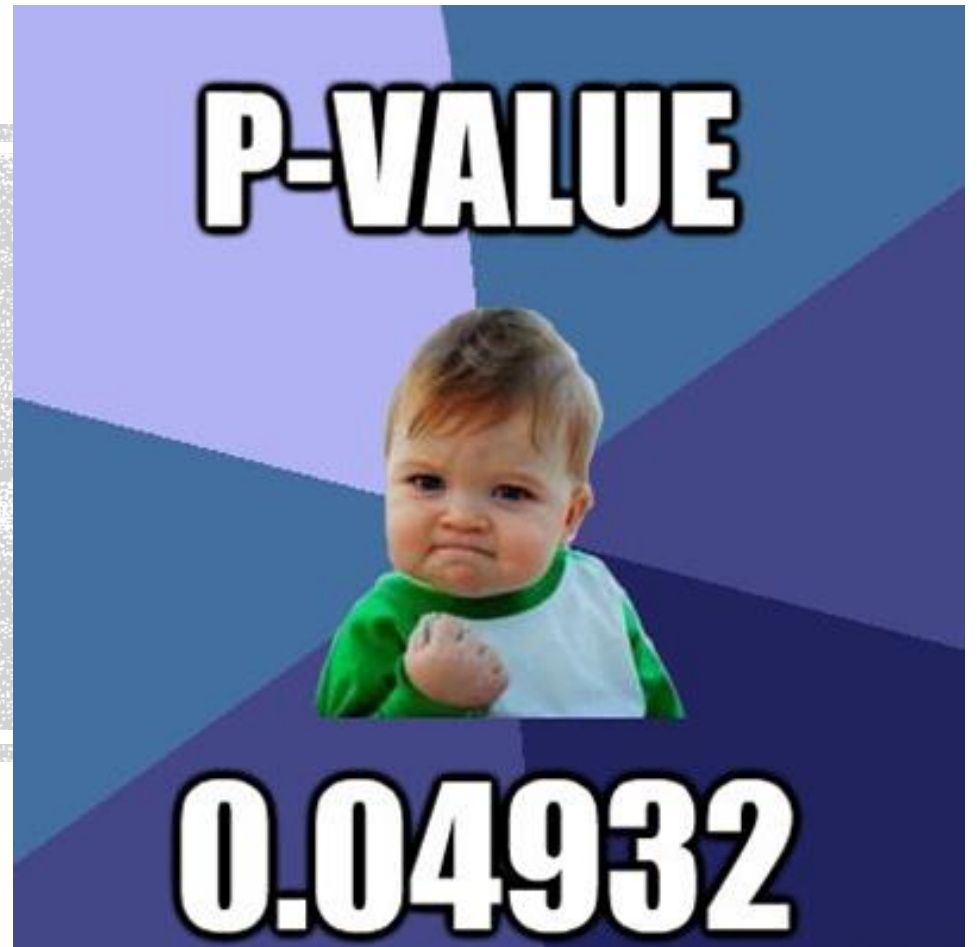


# PRUEBAS DE HIPÓTESIS



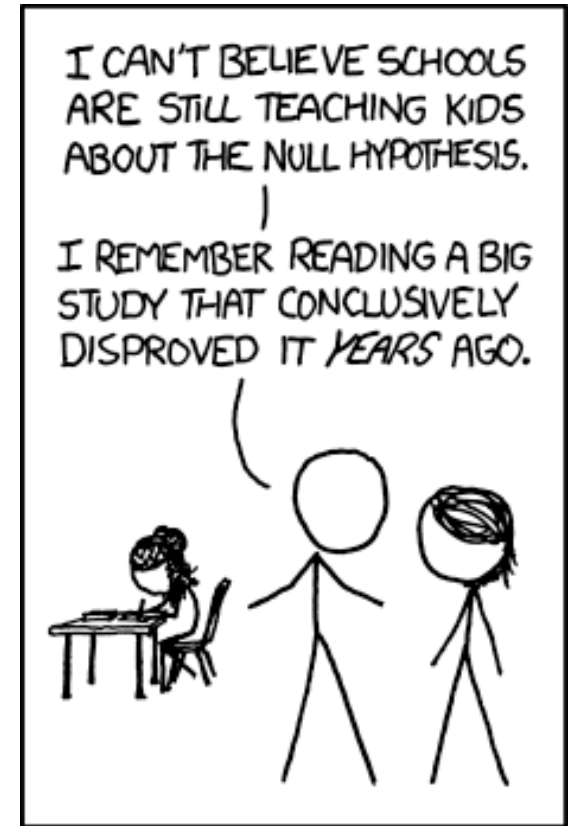
# PRUEBA DE HIPÓTESIS

- **Prueba de hipótesis:** proceso de sometimiento de las hipótesis investigativas a un escrutinio, basado en los datos disponibles
  - Herramienta de inferencia que aplica la estadística a problemas reales, usada cuando se necesita tomar decisiones sobre una **población**, basadas solamente en una **muestra** de la misma → **Generalización**
  - Las hipótesis se refieren a uno o varios parámetros de la **población**
  - Los datos de la **muestra** pueden aportar evidencia **a favor o en contra** de las hipótesis. Idealmente la muestra representa las características de la población, pero no todas las muestras aleatorias son ideales (→ **azar, incertidumbre, confianza**)
  - Las hipótesis pueden ser apoyadas o refutadas, mas no se puede establecer si son verdaderas o falsas
- ¿Que tanto puede diferenciarse una muestra de su población? Usamos **probabilidades** para cuantificarlo (→ el **valor-p** que se obtiene como resultado de la prueba)



