios de la población).

Dato: El valor que tiene cada elemento.



* Población: “Todos los trabajadores de una empresa”
* Muestra: 15 trabajadores de la empresa
* Individuo: un trabajador de la empresa
* Dato: El año de antigüedad de 1 trabajador

Variables Cualitativas:

Son aquellas que no se pueden medir con números, sino con cualidades.

Variables Cuantitativas

Se expresa con números y con ellas se pueden realizar operaciones algebraicas

* Discretas: Toma un número finito de valores ya que se trabaja solo con enteros.
* Continuas: Toma un número infinito de valores ya que se trabaja tanto con enteros como también con decimales.

**Tipos de variables y sus gráficos:**

+Acá vamos a trabajar con:

* **Primera columna: Valores que me toma la variable:**
  + Cualidades (cualitativas) o números (discretas) o subintervalos (continuas)
* **Segunda columna: Frecuencias absolutas:**
  + Me da la cantidad de veces que se repite cada una de las observaciones. Si las sumo me da el total de observaciones.
* **Tercera columna: Frecuencias relativas:**
  + Es la frecuencia absoluta, pero en una escala más pequeña. Para calcularla agarro la primera observación y la divido por el total de observaciones (todo esto de la frecuencia absoluta).
* **Cuarta columna: Frecuencia absoluta acumulada**
  + Si las sumamos nos da el total de observaciones también. Esta se encarga de llevar el porcentaje acumulado.
* **Quinta columna: Frecuencia relativa acumulada:** 
  + Es el porcentaje acumulado de la frecuencia relativa.

+Tenemos gráficos para variables cualitativas y cuantitativas

* Variables cualitativas (refiere a cualidades):

Gráfico de sector o torta

Este gráfico se presenta en forma redonda donde cada división de la torta representa la categoría y estas se escriben en porcentaje. Al sumar todas las porciones (las categorías) me da un total del 100%

Gráfico de barras

Son barras donde cada una tiene el mismo tamaño, ninguna se toca entre sí.

Este gráfico se puede presentar con las barras en vertical y en horizontal.

