

Prueba de Hipótesis Estadísticas

Prueba de Hipótesis Estadísticas

- El uso de la estadística es una parte del método científico, pero constituye una parte importante de nuestra investigación, especialmente en disciplinas relacionadas con los recursos naturales.
- Establecer hipótesis científicas (hipótesis de investigación), articular sus predicciones, diseñar y ejecutar el estudio, coleccionar, organizar y resumir nuestros datos ocurre antes de que usemos pruebas estadísticas.
- “Aceptar” o rechazar una hipótesis estadística es muy diferente a “aceptar” o rechazar una hipótesis científica.
- La hipótesis estadística nula es generalmente de “no efecto” o “no patrón” (no diferencia entre grupos, no relación entre variables). En contraste, la hipótesis alternativa es de que el efecto o patrón existe (existe una clara diferencia entre grupos, o una clara relación entre variables)

Prueba de Hipótesis Estadísticas

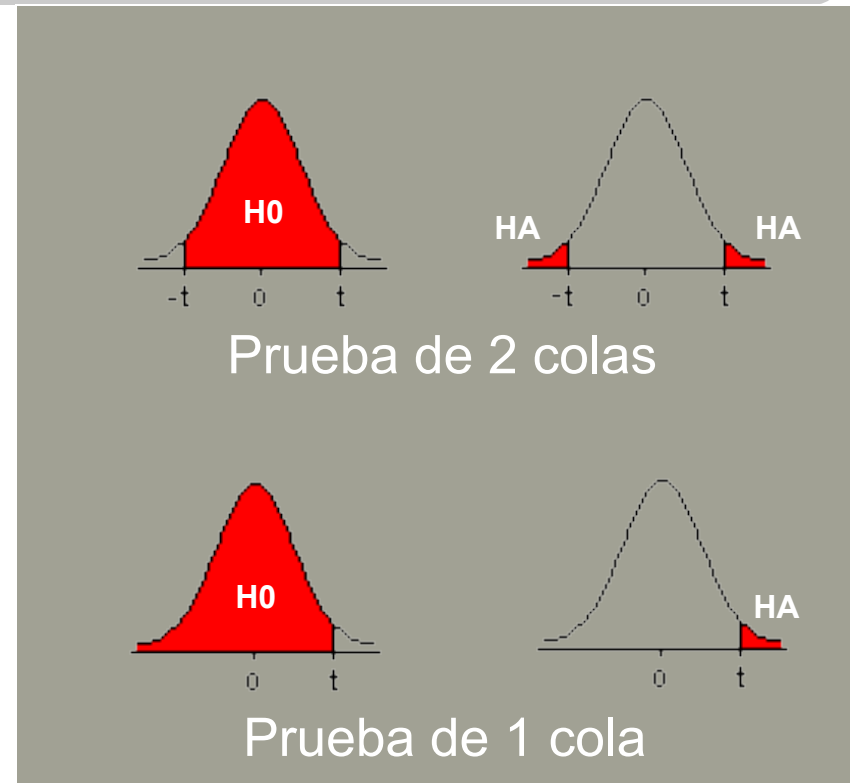
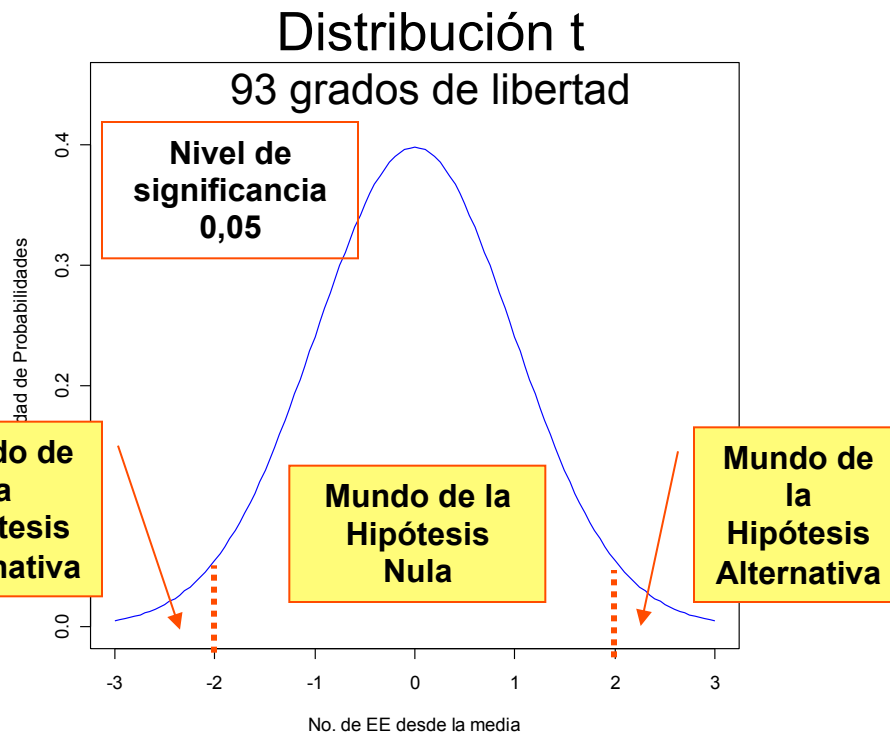
Prueba de Hipótesis Estadísticas, aproximación clásica

