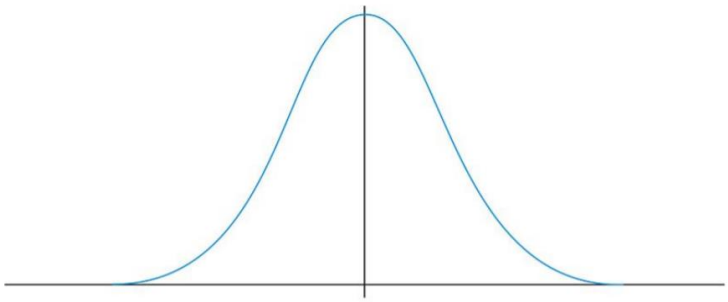
# CONCEPTOS ESTADÍSTICOS

## Distribución normal

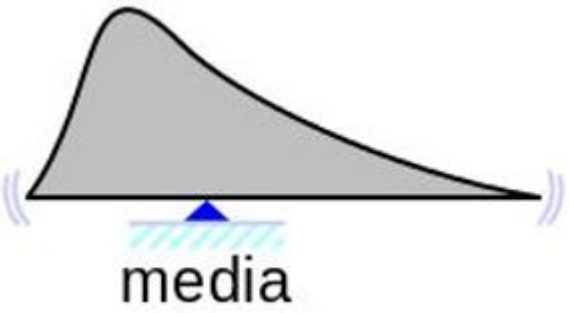
La distribución normal, también conocida como distribución gaussiana o **campana de Gauss**, es un concepto importante en estadísticas y matemáticas que describe la forma en que se distribuyen muchos fenómenos en la naturaleza y en la sociedad. Imagina una curva simétrica en forma de campana.



The Normal Distribution

## La media o promedio

La media, es una medida estadística que representa el valor central de un conjunto de datos.

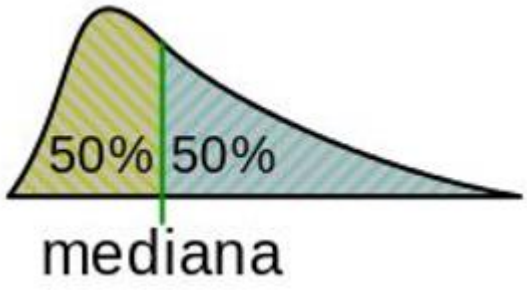



*Supongamos que tienes las edades de cinco personas: 20, 25, 30, 35, y 40 años.
Para calcular la media, sumarias todas estas edades y dividirías el resultado por la cantidad de personas.

(20+25+30+35+40) / 5=**30**

## La mediana

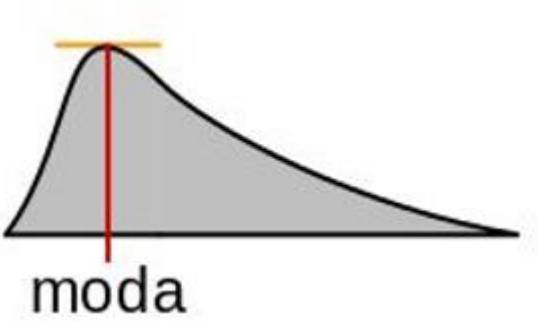
La mediana, describe la **posición central** de un conjunto de datos.




*Supongamos que tienes las edades de cinco personas: 20, 25, 30, 35, y 40 años.
Para calcular la mediaba, se ordenan los datos y el valor de la mediana es el que se encuentra en medio.
**30**

## La moda

La moda, representa el valor o los valores que **aparecen con mayor frecuencia** en un conjunto de datos. A diferencia de la media y la mediana, que se centran en la ubicación central de los datos, la moda se enfoca en identificar los valores más comunes o repetidos.




* Supongamos que tenemos el conjunto de datos: 10, 15, 20, 15, 25, 15, 30.
En este conjunto, el número 15 aparece con mayor frecuencia (tres veces), mientras que los demás valores aparecen solo una vez. Por lo tanto, la moda de este conjunto de datos es **15**.

## Desviación estándar

La desviación típica, describe cuanto **de agrupados o dispersos** están los datos alrededor de la media o del promedio en un conjunto de datos.

## La varianza

La varianza, representa **la dispersión o la extensión** de un conjunto de datos. En otras palabras, la varianza indica cuanto se alejan los valores individuales de la media del conjunto de datos. Es una medida de la variabilidad de los datos.

*Diferencia:

**Varianza ( $\sigma^2$):** Es una medida de cuanto se separan los valores individuales del promedio elevada al cuadrado.
**Desviación Estándar ( $\sigma$ ):** Es la raíz cuadrada de la varianza. Es una medida más fácil de entender y tiene la misma unidad que los datos originales.

Ejemplo:
Supongamos que estas midiendo la altura de los estudiantes en centímetros y calculas la varianza como 25. Eso significa que hay cierta variabilidad en las alturas, pero la varianza en sí misma no te da una idea clara de cuánto.