Cuenta con dos partes: Las categorías y la frecuencia absoluta

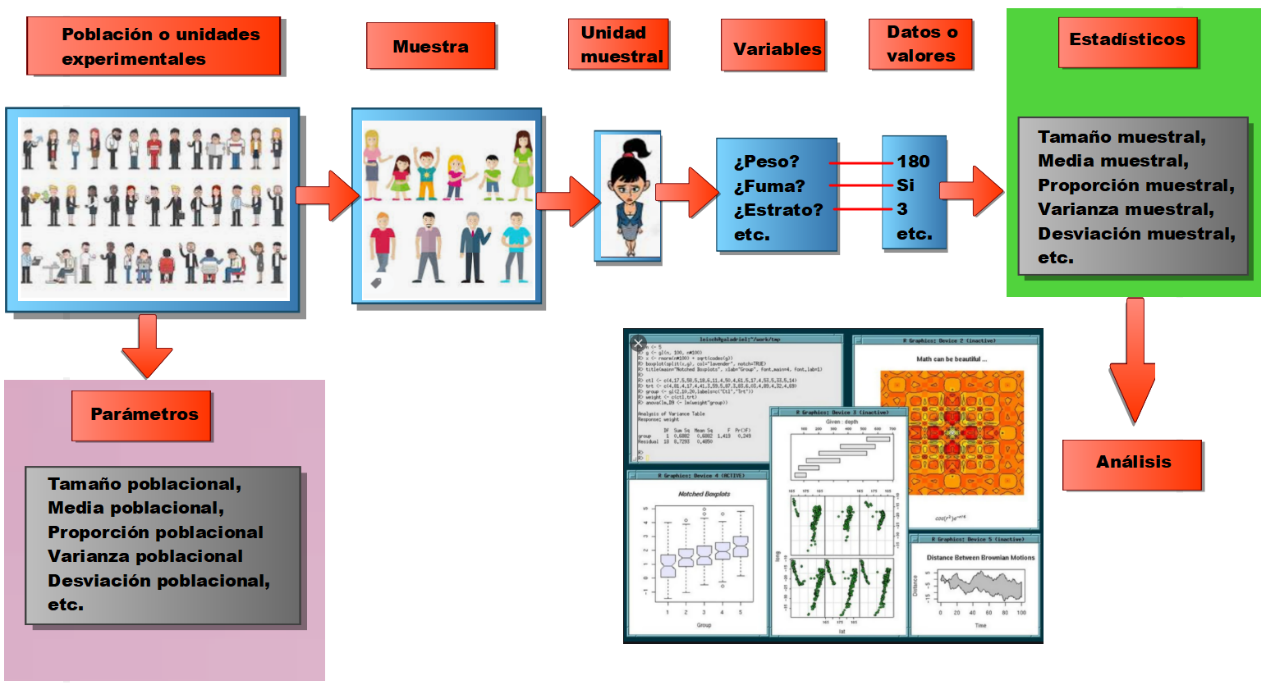
Gráfico de Pareto

Este se compone de 3 columnas: Las categorías, frecuencias absolutas y porcentaje acumulado

Las categorías deben estar ordenadas de mayor a menor

**Medidas de posicionamiento y dispersión:**

+Las medidas de posicionamiento y dispersión se utilizan para entender mejor los datos que tenemos. Trabajamos con medidas que **nos indican la ubicación de un valor dentro de un conjunto de datos**.



Términos fundamentales:

* **Población no es lo mismo que espacio muestral.**
  + **Población**: Se usa en el contexto de la estadística. Refiere a todos los elementos sobre algo en particular.
  + **Espacio Muestral**: Se utiliza en el contexto de la teoría de la probabilidad. Es el conjunto de todos los posibles resultados que puedan tocar en ese experimento.
  + **Suceso elemental**: Es un resultado posible de un experimento. Por ejemplo: En un dado el “3” es un resultado posible de que toque en el experimento, ese mismo es un suceso elemental. Como también el 4 es otro resultado posible.
* **Evento es lo mismo que suceso**
* **Observación y elemento no es lo mismo**
  + **La observación:** Es el resultado de un elemento. Por ejemplo: Si estas realizando una encuesta, cada respuesta de un encuestado será una observación.
  + **Elemento:** Se refiere a un individuo o ítem específico dentro de la población o muestra. Por ejemplo: Cada persona en una encuesta sería un elemento de la población o muestra.
* **Estadística y probabilidad:** 
  + **Estadística:** Sirve para recopilar, organizar y analizar datos, y en base a ello se toman conclusiones o estimaciones sobre la población completa.
  + **Probabilidad:** Mide la frecuencia con la que se obtiene un resultado al llevar a cabo un experimento aleatorio.

+Trabajamos con una serie de datos que cada uno tiene su significado. Todos estos datos que vamos a extraer son el espacio muestral ¿Qué es esto? Son cada resultado de cada elemento que tiene el experimento que recopilamos, y la muestra es una parte del experimento (no analizamos todo).

Para la **medida de posicionamiento** (Estas medidas indican la ubicación de un valor dentro de un conjunto de datos) trabajamos con:

* **Media**: Es el promedio de todo el experimento.
  + Si el elemento es muy grande generalmente se saca (es sensible a valores extremos).
* **Mediana**: Es el elemento del medio
  + Fundamental de que esté ordenado de menor a mayor.
  + Primero ordenamos el experimento de menor a mayor y vemos la cantidad de elementos que tiene. Si ese número es par o impar tiene sus acciones respectivas:
    - Par
    - Impar
* **Moda:** Es el valor de la variable que más se repite. La muestra puede ser unimodal, bimodal, trimodal.

Para la **medida de dispersión** trabajamos con: (Indican cuanto varían o se dispersan los datos)

* **Cuartil 1, 2 y 3:** Estos se encargan de dividir el experimento en 4 partes. El cuartil 1 es el 25% el 2 es el 50% y el 3 es el 75%.
  + Tengo que ordenarlos de menor a mayor.
* **Varianza:** Mide la dispersión de los datos alrededor de la media, pero elevados al cuadrado. A mayor dispersión de la media mayor es la varianza. Cuando los valores estén más cercanos a la media la varianza es menor.
  + Poblacional
  + Muestral
* **Desviación estándar:** Mide la dispersión de datos alrededor de la media
* **Rango**: Sirve para ver qué tan extendidos están los datos en el conjunto. Y es la diferencia entre la observación de mayor valor y la de menor valor.
* **Rango intercuartílico:** Sirve para ver qué tan dispersos están los datos en el 50% central del conjunto.Es la diferencia entre el cuartil 3 y el 1.
* **Coeficiente de variación:** Es la desviación estándar expresada como un porcentaje en la media. Este resultado siempre se multiplica por 100 y se representa en forma de porcentaje.

