Importamos las librerías necesarias para el análisis

```
pip install researchpy
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting researchpy
  Downloading researchpy-0.3.5-py3-none-any.whl (33 kB)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (from researchpy) (1.10.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from researchpy) (1.22.4)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from researchpy) (1.5.3)
Requirement already satisfied: statsmodels in /usr/local/lib/python3.10/dist-packages (from researchpy) (0.13.5)
Requirement already satisfied: patsy in /usr/local/lib/python3.10/dist-packages (from researchpy) (0.5.3)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas->researchpy) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->researchpy) (2022.7.1)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages (from patsy->researchpy) (1.16.0)
Requirement already satisfied: packaging>=21.3 in /usr/local/lib/python3.10/dist-packages (from statsmodels->researchpy) (23.1)
Installing collected packages: researchpy
Successfully installed researchpy-0.3.5
```

```
import pandas as pd
import numpy as np
import seaborn as sns
import researchpy as rp
import scipy.stats as st
from scipy.stats import f_oneway
from scipy.stats import ttest_ind
```

De nuevo, usaremos la base de datos Iris

```
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']
iris = pd.read_csv(url, names=names)
iris.head()
```

	sepal-length	sepal-width	petal-length	petal-width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

Tipos de pruebas:

Referencia: https://www.pythonfordatascience.org/independent-samples-t-test-python/

a) Distribución t de Student

Suposiciones de prueba paramétrica

- Las distribuciones de población son normales.
- Las muestras tienen varianzas iguales
- Las dos muestras son independientes.

1. $H_0 : \mu_1 - \mu_2 \leq D_o$
2. $H_0 : \mu_1 - \mu_2 \geq D_o$
3. $H_0 : \mu_1 - \mu_2 = D_o$

1. $H_A : \mu_1 - \mu_2 > D_o$
2. $H_A : \mu_1 - \mu_2 < D_o$
3. $H_A : \mu_1 - \mu_2 \neq D_o$

Típicamente D_0 se establece en 0 y se está probando la tercera hipótesis, es decir, no hay diferencia entre los grupos. La estadística de prueba es el valor t y se puede calcular usando la siguiente fórmula. El estadístico de prueba es el valor t y se puede calcular mediante la siguiente fórmula:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - D_0}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

Dónde s_p es la desviación estándar agrupada y se calcula como $$s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

Se rechaza la hipótesis nula,H_0, si el valor t calculado es mayor o menor que el valor t crítico. El valor t crítico está determinado por los grados de libertad y alfa,α, valor. Los grados de libertad se calculan como df=n_1+n_2 -2 y α seestablece comúnmente en 0,05. Rechazar si

1. $t \geq t_\alpha$
2. $t \leq -t_\alpha$
3. $|t| \geq t_{\frac{\alpha}{2}}$

PRUEBA T INDEPENDIENTE CON RESEARCHPY

```
rp.ttest(iris['sepal-length'],iris['sepal-width'])
```

```
/usr/local/lib/python3.10/dist-packages/researchpy/ttest.py:38: FutureWarning: The series.append method is deprecated and will be r
  groups = group1.append(group2, ignore_index= True)
(       Variable      N      Mean        SD        SE  95% Conf.  Interval
 0  sepal-length  150.0  5.843333  0.828066  0.067611   5.709732  5.976934
 1   sepal-width  150.0  3.054000  0.433594  0.035403   2.984044  3.123956
 2      combined  300.0  4.448667  1.544988  0.089200   4.273127  4.624206,
                          Independent t-test   results
 0  Difference (sepal-length - sepal-width) =     2.7893
 1                       Degrees of freedom =   298.0000
 2                                        t =    36.5482
 3                    Two side test p value =     0.0000
 4                   Difference < 0 p value =     1.0000
 5                   Difference > 0 p value =     0.0000
 6                               Cohen's d =     4.2202
 7                               Hedge's g =     4.2096
 8                           Glass's delta1 =     3.3685
 9                         Point-Biserial r =     0.9042)
```

```
tabla, results =rp.ttest(group1=iris['sepal-length'],group2=iris['sepal-width'])
print(tabla)
```

```
       Variable      N      Mean        SD        SE  95% Conf.  Interval
0  sepal-length  150.0  5.843333  0.828066  0.067611   5.709732  5.976934
1   sepal-width  150.0  3.054000  0.433594  0.035403   2.984044  3.123956
2      combined  300.0  4.448667  1.544988  0.089200   4.273127  4.624206
/usr/local/lib/python3.10/dist-packages/researchpy/ttest.py:38: FutureWarning: The series.append method is deprecated and will be r
  groups = group1.append(group2, ignore_index= True)
```

```
print(results)
```

```
                         Independent t-test   results
0  Difference (sepal-length - sepal-width) =     2.7893
1                       Degrees of freedom =   298.0000
2                                        t =    36.5482
3                    Two side test p value =     0.0000
4                   Difference < 0 p value =     1.0000
5                   Difference > 0 p value =     0.0000
6                               Cohen's d =     4.2202
7                               Hedge's g =     4.2096
8                           Glass's delta1 =     3.3685
9                         Point-Biserial r =     0.9042
```

