**Task 1 || From the table below, determine:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **X** | **65** | **63** | **67** | **64** | **68** | **62** | **70** | **66** | **68** | **67** | **69** | **71** |
| **Y** | **68** | **66** | **68** | **65** | **69** | **66** | **68** | **65** | **71** | **67** | **68** | **70** |

**Given that, a0 = 35.82, a1 = 0.476**

**(i) The standard error of estimate of Y on X.**

**(ii) Total variation, unexplained variation, and explained variation.**

**(iii) Coefficient of determination, and coefficient of correlation.**

**Solution:**

|  |
| --- |
| **Code:** |
| clc;  clear all;  % Mentioning given data points  X = [65 63 67 64 68 62 70 66 68 67 69 71];  Y = [68 66 68 65 69 66 68 65 71 67 68 70];  %Finding length  n =length(X);  Y1 = mean(Y);  Y\_est = 35.82 + .476\* X ;  S\_yx = sqrt(sum((Y - Y\_est).^2 ) / n) ;  t= Y - Y1;  total = sum(t.^2) ;  unexp= sum((Y - Y\_est).^2 );  exp= sum((Y\_est - Y1).^2 );  %Coefficient of determination  r2 = exp / total ;  %Coefficient of correlation  r = sqrt(exp / total) ;  %Printing Results...  fprintf("Standard error of estimate of Y on X = %g\n\n", S\_yx);  fprintf("Total Variation= %g\n", total);  fprintf("Unexpected variation = %g\n", unexp);  fprintf("Expected variation = %g\n\n", exp);  fprintf("Coefficient of determination = %g\n", r2);  fprintf("Coefficient of correlation = %g\n", r); |
| **Output:** |
| Standard error of estimate of Y on X = 1.28172  Total Variation= 38.9167  Unexpected variation = 19.7136  Expected variation = 19.1942  Coefficient of determination = 0.493214  Coefficient of correlation = 0.702292  >> |
| **Comment:** |
|  |

**Task 2 || From the given table below:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **X** | **1** | **3** | **4** | **6** | **8** | **9** | **11** | **14** |
| **Y** | **1** | **2** | **4** | **4** | **5** | **7** | **8** | **9** |

**(i) The coefficient of linear correlation.**

**(ii) The standard deviation of X and standard deviation of Y.**

**(iii) The variance of X and variance of Y.**

**(iv) The covariance of X and Y.**

**Solution:**

|  |
| --- |
| **Code:** |
| clc;  clear all;  % Mentioning given data points  X = [1 3 4 6 8 9 11 14];  Y = [1 2 4 4 5 7 8 9];  %Finding length  N =length(X);  x = X - mean(X);  y = Y - mean(Y);  %The coefficient of linear correlation  r = (sum(x.\*y) / sqrt(sum(x.^2) \* sum(y.^2)));  %The standard deviation of X and Y  Sx= sqrt(sum(x .^2 )/ N);  Sy= sqrt(sum(y .^2 )/ N);  %The variance of X and Y  Sx2= sum(x .^2 )/ N ;  Sy2= sum(y .^2 )/ N ;  %The covariance of X and Y  Sxy = (sum(x .\*y) / N);  %Printing Results...  fprintf('The coefficient of linear correlation, r =%g\n\n', r);  fprintf("The standard deviation of X, Sx= %g\n", Sx);  fprintf("The standard deviation of Y, Sy =%g\n\n", Sy);  fprintf("The variance of X, Sx^2= %g\n", Sx2);  fprintf("The variance of Y, Sy^2 =%g\n\n", Sy2);  fprintf("The covariance of X and Y, Sxy =%g\n", Sxy); |
| **Output:** |
| The coefficient of linear correlation, r =0.977008  The standard deviation of X, Sx= 4.06202  The standard deviation of Y, Sy =2.64575  The variance of X, Sx^2= 16.5  The variance of Y, Sy^2 =7  The covariance of X and Y, Sxy =10.5  >> |
| **Comment:** |
|  |

**Task 3 || From the table below, find the coefficient of rank correlation.**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Laboratory** | **8** | **3** | **9** | **2** | **7** | **10** | **4** | **6** | **1** | **5** |
| **Lecture** | **9** | **5** | **10** | **1** | **8** | **7** | **3** | **4** | **2** | **6** |

**Solution:**

|  |
| --- |
| **Code:** |
| clc;  clear all;  % Mentioning given data points  lab= [8 3 9 2 7 10 4 6 1 5];  lec= [9 5 10 1 8 7 3 4 2 6];  %Finding length  n= length(lab);  d = lab - lec; %difference of ranks  d\_2 = d.^2; % d^2  D2\_sum = sum(d\_2); %summation of d^2  %Formula implementation to find r\_rank  r\_rank = (1 - ((6\* D2\_sum)/(n\*(n^2) - 1)));  %Printing Results...  fprintf("Co-efficient of rank correlation = %g\n", r\_rank); |
| **Output:** |
| Co-efficient of rank correlation = 0.855856  >> |
| **Comment:** |
|  |

Correlation:

The degree of relationship between variables is called correlation. When the relationship between two variables is exact, then it is called perfect correlation. For example, the circumference CCC and radius rrr of a circle are perfectly related because C=2πr.

