

# Proceso de Inferencia Estadística

**Prueba de significación de la hipótesis nula** (*Null Hypothesis Significance Testing*, NHST) introducida por Fisher en 1925.



Larga tradición, muy utilizada pero también muchas críticas desde su aparición

En el **97%** de los artículos publicados en 10 revistas internacionales de Psicología se utiliza la prueba de significación de la hipótesis nula como la técnica para comprobar las hipótesis de investigación (Cumming, Fidler, Leonard, Kalinowski, Christiansen y cols. 2007).

## CRÍTICAS → CAMBIOS IMPORTANTES EN:

Normas de publicación científica desde finales del siglo XX

Ha dado lugar a la denominada “reforma estadística”



# DEBATE EN LA CIENCIA

-**Defensores** del procedimiento NHST.

-Mantener el procedimiento, pero **con cambios** como por ejemplo situar el punto de corte del alfa en **.005**.

-Debe **abandonarse** el procedimiento NHST. Hay algunas revistas que ya lo han hecho.

-**Acompañar** el procedimiento NHST **con otros estadísticos** como el Tamaño del Efecto y su intervalo de confianza o **reemplazarla** por la técnica bayesiana y el Factor Bayes (BF).

## Educación y reeducación estadística



## Tamaño del Efecto y su IC

-Acompañar el procedimiento NHST o contraste de hipótesis con el **Tamaño del Efecto y su intervalo de confianza** e interpretar ambos.

- *American Psychological Association (APA)*
- *American Educational Research Association (AERA)*
- *International Committee of Medical Journal Editors (ICMJE)*
  
- *Medicina Basada en la Evidencia (MBE) / Práctica Basada en la Evidencia (PBE)*
- *Recomendaciones o protocolos de revisión (CONSORT, STREND)...*



# Reforma Estadística

- La reforma estadística destaca la importancia de reflexionar sobre el **proceso del diseño** donde se requieren cuestiones de investigación precisas cuyo estudio se realizará con el diseño de investigación más apropiado.
- La identificación y operacionalización adecuada de las **variables** que se van a medir y los procedimientos de medida adecuados.
- Los **análisis estadísticos** más apropiados para analizar los datos obtenidos.
- Elaborar **informes** de los resultados estadísticos que sean relevantes para obtener conclusiones sustanciosas y que sean **transparentes**.
- La elaboración y uso de las **revisiones sistemáticas y meta-análisis**.
- La reforma estadística también destaca la importancia de la **educación y re-educación estadística** (Swaminathan y Rogers, 2007) y el desarrollo del **pensamiento meta-analítico** entre los investigadores para contextualizar los efectos.

**El cambio no es sencillo**



## EN CONCLUSIÓN

Comprender el proceso de contraste de hipótesis es fundamental no ya para el **investigador** o **investigadora** que por supuesto lo debe conocer y saber aplicar sino también para el **lector** o **lectora de informes y artículos de naturaleza empírica** ya que es el procedimiento más utilizado de análisis estadístico descrito en dichos informes. **Saber interpretar los valores  $p$  de probabilidad y los intervalos de confianza de un estadístico son competencias básicas del profesional de todas aquellas disciplinas donde se aplica la inferencia estadística.**

- Reflexionar sobre el proceso de contraste de hipótesis estadísticas: **FALACIAS ESTADÍSTICAS**
- SESGO DE PUBLICACIÓN**: repercute en la acumulación válida del conocimiento científico
- Necesidad de la **Re-educación estadística** de docentes y profesionales



# Proceso de contraste de hipótesis estadísticas

**Implica un proceso de inferencia estadística:**

La *inferencia estadística* es un proceso que supone obtener conclusiones sobre un valor poblacional (*parámetro*) a partir de un valor obtenido con una muestra aleatoria (*estadístico*). Proceso de inferencia.

