

Estadística inferencial – Parte II

A. Formas habituales de malinterpretar un único valor de p (Greenland et al., 2016)

1. El valor de p es la probabilidad de que la hipótesis principal sea cierta (Ej. $p=0,01$ es 1% de chance de que H_0 sea verdad).

No. El test asume que H_0 es verdadera. El valor de p indica el grado en el cual la data de la muestra conforma el modelo que predice el test. $P=0.01$ indica que los datos no se acercan al modelo y a la hipótesis sostenidas en el test.

2. El valor de p indica la probabilidad de que la asociación se deba al azar (chance).

3. Un resultado significativo ($p \leq 0,05$) significa que la H_0 es falsa o debe rechazarse.

No. Un valor bajo de p indica que la data que tenemos es inusual, dados los supuestos estadísticos del test. Puede ser un valor causado por un gran error aleatorio o por una violación de algún otro supuesto.

4. Un resultado no significativo ($p > 0,05$) indica que la H_0 es verdadera o debe ser aceptada.

No. Un valor alto de p indica que el valor no es inusual, dados los supuestos asumidos en el test. Puede ser un valor causado por un gran error aleatorio o por una violación de algún otro supuesto

5. Un valor alto de p es evidencia a favor de la H_0 .

No, salvo que p sea 1, cualquier p menor a 1 es indicio de que la hipótesis del test no es compatible con la data.

6. Un valor de p mayor a 0,05 significa que ningún efecto fue observado.

No. Significa que la H_0 sigue siendo una entre otras hipótesis sobre los datos. Salvo que el punto observado coincida exactamente con el punto nulo, es un error definir los resultados como “sin evidencia de un efecto”.

7. La significancia estadística indica que se encontró una relación importante a nivel científico.

No. Puede deberse a una violación de supuestos o a errores de una muestra grande. El hecho de que sea inusual no significa que sea clínicamente relevante.

8. La falta de significancia estadística indica que el tamaño del efecto es pequeño.

No. Especialmente en un estudio chico, grandes efectos pueden estar tapados por ruido.

9. El valor de p indica el porcentaje de chances de que ocurra nuestra observación si la H_0 es real.

No. Muestra la distribución de la frecuencia de la data si los supuestos son ciertos.

10. Si se rechaza H_0 porque $p < 0,05$ significa que la probabilidad del error de “falso positivo” es 5%.

No. Si rechazó H_0 y es real, el error es del 100%.

11. Un valor de $p=0,05$ y $p \leq 0,05$ significan lo mismo.

No. Uno es un punto determinado y el otro incluye los valores más extremos.

12. Una forma correcta de reportar p es a partir de un valor que no incluya el = (Ej: reportar $p < 0,02$ cuando $p = 0,015$)

No. No se pueden interpretar bien los resultados. Únicamente cuando los valores p son muy bajos pierden precisión (0.001)

13. La significancia estadística es una propiedad del fenómeno, entonces una prueba estadística detecta esta significancia,

Este error se manifiesta en la frase “se encontró evidencia de significancia estadística...”. La significancia estadística es una descripción del valor de p , no una propiedad de la población.

14. Siempre se debe usar un valor de p bilateral.

No. Es lo más habitual y se usa si la H_0 incluye un valor específico (ej. $=0$). Si la pregunta de investigación es unilateral, es válido usar un valor de p unilateral.

B. Formas habituales de malinterpretar muchos valores y comparaciones de p (Greenland et al., 2016)

15. Cuando se testea una hipótesis y todos, o casi todos, muestran $p > 0,05$, se asume que es una forma de verificar la H_0 .

No. Que un grupo de muestras individuales no den significancia estadística, no implica que sea una evidencia total de “falta de efecto”. Al contrario, si, por ejemplo, cinco muestras resultan en un valor de $p = 0,10$, se pueden combinar los valores de p y el valor total de esto daría $p = 0,01$.

16. Cuando los valores de p obtenidos de dos muestras se encuentran en diferentes lados de 0,05, los resultados se interpretan como opuestos.

No. Los test estadísticos son sensibles a las variaciones de las muestras. Eso se puede reflejar en p , pero pueden presentar las mismas asociaciones. Las diferencias de resultados pueden evaluarse con pruebas (llamadas análisis de la heterogeneidad, interacción o modificación).

17. Cuando la misma hipótesis es testeada en diferentes poblaciones y se obtienen p similares, los resultados concuerdan.

No. Dos estudios pueden mostrar p similares pero no mostrar las mismas asociaciones, debido a diferencias en las poblaciones.

18. Si se observa un valor de p muy bajo en un estudio, es altamente probable que en el próximo estudio se obtenga un resultado similar.

No. El valor de p indica justamente la probabilidad de obtener un valor similar o menor. Es decir, si $p = 0,03$, la chance de que sea igual o menor es 3%. La probabilidad de la réplica es exactamente el valor de p . Esto siempre que sean muestras independientes y que los supuestos sean ciertos en ambos estudios. Si no es así, el valor de p es muy sensible al tamaño de la muestra y a las violaciones de los supuestos.

C. Formas habituales de malinterpretar los intervalos de confianza (Greenland et al., 2016).

El problema empieza cuando se interpreta que $p > 0,05$ significa que la H_0 tiene 5% de chances de ser falsa y surgen las siguientes falacias:

19. El intervalo de confianza específico presentado en un estudio tiene 95% de probabilidad de contener el verdadero tamaño del efecto.

No. La frecuencia con la cual un intervalo contiene el efecto real es 100% si lo contiene y 0%, si no lo contiene. El 95% refiere a la cantidad de intervalos de confianza que contiene el valor real. Esto significa que el 95% de los intervalos de confianza creados por muestras del mismo tamaño de la misma población van a contener el parámetro real de la población, si todos los supuestos son reales.

20. Un resultado afuera del intervalo de confianza del 95% ha sido refutado o excluido por la data.

No. Lo que significa es que la combinación de los datos con los supuestos asumidos, en relación al criterio del 95%, es incompatible con los datos observados por algún motivo.

21. Si dos intervalos de confianza se superponen significa que las diferencias entre los valores estimados o de los estudios no es significativa.

No. Los intervalos pueden coincidir, pero sus diferencias pueden producir $p < 0,05$ para diferencias entre estudios.

22. Un intervalo de confianza de 95% observado predice que el 95% de los puntos estimados de los futuros estudios van a estar en dentro de ese intervalo.

No. El 95% refiere a la cantidad de otros intervalos no observados van a contener el verdadero efecto, no cuál es la frecuencia de futuros puntos estimados sobre ese intervalo.

23. Si un intervalo incluye el valor nulo y otro lo excluye, el primero es más preciso.

No. La precisión depende de la anchura del intervalo.

24. Si se rechaza la hipótesis nula porque $p > 0,05$ y el poder estadístico es de 90%, las chances de cometer un falso negativo son del 10%.

No. El 10% refiere a la cantidad de veces que se cometería este error en este test estadístico través de muchos estudios si la hipótesis alternativa es cierta, no refiere a un uso singular del test.

25. Si p supera el 0,05 y el poder es del 90%, los resultados apoyan la hipótesis nula y no la alternativa.

Si bien es un razonamiento intuitivo, hay contraejemplos en los cuales el valor de p para la hipótesis nula está entre 0,05 y 0,10 y, sin embargo, las alternativas tienen un valor de p que excede el 0,10 y el poder es del 90%.

Este tipo de cálculos pueden explicar la crisis de replicación, ya que definir los resultados de un estudio único de forma dicotómica, anticipa un problema.