

Desarrollo de herramienta para estimar la tasa de migración de flujos genéticos en diferentes poblaciones

Cinthya Guadalupe Villegas López¹, Raquel Muñiz-Salazar²,
Dora-Luz Flores Gutiérrez³

¹Departamento de Sistemas y Computación, Instituto Tecnológico de Durango, Victoria de Durango, Durango, México

²Escuela de Ciencias de la Salud, Universidad Autónoma de Baja California, Ensenada, Baja California, México

³Facultad de ingeniería, arquitectura y diseño, Universidad Autónoma de Baja California, Ensenada, Baja California, México

31 de julio de 2020

Índice

Introducción

Prueba t

La prueba t (o también llama t de student) es un tipo de estadística inferencial usada para determinar si existe diferencia significativa entre las medias de dos grupos, las cuales pueden tener relación en ciertas características. Esta prueba hace uso del estadístico t , la distribución t y los grados de libertad para determinar la significancia.

Para el cálculo de esta prueba se requieren tres datos clave:

- Las diferencias entre las medias de cada conjunto de datos (conocido como diferencia de medias).
- La desviación estándar de cada grupo.
- El total de datos en cada uno de los grupos.

Dentro de la prueba t , se encuentran dos clasificaciones: prueba t dependiente e independiente.

- Correlacionada o pareada: El cálculo de la prueba t pareada las medias de dos variables de un solo grupo. El procedimiento calcula las diferencias entre los valores de las dos variables de cada caso y contrasta si la media difiere de 0.

La fórmula para calcular el valor t para una prueba pareada es:

$$t = \frac{\mu_1 - \mu_2}{\frac{\sigma}{\sqrt{n}}} \quad (1)$$

donde:

μ_1 y μ_2 = Los valores promedio de cada uno de los conjuntos de datos. σ = La desviación estándar de la diferencia de los valores relacionados. n = El tamaño de la muestra.

Las pruebas restantes pertenecen a la categoría de independientes. Las muestras de esta clasificación se seleccionan independientemente unas de otras, es decir, los conjuntos de datos no contienen los mismos valores.

- No pareada con igual varianza: La prueba t con varianzas iguales es calculada cuando el número de muestras en cada grupo es el mismo, o la varianza de los dos conjuntos de datos es similar.

La fórmula para calcular el valor t para una prueba independiente con igual varianza es:

$$t = \frac{\mu_1 - \mu_2}{\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2 - 2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (2)$$

donde:

μ_1 y μ_2 = Los valores promedio de cada uno de los conjuntos de datos. σ_1^2 y σ_2^2 = La varianza de cada uno de los conjuntos de datos.

n_1 y n_2 = Cantidad de elementos en cada conjunto de datos.

n_1 y $n_2 - 2$ = Cálculo de los grados de libertad.

- No pareada con varianza desigual: La prueba t con varianza desigual se calcula cuando el número de muestras en cada grupo es diferente, y la varianza de los conjuntos de datos también es diferente. A esta prueba también se le conoce como t de Welch.

$$t = \frac{\mu_1 - \mu_2}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (3)$$

Intervalos de confianza entre pares

La estimación por intervalos de confianza consiste en determinar un posible rango de valores o intervalo (a,b), en el que, con una determinada probabilidad, sus límites contendrán el valor del parámetro poblacional que se está buscando. Para cada muestra se obtendrá un intervalo distinto que, para el $X\%$ de ellas, contendrá el verdadero valor del parámetro. A este intervalo se le denomina intervalo de confianza.

p Value

El valor p o nivel de significancia observado de una prueba estadística es el valor más pequeño de α para el cual H_0 se puede rechazar. Es el riesgo real de cometer un error tipo I, si H_0 es rechazada con base en el valor observado del estadístico de prueba. El valor p mide la fuerza de la evidencia contra H_0 .

Si el valor p es menor o igual a un nivel de significancia α asignado previamente, entonces la hipótesis nula puede ser rechazada y se puede informar que los resultados son estadísticamente significativos al nivel α .

Grados de libertad

El origen del término grados de libertad es teórico y se refiere al número de desviaciones independientes evaluadas al cuadrado en s^2 existentes para estimar σ^2 . Estos grados de libertad pueden cambiar para diferentes aplicaciones y, como especifica la distribución t correcta a usar, es necesario recordar que hay que calcular los grados de libertad correctos para cada aplicación.

Teorema del límite central

El teorema del límite central dice que, bajo condiciones más bien generales, las sumas y medias de muestras aleatorias de mediciones tomadas de una población tienden a tener una distribución aproximadamente normal.

Combinaciones y permutaciones

Una *combinación* de los objetos es una selección de estos sin importar el orden. Una *permutación* es un arreglo particular de los objetos que forman un conjunto.

Errores tipo I y II

Un error tipo *I* para una prueba estadística es el error de rechazar la hipótesis nula cuando es verdadera. El nivel de significancia para una prueba estadística de hipótesis es:

$$\alpha = P(\text{error tipo I}) = P(\text{rechazar falsamente } H_0) = P(\text{rechazar } H_0 \text{ cuando es verdadera}) \quad (4)$$

Este valor α representa el máximo riesgo tolerable de rechazar incorrectamente H_0 .

Una vez fijo este nivel de significancia, la región de rechazo se puede fijar para permitir que el investigador rechace H_0 con un grado fijo de confianza en la decisión. La probabilidad de cometer un error de tipo *I* se denota por el símbolo α .

Un error tipo *II* para una prueba estadística es el error de aceptar la hipótesis nula cuando es falsa y alguna hipótesis alternativa es verdadera. La probabilidad de cometer un error tipo *II* se denota por el símbolo β .

Materiales y Métodos

Resultado

Conclusión

Referencias