

## Desafío - Pruebas de hipótesis

- Para realizar este desafío debes haber estudiado previamente todo el material disponible correspondiente a la unidad.
- Una vez terminado el desafío, comprime la carpeta que contiene el desarrollo de los requerimientos solicitados y sube el `.zip` en el LMS.
- Desarrollo desafío: Individual.
- Para la realización del desafío necesitarás apoyarte del archivo *Apoyo Desafío - Pruebas de hipótesis*.

### Habilidades a evaluar

- Identificar la estandarización de variables mediante el cálculo de puntajes  $z$ .
- Aplicar funciones a columnas de datos mediante `ufuncs`,
- Conocer la distribución “ $t$ ” de Student y su aplicación.
- Aplicar pruebas de hipótesis simples en el contexto de la inferencia.

### Descripción

A partir del trabajo realizado en el Desafío anterior, la empresa para la cual trabajas, te ha solicitado continuar el análisis de datos pero, esta vez, evaluando juicios de hipótesis e implementando funciones que faciliten el análisis de datos.

Para ello, te entregan un material de apoyo el cual deberá servir como base para desarrollar los siguientes requerimientos.

### Requerimientos

A continuación revisaremos los requerimientos y acciones que la empresa a la cual postulas te pide realizar.

1. **Evaluar juicios de hipótesis (2 puntos):** para evaluar la pertinencia de diversas hipótesis, se te solicita discernir si es posible rechazar la hipótesis nula.
  - *75 individuos elegidos al azar fueron alcanzados por activistas pro-LGTB que se identificaron como homosexuales, y 75 alcanzados por activistas pro-LGTB que se identificaron como heterosexuales. El objetivo era medir actitudes frente a la adopción homoparental. La organización encargada de procesar los datos obtuvo los siguientes resultados:*

- El 67% de los encuestados por heterosexuales se mostró a favor de la adopción homoparental, mientras que un 72% de los encuestados por activistas que se identificaron como homosexuales se mostraron a favor de la adopción homoparental.
- Asumiendo que la hipótesis nula es que ambos porcentajes no son diferentes, y la hipótesis alternativa es que existe una diferencia substancial entre ambos.
- Con p-valor de 0.1183 ¿es posible rechazar la hipótesis nula en favor de la alternativa con una confianza de 95%?, ¿Cuanta es la confianza máxima con la que sería posible rechazar la hipótesis nula en favor de la alternativa con estos resultados?

## 2. Importar la base de datos utilizada la sesión presencial anterior (2 puntos):

- Agregue una serie de variables binarias por cada continente de la variable region.



**Tip:** Utilice `np.where` para ello.

- De manera similar a la sesión anterior, enfóquese en las siguientes variables:
  - Apellidos desde la A hasta la N: Enfocarse en las variables `chldmort`, `adfert` y `life`.
  - Apellidos desde la N hasta la Z: Enfocarse en las variables `femlab`, `literacy` y `school`.

## 3. Implementar una función de prueba de hipótesis a mano (3 puntos):

- La función debe ingresar los siguientes argumentos:
  - `df`: La tabla de datos.
  - `variable`: La variable a analizar.
  - `binarize`: El indicador binario a utilizar.



**Tip:**

- ❖ Separe la variable en dos, utilizando el indicador binario. Recuerde eliminar los perdidos con `dropna()`.
- ❖ Implemente `ttest_ind` de `scipy.stats` y guarde el valor `t` y `pval`.
- ❖ Reporte las medias para cada grupo (0 y 1).
- ❖ Reporte la diferencia de entre las medias.
- ❖ Reporte los valores `t` y `p`

4. Implementar una función que grafique los histogramas para ambas muestras (3 puntos)

- Genere una función que devuelva un gráfico donde visualice los dos histogramas cuando la variable es 1 y 0, respectivamente.



**Tip:** Refactorize la función incluyendo el método `hist` de `matplotlib.pyplot`. Incluya los argumentos `alpha` y `label`.

- Para las tres variables de interés acorde a su grupo, analice las diferencias de medias por cada continente, y posteriormente grafique. Concluya con los principales resultados al respecto.