# PRUEBA DE HIPÓTESIS

- **Definición de hipótesis:** una proposición que se puede probar.
- **Hipótesis nula ( $H_0$ ):** usualmente, el contrario de lo que se busca
  - Tratan situaciones de igualdad ( $=, \leq, \geq$ ), estabilidad, independencia
  - Ejemplo: “los resultados del modelo A son iguales o peores a los del modelo B”
- **Hipótesis alternativa ( $H_a$ ):** usualmente, lo que se busca
  - Tratan situaciones de cambio, diferencia ( $\neq, <, >$ ), inestabilidad, dependencia
  - Ejemplo: “los resultados del modelo A son mejores a los del modelo B”
- **Valor de significancia  $\alpha$ :** sirve como un umbral de probabilidad que se compara con el **valor-p** que resulta de la prueba de hipótesis. Es igual a 1-confianza de la prueba.



# PRUEBA DE HIPÓTESIS

- Se estudia la hipótesis **nula**  $H_0$ , se trata de comprobar su validez
- **Conclusión de una prueba de hipótesis:** las pruebas de hipótesis tienen como resultado un **valor-p**, que se compara con la **significancia**  $\alpha$  establecida para el test:
  - Si  $\alpha < \text{valor-p}$ , **se rechaza la hipótesis nula  $H_0$** , en favor de la hipótesis alternativa  $H_a$ . La evidencia permite contradecir  $H_0$  con un nivel de confianza de  $(1 - \alpha)$ .
  - Si  $\alpha > \text{valor-p}$ , **no se puede rechazar la hipótesis nula  $H_0$** . La evidencia no es suficiente para contradecir  $H_0$ , en caso tal, no se puede concluir nada al respecto de la hipótesis alternativa. Puede que con una muestra mas representativa de las poblaciones se pueda llegar a rechazar  $H_0$ . Esta conclusión también aporta conocimiento.
  - **Nunca se llega a una aceptación** de  $H_0$ .
- Las regiones de rechazo corresponden a regiones por fuera de intervalos de confianza de una distribución dada.



# PRUEBA DE HIPÓTESIS

- La conclusión de la prueba de hipótesis corresponde a la realidad?

		Realidad	
		H0 es verdadera	H0 es falsa
Conclusión de la prueba de hipótesis	Rechazar H0	Error Tipo I	Decisión correcta
	No rechazar H0	Decisión correcta	Error Tipo II

- Tipos de error** de las pruebas de hipótesis

- Tipo I (Falso positivo):  $\alpha$  = probabilidad de rechazar H0 cuando no se debería de haber rechazado. El valor utilizado por defecto es 5%, que implica una inferencia con 95% de confianza
- Tipo II (Falso negativo):  $\beta$  = probabilidad de no rechazar H0 cuando se debería de haber rechazado  
**No nos interesa en este momento**



# PRUEBA DE HIPÓTESIS

- Se quiere evaluar si una iniciativa educativa propuesta en una comunidad desfavorecida tuvo éxito o no

Estudiante	Notas evaluación	
	Antes	Después
1	20	23
2	23	34
3	35	33
4	14	31
5	18	22
6	23	20
7	21	32
8	32	42
9	54	45
10	35	30
11	29	25
12	37	38
13	31	32
14	21	23
15	28	28
Promedio	28,1	30,5

- Se realiza un examen de los estudiantes antes y después de la aplicación de la iniciativa educativa
- Se puede usar el promedio de las notas antes y después de la iniciativa: 28,1 y 30,5 respectivamente
- ¿Se puede decir que mejoraron las notas?
- ¿Con que nivel de confianza?
- Estadísticamente, con un nivel de confianza del 95%, no se podría decir que las notas de los estudiantes mejoraron (aunque si, con un nivel de confianza del 90%)
- Un objetivo de la iniciativa era mejorar en al menos 2 puntos en la evaluación, con un nivel de confianza del 90%, ¿se cumplió?



# PRUEBA DE HIPÓTESIS

Para evaluar la facilidad de acceso a una comunidad desfavorecida, se quiere establecer si la nueva carretera (B) es mejor que el antiguo camino (A). Se tomaron varias mediciones de los tiempos de los trayectos.

Tiempos ruta A	Tiempos ruta B
73	68
80	66
83	65
69	73
80	80
73	65
82	72
67	74
82	82
77	74
82	79
67	70
65	58
77	82
77	
80	
69	
75	
77	
75	

Promedio	
A	75,5
B	72,0

- ¿Tiene sentido comparar los tiempos de las dos rutas uno a uno?
- ¿Se puede decir que la ruta B permite reducir los tiempos en al menos 1 minuto con respecto a la ruta A, con un nivel de confianza del 90%?
- No



# PRUEBA DE HIPÓTESIS

Analizando las preferencias políticas de una población específica se encuentran diferencias profundas entre los hombres y las mujeres.

- ¿Se puede decir que las mujeres tienden a votar más por los liberales y los hombres por los conservadores?

Observado	Hombre	Mujer	Total
Conservador	40	58	98
Liberal	32	130	162
Total	72	188	260

- Si, con un nivel de confianza superior al 99%





# TALLER BRAIN SIZE + CHURN

- Desarrolle el taller de brain size
- Con los datos del Churn responda a la siguiente pregunta:
  - ¿Se puede decir que los clientes que se fueron ganan mas que los que se quedaron con una confianza del 95%?
- Establezca la prueba de hipótesis a realizar, hipótesis nula y alternativa y posibles conclusiones
- Ejecútela y concluya