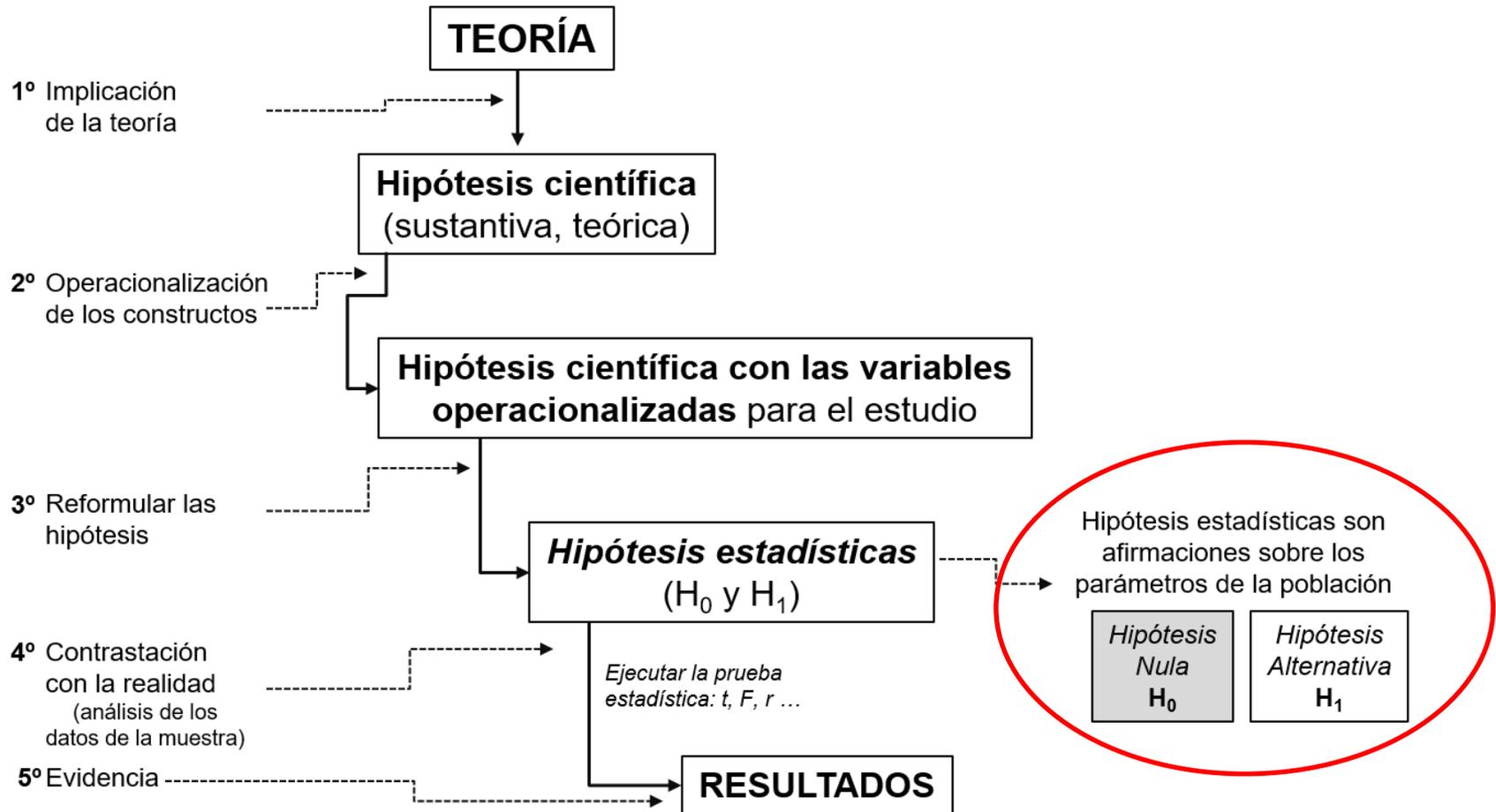
Errores estadísticos. [Manual \(2022\)](#): páginas 232-235 y **Anexo 1. Breve explicación de conceptos fundamentales de diseño de investigación**. Hay explicaciones cada vez que se realiza un contraste de hipótesis.



## Errores estadísticos:

# Error de Tipo I ( $\alpha$ ) y Error de Tipo II ( $\beta$ ). Nivel de confianza y potencia estadística

## Modelo tradicional de decisión estadística: Procedimiento de significación de la hipótesis nula (NHSP)



Al final del experimento se observa si la diferencia detectada en la variable dependiente medida Y entre el grupo experimental  $a_1$  y el grupo control o de comparación  $a_2$  **¿es o no es estadísticamente significativa?**

El **RESULTADO observado** puede deberse:

1. Al proceso del muestreo aleatorio: diferencias de muestreo, diferencias individuales...).
2. Un EFECTO real de la intervención o tratamiento.



**En definitiva**, en un contraste de hipótesis se compara **lo que se observa en el experimento** (el resultado o valor empírico obtenido con la ejecución de la prueba estadística) **con aquello que se esperaría si la hipótesis nula fuese cierta** (valor tabular o teórico del estadístico y que se recoge en las tablas de la distribución muestral de los estadísticos que se encuentran en los manuales de estadística y de diseños de investigación, en el SPSS).



Los resultados de las pruebas estadísticas facilitan al investigador o investigadora una **herramienta de decisión objetiva** (la **significación estadística o valor  $p$  de probabilidad del estadístico estimado**) y a partir de se deberá interpretar **qué causó probablemente el resultado de la prueba estadística aplicada**, valorando la calidad de su proceso de diseño de investigación, la validez de sus hallazgos y los resultados estadísticos obtenidos. Y, si el diseño se planificó y ejecutó correctamente entonces podrá atribuir el efecto o la relación detectada entre las variables a su hipótesis alternativa. *Porque... hipótesis alternativas que podrían explicar ese efecto o relación hay muchas.*



**-Resulta extraño,** pero el procedimiento de inferencia estadística tiene como objetivo comprobar que **el azar (el error muestral que se produce en la selección de las muestras o las diferencias individuales aleatorias) no es la posible causa de las diferencias encontradas;**



-Idea que desde el comienzo *no cree el investigador o investigadora* que suceda dado que para ellos, el tratamiento debe tener un efecto tal y como plantea en la hipótesis científica, teórica o sustantiva (diseño de superioridad).