Ahora, si calculas la desviación estándar, obtendrías 5. Esto significa que, en promedio, las alturas de los estudiantes se desvían aproximadamente 5 centímetros de la media. Es más fácil entender 5 centímetros que 25 centímetros cuadrados.

Entonces, la varianza es como medir cuanto se esparcen los valores, pero al cuadrado, mientras que la desviación estándar es una versión más fácil de entender porque es la raíz cuadrada de la varianza y tiene las mismas unidades que los datos originales.

## Distribución binomial

La distribución binomial es un modelo de probabilidad que describe el número de éxitos en un número fijo de ensayos independientes, donde cada ensayo tiene solo dos resultados posibles: éxito o fracaso.

Ejemplo:
Imagina que lanzas un dado de seis caras cinco veces. La distribución binomial te ayuda a entender cuántas veces obtendrás un número específico en esos lanzamientos.

## Distribución de Poisson

La Distribución de Poisson es una distribución de probabilidad discreta que modela el número de eventos raros que ocurren en un intervalo de tiempo fijo o en un área espacial fija, dado que estos eventos ocurren de manera independiente y a una tasa promedio constante. Es especialmente útil para describir situaciones donde los eventos son infrecuentes, pero ocurren de manera aleatoria.

Ejemplo:
Imagina que trabajas en una tienda pequeña y, en promedio, recibes 4 clientes por hora. La tasa promedio de llegada de clientes por hora ( $\lambda$ ) es 4. Esta probabilidad nos dirá qué tan probable es que exactamente 2 clientes entren en la tienda en una hora p.e.

## Distribución uniforme continua

Es la probabilidad de que una variable aleatoria caiga en cualquier subintervalo del intervalo total es constante.

Ejemplo:
Imagina que estas organizando una lotería para elegir a un ganador entre los participantes. La lotería implica asignar un número a cada participante, y el numero está en el rango de 1 a 100.

En este caso, podríamos modelar la distribución de probabilidad de la selección de un numero aleatorio utilizando una Distribución Uniforme Continua en el intervalo de 1 a 100.

Entonces, si alguien te preguntara: **"¿Cuál es la probabilidad de que el número ganador sea exactamente 42?**", podrías usar la distribución uniforme continua para decir que la probabilidad es igual para todos los números

## Distribución exponencial

Es una distribución de probabilidad continua que se utiliza para modelar el tiempo que debe transcurrir hasta que ocurra un evento especifico. Es particularmente útil para eventos que ocurren de manera aleatoria e independiente a una tasa constante por unidad de tiempo.

Ejemplo:
Imagina que estas monitoreando el tiempo entre llegadas de clientes a una tienda, y observas que, en promedio, llega un cliente cada 10 minutos ( $\lambda$ = 1/10).
La distribución exponencial podría modelar el tiempo que pasara hasta que llegue el próximo cliente. La probabilidad de que el próximo cliente llegue en menos de 5 minutos seria calculada usando la función de

densidad de probabilidad de la distribución exponencial. Este cálculo proporcionaría la probabilidad de que el próximo cliente llegue en menos de 5 minutos.

## Tabla de contingencia

Las tablas de contingencia son una forma de organizar y presentar datos en dos dimensiones. Se utilizan para mostrar como dos variables diferentes están relacionadas entre sí y como se distribuyen conjuntamente.

Imagina que estas interesado en entender si hay alguna relación entre dos cosas, por ejemplo, si hay una conexión entre el género de los estudiantes y la preferencia por ciertas asignaturas en una escuela. Podrías usar una tabla de contingencia para organizar y mostrar estos datos de manera clara.

Ejemplo:
Supongamos que tienes una lista de estudiantes en una escuela y quieres ver como se distribuye la preferencia por dos asignaturas diferentes (Matemáticas e Historia) según el género (chicos o chicas).

| | Matemáticas | Historia | Total |
|---|---|---|---|
| Chicos | 20 | 15 | 35 |
| Chicas | 25 | 30 | 55 |
| Total | 45 | 45 | 90 |

## Transformación de Box-Cox

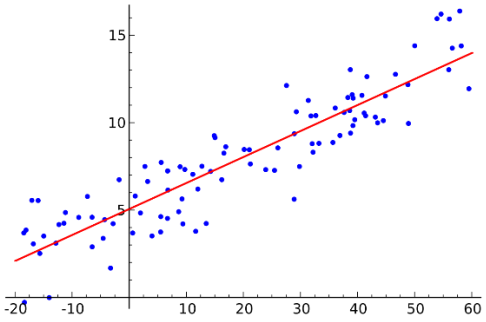
La transformación de Box-Cox es un método estadístico utilizado para estabilizar la varianza y hacer que **los datos se aproximen más a una distribución normal**. Es especialmente útil cuando los datos no cumplen con los supuestos de homocedasticidad (varianza constante) y normalidad en ciertos análisis estadísticos.

Ejemplo:
Supongamos que tienes un conjunto de datos sesgado positivamente (con muchos valores pequeños y pocos valores grandes). Aplicar la transformación de Box-Cox puede ayudar a hacer que los datos sean más simétricos y estabilizar su varianza, lo que facilita la aplicación de técnicas estadísticas que asumen normalidad.

### Cuando usar Transformación de Box-Cox

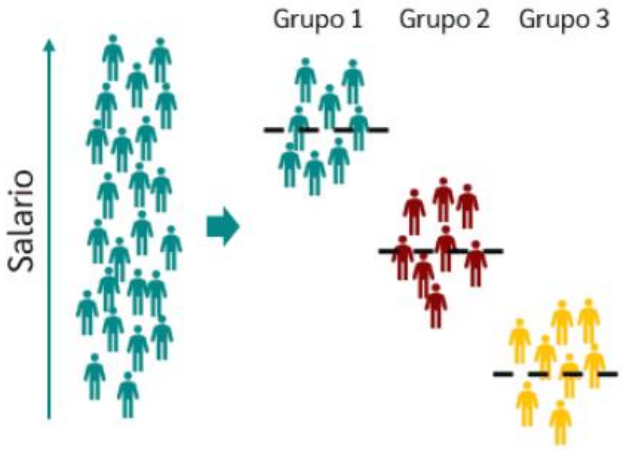
Regresión Lineal:

En la regresión lineal, la transformación de Box-Cox puede ser útil para estabilizar la varianza y cumplir con los supuestos de homocedasticidad y normalidad de los residuos.

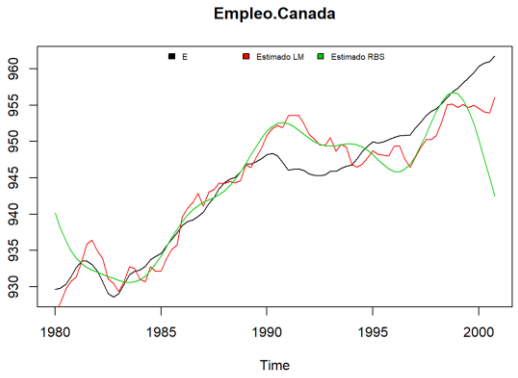



Análisis de Varianza (ANOVA):

ANOVA asume igualdad de varianzas entre los grupos. La transformación de Box-Cox puede ayudar a cumplir con este supuesto cuando las varianzas son heterogéneas.




Análisis de Series Temporales: es una técnica estadística que se utiliza para **analizar patrones temporales** en conjuntos de datos. En lugar de considerar observaciones independientes, como en muchos otros tipos de análisis estadísticos, el análisis de series temporales se centra en **datos recopilados a lo largo del tiempo, generalmente en intervalos regulares**.



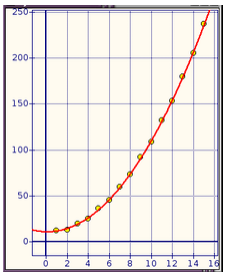

Control Estadístico de Procesos (CEP): Es una técnica utilizada en la gestión de la calidad para monitorear y controlar la variabilidad de un proceso de producción. El objetivo principal del CEP es asegurar que un

proceso se mantenga dentro de límites específicos y produzca productos o servicios consistentes y de alta calidad. Este enfoque se basa en la recopilación y análisis de datos a lo largo del tiempo.

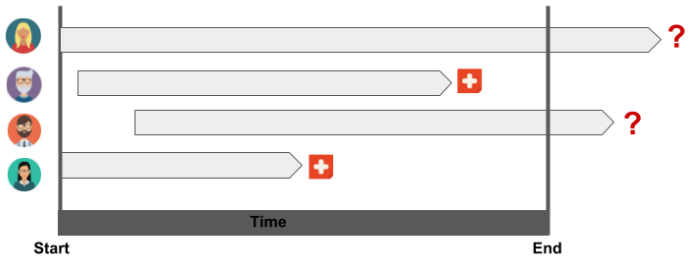
En el CEP, donde se monitorean y controlan procesos industriales, la transformación de Box-Cox puede ser útil para estabilizar la variabilidad del proceso.

Modelos de Regresión No Lineal:

Puede ser aplicada cuando se ajustan modelos de regresión no lineales y los supuestos de normalidad y homocedasticidad no se cumplen.



Análisis de Supervivencia: Es una técnica estadística utilizada para analizar el tiempo hasta que ocurre un evento especifico. Este tipo de análisis es comúnmente aplicado en estudios médicos, biológicos, sociales y de ingeniería para examinar el tiempo hasta que ocurre un resultado de interés, como la muerte, la falla de un dispositivo, o cualquier otro evento cuyo tiempo de ocurrencia sea relevante.




En el análisis de supervivencia, la transformación de Box-Cox puede ayudar a cumplir con los supuestos relacionados con la distribución de los tiempos hasta el evento.