Cap 5

Error tipo I: rechazar H0 en favor de HA cuando H0 es en realidad verdadera.

Error tipo II: no rechazar H0 en favor de HA cuando HA es en realidad verdadera

Se define \_ como la probabilidad de cometer errores de tipo II. alpha y beta están relacionados: para un tamaño constante de la muestra, al reducir beta, alpha aumenta, fenómeno que se evidencia con mayor fuerza mientras más pequeña sea la muestra. No obstante, en la práctica resulta más interesante conocer la probabilidad de no cometer errores de tipo II. Esto conduce a un nuevo concepto: el poder de una prueba de hipótesis, dado por 1 - beta, que puede definirse como la probabilidad de correctamente rechazar H0 cuando es falsa. Otra forma de entender la noción de poder de una prueba es qué tan propensa es esta para distinguir un efecto real de una simple casualidad. Esto lleva a la noción de tamaño del efecto, que corresponde a una cuantificación de la diferencia entre dos grupos o la diferencia real entre dos medias.

Cap 6

Ejemplo estudiado:

Al usar la distribución t de Student para la diferencia de medias se deben cumplir los siguientes requisitos:

1. Cada muestra cumple las condiciones para usar la distribución t.

2. Las muestras son independientes entre sí.

Un equipo médico desea determinar si una nueva vacuna A es más efectiva que otra vacuna B, a fin de

inmunizar a la población mundial contra una terrible enfermedad. Para ello, se reclutó a un grupo de 28

voluntarios en diferentes países, 15 de los cuales (seleccionados al azar) recibieron la vacuna A y los 13 restantes,

la vacuna B. Las concentraciones de anticuerpos (en microgramos por cada mililitro) al cabo de un

mes, para cada voluntario que recibió la vacuna A fueron: 9,534639; 4,458274; 9,961477; 7,831859; 8,587947; 4,233574; 0,957678; 10,503414; 10,009914; 23,643719; 4,793084; 13,441451; 7,042911; 5,303227; 9,557578. Asimismo, para quienes recibieron la vacuna B fueron: 16,178461; 15,687254; 4,879188; 13,768377; 15,591773; 4,493893; 17,849865; 10,549353; 29,092156; 7,951576; 15,851443; 9,418612; 18,309343.

Las hipótesis a formular en este caso son:

H0: No hay diferencia entre la efectividad promedio de ambas vacunas.

HA: La vacuna A es, en promedio, más efectiva que la B.

En lenguaje matemático:

H0: μA = μB HA: μA > μB

Ambas muestras son independientes entre sí, pues son diferentes voluntarios y fueron designados aleatoriamente a cada grupo. Además, se puede asumir que las observaciones son independientes, pues cada muestra es significativamente menor a la población total a vacunar. En cuanto al supuesto de normalidad para cada muestra, la figura 6.4 muestra que, en general, ambas distribuciones se acercan al supuesto de normalidad, aunque cada una de ellas presenta un valor atípico. En consecuencia, se debe proceder con algo más de cautela pues, además, las muestras son pequeñas. Dada la presencia de valores atípicos y lo reducido de las muestras, se debería considerar un nivel de significación más exigente. Además, en este escenario, un error tipo I (rechazar H0 cuando es verdadera) implicaría reducir innecesariamente la cantidad de vacunas disponibles y retrasar el proceso de vacunación, poniendo en riesgo a todos los habitantes del planeta. Un error tipo II, en cambio, podría causar que se continúe el uso indistinto de ambas vacunas retrasando el efecto inmune en la población. En consecuencia, el error tipo I es más grave, por lo que el nivel de significación debiese ser aún más exigente. En consecuencia, se opta por alpha = 0, 01.

se obtiene que la diferencia entre las medias es −5, 15964 y que el intervalo de confianza es [−10, 82697;∞). Además, el valor p es p = 0.9838. Esto significa que hay muy poca evidencia en favor de HA, por lo que se falla al rechazar la hipótesis nula y no es posible afirmar que la vacuna A sea, en promedio, mejor que la vacuna B.

