

Estadística inferencial

La estadística inferencial está interesada en dos tipos de problemas:

- La *estimación* de los parámetros de la población
- Las pruebas de hipótesis

Técnicas Paramétricas vs. técnicas no paramétricas

Técnicas paramétricas

En los métodos estadísticos, las técnicas que hacen suposiciones acerca de la naturaleza de las poblaciones de las cuales se derivan las observaciones y los datos se denominan *paramétricas*.

Estas técnicas permiten plantear conclusiones de la siguiente forma: “Si las suposiciones acerca de la forma de la distribución de la población son válidas, entonces se puede concluir que...”

Un parámetro es una cantidad numérica que mide algún aspecto de unos puntajes de una población. Los parámetros son de dos tipos: centralizados

A) centralizados: media

μ [Notación con letras griegas para indicar el parámetro de la población] O

M [Notación con letras romanas para el estadístico de una muestra]

B) de dispersión: desviación estándar,

σ [Notación con letras griegas para indicar el parámetro de la población] O

S [Notación con letras romanas para el estadístico de una muestra]

Técnicas no paramétricas

Hay, por otro lado, un gran número de técnicas de inferencia que no hacen suposiciones numerosas o rigurosas acerca de la población de la cual se han muestreado los datos. Estas técnicas se llaman *no paramétricas* o *de distribución libre*.

Estas técnicas dan como resultado conclusiones que requieren menos calificaciones, entonces se puede concluir en los siguientes términos: “Sin considerar la(s) forma(s) de la (s) población(es), se puede concluir que...”

Las técnicas o métodos no paramétricos están disponibles para tratar datos categóricos, es decir, que son medidos en una escala nominal u ordinal. Ninguna técnica paramétrica se aplica a estos datos.

Las pruebas de hipótesis

En muchas investigaciones el propósito es probar cierta hipótesis que se deriva de una determinada teoría.

Se recaban los datos que permitan decidir acerca de esa hipótesis y los resultados pueden conducir al investigador a sostener, revisar o rechazar la hipótesis (y muy probablemente la teoría de la cual se originó).

Para saber si la hipótesis específica es confirmada por el conjunto de datos recabados, se debe seguir un procedimiento que permita rechazar o aceptar dicha hipótesis. Este procedimiento tiene que poder ser repetido por otros investigadores que quieran corroborar las conclusiones a las que se llegaron en una investigación previa.

Primeros pasos del procedimiento:

1. Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1)
2. Decidir qué datos se van a recabar y en qué condiciones.
3. Seleccionar una prueba estadística para intentar rechazar la H_0 .

La hipótesis nula (H_0) y la hipótesis alternativa (H_1)

La hipótesis nula (H_0) es una hipótesis de no efecto y, generalmente, se formula con el propósito de ser rechazada. En otras palabras, es la negación del punto que se está tratando de probar. Si es rechazada, entonces se acepta la hipótesis alternativa.

La hipótesis alternativa (H_1) es la declaración operacional de la hipótesis de investigación. La hipótesis de investigación es la predicción derivada de la teoría sometida a prueba.

En lo que respecta a la H_0 se plantea que la variable explicativa y la variable de respuesta son en realidad independientes, es decir no hay una asociación entre la primera y la segunda. Si los resultados del análisis estadístico arrojan resultados para rechazarla, entonces se puede aceptar la H_1 , la cual permite sostener que sí hay una asociación entre la variable explicativa y la variable de respuesta.

ERRORES COMUNES

1. Confundir la población de interés (población diana) con la población estadística. La primera es la población a la que se quieren extrapolar o generalizar los resultados de un estudio de investigación. La segunda, la población estadística, es aquella en la que se han estudiado todos sus elementos, o que se ha muestreado aleatoriamente mediante técnicas estadísticas. Para que se considere que una población ha sido muestreada de manera estadísticamente correcta tienen que cumplirse al menos dos condiciones:

- a. que el muestreo sea aleatorio y,
- b. que todos los integrantes de la población tengan una probabilidad mayor a cero de estar en la muestra.

Es un error intentar demostrar que una muestra no aleatoria refleja los valores de la población. Si resulta que dicha muestra refleja los valores de la población, esto no refleja nada. Lo importante es que la inferencia estadística es cierta bajo el supuesto de la *distribución al azar*.