En términos de Shaver (1993), la **prueba de significación estadística** sólo responde a la cuestión de **si un resultado es probable que ocurra bajo la hipótesis nula.**

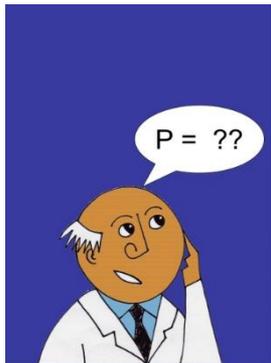


Tabla 2. *Hipótesis Nula/Hipótesis Alternativa*

<i>HIPÓTESIS NULA</i>	<i>HIPÓTESIS ALTERNATIVA</i>
<p>Asume que no hay diferencias entre las poblaciones de los grupos. Es decir, el efecto en la población es cero.</p> <p>Por lo tanto:</p> <ol style="list-style-type: none"><li>1) Las pequeñas diferencias encontradas en las poblaciones serán debidas al azar</li><li>2) Las grandes diferencias ocurrirán en muy pocas ocasiones y serán también por azar</li></ol>	<p>Hay diferencias entre las poblaciones de los grupos. El efecto en la población es diferente de cero.</p> <p>Por lo tanto:</p> <ol style="list-style-type: none"><li>1) Las pequeñas o grandes diferencias están provocadas por el efecto del tratamiento</li></ol>



# Probability & Statistical Significance Explained

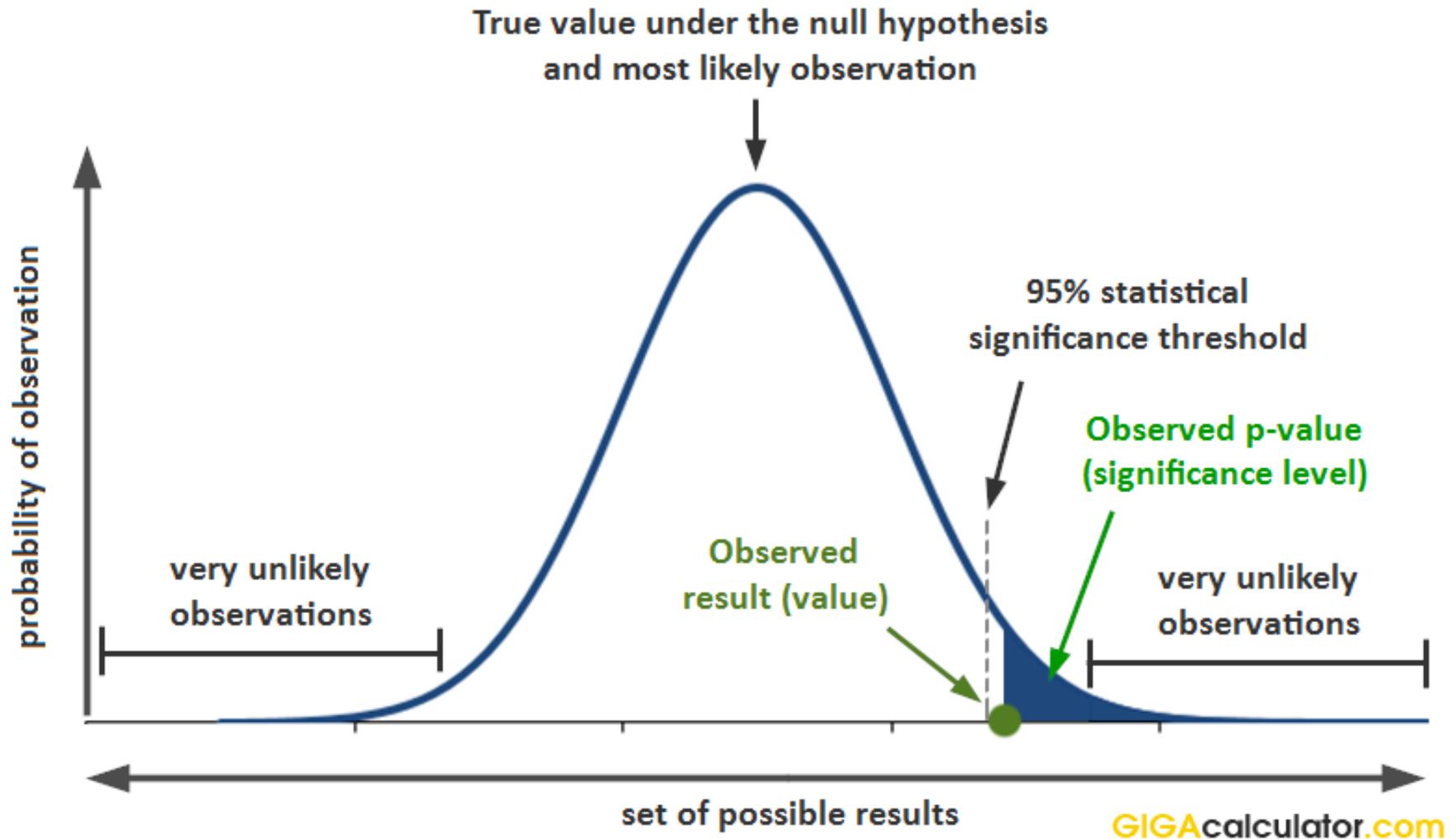


Tabla 10. *Hipótesis Nula/Hipótesis Alternativa*

<i>HIPÓTESIS NULA</i>	<i>HIPÓTESIS ALTERNATIVA</i>
<p>Asume que no hay diferencias entre las poblaciones de los grupos. Es decir, el efecto en la población es cero. Por lo tanto:</p> <ol style="list-style-type: none"> <li>1) Las pequeñas diferencias encontradas en las poblaciones serán debidas al azar</li> <li>2) Las grandes diferencias ocurrirán en muy pocas ocasiones y serán también por azar</li> </ol>	<p>Hay diferencias entre las poblaciones de los grupos. El efecto en la población es diferente de cero. Por lo tanto:</p> <ol style="list-style-type: none"> <li>1) Las pequeñas o grandes diferencias están provocadas por el efecto del tratamiento</li> </ol>