Cap 7

El poder de la prueba aumenta mientras mayor es el tamaño del efecto. A medida que el tamaño del efecto disminuye (es decir, el estimador se acerca al valor nulo), el poder se aproxima al nivel de significación. Usar un valor de más exigente (menor) manteniendo constante el tamaño de la muestra hace que la curva de poder sea más baja para cualquier tamaño del efecto (lo que verifica la relación entre alpha y beta ). Usar una muestra más grande aumenta el poder de la prueba para cualquier tamaño del efecto distinto de 0.

el tamaño del efecto en la inferencia con dos muestras, ya sea diferencia de medias o muestras pareadas, es la llamada d de Cohen. se considera que d = 0; 2 es un efecto pequeño (imperceptible a simple vista), d = 0; 5 es un efecto mediano (probablemente perceptible a simple vista) y d = 0; 8, un efecto grande (definitivamente perceptible a simple vista). Una gran ventaja del poder estadístico es que es de gran ayuda para determinar el tamaño adecuado de la muestra para detectar un tamaño del efecto dado.

Sobre el poder:

el poder es la probabilidad de correctamente rechazar H0 cuando es falsa, lo que equivale a la probabilidad de distinguir un efecto real de una mera casualidad.

Cap 8

Ejemplo prueba de hipótesis para una proporción

n = 150 y ^p = 0; 64. El desarrollador del programa afirma que más del 70% de las instancias se ejecutan en menos de 25 segundos, pero el ingeniero no está seguro, por lo que decide comprobarlo mediante una prueba de hipótesis con un nivel de significación alpha= 0, 05:

H0: el 70% de las instancias se ejecutan en menos de 25 segundos.

HA: más del 70% de las instancias se ejecutan en menos de 25 segundos. El valor nulo es entonces p0 = 0; 7, con lo que las hipótesis anteriores pueden formularse matemáticamente

como:

H0: p = p0

HA: p >p0

Ya se había comprobado anteriormente que se verifica la independencia de las observaciones. Además, se

espera encontrar np0 = 150 \* 0, 7 = 105 éxitos y n(1 - p0) = 150 \* (1 – 0, 7) = 45 fracasos, ambos valores

mayores que 10, por lo que la condición de éxito-fracaso se verifica.

El valor p asociado, calculado en R mediante la llamada a la función pnorm(-1,6036, lower.tail=FALSE),

obteniéndose como resultado p = 0; 9456. En consecuencia, la evidencia no es suficiente para rechazar la hipótesis nula, por lo que se concluye, con 95% de confianza, que no es cierto que el algoritmo se ejecute en

menos de 25 segundos para más del 70 %. La prueba de hipótesis anterior puede efectuarse en R como se muestra en el script 8.1, usando para ello la función prop.test(), cuyos principales parámetros son:

x: cantidad de éxitos en la muestra.

n: tamaño de la muestra.

p: valor nulo (por defecto, p=0).

alternative: tipo de hipótesis alternativa, por defecto bilateral (alternative=“two.sided”), y

valores “less” y “greater” para hipótesis unilaterales.

conf.level: nivel de confianza (conf.level=0.95 por defecto).

Cap 9

Prueba exacta de Fisher

Esta prueba permite determinar si dos variables categóricas (con dos niveles cada una) son independientes,

vale decir, si la variación de una de ellas no afecta a la otra. Así,

H0: las variables son independientes.

HA: las variables no son independientes.

La prueba exacta de Fisher calcula las probabilidades para todas las posibles tablas que puedan tener los

totales por fila y por columna observados en la muestra, y calcula el p-valor como la suma de las probabilidades de todas las tablas con probabilidad menor o igual que la tabla dada.

Prueba de McNemar

Esta prueba resulta apropiada cuando una misma característica se mide en dos ocasiones diferentes para los

mismos sujetos (muestras pareadas) y se desea determinar si se produce o no un cambio significativo entre

ambas mediciones. En consecuencia, la hipótesis nula establece que no hay diferencia entre las proporciones de ambas mediciones de la variable de interés.

Prueba de homogeneidad

Esta prueba resulta adecuada para determinar si dos poblaciones (variable dicotómica) presentan las

mismas proporciones en los diferentes niveles de una variable categórica.

EJ:Suponga que la sociedad científica de computación ha realizado una encuesta a programadores con más de 3 años de experiencia, y les ha preguntado cuál es su lenguaje de programación favorito. La tabla 9.6 muestra las preferencias para cada lenguaje separados en programadores (varones) y programadoras (mujeres). ¿Son distintas las preferencias de lenguaje de programación entre hombres y mujeres?

