

# Hoja informativa: Prueba de hipótesis

## Práctica

```
# probar una hipótesis sobre la media de la población igualando un valor dado
# matriz es la muestra
# interested_value es la media propuesta para la prueba

from scipy import stats as st

results = st.ttest_1samp(
    array,
    interested_value)

print('p-value: ', results.pvalue)

# Probar una hipótesis sobre las medias de dos poblaciones estadísticas en función de mues
tras tomadas de ellas
# sample_1 es la muestra de la primera población estadística
# sample_2 es la muestra de la segunda población estadística
# equal_var define si las varianzas de las muestras se consideran iguales o no; el valor p
redeterminado es Verdadero

from scipy import stats as st

sample_1 = [...]
sample_2 = [...]

results = st.ttest_ind(
    sample_1,
    sample_2,
    equal_var = True)

print('p-value: ', results.pvalue)

# Probar una hipótesis sobre las medias de dos poblaciones estadísticas que son iguales pa
ra muestras dependientes (pareadas)
# pair_1 es la primera muestra pareada
# pair_2 es la segunda muestra pareada

from scipy import stats as st

results = st.ttest_rel(
    before,
    after)
```

```
print('p-value: ', results.pvalue)
```

## Teoría

**Población estadística** es un gran conjunto de datos para estudios estadísticos.

Una **muestra** es una porción seleccionada de la población estadística.

Una **muestra representativa** es una porción de los datos que representa a la población estadística completa.

Una **muestra aleatoria** es una porción de la población estadística seleccionada aleatoriamente.

**Estratos** son grupos en una población estadística que están unidos por una característica común.

Una **muestra estratificada** es una muestra compuesta por muestras proporcionales de diferentes estratos.

La **media muestral** es la media de una muestra.

La **varianza muestral** es la varianza de una muestra.

El **error estándar estimado** es la desviación estándar de la media muestral con respecto a la media real de la población estadística (**S** es la desviación estándar estimada de la población estadística; **n** es el tamaño de la muestra):

$$E.S.E. = \frac{S}{\sqrt{n}}$$

La **hipótesis nula (H<sub>0</sub>)** es una hipótesis que se prueba con la muestra.

La **hipótesis alternativa (H<sub>1</sub>)** es la hipótesis de significado opuesto a la hipótesis nula.

La **significación estadística** es la probabilidad total de que un valor medido empíricamente se encuentre alejado de la media.

La **estadística de diferencia** es el número de desviaciones estándar entre los valores comparados, si ambas distribuciones se convierten en una distribución normal estándar con media 0 y desviación estándar 1.

El **valor p** es la probabilidad de obtener un resultado al menos tan extremo como el que estás considerando, suponiendo que la hipótesis nula sea correcta.

Una **muestra pareada** es una muestra utilizada para medir variables para las mismas unidades.