2. Realizar conclusiones inferenciales en muestras no aleatorias. Ahora bien, en caso de que no se cumplan las dos condiciones arriba mencionadas no es posible hacer inferencia estadística, sin embargo, se pueden extrapolar los resultados mediante consideraciones técnicas, pero no es correcto hacer inferencia estadística.

3. La certeza sobre las muestras representativas. Una muestra representativa es la que refleja los parámetros de la población objetivo del estudio. Sin embargo, aunque el muestreo sea estadísticamente correcto no se puede estar seguro de que la muestra sea representativa. Por efectos del azar, puede suceder que los resultados muestrales se alejen de los valores poblacionales. La única manera de saber si una muestra es representativa de una población es estudiar la población entera (algo muy difícil y en muchas ocasiones imposible) y comprobar que los parámetros poblacionales y muestrales coinciden. Los métodos estadísticos sólo permiten conocer la *probabilidad* de que un determinado parámetro esté dentro un determinado intervalo en referencia a la población muestreada.

4. Confundir los **errores tipo I** (falso positivo) y **tipo II** (falso negativo) con sus probabilidades. El error tipo I es la probabilidad de rechazar la hipótesis nula por azar, cuando en realidad es cierta. A la probabilidad de cometerlo se le denomina α . El error tipo II es la probabilidad de no rechazar la hipótesis nula cuando es cierta la hipótesis alternativa. A la probabilidad de cometerlo se denomina β . Los errores tipo I y error tipo II son conceptos y α y β son probabilidades. Por lo tanto, no se les puede denominar errores tipo α y β .

<i>¿Qué opciones tenemos?</i>	H_0 no se puede rechazar	H_0 se puede rechazar
Conservar H_0	Decisión acertada	Error tipo II
Rechazar H_0	Error tipo I	Decisión acertada

¿Cuál es el factor que está en juego? Cuando rechazamos la H_0 nunca sabemos con toda certeza que ésta sea falsa. Sin embargo, cuanto menos sea el *azar* la causa de los resultados, mayor confianza tendremos que no haber cometido un error al haber rechazado la H_0 . Por ejemplo, cuando la probabilidad de obtener los resultados es una en un millón ($p = 0.000001$) bajo el supuesto del azar, consideramos o tenemos más confianza en suponer que la H_0 es falsa que cuando la probabilidad es una en 10 ($p = 0.10$).

De aquí que el nivel de significación (nivel α) que se escoja al inicio del estudio es el que limita la probabilidad de cometer un error del tipo I. Si el investigador decide escoger $\alpha = 0.05$ está diciendo que rechazará la H_0 , bajo el supuesto del azar, la probabilidad obtenida sea igual o menor que 5 veces en 100. Esto es, restringe la probabilidad de cometer el error I a 0.05.

Pero al seleccionar un criterio más estricto, un nivel α más conservador, incrementamos la probabilidad de cometer un error del tipo II.

Nivel α	p obtenida ($\leq \alpha$)	¿Qué hacemos?	H_0 no se puede rechazar	H_0 se puede rechazar
0.05	0.02	Rechazar H_0	Error tipo I	Decisión acertada
0.01	0.02	Conservar H_0	Decisión acertada	Error tipo II

5. Considerar que se ha demostrado la hipótesis nula cuando no se ha rechazado. Si en un contraste de hipótesis no se rechaza la hipótesis nula, no se puede considerar que ha sido demostrada, ni siquiera es una prueba a favor de que sea cierta. En otras palabras, la no evidencia no es una evidencia de no evidencia. Un contraste de hipótesis no implica la demostración de una de las dos hipótesis.

6. Considerar que la significación estadística es el parámetro para evaluar las conclusiones de un estudio. Lo importante es la significación técnica de los valores obtenidos y de las diferencias observadas y como dato secundario la significación estadística.

Inferencia estadística. La extrapolación de los resultados experimentales obtenidos a partir de la muestra a la población estadística se denomina inferencia estadística.

Inferencia técnica. La extrapolación de los resultados experimentales a poblaciones distintas de la población estadística, se denomina inferencia técnica, ya que la posibilidad de hacerlo depende de los conocimientos y de la lógica de la disciplina científica implicada en el estudio.