

Prueba de hipótesis: prueba de una asociación

Prueba T de dos muestras

Podemos probar una asociación entre una variable cuantitativa y una variable categórica binaria utilizando una prueba t de dos muestras. La hipótesis nula para una prueba t de dos muestras es que la diferencia en las medias de los grupos es igual a cero. Se puede implementar una prueba t de dos muestras en Python usando la función `ttest_ind()` de `scipy.stats`. El código de ejemplo muestra una prueba t de dos muestras para probar una asociación entre la longitud de las garras y la especie de oso (grizzly o negro).

```
de scipy.estadísticas importar ttest_ind

#separar la longitud de las garras de dos especies
oso_grizzly = datos.longitud_garra
[datos.especie==] 'grizzly'
oso_negro = datos.longitud_garra
[datos.especie==] 'negro'

#ejecute la prueba t aquí:
tstat, pval = ttest_ind
(oso_grizzly,oso_negro)
```

Pruebas T múltiples de dos muestras

Para probar una asociación entre una variable cuantitativa y una variable categórica no binaria, uno *podría* utilizar múltiples t- pruebas. Sin embargo, ejecutar varias pruebas aumenta la probabilidad de un falso positivo (error tipo I) de modo que sea mayor que el umbral de significancia para cada prueba. Para evitar este problema, una mejor solución es ejecutar un ANOVA; luego, si el valor p del ANOVA es significativo, ejecute la prueba de rango de Tukey.

Análisis de variación

Se utiliza un análisis de varianza (ANOVA) para probar una asociación entre una variable categórica no binaria y una variable cuantitativa al tiempo que se limita la probabilidad de un error de tipo I. La hipótesis nula para ANOVA es que las medias del grupo son todas iguales. La hipótesis alternativa es que al menos un par de medias grupales son diferentes. Se puede implementar un ANOVA en Python usando la función `f_oneway()` de `scipy.stats`. El código de ejemplo muestra una prueba ANOVA para una asociación entre la altura del árbol y las especies de árboles (pino, roble o abeto).

```
de scipy.estadísticas importar f_oneway
fstat, pval = f_oneway
(alturas_pino, alturas_roble,
alturas_abeto)
```

Prueba de alcance de Tukey

La prueba de rango de Tukey debe usarse después del ANOVA (si el valor p es significativo) para comparar simultáneamente las medias de los grupos para todos los pares posibles de grupos mientras se mantiene alguna probabilidad preseleccionada de un error de tipo I. Para cada par de grupos, la prueba de rango de Tukey indicará si se debe “rechazar el resultado nulo” y concluirá que esos dos grupos son significativamente diferentes. La prueba de alcance de Tukey se puede implementar con la función `pairwise_tukeyhsd()` de `statsmodels.stats.multicomp`. El código de ejemplo muestra cómo utilizar esta función para examinar una asociación entre la altura de los árboles y las especies de árboles utilizando una tasa de error general de tipo I de 0,05.

```
# Prueba de alcance de Tukey
de modelos de estadísticas.estadísticas.
multicomp importar pairwise_tukeyhsd
tukey_results =
pairwise_tukeyhsd(tree_data .altura,
datos_árbol. especie, 0, 05)
```

Supuestos de prueba de hipótesis

Antes de utilizar pruebas t para dos muestras, ANOVA o prueba de rango de Tukey, es importante comprobar si los supuestos de las pruebas son verdaderos:

- (1) Todas las observaciones deben ser muestreadas de forma independiente y aleatoria.
- (2) Las desviaciones estándar de los grupos deben ser iguales.
- (3) Los datos deben tener una distribución normal o el tamaño de la muestra debe ser grande
- (4) Los grupos creados por la variable categórica deben ser independientes.

Prueba de chi-cuadrado

Para probar una asociación entre dos variables categóricas, podemos usar una prueba de Chi-Cuadrado. La hipótesis nula para una prueba de Chi-Cuadrado es que no existe asociación entre las variables y la hipótesis alternativa es que sí existe asociación entre las variables. Se puede implementar una prueba de Chi-Cuadrado en Python usando la función `chi2_contingency()` de `scipy.stats`. El código de ejemplo muestra cómo implementar una prueba de Chi-Cuadrado para investigar una asociación entre la versión de un sitio web que alguien vio y si se suscribió o no.

```
importar pandas como pd
de scipy.estadísticas importar
chi2_contingencia

# crear tabla de contingencia
ab_contingencia = pd.tabulación cruzada
(datos.Versión_web, datos.Suscrito)

# ejecutar una prueba de chi-cuadrado
chi2, pval, gld , esperado =
chi2_contingencia( ab_contingencia)
```

Supuestos de chi-cuadrado

El uso adecuado de la prueba Chi-Cuadrado requiere que se cumplan ciertos supuestos. El primer supuesto es que las observaciones son independientes y aleatorias para garantizar que la muestra represente adecuadamente a la población. El siguiente supuesto es que las categorías de ambas variables son mutuamente excluyentes; esto es así, las observaciones sólo pueden caer en una categoría u otra, pero no en ambas. Finalmente, los grupos creados por las variables categóricas deben ser independientes; Ninguno de los grupos debería tener ninguna influencia sobre el otro.

 **Imprimir**  **Compartir** ▼