Pág. 104

↑  
**H<sub>0</sub>**

↑  
**H<sub>1</sub>**

Cohen (1994) llamó a la hipótesis nula de efectos cero la 'nil hypothesis' frente a otras hipótesis nulas de efectos diferentes de cero ('non-nil hypothesis'). Una 'nil hypothesis' formula que la correlación entre las medias ( $\rho = 0$ ) o su diferencia ( $\mu_1 - \mu_2 = 0$ ) es literalmente cero y es el tipo de hipótesis que se plantea de forma generalizada en los diseños tradicionales

La meta del procedimiento de decisión estadística es por lo tanto evaluar si se puede o no se puede rechazar la hipótesis nula ( $H_0$ ) con un cierto grado de plausibilidad. El proceso de contraste estadístico implica los siguientes pasos (ver Figura 24):

## Errores estadísticos

Pero qué criterio utilizar para decir que el valor  $p$  de probabilidad obtenido en un experimento es 'pequeño'. Estamos hablando del nivel de significación estadística.

Dentro del modelo de decisión estadística se trabaja con un cierto grado de incertidumbre estadística elegido por el investigador a priori mediante un criterio de decisión probabilística conocido como alfa ( $\alpha$ ) o error de Tipo I ('probabilidad de rechazar la hipótesis nula siendo realmente verdadera', es decir, cuando la diferencia encontrada entre las poblaciones está realmente provocada por el azar) (ver Tabla 11).

Tabla 11. *Errores estadísticos y decisiones correctas*

<i>Decisión estadística</i>	<i>Realidad del fenómeno en la población</i>	
	$H_0$ cierta $p(H_0) = 1$	$H_1$ cierta $p(H_0) = 0$
Mantener la Hipótesis Nula	<u>Nivel de confianza</u> (1 - alfa)	<u>Error de Tipo II</u> (beta, $\beta$ )
Rechazar la Hipótesis Nula	<u>Error de Tipo I</u> (alfa, Nivel de significación)	<u>Potencia estadística</u> (1 - beta)

El valor de alfa es el nivel de significación estadística. Tradicionalmente el nivel de alfa máximo se ha fijado en .05 ó en el 5%, es decir, cinco veces de cada cien se rechazará la hipótesis nula siendo realmente cierta. Por lo tanto alfa (falso positivo) es igual a:

$$\alpha = \Pr[\text{rechazar } H_0 | H_0 \text{ es cierta}] \quad [6]$$

El resultado de 1 – alfa es la ‘probabilidad de mantener la hipótesis nula cuando realmente es verdadera’ (por ejemplo es el nivel de confianza de .95 cuando alfa es .05). Se trata de una decisión correcta. El nivel de confianza es igual a:

$$1 - \alpha = \Pr[\text{mantener } H_0 | H_0 \text{ es cierta}] \quad [7]$$

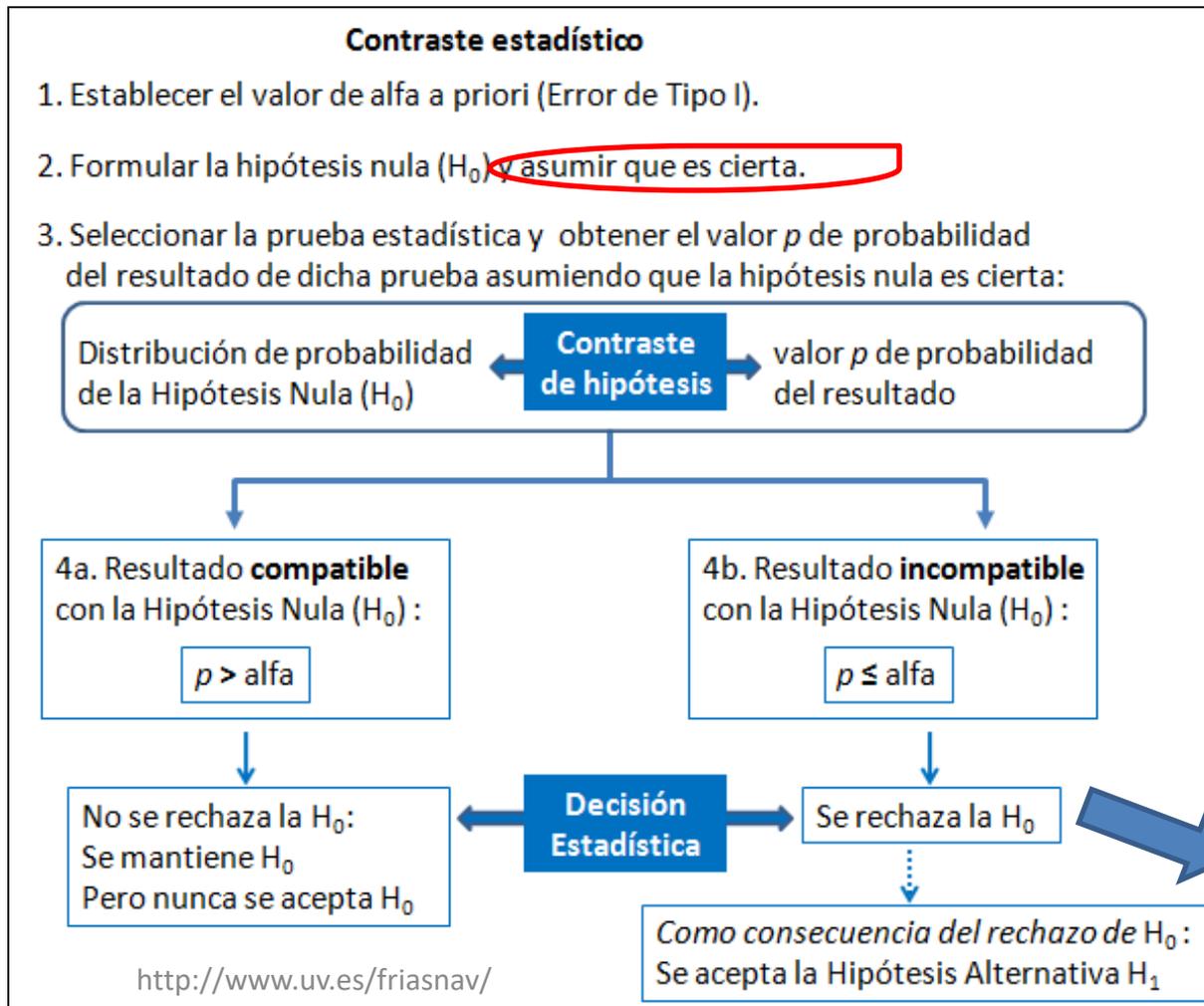
El riesgo de no rechazar la hipótesis nula cuando realmente es falsa se conoce como beta ( $\beta$ ) o error de Tipo II (‘probabilidad de mantener la hipótesis nula cuando realmente es falsa’, es decir si existe una diferencia real entre las poblaciones) (falso negativo) y también se puede estimar evaluando la denominada ‘potencia estadística’ de la prueba estadística o la ‘probabilidad de rechazar la hipótesis nula cuando realmente es falsa’. Por lo tanto, beta es igual a:

$$\beta = \Pr[\text{mantener } H_0 | H_0 \text{ es falsa}] \quad [8]$$

La potencia estadística es 1 – beta y representa una decisión correcta donde se rechaza la hipótesis nula siendo realmente falsa, es decir, existe un efecto y la prueba estadística lo ha detectado:

$$1 - \beta = \Pr[\text{rechazar } H_0 | H_0 \text{ es falsa}] \quad [9]$$

La meta del procedimiento de decisión estadística es por lo tanto evaluar **si se puede o no se puede rechazar la hipótesis nula ( $H_0$ ) con un cierto grado de plausibilidad. NUNCA SE ACEPTA.** El proceso de contraste estadístico



P-Value



**¿cuál es la probabilidad de ese resultado o un resultado más extremo si la hipótesis nula fuese cierta?**

**DECISIÓN DICOTÓMICA**

**¿ERROR TIPO I o II?**

En general, desde un punto de vista de planificación de la investigación y de validez de conclusión estadística, el valor máximo de beta se sitúa en .20 (por lo tanto potencia estadística de .80) (Hallahan y Rosenthal, 1996). Sin embargo, a diferencia del valor de alfa (fijado a priori por el investigador) el investigador no tiene control directo del valor exacto de beta (o de la potencia estadística) hasta que se ejecuta la prueba estadística y sólo puede planificarlo en función de una serie de parámetros como tamaño del efecto que se espera detectar, número de observaciones y alfa elegido a priori. De ahí que cualquier cambio que se produzca en los datos respecto a tamaño del efecto y número de observaciones afectará al valor de potencia estadística a posteriori, es decir el obtenido u observado con los datos de la investigación. Dependiendo del contexto de investigación, el tamaño del efecto puede ser la diferencia entre medias de dos grupos o la correlación entre dos variables como veremos después.

Como se observa, en el proceso de decisión estadística se admite de forma consensuada un mayor grado de error de Tipo II (.20) que de Tipo I (.05), en concreto cuatro veces más de

Por lo tanto, el valor  $p$  es la probabilidad del resultado obtenido (o un resultado más extremo) con la prueba estadística asumiendo que la hipótesis nula es cierta (es decir, cualquier diferencia entre las puntuaciones es debida al azar). Así:

- Si  $p = 1$  entonces ‘es completamente cierto’ que el dato obtenido pertenece a una distribución que plantea la igualdad de las poblaciones (hipótesis nula).
- Si  $p = 0$  entonces ‘es absolutamente imposible’ que el dato obtenido pertenezca a la distribución que plantea la igualdad de las poblaciones (hipótesis nula) y por lo tanto pertenecerá a la hipótesis alternativa que mantiene la diferencia entre las poblaciones.