1. Especificar H_0 , H_A y una prueba estadística apropiada
2. Especificar a priori el nivel de significancia (ejemplo, 0.05 (5%)) la cual es la frecuencia de la corrida larga del error de tipo I que estamos dispuestos a aceptar (1 error en 20 ensayos).
3. Colectar los datos (utilizando un método probabilístico) y calcular el estadístico de la prueba (ejemplo, t, F, Chi-cuadrado, etc).
4. Comparar el valor del estadístico con su distribución probabilística, asumiendo que H_0 es verdadera.
5. Decisión:
 - Si la probabilidad de obtener este valor, o uno mayor, es menor que el nivel de significancia especificada (ejemplo, 0,05), entonces concluimos que la H_0 es falsa y la rechazamos (resultado “estadísticamente significativo”)
 - Si la probabilidad de obtener este valor es mayor o igual al nivel de significancia especificado (ejemplo, 0,05) entonces concluimos que no hay evidencia de que la H_0 es falsa y la retenemos (resultado “estadísticamente no significativo”)

Prueba de Hipótesis Estadísticas

Prueba de Hipótesis Estadísticas, aproximación clásica

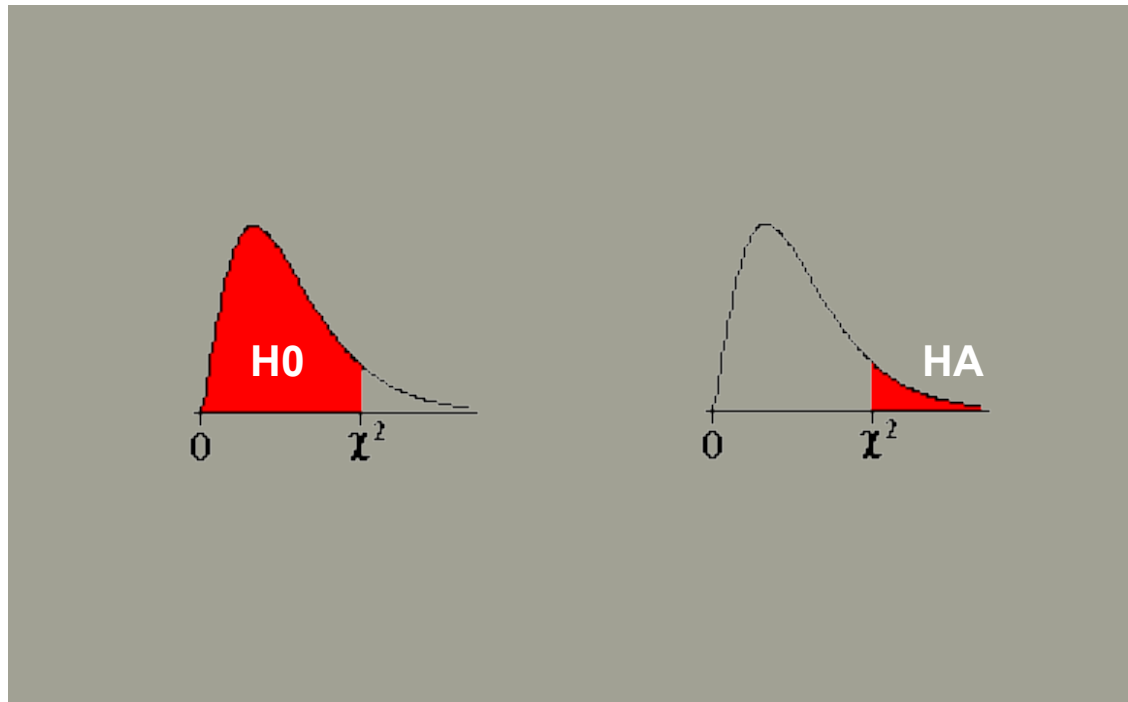
Por ejemplo, si efectuamos una prueba de t (t de student), la prueba nos calcula un valor t que debemos ubicar en una distribución t (para los grados de libertad específicos) y determinar si la prueba nos ha dado un resultado estadísticamente significativo o no.



Prueba de Hipótesis Estadísticas

Prueba de Hipótesis Estadísticas, aproximación clásica

Por ejemplo, si efectuamos una prueba de chi-cuadrado, la prueba nos calcula un valor chi-cuadrado que debemos ubicar en una distribución chi-cuadrado (para los grados de libertad específicos) y determinar si la prueba nos ha dado un resultado estadísticamente significativo o no.



Prueba de Hipótesis Estadísticas

Comparación de las hipótesis nula y alternativa

Variable aleatoria: Largo total de machos y hembras de zorro pelón.