**IMPORTANTE: ¿A qué se refiere mayor dispersión?**

**Rta:** En el conjunto de datos 5,10,15,20,25 el rango es 20 (25 - 5), lo que indica una mayor dispersión que en el conjunto 10,12,14,16,18 donde el rango es solo 8 (18 - 10).

**Diagrama de caja (Box-plot):**

+Trabajamos con 3 medidas: La observación mas chica, la mas grande y los 3 cuartiles.

+Podemos detectar valores extremos (outliers).

* Para ella deben primero calcularse la barrera inferior y la superior, y si se llega a superar esa barrera hablamos de outlier.
* Esta se marca solo con un asterisco
* Puedo tener más de un asterisco.

+Las patillas del barbijo van a llegar siempre a la observación más grande que no supere las barreras.

**Análisis combinatorio:**

+El análisis combinatorio se utiliza para

+Tenemos tres tipos:

* Arreglo:
  + Me interesan todos los elementos
  + No me interesa el orden
* Combinación:
* Permutación:

Estos 3 pueden ser simples (que el elemento a analizar no se repite), o también pueden ser con repetición (el elemento a analizar se puede repetir).

**Estadística:**

+La estadística se utiliza para recopilar, organizar y analizar datos, y en base a ello se toman conclusiones o estimaciones sobre la población completa.

**Probabilidad:**

+La probabilidad se utiliza para medir la frecuencia con la que se obtiene un resultado al llevar a cabo un experimento aleatorio.

+Cando pienses que significa la probabilidad pregúntate ¿Cuál es la probabilidad de…? Entonces piensa a qué refiere con eso: Refiere a cuál es la frecuencia de obtener un resultado que nosotros queramos.

+Tenemos 2 tipos de experimentos:

* Determinísticos: de antemano podemos determinar cuál va a ser el resultado. Si tengo una cubierta desinflada y le hecho aire va a inflarse, lo mismo que a otra cubierta desinflada.
* Aleatorios: No podemos predecir de antemano el resultado.

Tenemos dentro de la probabilidad **los sucesos aleatorios**:

Y estos son 3 tipos:

* Suceso seguro: Que en tenga en una urna 6 pelotas azules. Es un suceso seguro de que si necesito sacar una pelotita azul la saco. Otro suceso seguro es que si voy a tirar un dado al aire es un suceso seguro que vaya a sacar un número mayor que 0 y menor que 7.
* Suceso imposible: sacar una pelotita de color amarillo de la urna cuando todas son azules. Sacar el número 7 en el dado.
* Suceso contrario: tenemos el ejemplo de un dado y nosotros definimos que en el suceso salga un número par cuando tiramos el dado, lo contrario es que termine saliendo un número impar. Acordate que el complemento es lo que me falta a mí para hacer el espacio muestral.

**Probabilidad clásica:**

Es una forma de calcular la probabilidad. Esta probabilidad se basa en que todos los resultados posibles de un experimento son igualmente probables.

Por ejemplo: Tengo una bolsa con 10 chupetines rojos y 10 chupetines azules. ¿Cuál es la probabilidad de que saque dos chupetines azules a la primera? Primero tengo que calcular cuál es la probabilidad de sacar un chupetín azul. Utilizo la formula

0.5 (o del 50%) y luego calculo la probabilidad para el segundo chupetín (ahora hay uno menos en la bolsa). 0.474 (o 47.4%).

Ahora. El mismo ejemplo, pero quiero sacar primero 1 chupetín azul, uno rojo y otro azul nuevamente.

La probabilidad de sacar el primer chupetín azul es = 50% (quedan 9 azules), luego sacar uno de color rojo es = 0,53 (53%) y por último para sacar un chupetín azul nuevamente es 0.5 (50%)

FUNDAMENTAL: Ver si saco la probabilidad conjunta o no.

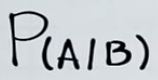
**Tenemos los axiomas también:**

Estas son las propiedades de la probabilidad.

**Probabilidad condicionada:**

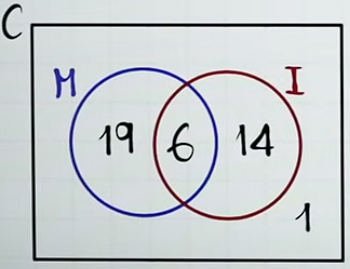
Intervienen 2 sucesos por un lado uno de los sucesos ya ocurrió.