El valor  $p$  es una probabilidad e indica el grado de probabilidad que tiene un resultado de ocurrir. Representa la fuerza de la evidencia proporcionada por los datos de la muestra a favor de la hipótesis nula. Un valor alto de  $p$  sugiere que los datos de la muestra apoyan a la hipótesis nula mientras que un valor pequeño de  $p$  señala escasa evidencia para apoyar la hipótesis nula. El punto de corte entre valor ‘alto’ o un valor ‘bajo’ es generalmente .05 (alfa).

-Si  $p \leq$  alfa entonces entra en la categoría de improbable bajo el modelo de la hipótesis nula, asumiendo cierto error de Tipo I. Rechazamos la hipótesis nula y por lo tanto existe un efecto diferente de cero. Si el valor de alfa es igual a .05 entonces diremos que el efecto es estadísticamente significativo al 95%. Si se ha cometido un error estadístico solamente podrá ser el error de Tipo I.

-Si  $p >$  alfa entonces no hay suficiente evidencia para rechazar la hipótesis nula y por lo tanto mantenemos que el resultado pertenece a la distribución de dicha hipótesis hasta que se demuestre lo contrario. El valor  $p$  de probabilidad del resultado apoya a la hipótesis nula. La diferencia encontrada es muy probable que se deba al azar y no a un efecto sistemático. Sólo podemos concluir que se trata de un resultado no concluyente, siendo incorrecto realizar inferencias de igualdad de los grupos o ausencia de relación entre las variables. Se mantienen la hipótesis nula, es decir, el efecto encontrado no es

estadísticamente significativo al 95%. Si se ha cometido un error estadístico solamente podrá ser el error de Tipo II.

## **Errores humanos implicados en el proceso de significación estadística**

Creencia errónea: *el valor  $p$  de probabilidad es el valor de alfa*

Creencia errónea: *un valor  $p > \alpha$  señala que la hipótesis nula es cierta*

Creencia errónea: *el valor  $p$  de probabilidad señala importancia del efecto*

El valor  $p$  **no** indica ni la probabilidad de la hipótesis nula, ni la probabilidad de la hipótesis alternativa.

El valor  $p$  **no** indica la verdad o falsedad de la hipótesis nula, ni la verdad o falsedad de la hipótesis alternativa.

El valor  $p$  **no** indica la probabilidad de la replicación ni tampoco señala la importancia del efecto.

El valor  $p$  **indica** la probabilidad de los datos asumiendo una hipótesis como cierta ( $p(D | H)$ ) y **NO** la probabilidad de la hipótesis nula dados ciertos datos ( $p(H | D)$ ).

-Y cuando mantenemos la  $H_0$  no podemos rechazar la explicación del azar como posible explicación de las diferencias. (con riesgo de equivocarnos)

-Y cuando rechazamos la  $H_0$  podemos concluir que hubo algo más que el azar en la formación de las diferencias. (con riesgo de equivocarnos)

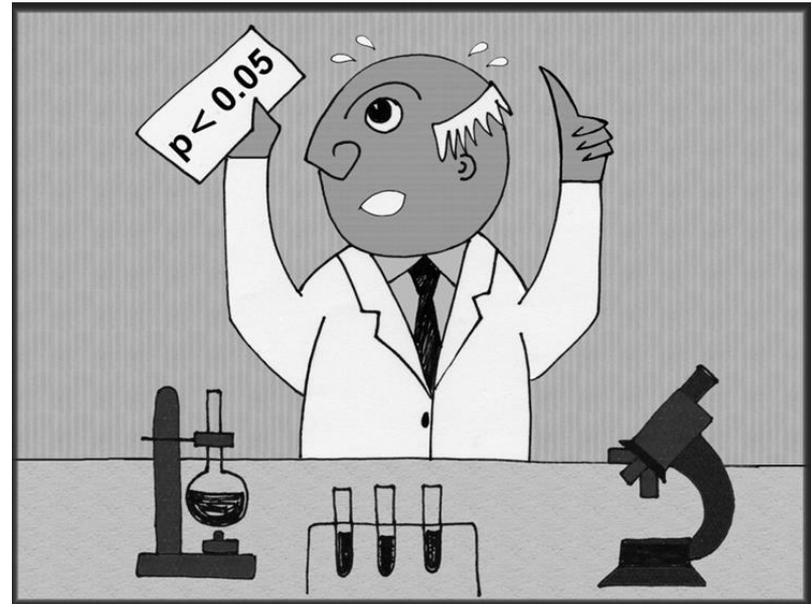
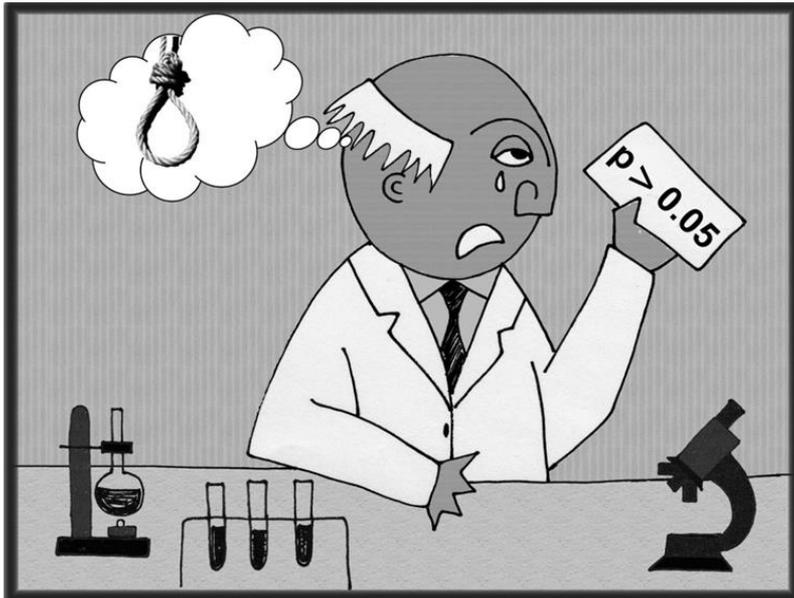
PERO,