Hipótesis Nula

$$H_0: \mu_{\text{machos}} = \mu_{\text{hembras}}$$

Hipótesis Alternativa

$$H_A: \mu_{\text{machos}} \neq \mu_{\text{hembras}}$$

$$H_A: \mu_{\text{machos}} > \mu_{\text{hembras}}$$

$$H_A: \mu_{\text{machos}} < \mu_{\text{hembras}}$$

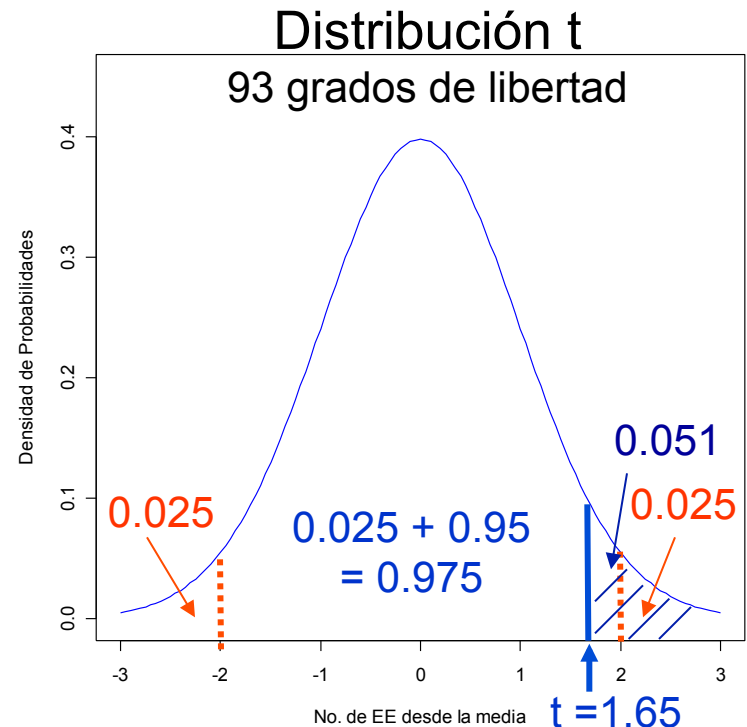
Prueba de Hipótesis Estadísticas

Ejemplo de la comparación del largo total de los machos y hembras de zorro pelón:

Se realiza una prueba de t, y por lo tanto, se obtiene un valor de t.

$$t = \frac{\text{largo total machos} - \text{largo total hembras}}{\text{Error estándar de la diferencia}}$$

$$t = 1,65$$



Valor de $t = 1.65$, queda dentro del área de no significancia

Valor de $p = 0.051$ ($0.102/2 = 0.051$)

Conclusión: no hay diferencia significativa entre el largo total de machos y hembras de zorro pelón

Prueba de Hipótesis Estadísticas

Un caso hipotético: Comparando medias

- Evaluar si el uso de cuatriciclos resulta en un incremento del estrés en poblaciones de venados de campo en un área de Guanacaste.
- El estrés será evaluado por niveles de hormonas glucocorticoides (GC) en la sangre o heces de los venados.
- Venados no expuestos al paso de cuatriciclos tienen una media de 872 (y su respectivo ee) ng GC/g, mientras venados expuestos al paso de cuatriciclos tienen una media de 1468 (y su respectivo ee) ng GC/g (Diferencia= 596 (y su respectivo ee) ngGC/g).
- Como decidir si la diferencia entre estas 2 medias es suficientemente grande para atribuírsela al efecto de los cuatriciclos.



Prueba de Hipótesis Estadísticas

La Hipótesis Estadística Nula

- Científicos tienden a favorecer explicaciones parsimoniosas o simples. La explicación más simple es que la diferencia representa la variación aleatoria entre los grupos y no refleja el efecto sistemático de los cuatriciclos.
- Aunque se debe recordar que es improbable que las medias de 2 muestras sean iguales, incluso si se extraen dos muestras diferentes de los venados no expuestos a los cuatriciclos.
- El nivel de los GC difieren entre individuos por muchas razones, estas variaciones (incluyendo la variación por errores de medición) es lo que se llama variación aleatoria.
- Nosotros queremos saber si la diferencia entre medias de los grupos es mayor a la que esperaríamos además de variación aleatoria entre individuos.
- Una típica hipótesis nula estadística es que “la diferencia entre los grupos no es mayor a la que esperaríamos debido a la variación aleatoria”.
- Así, la H_0 establece que no opera otro mecanismo más que la variación debida al azar (no hay efecto de los cuatriciclos).

Prueba de Hipótesis Estadísticas

La Hipótesis Alternativa

- Una vez establecida la hipótesis nula, podemos definir una o más hipótesis alternativas a la hipótesis nula.
- En nuestro ejemplo, la hipótesis alternativa natural es que la diferencia observada entre las medias de los dos grupos es demasiado grande para ser explicada por variación aleatoria.
- Esto no significa que la hipótesis alternativa es que los cuatriciclos son responsables por un aumento en el nivel de GC.
- La hipótesis alternativa se focaliza en el patrón presente en los datos.
- El investigador puede inferir mecanismos detrás del patrón, pero esa inferencia es un paso separado del proceso de prueba de hipótesis estadísticas.
- La prueba estadística sólo revela si el patrón de los datos es probable o improbable, dado que la hipótesis nula es verdadera.
- La habilidad de asignar mecanismos causales a los patrones estadísticos depende de la calidad de nuestro diseño de estudio y de nuestras observaciones, y no exclusivamente de nuestro resultado estadístico.

