```
In [1]:
         import pandas as pd
         import numpy as np
 In [3]: data=pd.read csv("C:/Users/USER/Desktop/Kisii University 2023 students projects/Pauline group/survey lung cance
In [20]: data.head
         <bound method NDFrame.head of</pre>
                                           GENDER
                                                   AGE
                                                        SMOKING YELLOW_FINGERS ANXIETY PEER_PRESSURE \
Out[20]:
                  Μ
                      69
                                                2
                                                         2
                                1
                  М
                                2
                                                1
                      74
                                                         1
                                                                         1
         1
                  F
         2
                      59
                                1
                                                1
                                                         1
                                                                         2
         3
                  Μ
                      63
                                2
                                                2
         4
                  F
                      63
                                1
                                                2
                                                         1
                                                                         1
         304
                      56
                                1
                                                1
         305
                      70
                  Μ
                                2
                                                1
                                                         1
                                                                         1
                                2
         306
                                                1
                  М
                      58
                                                         1
                                                                         1
         307
                  М
                      67
                                2
                                                1
                                                         2
                                                                         1
         308
                  М
                      62
                                                1
                                                                         2
              CHRONIC DISEASE
                               FATIGUE
                                         ALLERGY
                                                   WHEEZING
                                                             ALCOHOL CONSUMING
         0
                                      2
                                                1
         1
                            2
                                      2
                                                2
                                                                              1
         2
                                                          2
                            1
                                      2
                                                1
                                                                              1
         3
                            1
                                      1
                                                1
                                                          1
                                                                              2
         4
                            1
                                      1
                                                1
                                                          2
                                                                              1
                            2
                                      2
                                                                              2
         304
                                                1
                                                          1
         305
                            1
                                      2
                                                2
                                                          2
                                                                              2
         306
                            1
                                      1
                                                2
                                                          2
                                                                              2
                                                                              2
                                      2
                                                2
         307
                            1
                                                          1
         308
                            1
                                      2
                                                2
                                                          2
                                                                              2
              COUGHING SHORTNESS OF BREATH SWALLOWING DIFFICULTY
                                                                    CHEST PAIN
         0
                     2
         1
                     1
                                          2
                                                                  2
                                                                              2
                                          2
                                                                              2
         2
                     2
                                                                              2
         3
                     1
                                          1
                                                                 2
         4
                     2
                                          2
                                                                 1
                                                                              1
         304
                     2
                                          2
                                                                  2
                                                                              1
                     2
                                          2
         305
                                                                              2
                                                                 1
         306
                     2
                                          1
                                                                 1
                                                                              2
         307
                     2
                                          2
                                                                              2
         308
                     1
                                          1
                                                                              1
              LUNG CANCER
         0
                        1
         1
                        1
         2
                        0
         3
                        0
         4
                        0
         304
                        1
         305
                        1
         306
                        1
         307
                        1
         308
         [309 rows x 16 columns]>
 In [6]: print(data.columns)
         'SWALLOWING DIFFICULTY', 'CHEST PAIN', 'LUNG CANCER'],
               dtype='object')
 In [7]: # Define the independent variables
         X diet = data[['AGE', 'SMOKING', 'YELLOW FINGERS', 'ANXIETY', 'PEER PRESSURE', 'CHRONIC DISEASE', 'FATIGUE',
         # Define the dependent variable
         y cardio = data['LUNG CANCER']
         from sklearn.preprocessing import LabelEncoder
In [13]:
         # Initialize label encoder
         label_encoder = LabelEncoder()
         # Encode the 'LUNG_CANCER' column
         data['LUNG CANCER'] = label encoder.fit transform(data['LUNG CANCER'])
In [16]: data.columns
```

```
'SWALLOWING DIFFICULTY', 'CHEST PAIN', 'LUNG_CANCER'],
      dtype='object')
```

Objective 1: Investigate the Relationship Between Diet and Cardiovascular Health

```
In [18]: import statsmodels.api as sm
                           import pandas as pd
                           # Assuming 'data' is your DataFrame with the provided column names
                           # Define the independent variables
                          'SWALLOWING DIFFICULTY', 'CHEST PAIN']]
                          # Define the dependent variable
                          y cardio = data['LUNG CANCER']
                           # Add a constant term to the independent variables
                          X diet = sm.add constant(X diet)
                           # Fit the multiple linear regression model for Objective 1
                          model obj1 = sm.OLS(y cardio, X diet).fit()
                           # Print the summary of the linear regression model for Objective 1
                           print(model_obj1.summary())
                                                                                                    OLS Regression Results
                          _____
                         Dep. Variable:

Dep. Variable:

Model:

Mothod:

Date:

Time:

Date:

Time:

Date:

Time:

Date:

Da
                                                                                                                                                                                                                                    0.371
13.98
                                                                                                                                                                                                                       1.83e-25
                                                                                                                                                                                                                                        67.94
                          Df Residuals:
                                                                                                                                                                                                                                         123.9
                          Df Model:
                                                                                                                            14
                          Covariance Type:
                                                                                                     nonrobust
                                                                                                    coef std err t P>|t| [0.025 0.975]
                         CONST -1.2119 0.194 -6.258 0.000 -1.593 -0.831
AGE 0.0014 0.002 0.758 0.449 -0.002 0.005
SMOKING 0.0732 0.032 2.270 0.024 0.010 0.137
YELLOW_FINGERS 0.1197 0.041 2.945 0.003 0.040 0.200
ANXIETY 0.0768 0.043 1.800 0.073 -0.007 0.161
PEER_PRESSURE 0.0880 0.035 2.485 0.014 0.018 0.158
CHRONIC DISEASE 0.0912 0.032 2.865 0.004 0.029 0.154
FATIGUE 0.1503 0.039 3.891 0.000 0.074 0.226
ALLERGY 0.1488 0.033 4.453 0.000 0.074 0.226
ALLERGY 0.1488 0.033 4.453 0.000 0.083 0.215
WHEEZING 0.0600 0.035 1.714 0.088 -0.009 0.129
ALCOHOL CONSUMING 0.1945 0.038 5.125 0.000 0.120 0.269
ALCOHOL CONSUMING 0.1945 0.038 5.125 0.000 0.120 0.269
SHORTNESS OF BREATH 0.0445 0.039 1.138 0.256 -0.032 0.121
SWALLOWING DIFFICULTY 0.0997 0.038 2.609 0.010 0.024 0.175
CHEST PAIN 0.0279 0.033 0.836 0.404 -0.038 0.093
                           _____
                                                                                                          30.983 Durbin-Watson:
0.000 Jarque-Bera (JB):
-0.743 Prob(JB):
                          Omnibus:
Prob(Omnibus):
Skew:
                          Omnibus:
                                                                                                                                                                                                                                       1.742
```