El valor  $p$  **NO** informa de la probabilidad del azar.

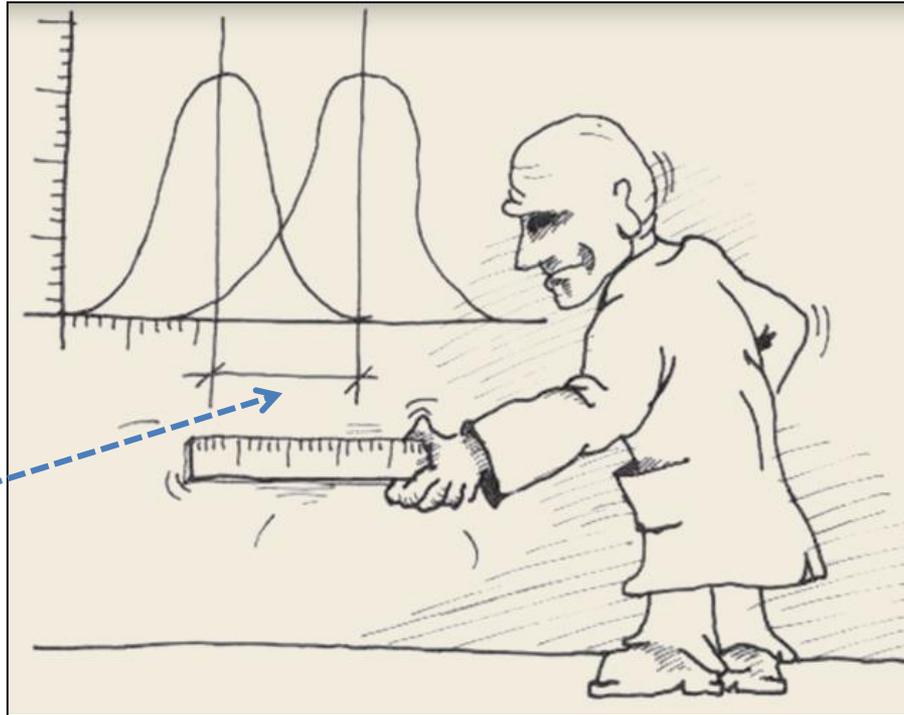
Es decir, no informa de la probabilidad de la hipótesis nula

Y, por supuesto, el valor  $p$  no es el oráculo de la verdad

En la siguiente ilustración queda reflejada la excesiva confianza que el investigador otorga al valor  $p$  de probabilidad, alentando su poder como oráculo de la verdad y éxito de la investigación.



# Tamaño del efecto



**¿Qué magnitud tiene la diferencia entre las medias de las dos distribuciones?**

**En definitiva, el investigador debe responder a las siguientes cuestiones en la fase de planificación de su estudio:**

**Tamaño del Efecto (ej.  $d$ ,  $\eta^2$ ,  $f$ ,  $\phi$ ...):** ¿qué magnitud del efecto se trata de detectar?

**Número de Observaciones ( $N$ ):** ¿cuántas observaciones serán necesarias para detectar el efecto?

**$n$ :** número de observaciones por condición o grupo

**$S$ :** número de sujetos que tiene el estudio

**Error de Tipo I,  $\alpha$  (Nivel de significación, error estadístico):** ¿qué cantidad de riesgo o probabilidad se asume para decir que hay un efecto (por ejemplo una diferencia) cuando realmente no la hay? (*FALSOS POSITIVOS*)

**Nivel de confianza,  $1 - \alpha$  (decisión correcta):** ¿qué probabilidad tiene la prueba estadística de no detectar el efecto cuando realmente no existe?

**Error de tipo II,  $\beta$  (error estadístico):** ¿qué cantidad de riesgo o probabilidad se asume para decir que no hay un efecto (por ejemplo una diferencia) cuando realmente sí la hay? (*FALSOS NEGATIVOS*)

**Potencia estadística,  $1 - \beta$  (decisión correcta):** ¿qué probabilidad tiene la prueba estadística de detectar el efecto cuando realmente existe?