Prueba de Hipótesis Estadísticas

Significancia Estadística y Valores de P

- En muchos análisis estadísticos la pregunta es si la hipótesis nula de variación debido al azar puede ser rechazada.
- El valor de P es la guía para alcanzar una decisión.
- El valor de P mide la probabilidad de observar, en estudios idénticos repetidos, la diferencia encontrada o valores aún más extremos, asumiendo que la hipótesis nula es verdadera.
- Lo que estamos analizando es $P(\text{datos observados o más extremos} | H_0)$ y no $P(H_0 | \text{datos})$.
- Esto no es lo mismo que decir: la probabilidad de que H_0 sea verdadera dado los datos observados [$P(H_0 | \text{datos})$].
- Si quisiéramos establecer la probabilidad de que H_0 es verdadera deberíamos realizar la prueba de hipótesis en el marco de la estadística Bayesiana.

Prueba de Hipótesis Estadísticas

Significancia Estadística y Valores de P

- ❖ Si el valor de P es pequeño (cercano a 0.0), entonces es improbable que las diferencias observadas entre grupos de venados haya sido obtenida por azar asumiendo que la hipótesis nula fuera verdadera.
- ❖ Si el valor de P es grande (cercano a 1.0), entonces es probable que la diferencia entre grupos de venados pueda haber ocurrido al azar asumiendo que la hipótesis nula es verdadera.
- ❖ Pero, se debe mantener en mente de que cuando se calcula el valor de P, estamos viendo los datos a través del lente de la hipótesis nula.

Prueba de Hipótesis Estadísticas

Significancia Estadística y Valores de P

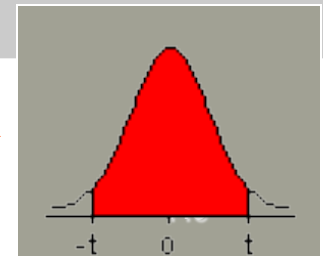
3 elementos determinan el valor de P

- El tamaño de muestra (n)
 - A mayor tamaño de muestra menor el valor de P. Con un tamaño de muestra suficiente se pueden detectar diferencias triviales entre los grupos en comparación.
- La diferencia entre medias observada ($\text{media}_{\text{grupoA}} - \text{media}_{\text{grupoB}}$)
 - A mayor diferencia entre las medias menor el valor de P.
- El nivel de variación entre individuos (variabilidad de los datos)
 - A menor variabilidad en los datos menor el valor de P. A menor variabilidad en los datos será más fácil detectar una diferencia entre grupos.

$$t = \frac{\text{largo total machos} - \text{largo total hembras}}{\text{Error estándar de la diferencia}}$$



Al aumentar t
disminuye P



Prueba de Hipótesis Estadísticas

Significancia Estadística y Valores de P

El valor de P es interpretado erróneamente de 3 maneras:

1. La probabilidad de que el resultado de un análisis es debido al azar. Si P es bajo (ejemplo, 0,03), se interpreta (erróneamente) de que existe un 3% de que el resultado obtenido se deba al zar.
2. $1-P$ es considerado la probabilidad de obtener el mismo resultado si el estudio fuera repetido en exactamente las mismas condiciones. Ejemplo, $P=0,03$, $1-0,03 = 0.97$, 97% de confianza de que el mismo resultado será obtenido si el estudio fuera repetido.
3. La probabilidad de que la hipótesis nula (H_0) es verdadera. Ejemplo, $P=0,03$ significa que la probabilidad de que H_0 es verdadera es de 3%.

- Sin embargo, el valor de P es la probabilidad de que un resultado o resultados más extremos ocurran por azar en la corrida larga (long-run), condicional a que la H_0 es verdadera.

Prueba de Hipótesis Estadísticas

Prueba de Hipótesis Estadísticas, aproximación clásica

- Un resultado no significativo (ejemplo, $P = 0,64$) no implica que H_0 es verdadera.
- Para algunos investigadores un resultado no significativo implica que no se tiene evidencia para rechazar H_0 .
- La excepción sería si tenemos una potencia alta de la prueba estadística para detectar una hipótesis alternativa, entonces podríamos afirmar que el verdadero efecto es probablemente menor que el efecto específico que estamos buscando.

Prueba de Hipótesis Estadísticas

Prueba de Hipótesis Estadísticas, aproximación clásica

Valor de P	Tradicional	Recomendado
$< 0,05$	Rechazar la hipótesis nula	Existe una fuerte evidencia en contra de la hipótesis nula
$0,05 < P < 0,15$	Apenas se falla en rechazar la hipótesis nula	Evidencia es inconclusa y se necesitan más datos
$P > 0,15$	Se falla en rechazar la hipótesis nula	No hay evidencia en contra de la hipótesis nula

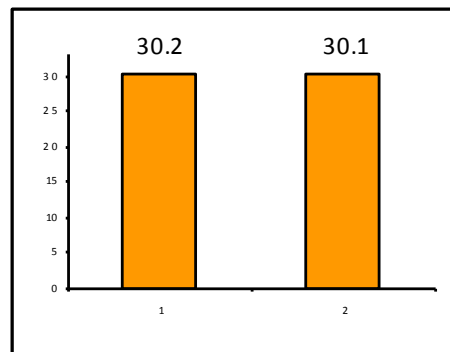
- No es apropiado decir:
 - Se acepta la H_0
 - Se acepta la hipótesis alternativa (H_A)
 - La H_0 es verdadera
 - La H_0 es falsa

Prueba de Hipótesis Estadísticas

Críticas a las Pruebas de Significancia Estadística

1. La significancia estadística depende del tamaño de la muestra

- Cualquier H_0 puede ser rechazada con un tamaño de muestra suficientemente grande. Por lo tanto, si una H_0 no es rechazada se debe a que el tamaño de muestra no fue suficiente.
- En el ejemplo del estrés de los venados una diferencia de 1 nanogramo puede ser detectada como estadísticamente significativa si el tamaño de muestra es suficientemente grande (más venados son analizados).



