

```
df =
pd.read_csv("https://raw.githubusercontent.com/Eowin2210/Estadistica-
inferencial/refs/heads/main/mental_health.csv")
df.dropna(inplace = True)
df
```

	Entity	Code	Year	\
0	Afghanistan	AFG	1990	
1	Afghanistan	AFG	1991	
2	Afghanistan	AFG	1992	
3	Afghanistan	AFG	1993	
4	Afghanistan	AFG	1994	
...	
6415	Zimbabwe	ZWE	2015	
6416	Zimbabwe	ZWE	2016	
6417	Zimbabwe	ZWE	2017	
6418	Zimbabwe	ZWE	2018	
6419	Zimbabwe	ZWE	2019	

Schizophrenia disorders (share of population) - Sex: Both - Age:

Age-standardized	\
0	0.223206
1	0.222454
2	0.221751
3	0.220987
4	0.220183
...	...
6415	0.201042
6416	0.201319
6417	0.201639
6418	0.201976
6419	0.202482

Depressive disorders (share of population) - Sex: Both - Age:

Age-standardized	\
0	4.996118
1	4.989290
2	4.981346

3	4.976958
4	4.977782
...	...
6415	3.407624
6416	3.410755
6417	3.411965
6418	3.406929
6419	3.395476

Anxiety disorders (share of population) - Sex: Both - Age: Age-standardized \

0	4.713314
1	4.702100
2	4.683743
3	4.673549
4	4.670810
...	...
6415	3.184012
6416	3.187148
6417	3.188418
6418	3.172111
6419	3.137017

Bipolar disorders (share of population) - Sex: Both - Age: Age-standardized \

0	0.703023
1	0.702069
2	0.700792
3	0.700087

4	0.699898
---	----------

...	...
-----	-----

6415	0.538596
------	----------

6416	0.538593
------	----------

6417	0.538589
------	----------

6418	0.538585
------	----------

6419	0.538580
------	----------

Eating disorders (share of population) - Sex: Both - Age: Age-standardized	
--	--

0	0.127700
---	----------

1	0.123256
---	----------

2	0.118844
---	----------

3	0.115089
---	----------

4	0.111815
---	----------

...	...
-----	-----

6415	0.095652
------	----------

6416	0.096662
------	----------

6417	0.097330
------	----------

6418	0.097909
------	----------

6419	0.098295
------	----------

[6150 rows x 8 columns]

#a) Hipótesis general: Los trastornos depresivos (X) tienen una considerable influencia significativa en el desarrollo de trastornos de ansiedad (Y), de tal manera que un aumento

en la presencia o gravedad de los trastornos depresivos podría estar relacionado con un incremento en la probabilidad de llegar a desarrollar trastornos de ansiedad.

#H0: Los trastornos depresivos pueden aumentar la probabilidad de

desarrollar trastornos de ansiedad,
esto nos quiere decir que existe una relación causal en la que los síntomas depresivos contribuyen al desarrollo o agravamiento de los trastornos de ansiedad.

H1: No existe una relación significativa entre los trastornos depresivos y los trastornos de ansiedad,
la existencia de un trastorno depresivo no influye en el desarrollo de un trastorno de ansiedad.

#Dado el bajo coeficiente de correlación ($r=0.1204$) y el coeficiente bajo de la determinación ($R^2=0.0145$), la relación entre ambas variables parece ser débil,
esto sugiere que aunque puede llegar haber una conexión entre la depresión y la ansiedad también
hay otros factores aun más relevantes que nos explican la variabilidad de los trastornos de ansiedad. Así que en pocas palabras se puede rechazar H_0 .