+Así se escribe la probabilidad condicionada:

 y se lee “la probabilidad de que haya sucedido el evento A dado que ya ocurrió el evento B”.

Por ejemplo: De los 40 estudiantes de mi curso a 25 les gusta las matemáticas. A 20 les gusta el inglés y a 6 les gusta las 2 materias. Si se selecciona un estudiante que le gustan las matemáticas ¿Cuál es la probabilidad de que le guste el inglés?

Paso 1:



Paso 2: Ahora escribimos esto:

Esto quiere decir ¿Cuál es la probabilidad de que al alumno le guste el inglés dado que le gustan las matemáticas? Ese “/” significa dado que.

Paso 3: Y ahora aplicamos la formula de la probabilidad normal:



Si nos dijeras cuál es la probabilidad de que al alumno seleccionado le gusten las matemáticas hacemos entonces:

Pero cómo nos piden cuál es la probabilidad de que al alumno le guste el inglés sabiendo que le gustan las matemáticas, entonces actuamos directamente con los alumnos del conjunto de matemáticas (Mira el gráfico del paso 1 sino).

** =** 6/25 (o si elijo un alumno de matemática tengo una probabilidad del 24% de que le guste el inglés).

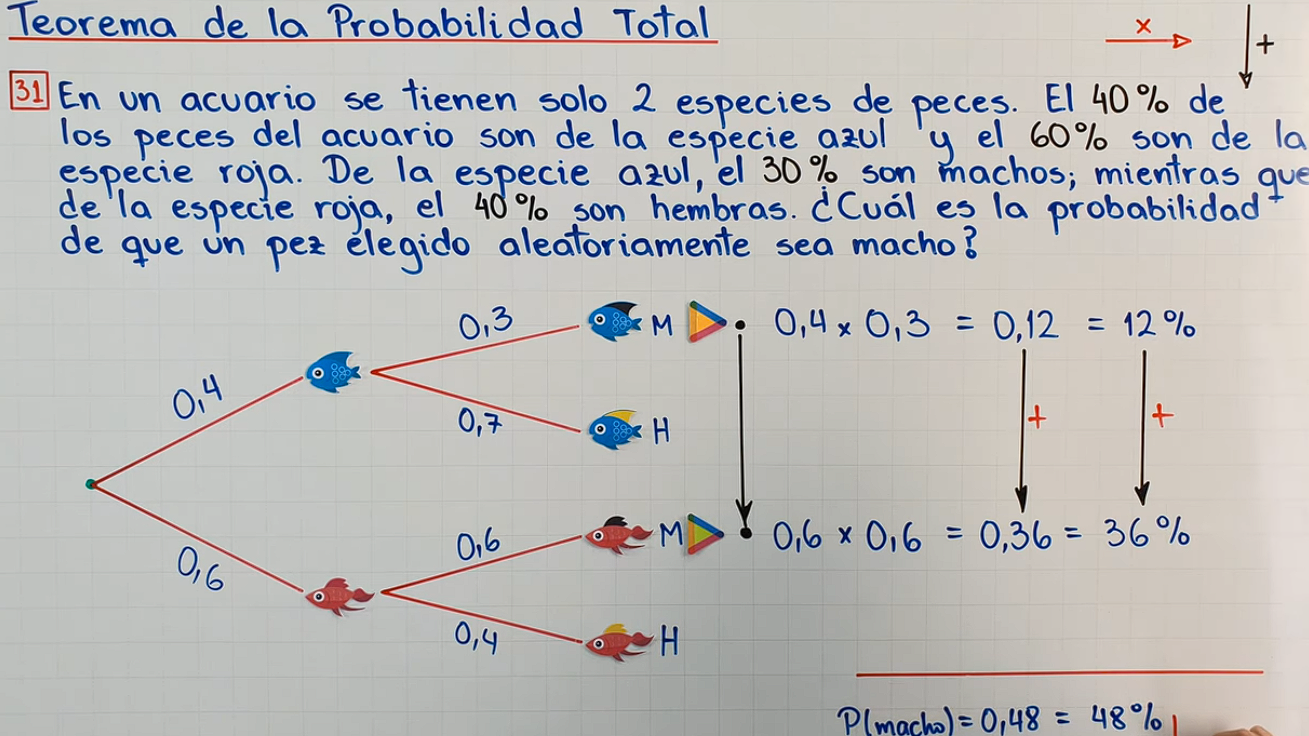
**¿Qué tiene este ejercicio para que sea de probabilidad condicionada?**

Porque dice “Si se selecciona un estudiante que le gustan las matemáticas, ¿Cuál es la probabilidad de que le guste el inglés?” uno de los sucesos ya ocurrió (No se está seleccionando un estudiante cualquiera, se está seleccionando a uno que le gustan las matemáticas, esa es la condición), al alumno le gusta las matemáticas, ahora hay que ver el siguiente suceso, si le gusta el inglés.