30.2 no es diferente estadísticamente de 30.1 cuando $n = 30$, pero son estadísticamente diferentes cuando $n = 100$.

Prueba de Hipótesis Estadísticas

Críticas a las Pruebas de Significancia Estadística

2. Las H_0 son generalmente falsas, por lo tanto, no tiene mucho valor el rechazar H_0

- Dos medias poblacionales nunca van a ser exactamente iguales. Por lo tanto su diferencia nunca va a ser exactamente 0.
- Ejemplos donde se rechazan hipótesis nulas de manera trivial:
 - La densidad de árboles fue mayor en áreas no deforestadas que en áreas deforestadas ($P = 0.02$)
 - La masa corporal de la población de mapachines del PN Manuel Antonio difiere de la población de mapachines del PN Santa Rosa ($P = 0.03$, $n = 456$)

Prueba de Hipótesis Estadísticas

Críticas a las Pruebas de Significancia Estadística

3. El valor de significancia es arbitrario

- El uso de $\alpha = 0.05$ como el criterio para rechazar o no rechazar una hipótesis es arbitrario.
- Fisher sugirió primero 0.05, pero más tarde sugirió usar un nivel de significancia específico para cada ocasión.
- Neyman-Pearson no establecen un alfa específico, simplemente un valor escogido a priori.
 - Aunque cabe notar, que el problema de la arbitrariedad del nivel de la significancia no es un elemento para descalificar las pruebas de significancia estadística.

Prueba de Hipótesis Estadísticas

Críticas a las Pruebas de Significancia Estadística

4. Un valor de P no es informativo

- Es silencioso con respecto al tamaño del efecto.

$$P = 0.001$$

Pero no dice cual es la diferencia entre las medias o el parámetro que se esté estimando

5. Los valores de P no pueden ser comparados si provienen de tamaños de muestras diferentes

$$P_1 = 0.05 \text{ (n=30) no es comparable a } P_2 = 0.001 \text{ (n=35)}$$

6. Producen un decisión dicotómica

Para un nivel fijo de alfa (α) = 0.05,
significativo si $P = 0.049$ y no significativo si $P = 0.051$

Prueba de Hipótesis Estadísticas

Alternativas a las Pruebas de Significancia Estadística

- ❖ Tamaño del efecto y su respectivo intervalo de confianza (estimado de punto e intervalo)
- ❖ Métodos basados en la Teoría de la Información
- ❖ Inferencia Bayesiana

Significancia Estadística y Significancia Biológica o Práctica

- No significancia no significa “no efecto”.
 - Estudios pequeños a menudo culminan con resultados no significativos aún cuando hay importantes efectos (no tienen el tamaño de muestra suficiente para encontrar el efecto).
 - Ausencia de evidencia no significa evidencia de ausencia.
- Significancia estadística no significa que el efecto es real.
 - Sólo por azar, para un 0.05%, 1 de cada 20 resultados estadísticamente significativos es erróneo.
- Significancia estadística no necesariamente significa biológicamente importante.
 - Es el tamaño del efecto lo que determina la importancia biológica o práctica, no la presencia de la significancia estadística.

Prueba de Hipótesis Estadísticas

Comparación de las hipótesis nula y alternativa

Variable aleatoria: Distancia de dispersión de mapachines machos y hembras de su madriguera natal.



Hipótesis Nula

$$H_0: \mu_{\text{machos}} = \mu_{\text{hembras}}$$

Hipótesis Alternativa

$$H_A: \mu_{\text{machos}} \neq \mu_{\text{hembras}}$$

$$H_A: \mu_{\text{machos}} > \mu_{\text{hembras}}$$

$$H_A: \mu_{\text{machos}} < \mu_{\text{hembras}}$$

Prueba de Hipótesis Estadísticas

Significancia estadística y significancia biológica

Ejemplo

Madriguera
natal

Mean = 2,8 km (IC 95% = 2-3,6 km)

Mean = 1,5 km (IC 95% = 0,9-2,1 km)