Kurtosis:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

3.907 Cond. No.

Objective 2: Examine the Influence of Physical Activity on Cardiovascular Health

```
In [19]: import statsmodels.api as sm
        # Assuming 'data' is your DataFrame with the provided column names
       'SWALLOWING DIFFICULTY', 'CHEST PAIN']]
       # Define the dependent variable
       y_cardio = data['LUNG CANCER']
        # Add a constant term to the independent variables
       X physical activity = sm.add constant(X physical activity)
        # Fit the multiple linear regression model for Objective 2
        model obj2 = sm.OLS(y cardio, X physical activity).fit()
```

39.039 3.33e-09

826.

## OLS Regression Results

OLS Regression results										
Dep. Variable: Model: Method: Date: Time: No. Observations: Df Residuals: Df Model:	LUNG_CA Least Squ Tue, 05 Dec 23:0	NCER OLS ares 2023 0:11 309 294 14	R-sq Adj. F-st Prob	uared: R-squared: atistic: (F-statistic) Likelihood:	:	0.400 0.371 13.98				
Covariance Type:										
	coef	std	err	t	P> t	[0.025	0.975]			
	-1.2119 0.0014 0.0732 0.1197 0.0768 0.0880 0.0912 0.1503 0.1488 0.0600 0.1945 0.1058 0.0445 Y 0.0997	0 0 0 0 0 0 0 0 0 0	.002 .032 .041 .043 .035 .032 .039 .033 .035 .038	2.485 2.865 3.891 4.453 1.714 5.125 2.791	0.449 0.024 0.003 0.073 0.014 0.004 0.000 0.008 0.008 0.000 0.088 0.000	0.010 0.040 -0.007 0.018 0.029 0.074 0.083 -0.009 0.120 0.031	0.005 0.137 0.200 0.161 0.158 0.154 0.226 0.215 0.129 0.269 0.180 0.121			
CHEST PAIN						========	0.093			
Omnibuci	20	002	Durh	in Watson:		1 7/12				

Omnibus: 30.983 Durbin-Watson: 1.742 0.000 Jarque-Bera (JB): 39.039 Prob(Omnibus): -0.743 Prob(JB): 3.33e-09 Skew: Cond. No. Kurtosis: 3.907 826.

Notes

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Objective 3: Assess the Impact of Stress Levels on Cardiovascular Health

## OLS Regression Results

	LUNG_CANCER OLS Least Squares Tue, 05 Dec 2023 23:06:40 309 294 14 nonrobust	Adj. F-sta Prob Log-I AIC: BIC:	uared: R-squared: atistic: (F-statistic) ikelihood:	:	0.400 0.371 13.98 1.83e-25 -18.971 67.94 123.9	
		td err	t	P> t	[0.025	0.975]
CONST AGE SMOKING YELLOW_FINGERS ANXIETY PEER_PRESSURE CHRONIC DISEASE FATIGUE ALLERGY WHEEZING ALCOHOL CONSUMING COUGHING SHORTNESS OF BREATH SWALLOWING DIFFICULT CHEST PAIN	0.0279	0.041 0.043 0.035 0.032 0.039 0.033 0.035 0.038 0.038 0.039 0.038	-6.258 0.758 2.270 2.945 1.800 2.485 2.865 3.891 4.453 1.714 5.125 2.791 1.138 2.609 0.836	0.024 0.003 0.073 0.014 0.004 0.000	-1.593 -0.002 0.010 0.040 -0.007 0.018 0.029 0.074 0.083 -0.009 0.120 0.031 -0.032 0.024 -0.038	-0.831 0.005 0.137 0.200 0.161 0.158 0.154 0.226 0.215 0.129 0.269 0.180 0.121 0.175 0.093
Omnibus: Prob(Omnibus): Skew: Kurtosis:	30.983 0.000 -0.743 3.907	Durb: Jarqu Prob Cond	No.		1.742 39.039 3.33e-09 826.	

## Notes:

Loading [MathJax]/jax/output/CommonHTML/fonts/TeX/fontdata.js

<sup>[1]</sup> Standard Errors assume that the covariance matrix of the errors is correctly specified.