Kernel: SageMath 10.3

NAME : SUMIT HELONDE

ROLL NO : 58

EXPERIMENT NO : 3

AIM : TO FIND CORRELATION COEFFICIENT STUDY THE LINEAR REGRESSION MULTIPLE REGRESSION ANALYSIS

COEFFICIENT CORRELATION

Q.1A

In [23]:
```
import numpy as np
```

In [24]:
```
x = [1, 2, 3, 4, 5]
```

In [25]:
```
y = [2, 4, 1, 3, 5]
```

In [26]:
```
correlation_coefficient = np.corrcoef(x,y)[0,1]
```

In [27]:
```
print("correlation coefficient :",correlation_coefficient)
```

Out[27]:
```
correlation coefficient : 0.49999999999999994
```

Q.1b

In [28]:
```
a = [7, 8, 1, 3, 4]
```

In [29]:
```
b = [3, 4, 6, 7, 8]
```

In [30]:
```
correlation_coefficient = np.corrcoef(a,b)[0,1]
```

In [31]:
```
print("correlation coefficient :",correlation_coefficient)
```

Out[31]:
```
correlation coefficient : -0.7030334943258522
```

MULTIPLE REGRESSION

Q.2a

In [32]:
```
import numpy as np
```

In [33]:
```
import statsmodels.api as sm
```

In [34]:
```
x1 = [1, 2, 3, 4, 5]
```

In [35]:
```
x2 = [2, 3, 4, 5, 6]
```

In [36]:
```
x3 = [3, 4, 5, 6, 7]
```

In [37]:
```
y = [2, 4, 1, 3, 5]
```

In [38]:
```
X = np.column_stack((x1, x2, x3))
```

In [39]:
```
X = sm.add_constant(X)
```

In [40]:
```
model = sm.OLS(y, X).fit()
```

In [41]:
```
print(model.summary())
```

Out[41]:
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.250
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     1.000
Date:                Tue, 26 Mar 2024   Prob (F-statistic):              0.391
Time:                        09:54:31   Log-Likelihood:                -8.1084
No. Observations:                   5   AIC:                             20.22
Df Residuals:                       3   BIC:                             19.44
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.3333      0.707      0.471      0.670      -1.917       2.584
x1            -0.1667      0.866     -0.192      0.860      -2.923       2.589
x2             0.1667      0.167      1.000      0.391      -0.364       0.697
x3             0.5000      0.553      0.905      0.432      -1.259       2.259
==============================================================================
Omnibus:                        nan   Durbin-Watson:                   2.533
Prob(Omnibus):                  nan   Jarque-Bera (JB):                0.361
Skew:                        -0.408   Prob(JB):                        0.835
Kurtosis:                     1.967   Cond. No.                     1.25e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.81e-32. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

```
/ext/sage/10.3/local/var/lib/sage/venv-python3.11.1/lib/python3.11/site-
packages/statsmodels/stats/stattools.py:74: ValueWarning: omni_normtest is not valid with less than 8
observations; 5 samples were given.
  warn("omni_normtest is not valid with less than 8 observations; %i "
```

Q.2b

In [47]:
```
ax1 = [21, 42, 43, 51, 88]
```

In [48]:
```
x2 = [42, 43, 52, 88, 73]
```

In [49]:
```
x3 = [43, 52, 88, 73]
```

In [55]:
```
y = [ 3, 5, 7, 2, 2]
```

In [56]:
```
X = np.column_stack((x1, x2, x3))
```

In [57]:
```
X = sm.add_constant(X)
```

In [59]:
```
model = sm.OLS(y, X).fit()
```

In [60]:
```
print(model.summary())
```

Out[60]:
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.133
Model:                            OLS   Adj. R-squared:                 -0.156
Method:                 Least Squares   F-statistic:                    0.4601
Date:                Tue, 26 Mar 2024   Prob (F-statistic):              0.546
Time:                        10:03:12   Log-Likelihood:                -10.049
No. Observations:                   5   AIC:                             24.10
Df Residuals:                       3   BIC:                             23.32
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.9333      1.042      1.855      0.161      -1.384       5.251
x1            -2.1000      1.277     -1.645      0.199      -6.163       1.963
x2            -0.1667      0.246     -0.678      0.546      -0.949       0.615
x3             1.7667      0.815      2.168      0.119      -0.827       4.360
==============================================================================
Omnibus:                        nan   Durbin-Watson:                   2.025
Prob(Omnibus):                  nan   Jarque-Bera (JB):                0.702
```

```
Skew:                          0.835   Prob(JB):                     0.704
Kurtosis:                      2.237   Cond. No.                  1.25e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.81e-32. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

```
/ext/sage/10.3/local/var/lib/sage/venv-python3.11.1/lib/python3.11/site-
packages/statsmodels/stats/stattools.py:74: ValueWarning: omni_normtest is not valid with less than 8
observations; 5 samples were given.
  warn("omni_normtest is not valid with less than 8 observations; %i "
```

LINEAR REGRESSION

Q.3A

In [61]:
```
import numpy as np
```

In [62]:
```
import scipy.stats as stats
```

In [63]:
```
x_data = np.array([1, 2, 3, 4, 5])
```

In [64]:
```
y_data = np.array([2, 3, 5, 6, 8])
```

In [70]:
```
slope, intercept, r_value, p_value, std_err = stats.linregress(x_data, y_data)
```

In [71]:
```
print(f"slope :{slope}")
```

Out[71]:
```
slope :1.5
```

In [72]:
```
print(f"intercept :{intercept}")
```

Out[72]:
```
intercept :0.2999999999999998
```

In [74]:
```
print(f"r_squared :{r_value**2}")
```

Out[74]:
```
r_squared :0.9868421052631577
```

conclusion :