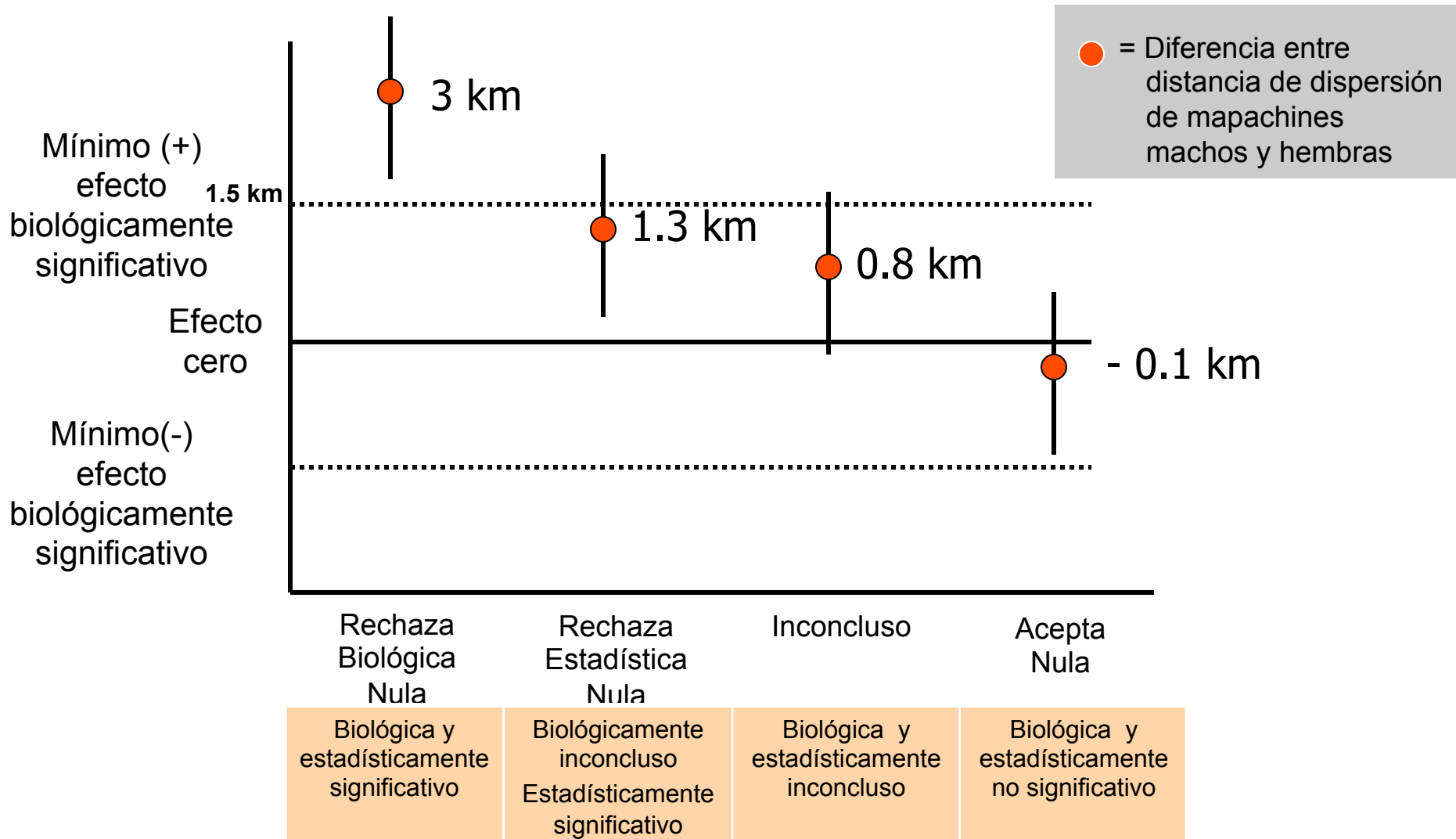
Tamaño del Efecto = 1.3 km (IC 95%: 0.13 – 2.5 km)

Aproximación clásica: los mapachines machos (media = 2,8 ; EE=0,4 km) se dispersaron más lejos que las hembras (media = 1,5 ; EE=0,3 km) ($P=0,04$).

Aproximación alternativa: la diferencia entre las distancias de dispersión entre mapachines machos (media=2,8; EE=0,4 km) y hembras (media=1,5; EE=0,3 km) fue de 1,3 km (95%CI: 0,13 – 2,5 km).

Prueba de Hipótesis Estadísticas

Estimado de punto e intervalo de confianza



Alternativa a las Prueba de Hipótesis Nulas (NHST)

Pruebas de significancia y significancia biológica

Estimado de punto e intervalo de confianza

Ventajas

- ❖ Provee más información que un valor de P.
- ❖ Provee un tamaño del efecto y una medida de precisión del tamaño del efecto.
- ❖ Provee la información necesaria para evaluar de manera confiable que no hay evidencia en contra de la hipótesis nula.
- ❖ Permite evaluar significancia biológica (o práctica).

Errores Tipo I y II

- El Error Tipo I ocurre cuando rechazamos una hipótesis nula verdadera. Decimos que hay un efecto o relación pero en realidad no lo hay. La probabilidad de un Error de Tipo I está dada por α (alfa).
- El Error Tipo II ocurre cuando aceptamos una hipótesis nula falsa. Fallamos en rechazar una hipótesis nula falsa. Hay un efecto o relación pero no lo detectamos. La probabilidad de un Error de Tipo II está dada por β (beta)

	Rechazamos H_0	No Rechazamos H_0
Hipótesis Nula Verdadera (No diferencia)	Error Tipo I (α) Diferencia detectada, pero en realidad no existe	Decisión Correcta Diferencia no existe y por lo tanto, no es detectada
Hipótesis Nula Falsa (Diferencia)	Decisión Correcta Diferencia existe y es detectada	Error Tipo II (β) Diferencia no detectada, pero en realidad existe

Errores Tipo I y II

- El Error Tipo I se considera la probabilidad de obtener un positivo falso.
- Cuando uno calcula un valor de P está en realidad estimando α .
- El Error Tipo II se considera la probabilidad de obtener un negativo falso.
- Idealmente, se deberían minimizar los dos tipos de errores. Sin embargo, las estrategias para disminuir un error de tipo I incrementan el riesgo de cometer un error de tipo II y viceversa.
- Aunque estos dos tipos de errores están inversamente relacionados no hay una simple relación matemática entre ellos, ya que la probabilidad del error de tipo II depende en parte de cual es la hipótesis alternativa, cuan largo es el efecto que uno intenta detectar, el tamaño de muestra, y nuestro diseño de estudio.
- La probabilidad de cometer un Error de Tipo II no es comúnmente calculada o reportada.
- Para calcular la probabilidad de un Error de Tipo II se deben especificar las Hipótesis Alternativas.