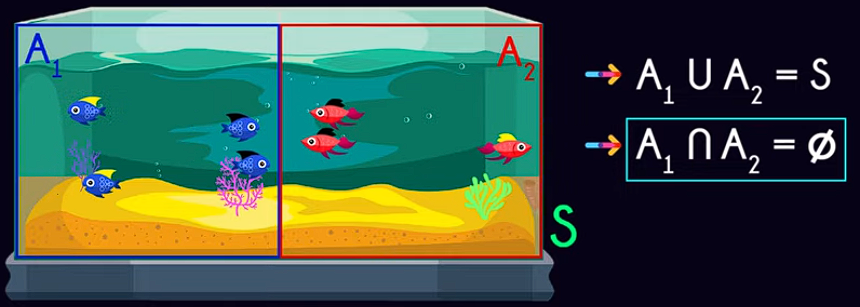
**Teorema de la probabilidad total:**

+Este tipo de probabilidad nos ayuda a calcular la probabilidad de que ocurra un evento considerando todas las formas diferentes (o caminos) en las que ese evento se pueda dar.

+Si empleamos el diagrama de árbol en el teorema de la probabilidad total es mucho más fácil hacer el análisis:



Esta es la parte teórica (sin diagrama de árbol)



Si junto A1 Y A2 nos da el espacio muestral “S” y si interceptamos A1 Y A2 nos da vacío porque no podemos tener un pez de forma azul y rojo. FUNDAMENTAL: si tiene el signo de vacío (que es un suceso imposible) se dirá que son disyuntos.

A1 y A2 son **mutuamente excluyentes** porque se produce uno o se produce el otro. Cuando extraemos un pez este es de la especie azul o de la especie roja, no puede ser de las 2 especies.

Si se cumple que si A1 y A2 son colectivamente exhaustivos (Que unidos forman el espacio muestral “s”) y que son mutuamente excluyentes (es decir que interceptados dan nulo, mira la segunda formula) entonces podemos trabajar con el teorema de la probabilidad total.

**¿Qué tiene este ejercicio para que sea del teorema de la probabilidad total?**

Este ejercicio lo que tiene es que

* Múltiples caminos para el evento: El evento “elegir un pez macho” puede suceder a través de dos caminos diferentes, ya sea por los peces azules o los rojos.
* Eventos mutuamente excluyentes: Los caminos son mutuamente excluyentes ya que el pez solo pertenece a una especie (No existe un pez de las 2 especies).