Así, las hipótesis a contrastar son:

H0: programadores hombres y mujeres tienen las mismas preferencias en lenguaje de programación favorito

(ambas poblaciones muestras las mismas proporciones para cada lenguaje estudiado).

HA: programadores hombres y mujeres tienen preferencias distintas en lenguajes de programación favorito.

Prueba de bondad de ajuste

Esta prueba permite comprobar si una distribución observada se asemeja a una distribución

esperada. Usualmente se emplea para comprobar si una muestra es representativa de la población

ej: suponga que una empresa de desarrollo de software cuenta con una nómina

de 110 programadores, especialistas en diferentes lenguajes de programación. El gerente ha seleccionado una muestra aleatoria de 20 programadores para enviarlos a cursos de perfeccionamiento en sus respectivos lenguajes, pero desea verificar que la muestra sea representativa a fin de asegurar una mejora en la productividad de toda la empresa.

Para la prueba de bondad de ajuste, deben considerarse las mismas condiciones que para la prueba de

homogeneidad, es decir:

1. Las observaciones deben ser independientes.

2. Debe haber a lo menos 5 observaciones esperadas en cada grupo.

En este ejemplo, las hipótesis a contrastar son:

H0: las proporciones de especialistas en cada lenguaje son las mismas para la nómina y la muestra.

HA: las proporciones de especialistas en cada lenguaje son diferentes en la nómina que en la muestra.

En este caso, se puede proceder de igual manera que para la prueba de bondad de ajuste, Para este ejemplo, el valor p resultante es p = 0; 823, por lo que se falla al rechazar la hipótesis nula con niveles de significación típicos (como \_ = 0; 05 o \_ = 0; 01. En consecuencia, se concluye con 95% de confianza que la muestra seleccionada es, en efecto, representativa de la nómina de programadores de la empresa.

Prueba de independencia

Esta prueba permite determinar si dos variables categóricas, de una misma población, son estadísticamente

independientes o si, por el contrario, están relacionadas. Suponga que un micólogo desea determinar si existe relación entre la forma del sombrero de los hongos y si éstos son o no comestibles. Para ello, tras recolectar una muestra de 8.120 hongos. En este caso, las hipótesis a docimar son:

H0: las variables clase y forma del sombrero son independientes.

HA: las variables clase y forma del sombrero son dependientes.

Al ejecutar la prueba en R (script 9.5) se obtiene que el valor para el estadístico de prueba es \_2 = 485; 64,

con \_ = 4 grados de libertad y un p-valor < 2; 2 \_ 10􀀀16.

Cap 10

el procedimiento ANOVA requiere que se cumplan algunas condiciones:

1. La escala con que se mide la variable dependiente tiene las propiedades de una escala de intervalos

iguales.

2. Las k muestras son obtenidas de manera aleatoria e independiente desde la(s) población(es) de origen.

3. Se puede suponer razonablemente que la(s) población(es) de origen sigue(n) una distribución normal.

4. Las k muestras tienen varianzas aproximadamente iguales.

La variabilidad total puede descomponerse en dos componentes: uno de ellos corresponde a la variabilidad

existente al interior de cada uno de los grupos (o variabilidad intra-grupos), within groups en inglés, denotada por SSwg; el otro corresponde a la variabilidad entre los diferentes grupos, between groups en inglés, denotada como SSbg . La ecuación presenta una identidad importante que relaciona ambas componentes. SST = SSbg + SSwg

La variabilidad entre grupos permite medir de manera agregada la magnitud de las diferencias entre las

distintas medias muestrales.

La variabilidad intra-grupos corresponde a la suma total de las desviaciones cuadradas al interior de cada

grupo, por lo que representa la variabilidad aleatoria de cada uno de los diferentes grupos.

En el contexto de análisis de varianza, se denota como MS, del inglés mean square, a la media de las

desviaciones cuadradas. Para el caso de la variabilidad entre grupos, se tienen Vbg = k-1 grados de libertad,

donde k corresponde a la cantidad de grupos (para el ejemplo, Vbg = 3-1 = 2).