Errores Tipo I y II

- Desde una perspectiva filosófica, se argumenta que en ciencia un Error de Tipo I es más serio que un Error de Tipo II.
 - El cometer un error de Tipo I se considera una falsedad, ya que rechazamos la H_0 y se argumenta a favor de un mecanismo (que en realidad no existe).
- En cambio, un Error de Tipo II es un error de ignorancia, que se corregirá en el futuro con un estudio mejor.
- Sin embargo, en problemas aplicados, tales como el monitoreo ambiental o diagnóstico de enfermedades, el Error de Tipo II puede tener consecuencias más serias.
 - Implica que cambios ambientales no deseados, contaminación, o enfermedades no fueron detectados.

Balance entre Errores de Tipo I y II

Error Tipo II



Error Tipo I



Error Tipo I



Error Tipo II



Aumentando el nivel de
significancia

(Haciendo alfa más pequeño)
Ejemplo: $0.05 \rightarrow 0.01$

Disminuyendo el nivel de
significancia

(Haciendo alfa más grande)
Ejemplo: $0.05 \rightarrow 0.10$

Potencia Estadística

Potencia de una prueba estadística

- La capacidad de una prueba de rechazar una hipótesis nula cuando ésta es falsa (La capacidad de no cometer un Error de Tipo II).
- La probabilidad de detectar un efecto dado cuando este realmente existe.
- En términos más prácticos, es una medida de nuestra confianza en que hemos detectado un efecto importante si este existe.

Si β es el riesgo (probabilidad) de cometer un Error de Tipo II,

$$\text{Poder} = 1 - \beta \text{ (Error tipo II)}$$

A mayor poder disminuye la probabilidad de cometer un Error de Tipo II.

Significa que pruebas robustas probablemente no resulten en errores de tipo II

Potencia Estadística

Para determinar la potencia de un análisis, se necesita determinar la hipótesis alternativa o el tamaño del efecto.

Factores que influyen en la potencia estadística:

- Tamaño del efecto (TE). Cuan grande es el cambio o diferencia que nos interesa. Es mas probable detectar cambios grandes.
- Varianza (σ^2) entre unidades de muestreo o experimentales. Generalmente expresada como error estándar (σ)
- Nivel de alfa. El poder varía con alfa (α).
- Tamaño de muestra (n). Un efecto dado es más fácil de detectar con un tamaño de muestra grande.

$$Potencia \propto \frac{TE \alpha \sqrt{n}}{\sigma}$$

↑
Proporcional

El vínculo de estos parámetros al poder estadístico depende de la prueba estadística.

Potencia Estadística

La potencia de una prueba es alta cuando:

1. La diferencia que estamos buscando es grande.
2. La distribución probabilística de la muestra tiene una varianza pequeña.
3. La prueba está basada en un tamaño de muestra grande.

Generalmente los investigadores intentan lograr un poder de 0.80 en sus análisis, aunque no hay una medida establecida formalmente.

Una estrategia para protegerse contra los dos tipos de errores es establecer un nivel de significancia alto (alfa pequeño, para controlar error de tipo I) y tener una muestra de buen tamaño (para controlar el error de tipo II)

Análisis de Potencia Estadístico a priori: Usando la Potencia Estadística para planear la investigación (Análisis Prospectivo)

Determinando el Tamaño de Muestra

- Dado el Tamaño del Efecto
- Dado α (alpha) (Error Tipo I)
- Dada la Potencia ($1 - \text{Error Tipo II}$)
- Dada la varianza

Tamaño de muestra
(n)

Es el más uso más común del análisis de potencia estadística

Análisis de Potencia Estadístico a priori: Usando la Potencia Estadística para planear la investigación (Análisis Prospectivo)

Determinando el Tamaño del Efecto

- Dado el Tamaño de Muestra
- Dado α (alpha) (Error Tipo I)
- Dada la Potencia ($1 - \text{Error Tipo II}$)
- Dada la varianza

Tamaño del Efecto
(TE)

Si factores externo restringen el tamaño de la muestra a niveles relativamente bajos, se puede calcular cual es el cambio o tamaño del efecto más pequeño que se podría detectar.