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df =
pd.read_csv("https://raw.githubusercontent.com/Aleman2202/Estadistica-Inferencial-II/refs/heads/main/mental_health.csv")
df.dropna(inplace = True)
df

X= df["Depressive disorders (share of population) - Sex: Both - Age: Age-standardized"]
Y= df["Anxiety disorders (share of population) - Sex: Both - Age: Age-standardized"]

plt.scatter(X, Y, color = "purple")
plt.xlabel("Depresión")
plt.ylabel("Ansiedad")
ax = plt.gca()
ax.spines["top"].set_visible(False)
ax.spines["right"].set_visible(False)

from scipy.stats import pearsonr
r, _ = pearsonr(X, Y)
print(f"Coeficiente de correlación: {r: 0.4f}")

print(f"Coeficiente de determinación: {r ** 2: 0.4f}")

import statsmodels.api as sm
x_constante = sm.add_constant(X)
modelo = sm.OLS(Y, x_constante).fit()
```

```

b0, b1 = modelo.params

fun = lambda x: b0 + b1 * x

Yc = fun(X)

plt.plot(X, Yc, color = "black")

nivel_de_confianza = 0.95
intervalo_de_confianza = modelo.conf_int(alpha = 1 -
nivel_de_confianza)
intervalo_de_confianza_b1 = intervalo_de_confianza.iloc[1]
print(f"Intervalo de confianza para b1 de {nivel_de_confianza: 0.0%}")
print(f"{intervalo_de_confianza_b1[0]: 0.4f} < b1 <
{intervalo_de_confianza_b1[1]: 0.4f}")

residuales = modelo.resid
plt.figure()
plt.scatter(X, residuales, color = 'green')
plt.xlabel("Depresión")
plt.ylabel("Ansiedad")
ax = plt.gca()
ax.spines["top"].set_visible(False)
ax.spines["right"].set_visible(False)
plt.axhline(y = 0, color = "black",)

from scipy.stats import shapiro
_, valor_p_sh = shapiro(residuales)
print(f"valor-p de Shapiro: {valor_p_sh: 0.4f}")

from statsmodels.stats.api import het_breuschpagan
_, valor_p_bp, _, _ = het_breuschpagan(residuales, x_constante)
print(f"valor_p de Breusch-Pagan: {valor_p_bp: 0.4f}")

# e) En este caso, 0.1204 nos indica una correlación muy débil y
positiva esto sugiere que hay una ligera tendencia
# de que una variable aumente cuando la otra también lo hace pero la
relación insignificante.

# e) Un R2 de 0.0145 (o 1.45%) significa que solo el 1.45% de la
variabilidad de la variable dependiente es explicada por la variable
independiente,
# mientras que el 98.55% se debe a otros factores no considerados en
el modelo.

# f) Dado que en el intervalo de confianza no se contiene el valor 0
se puede rechazar la hipótesis nula  $H_0: B_1 = 0$  (esto indicaría que no
hay una relación entre las variables).

```

pero aunque la relación es significativa su magnitud es pequeña, lo que sugiere que otros factores pueden influir en los trastornos de ansiedad además de los trastornos depresivos.

g)

Linealidad: No podemos evaluarla directamente sin un gráfico de residuos,

pero el bajo coeficiente de determinación ($R^2 = 0.0145$) nos dice que la relación entre las variables puede no ser lineal.

#Normalidad: No se cumple, los residuos no siguen una distribución normal.

#Homoscedasticidad: No se cumple, los residuos tienen una varianza no constante.

Coeficiente de correlación: 0.1204

Coeficiente de determinación: 0.0145

Intervalo de confianza para b_1 de 95%

0.1079 < b_1 < 0.1639

valor-p de Shapiro: 0.0000

valor_p de Breusch-Pagan: 0.0007

C:\Users\prisg\AppData\Local\Packages\

PythonSoftwareFoundation.Python.3.13_qbz5n2kfra8p0\LocalCache\local-

packages\Python313\site-packages\scipy\stats_axis_nan_policy.py:586:

UserWarning: scipy.stats.shapiro: For $N > 5000$, computed p-value may not be accurate. Current N is 6150.

res = hypotest_fun_out(*samples, **kws)

