

Tarea Test Hipótesis

Carlos Saquel

En nuestra primera y segunda tarea analizamos el data set de canciones de Spotify. En ese contexto sería interesante analizar si la popularidad de una canción en Spotify tiene relación con su carácter explícito, para ello es necesario plantear un test de hipótesis adecuado para comparar la popularidad entre canciones explícitas y no explícitas. A continuación se explica el planteamiento paso a paso:

Paso 1: Planteamiento del Problema

Queremos investigar si existe una diferencia significativa en la popularidad de las canciones en Spotify según si la canción es explícita o no. Esto se puede traducir en términos de hipótesis de la siguiente manera:

- **Variable numérica:** Popularidad de la canción en Spotify.
- **Variable categórica:** Carácter explícito de la canción (dos categorías: explícita y no explícita).

Paso 2: Formulación de Hipótesis

Planteamos las hipótesis de nuestro test:

- **Hipótesis nula (H_0):** La popularidad promedio de las canciones explícitas y no explícitas es igual. En otras palabras, no hay diferencia significativa en la popularidad entre los dos

$$H_0 : \mu_{\text{explícita}} = \mu_{\text{no explícita}}$$

- **Hipótesis alternativa (H_1):** La popularidad promedio de las canciones explícitas es diferente de la de las canciones no explícitas. Esto implica que el carácter explícito podría estar asociado con una diferencia en la popularidad.

$$H_1 : \mu_{\text{explícita}} \neq \mu_{\text{no explícita}}$$

Paso 3: Selección del Test Estadístico

Como la distribución de la popularidad de las canciones en Spotify presenta una distribución normal utilizaremos la **prueba t de Student**.

Paso 4: Verificación de Supuestos

Antes de elegir el test definitivo, verificamos los siguientes puntos:

- **Normalidad:** Utiliza una prueba de normalidad (como Shapiro-Wilk) para verificar si la popularidad sigue una distribución normal en ambos grupos.
- **Homogeneidad de Varianzas:** Verifica si las varianzas de popularidad en ambos grupos son aproximadamente iguales con una prueba de homogeneidad. Si no son iguales, buscar un test alternativo..

Paso 5: Ejecución del Test

Se aplica la prueba t de Student para las muestras. Calcula el valor de t y su p-value asociado. Si el p-value es menor que el umbral de significancia (en este caso 0.05), se rechaza la hipótesis nula y se concluye que hay una diferencia significativa en la popularidad entre canciones explícitas y no explícitas.

Paso 6: Interpretación de Resultados

- **Si se rechaza H_0 :** Esto sugiere que la popularidad de las canciones en Spotify sí varía significativamente según si la canción es explícita o no.
- **Si no se rechaza H_0 :** Esto indica que, con el nivel de significancia seleccionado, no existe suficiente evidencia para afirmar que la popularidad de las canciones esté relacionada con su carácter explícito.

Este análisis permite investigar la relación entre la popularidad y el carácter explícito de las canciones, arrojando evidencia estadística para tomar una decisión informada sobre si este atributo influye en la popularidad de una canción en Spotify.