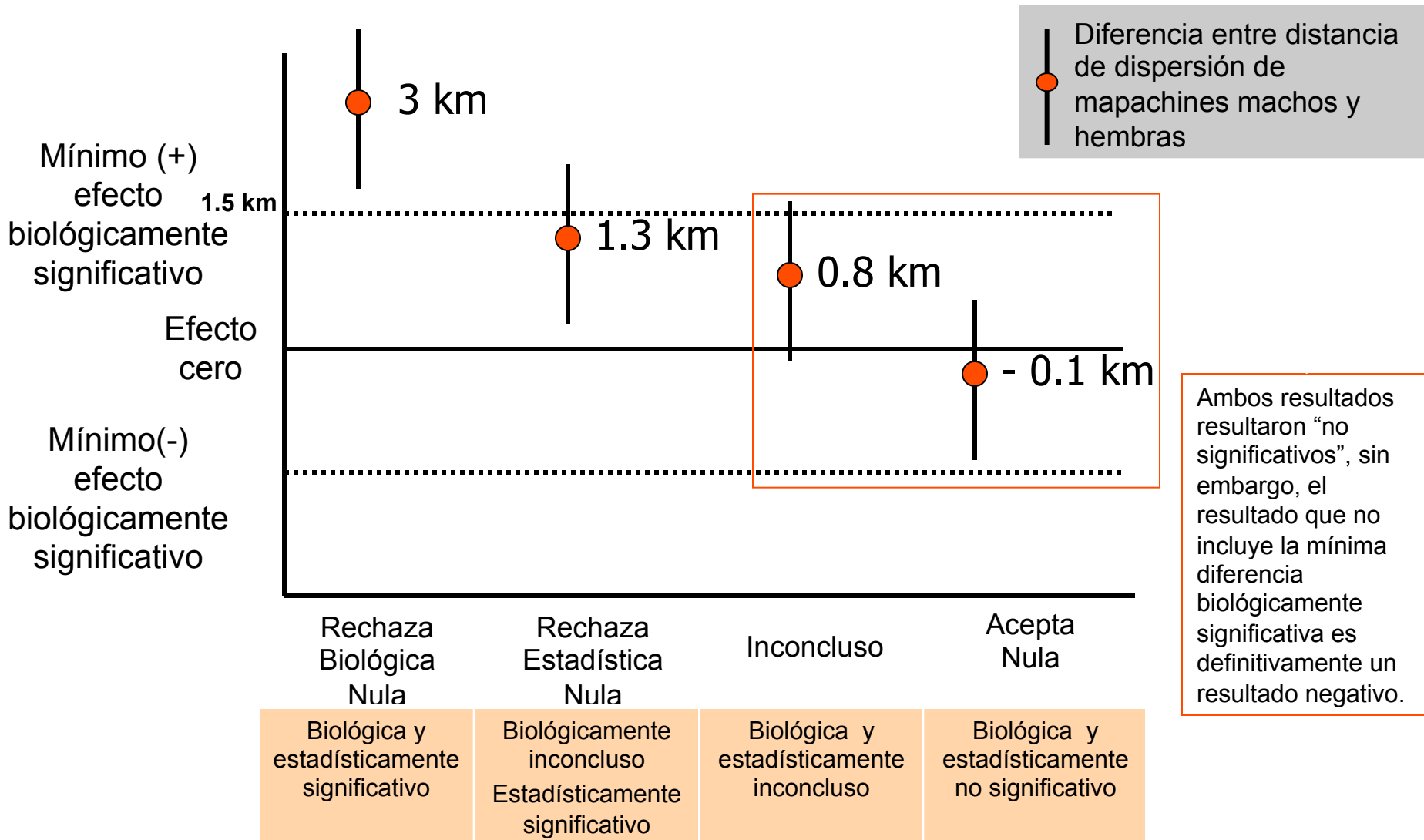
Análisis de Potencia Estadístico a posteriori

Este tipo de análisis se denomina también:

- Potencia observada
- Potencia Post hoc
- Análisis de Potencia Retrospectivo

- Este tipo de análisis de potencia se realiza cuando la prueba estadística usado no reveló significancia estadística.
- El objetivo de este tipo de análisis es ver si la prueba estadística fue suficientemente poderosa o no.
- Sin embargo, el análisis retrospectivo carece de sentido. Por supuesto que la prueba no fue lo suficientemente poderosa, por ello el resultado resultó no significativo.
- Los cálculos de potencia son útiles para diseñar, no para analizar.
- Como alternativa al análisis de potencia a posteriori o retrospectivo se debe presentar el intervalo de confianza.

Usar Intervalos de Confianza en lugar de Análisis de Potencia Retrospectivo



Ejemplo de un cálculo de tamaño de muestra

Tamaño de muestra requerido para una diferencia de medias

La práctica convencional es determinar un tamaño de muestra con un poder de 80% al nivel de significancia 0.05.

El tamaño de muestra puede ser determinado por:

$$n = \frac{16\sigma^2}{\Delta^2} + 1$$

n , tamaño de muestra requerido por grupo

Δ , diferencia entre medias esperada ($\mu_1 - \mu_2$, una diferencia que vale la pena detectar).

σ , representa el desvío estándar de la variable que es estimada por DE (desvío estándar agrupado).

Ejemplo: si se quiere detectar una diferencia entre medias de 18, para una variable con un desvío estándar de 30, el tamaño de muestra por grupo requerido:

$$n = \frac{(16)(30)^2}{18^2} + 1 \cong 46$$

Conclusión y mensaje

El tamaño del efecto (que incluye su intervalo de confianza) es preferido a la prueba de hipótesis estadística