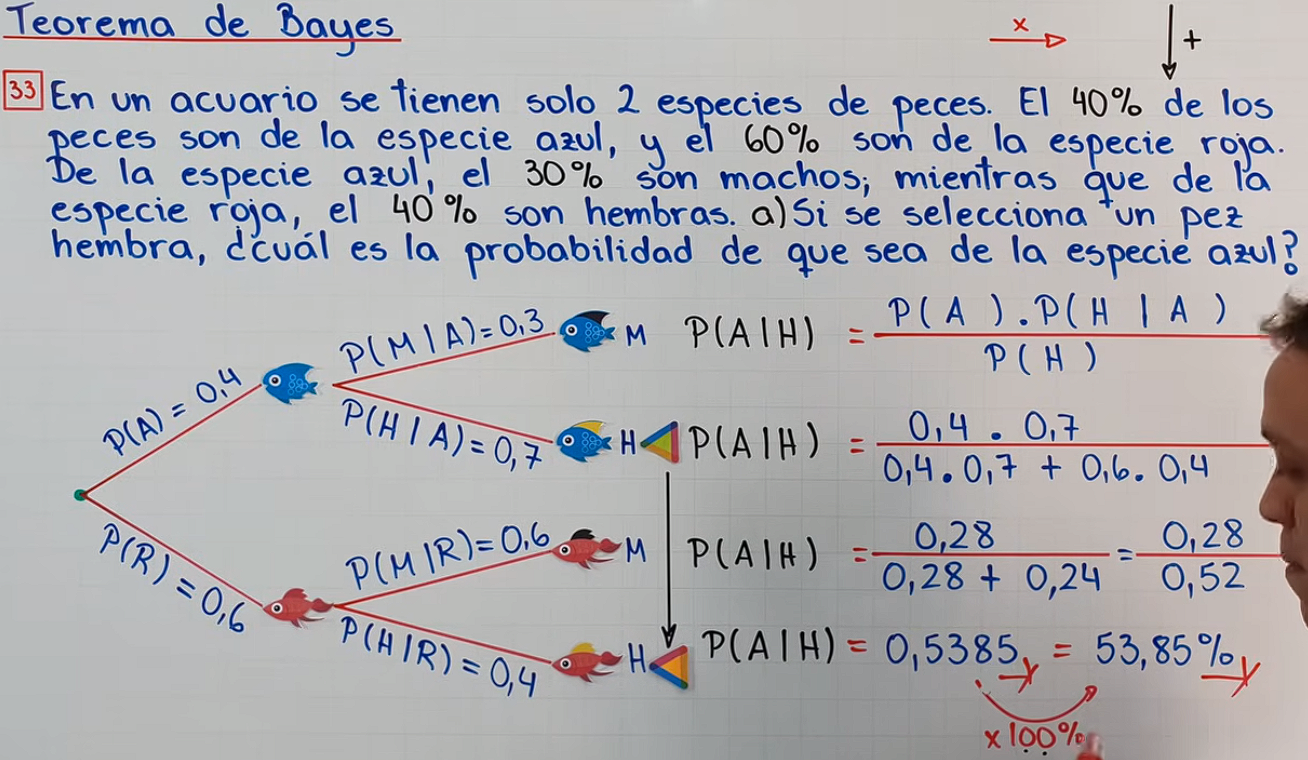
**IMPORTANTE: Ver donde ubico los eventos dependientes o independientes.**

**Teorema de bayes:**

+Si sé probabilidad condicionada sé teorema de bayes.

+Tiene su fórmula y es esta:



+ 

**Distribución de variables:**

+La distribución de variables es una herramienta que nos ayuda a entender cómo se comportan los datos en un conjunto. IMPORTANTE Se le puede agregar un poco más de condimento.

+Para variables discretas usamos:

* Distribución de Bernoulli
  + Realizamos una vez el experimento
  + Tenemos una probabilidad de éxito y una de fracaso
* Distribución binomial
  + Realizamos más de una vez el experimento
  + La probabilidad de éxito y de fracaso siempre es la misma en cada muestra.
  + Existe el criterio de aproximación por si x se distribuye binomial y n es mayor que 30. Esta misma puede distribuirse en forma de distribución normal o distribución de poisson (todo dependiendo del valor de p si está entre 0,10 sea menor que 0,10) ¿Y para qué se hace esto? Esto lo hacemos por si queremos utilizar la tabla de distribución binomial Porque si no aproximamos tenemos que calcular la integrada.
* Distribución Hipergeométrica
  + Realizamos más de una vez el experimento
  + La probabilidad de éxito y de fracaso va variando en cada muestra.
  + Tiene su propia formula.
* Distribución de poisson
  + Son pocas las probabilidades de éxito en esta distribución
  + Si trabajamos con ejercicios de “tiempo”

+Para variables continuas usamos:

* Distribución normal o gaussiana.
  + Tenemos que usar el paso de estandarización si queremos utilizar la tabla de la normal acumulada.

**Test de hipótesis**

+El test de hipótesis se usa cuando tenemos un dato que es certero, viene alguien y nos trata de reformular ese dato certero. El dato certero se lo denomina hipótesis nula y el dato refutado se llama hipótesis alternativa. Es acá donde tenemos que hacer una serie de pasos a través del **promedio** o **la proporción** ¿a qué se refiere esto?

**Intervalos de confianza**

+Los intervalos de confianza se usan para

+No es lo mismo que el parámetro esté dentro del intervalo, (Ya que el parámetro puede ser cualquier valor dentro de ese intervalo y el parámetro es un valor fijo) por eso se dice que el intervalo cubre al parámetro.

**Regresión Lineal:**

+La regresión lineal se utiliza para

Imágenes importantes:

