



PI DATA STRATEGY & CONSULTING

“Reject, or fail to reject... There is no accept”

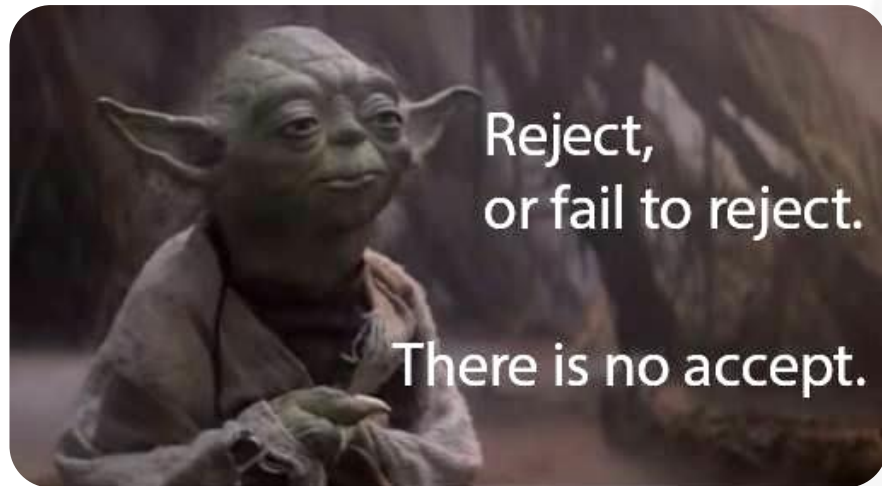


Nindirí Armenta

Data Scientist, Pi Data Strategy & Consulting



“Reject, or fail to reject... There is no accept”



Estadística para Data Science
3er encuentro

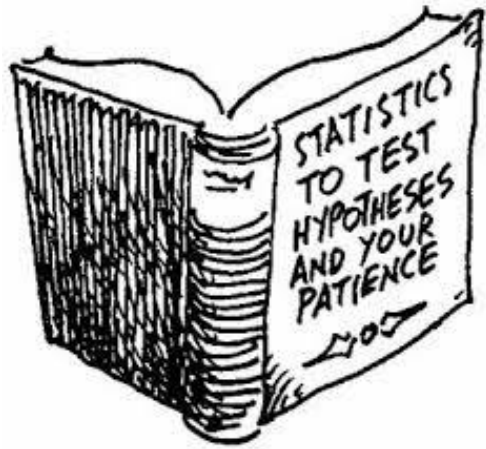


Agenda

- Estadística No Paramétrica
- Métodos para probar una hipótesis (review)
- Pasos para realizar un test estadístico (review)
- Hipótesis: H_0 y H_1
- El P-value
- ¿Cuándo aplicar test no paramétricos?
- Tipos de test estadísticos (review)
- Test estadísticos no paramétricos
- Selección del test estadístico no paramétrico
- Ventajas y desventajas
- Ejemplos con Python en Jupyter Notebook



Estadística No Paramétrica



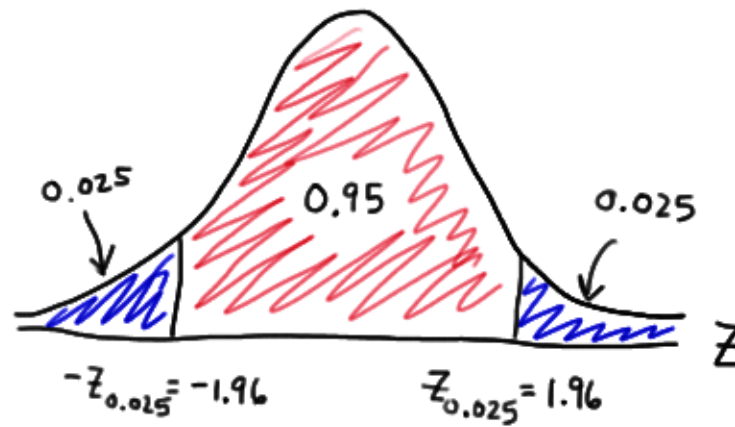
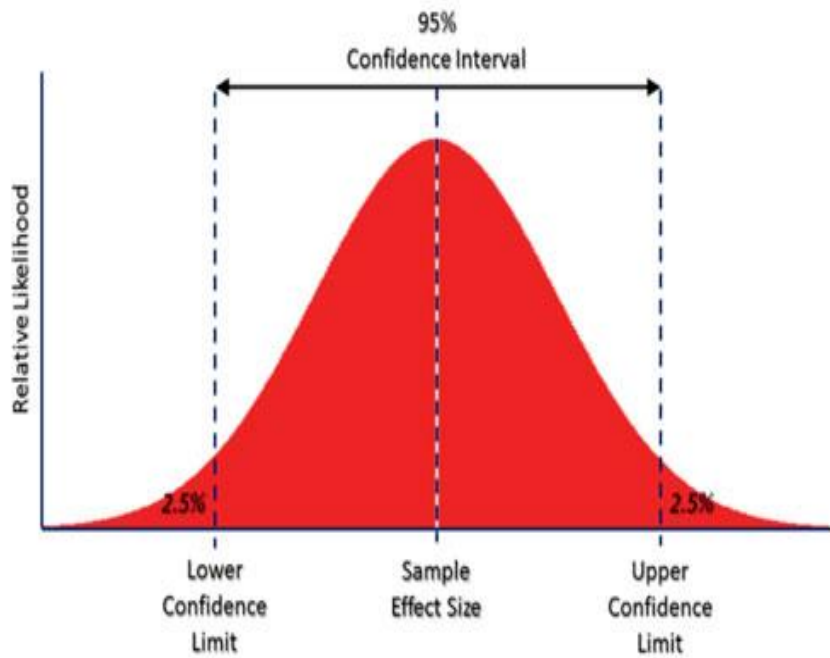
Las estadísticas no paramétricas son aquellos métodos que no asumen una distribución específica de los datos.

Fueron desarrollados para su uso con datos ordinales o de intervalo, pero en la práctica también se pueden usar con una clasificación de observaciones de valor real en una muestra de datos en lugar de los valores de observación en sí.



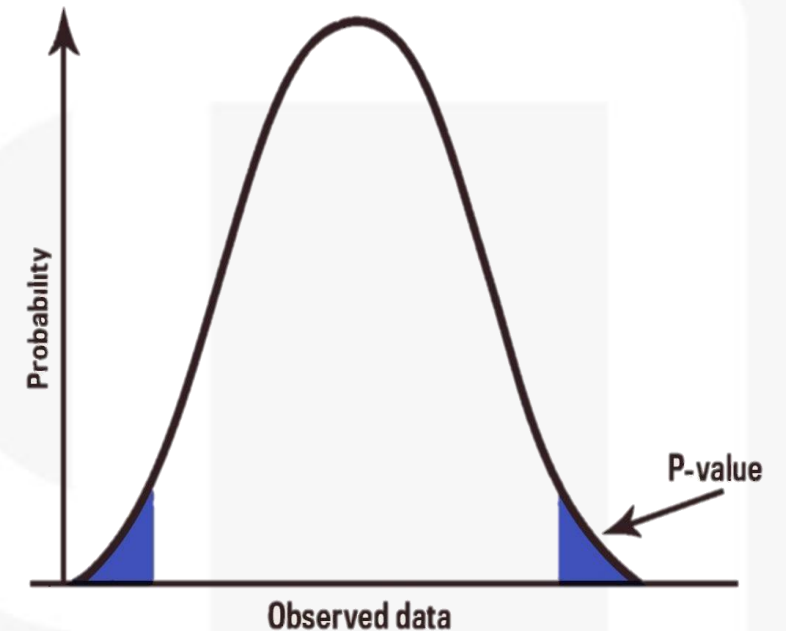
Métodos para probar una hipótesis

Intervalo de confianza



Estadístico de Prueba

P-valor



Pasos para realizar un test estadístico

Planteamiento de hipótesis:

Ho: Hipótesis nula - H1: Hipótesis alternativa

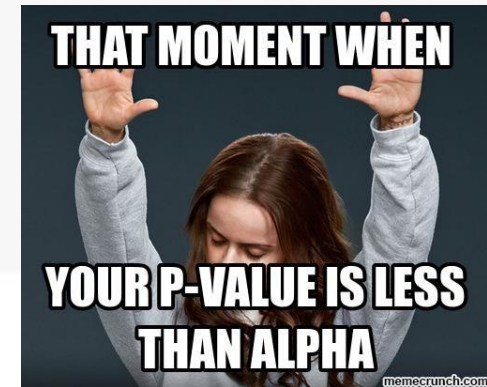
Elegir un nivel de significancia (α : alpha)

Selección del test estadístico

(selección del estadístico de prueba)

Comparar el p-valor con α

Realizar la conclusión correspondiente



Hipótesis H0 y H1

- $H_0: \theta = \theta_0 \rightarrow$ La media del diámetro de la rosca de los bulones es de 3/8"
- $H_0: \theta \leq \theta_0 \rightarrow$ La estrategia de marketing A tuvo mayor impacto que la estrategia B
- $H_0: \theta \geq \theta_0 \rightarrow$ Existe una menor proporción de estadounidenses que de italianos que prefieren beber champaña en las bodas

- $H_1: \theta \neq \theta_0 \rightarrow$ La medición de la contaminación atmosférica con ambos instrumentos es distinta
- $H_1: \theta < \theta_0 \rightarrow$ Un nuevo aditivo reducirá el tiempo de secado de una pintura
- $H_1: \theta > \theta_0 \rightarrow$ El contenido de alquitrán de la marca de cigarrillos B es mayor a la de A



El P-value

Conocido también como:

p-valor / valor p / p-value

Por ejemplo:

H_0 : dos muestras fueron extraídas de una población con la misma distribución

P-value: Probabilidad de observar H_0 dado que esta es cierta

$p < \alpha$: Rechazo $H_0 \rightarrow$ Las muestras provienen de distintas distribuciones

$p > \alpha$: No Rechazo $H_0 \rightarrow$ Tienen distribución similar



¿Cuándo aplicar test no paramétricos?



Las pruebas no paramétricas se pueden aplicar a situaciones en las que:

- No se conoce la distribución de probabilidad de los datos.
- Los datos constituyen valores de tipo ordinales o rangos.
- Existe valores atípicos en los datos.

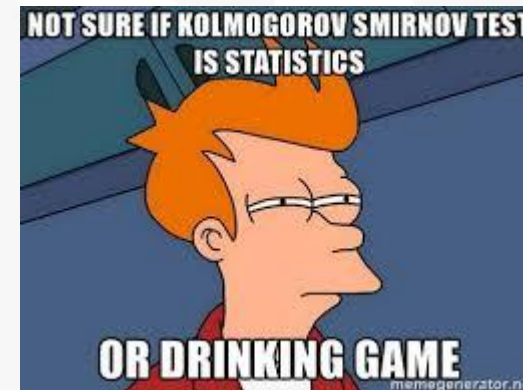
Tipos de test estadísticos

Paramétricos

- Media
- Varianza
- Proporción
- Comparación de medias
- Comparación de varianzas
- Comparación de proporciones

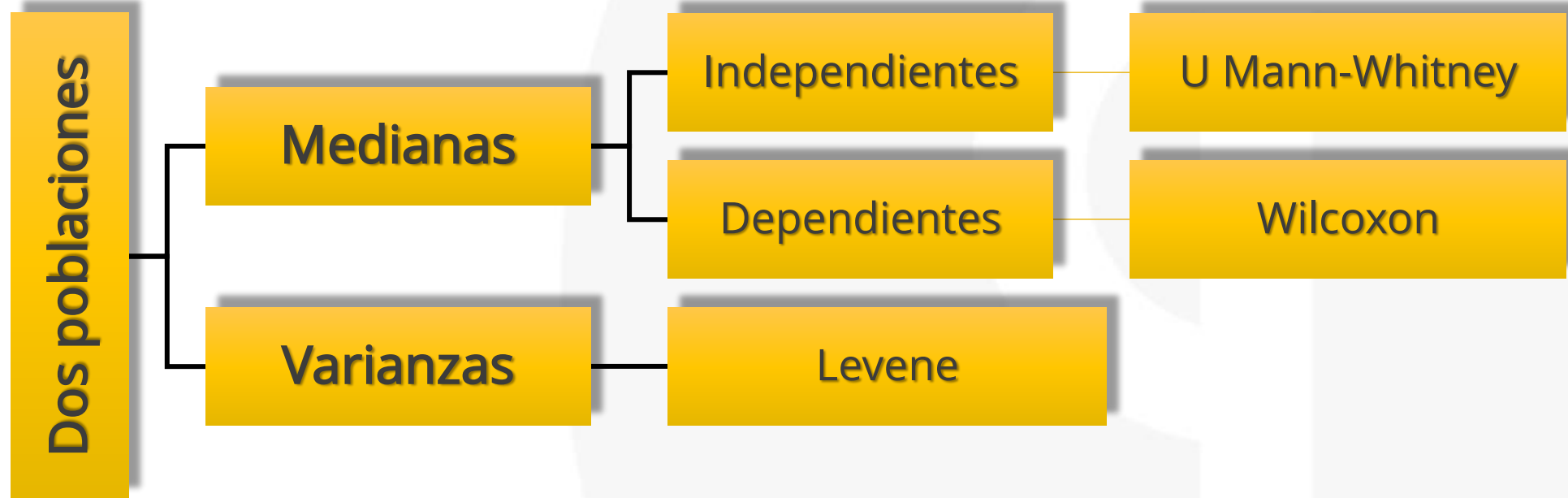
No Paramétricos

- Comparación de medias
- Comparación de varianzas
- Bondad de ajuste



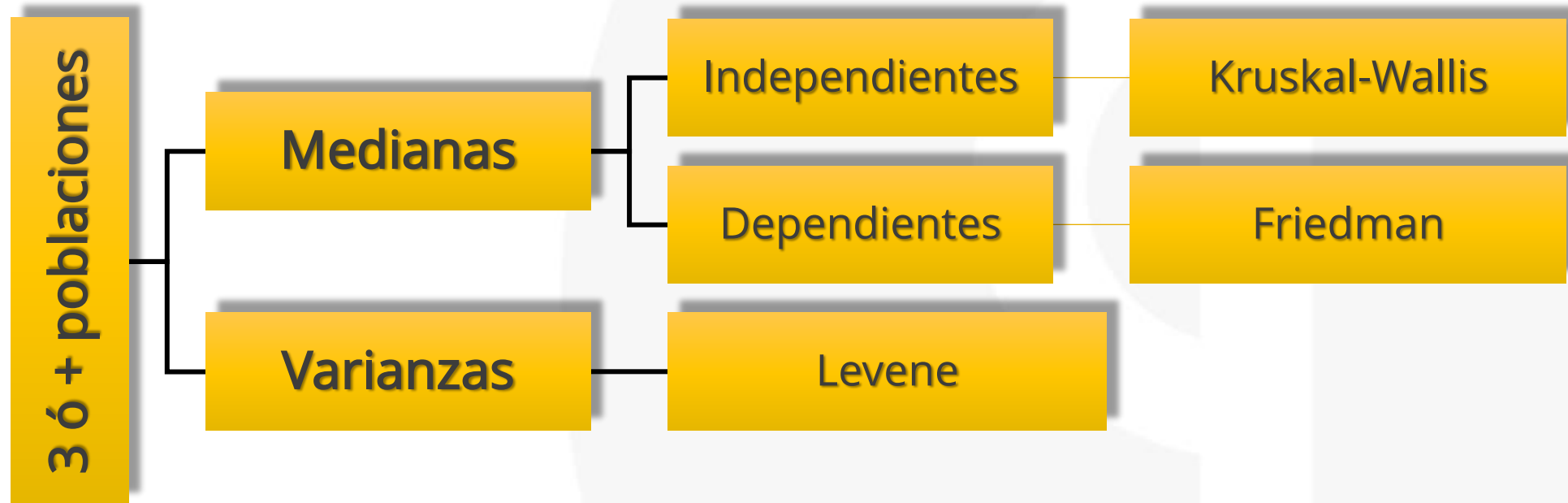
Test estadísticos no paramétricos

Comparación entre 2 parámetros
poblacionales

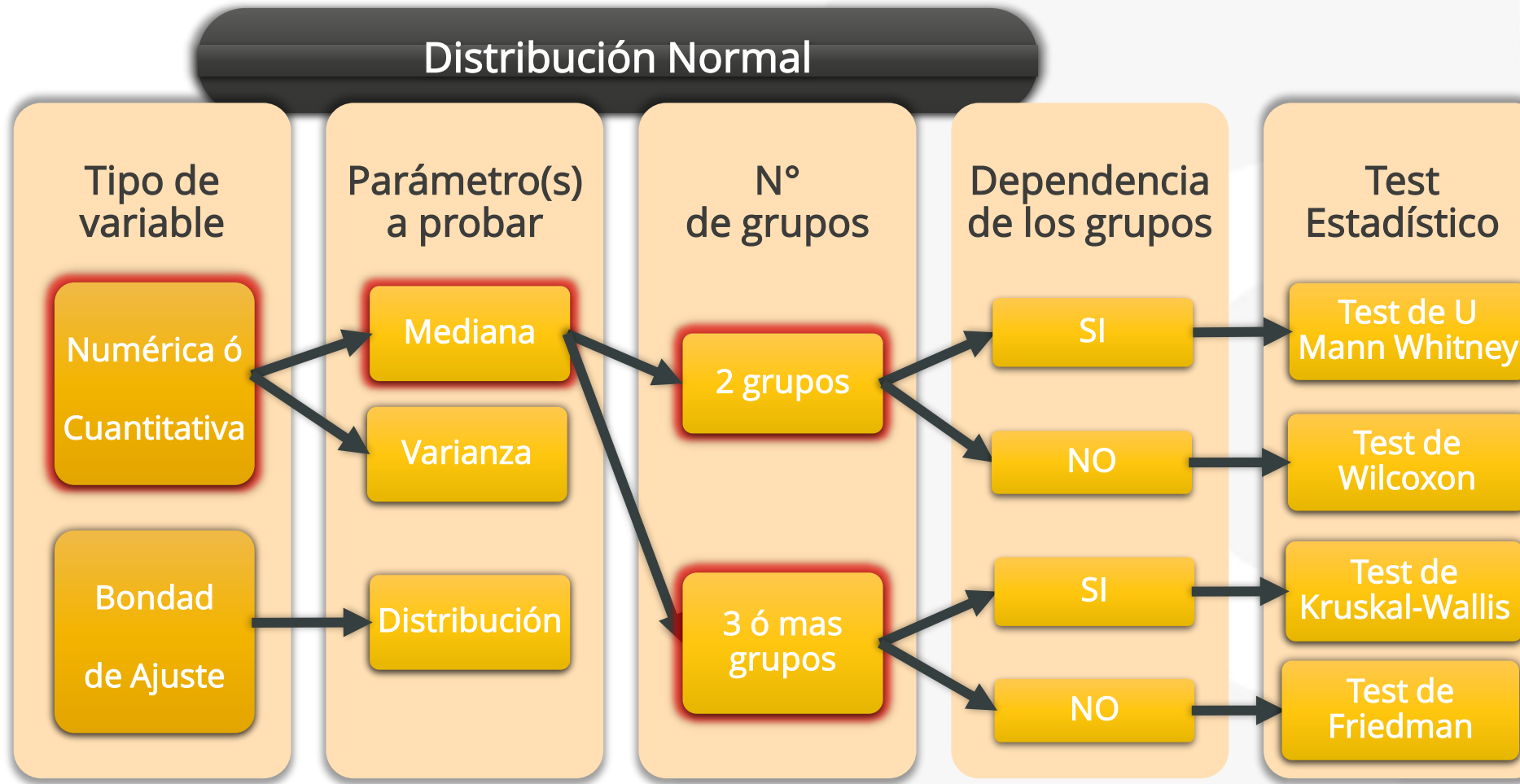
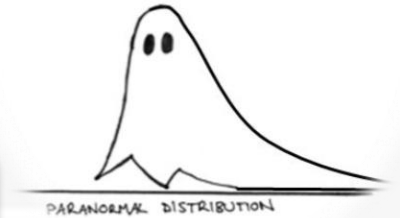
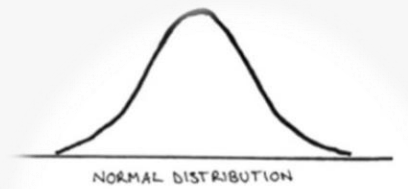


Test estadísticos no paramétricos

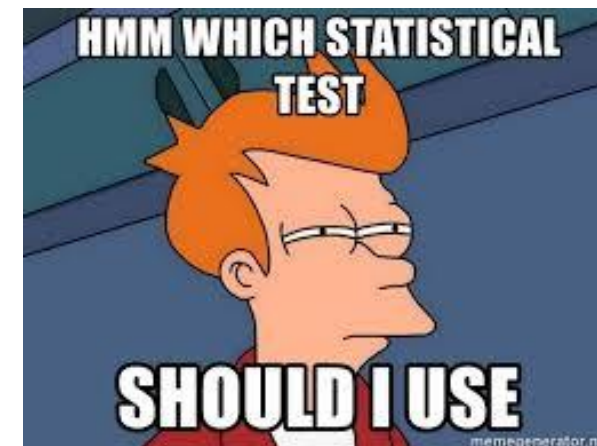
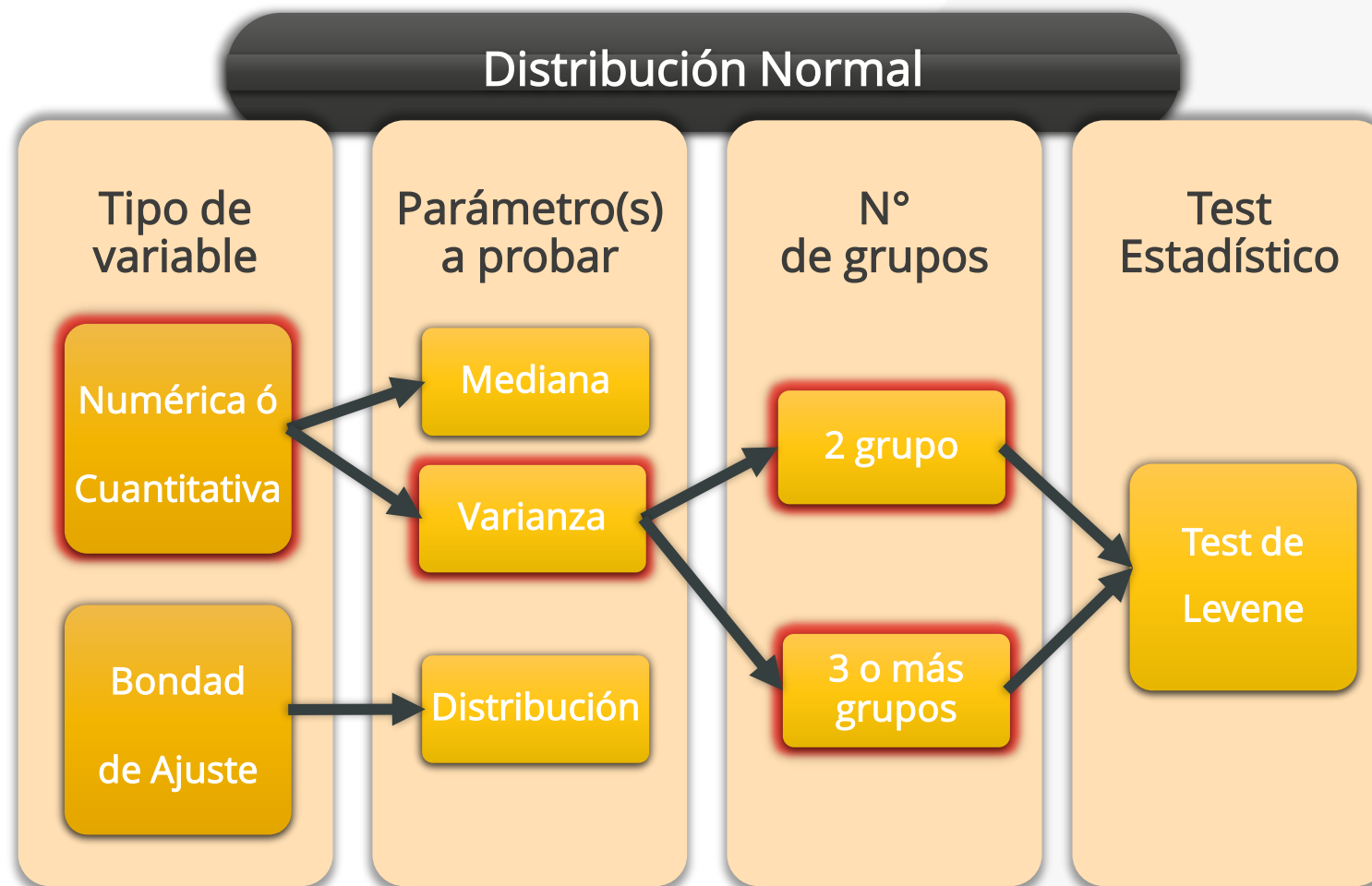
Comparación de 3 o más parámetros poblacionales



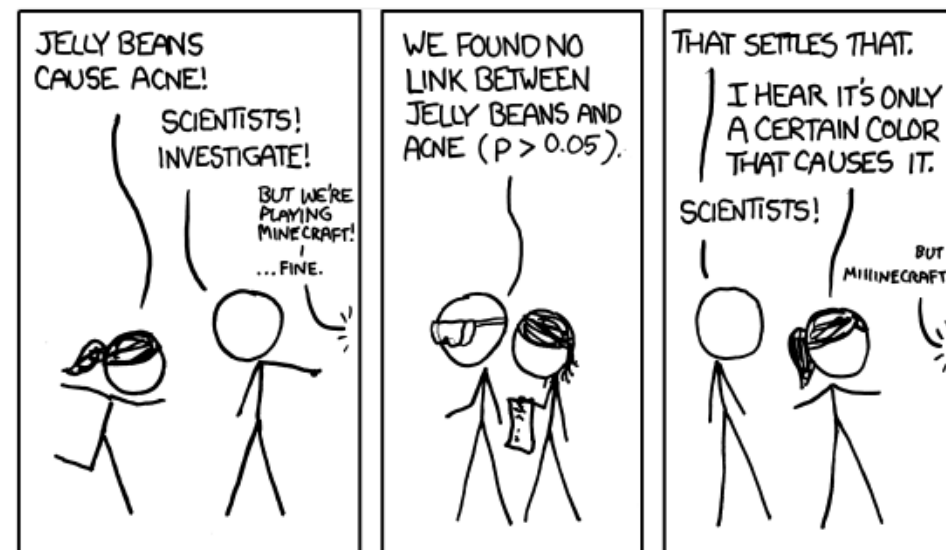
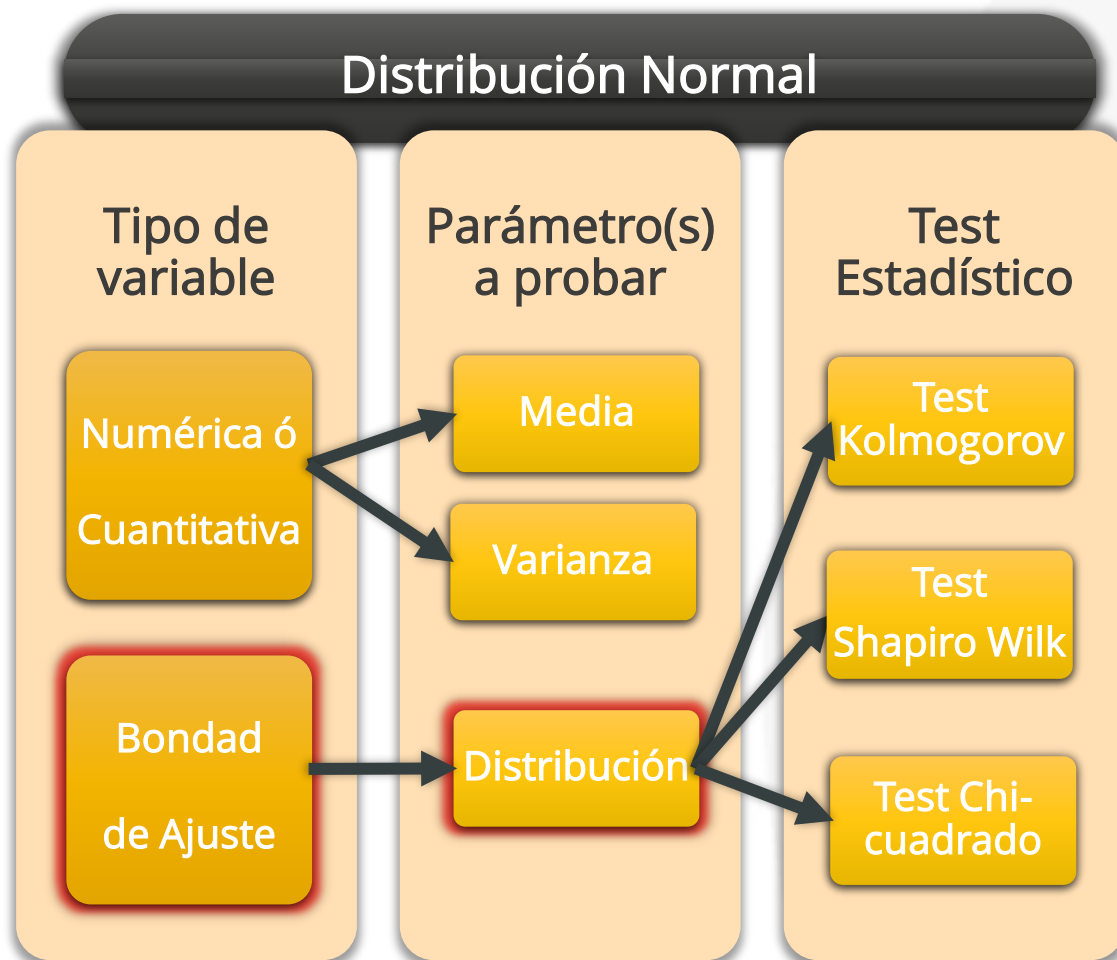
Selección del test estadístico



Selección del test estadístico



Selección del test estadístico



Ventajas

- Ofrecen resultados precisos incluso cuando el tamaño de la muestra es pequeño.
- Son más poderosas que las pruebas paramétricas cuando no se cumple el supuesto de normalidad.
- Son adecuados para todos los tipos de datos, como nominal, ordinal, intervalo o los datos que tienen valores atípicos.



Desventajas

- Si existe una prueba paramétrica para un problema, entonces el uso de pruebas no paramétricas podría arrojar respuestas inexactas.
- Si realmente existe un efecto sobre la media de una distribución de valores, es más probable que no lo detecte.



■ Algunas consideraciones finales

- Si la media es una mejor medida y tiene un tamaño de muestra suficientemente grande, una prueba paramétrica suele ser la mejor opción y la más poderosa.
- Si la mediana es una mejor medida, considere una prueba no paramétrica independientemente del tamaño de su muestra.
- Los análisis paramétricos pueden analizar distribuciones no normales para tamaños de muestras grandes.
- Las pruebas no paramétricas son válidas cuando nuestro tamaño de muestra es pequeño y sus datos son potencialmente no normales.



Ejemplos con Python en Jupyter Notebook



Link de referencia para ampliar el tema:

https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Nonparametric/BS704_Nonparametric_print.html