Con el objetivo de minimizar un resultado de falso-negativo (error Tipo II, beta), **un análisis de potencia debe ser ejecutado para determinar el tamaño de la muestra durante la fase de planificación del método o diseño de la investigación.**

-La potencia estadística es la capacidad que tiene el estudio para detectar una diferencia (efecto) si la diferencia realmente existe.

-Un error de Tipo II o beta ocurre cuando existe una diferencial real entre las poblaciones de los grupos estudiados pero no hay suficiente número de participantes para detectar dicha diferencia.

-**Potencia**: probabilidad de correctamente rechazar una hipótesis nula, (cuando la hipótesis alternativa es verdadera). La potencia es  $1 - \beta$

-Cuanto más potencia tenga la prueba estadística más probabilidad de rechazar una hipótesis nula falsa

-Es decir, cuanta más potencia estadística menor será la probabilidad del Error de Tipo II.

- Valor de **alfa** a priori:
- ¿**Potencia** deseada?
- ¿**Tamaño del efecto** previsto?



**Planificar el N**  
**Tablas de Jacob Cohen**



Tabla 19. Valores de N para tamaños del efecto pequeño, mediano y grande asumiendo una potencia estadística de .80 y un alfa de .01, .05 y .10

Prueba Estadística	alfa								
	.01			.05			.10		
	Pequeño d=0.2	Mediano d=0.5	Grande d=0.8	Pequeño d=0.2	Mediano d=0.5	Grande d=0.8	Pequeño d=0.2	Mediano d=0.5	Grande d=0.80
Diferencia de medias (t de Student)	586	95	38	393	64	26	310	50	20
	Pequeño r=.1	Mediano r=.3	Grande r=.5	Pequeño r=.1	Mediano r=.3	Grande r=.5	Pequeño r=.1	Mediano r=.3	Grande r=.5
Correlación r	1163	125	41	783	85	28	617	68	22
ANOVA									
A = 2	586	95	38	393	64	26	310	50	20
A = 3	464	76	30	322	52	21	258	41	17
A = 4	388	63	25	274	45	18	221	36	15
A = 5	336	55	22	240	39	16	193	32	13
A = 6	299	49	20	215	35	14	174	28	12
R múltiple									
VI = 2	698	97	45	481	67	30			
VI = 3	780	108	50	547	76	34			
VI = 4	841	118	55	599	84	38			
VI = 5	901	126	59	645	91	42			
VI = 6	953	134	63	686	97	45			

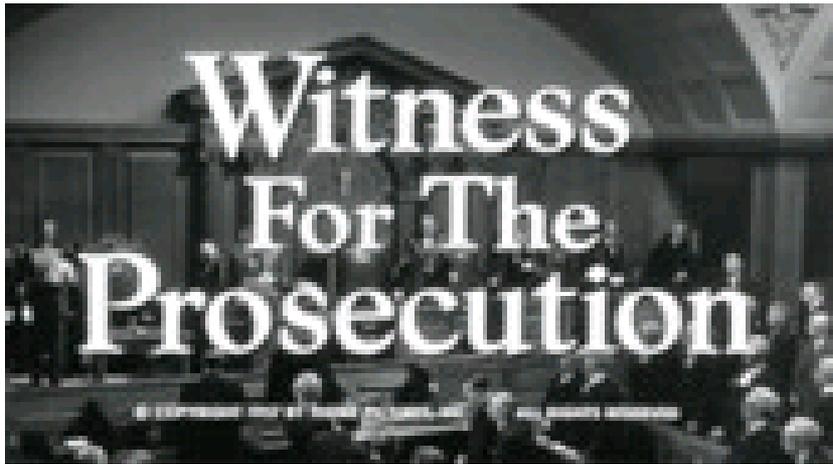
El tamaño de la muestra es por grupo o condición

# ERRORES ESTADÍSTICOS Y DECISIÓN DEL JURADO

**Tabla.** Sentencia y condición del acusado

## Analogía

<i>Sentencia del juez</i>	<i>Condición real del acusado</i>	
	<b>Inocente</b>	<b>Culpable</b>
<b>No Culpable</b>	Decisión correcta	Decisión incorrecta
<b>Culpable</b>	Decisión incorrecta	Decisión correcta



## Testigo de cargo

1958

Director **Billy Wilder**

Tabla 14. Comparación entre el proceso de juicio y el de decisión estadística

	PROCESO DE JUICIO	PROCESO DE DECISIÓN ESTADÍSTICA
Inicio	Asunción de inocencia del acusado	Asunción de que la hipótesis nula es cierta
El objetivo del juicio		
El criterio para determinar la culpabilidad		
Previo al proceso		
Durante el proceso		
Veredicto/Decisión		
Decisión correcta		
Decisión incorrecta		
Calidad de la decisión		
Nuevo estudio	Apelación	Replicación

31/10/2023



- Completar la analogía y realizar el ejercicio que se depositará en el aula virtual.
- El trabajo se subirá al aula virtual y se entregará impreso en el buzón o en la clase
- En clase se pasará un breve cuestionario de 5-10 preguntas.