La logitud media del sepalo es 5,84333 y el intervalo es de (5.709,5.97) con 95% de confiabilidad, fue significativamente mayor que para el ancho del sepalo donde la media del ancho es 3.054 y el intervalo es de (2.984044,3.123956); t(209)=36.54 , p_V=0. Como el |p_V|< $t_{\alpha/2}$ entoces rechazamos la Ho.

PRUEBA T INDEPENDIENTE CON SCIPY.STATS

```
t,p = ttest_ind(iris['sepal-length'], iris['sepal-width']) # equal_var indica que las varianza no son iguales
```

```
print('t=',t)
print('p=',p)
```

```
t= 36.548157693982006
p= 3.987838114848222e-112
```

Actividad

Investigar la relación que hay entre la longitud de sepalo y del pétalo. Recuende que la hipótesis nula es que son similares y la altenativa es que no estan relacionadas

```
t,p = ttest_ind(iris['sepal-length'], iris['petal-length'])
print('t=',t)
print('p=',p)
```

```
t= 13.099504494510061
p= 2.8297338637366177e-31
```

```
print("================================")
rp.ttest(iris['sepal-length'],iris['petal-length'])
```

```
================================
/usr/local/lib/python3.10/dist-packages/researchpy/ttest.py:38: FutureWarning: The series.append method is deprecated and will be r
  groups = group1.append(group2, ignore_index= True)
(       Variable      N      Mean        SD        SE  95% Conf.  Interval
 0  sepal-length  150.0  5.843333  0.828066  0.067611   5.709732  5.976934
 1  petal-length  150.0  3.758667  1.764420  0.144064   3.473994  4.043340
 2      combined  300.0  4.801000  1.727187  0.099719   4.604760  4.997240,
                            Independent t-test   results
 0  Difference (sepal-length - petal-length) =     2.0847
 1                          Degrees of freedom =   298.0000
 2                                       t =    13.0995
 3                       Two side test p value =     0.0000
 4                      Difference < 0 p value =     1.0000
 5                      Difference > 0 p value =     0.0000
 6                                  Cohen's d =     1.5126
 7                                  Hedge's g =     1.5088
 8                           Glass's delta1 =     2.5175
 9                         Point-Biserial r =     0.6045)
```

```
tabla, results =rp.ttest(group1=iris['sepal-length'],group2=iris['petal-length'])
print(tabla)
```

```
       Variable      N      Mean        SD        SE  95% Conf.  Interval
0  sepal-length  150.0  5.843333  0.828066  0.067611   5.709732  5.976934
1  petal-length  150.0  3.758667  1.764420  0.144064   3.473994  4.043340
2      combined  300.0  4.801000  1.727187  0.099719   4.604760  4.997240
/usr/local/lib/python3.10/dist-packages/researchpy/ttest.py:38: FutureWarning: The series.append method is deprecated and will be r
  groups = group1.append(group2, ignore_index= True)
```

```
print(results)
```

```
                           Independent t-test   results
0  Difference (sepal-length - petal-length) =     2.0847
1                        Degrees of freedom =   298.0000
2                                     t =    13.0995
3                     Two side test p value =     0.0000
4                    Difference < 0 p value =     1.0000
5                    Difference > 0 p value =     0.0000
6                                Cohen's d =     1.5126
7                                Hedge's g =     1.5088
8                         Glass's delta1 =     2.5175
9                       Point-Biserial r =     0.6045
```

✓ 0 s completado a las 18:54