Regression:

Regression is a method to predict one variable based on the value of another. For example, prediction of someone’s height based on their parent’s height.

The line that best fits the data is called the regression line. It’s calculated using the least-squares method, which minimizes the distance between the line and all the data points.

Regression Line Formula (for predicting Y based on X):

Y = a0 + a1X

Where, ​a0 is the intercept (the point where the line crosses the Y-axis) and a1 is the slope (how much Y changes when X changes by 1 unit).

Standard Error of Estimate

This measures how far the actual data points are from the regression line, on average. I other words, a measure of scatter about the regression line of Y on X is supplied by the quantity.

Formula:

SY.X  
Where, SY.X ​ is the standard error of estimate, Yest​ is the estimated value of Y from the regression line and N is the number of data points.

Total Variation: How much all the YYY values differ from their average Yˉ\bar{Y}Yˉ.

* + Formula:  
    Total Variation=∑(Y−Yˉ)2\text{Total Variation} = \sum (Y - \bar{Y})^2Total Variation=∑(Y−Yˉ)2
* Explained Variation: The part of the total variation that can be explained by the regression line.
  + Formula:  
    Explained Variation=∑(Yest−Yˉ)2\text{Explained Variation} = \sum (Y\_{\text{est}} - \bar{Y})^2Explained Variation=∑(Yest​−Yˉ)2
* Unexplained Variation: The part of the total variation that cannot be explained by the regression line (the random scatter around the line).
  + Formula:  
    Unexplained Variation=∑(Y−Yest)2\text{Unexplained Variation} = \sum (Y - Y\_{\text{est}})^2Unexplained Variation=∑(Y−Yest​)2

5. Coefficient of Determination (R2R^2R2)

This tells us what percentage of the variation in YYY is explained by XXX.

* Formula:  
  R2=Explained VariationTotal VariationR^2 = \frac{\text{Explained Variation}}{\text{Total Variation}}R2=Total VariationExplained Variation​  
  It ranges from 0 to 1. For example, R2=0.8R^2 = 0.8R2=0.8 means 80% of the variation in YYY is explained by XXX.

6. Coefficient of Correlation (rrr)

This is the square root of R2R^2R2 and measures the strength and direction of the relationship between XXX and YYY.

* Formula:  
  r=±R2r = \pm \sqrt{R^2}r=±R2​  
  It ranges from -1 to +1:
  + r=1r = 1r=1: Perfect positive relationship.
  + r=−1r = -1r=−1: Perfect negative relationship.
  + r=0r = 0r=0: No relationship.

7. Coefficient of Linear Correlation

This formula measures how two variables move together:

* Formula:  
  r=∑(x⋅y)∑x2⋅∑y2r = \frac{\sum (x \cdot y)}{\sqrt{\sum x^2 \cdot \sum y^2}}r=∑x2⋅∑y2​∑(x⋅y)​  
  Where:
  + x=X−Xˉx = X - \bar{X}x=X−Xˉ and y=Y−Yˉy = Y - \bar{Y}y=Y−Yˉ.
  + ∑x2\sum x^2∑x2 and ∑y2\sum y^2∑y2 represent the sum of squared deviations from the means.

8. Standard Deviation and Variance

* Standard Deviation: A measure of how spread out the values are.
  + Formula for XXX:  
    sX=∑(X−Xˉ)2Ns\_X = \sqrt{\frac{\sum (X - \bar{X})^2}{N}}sX​=N∑(X−Xˉ)2​​
  + Formula for YYY:  
    sY=∑(Y−Yˉ)2Ns\_Y = \sqrt{\frac{\sum (Y - \bar{Y})^2}{N}}sY​=N∑(Y−Yˉ)2​​
* Variance: The square of the standard deviation.
  + Formula for XXX:  
    sX2=∑(X−Xˉ)2Ns^2\_X = \frac{\sum (X - \bar{X})^2}{N}sX2​=N∑(X−Xˉ)2​

9. Covariance

This measures how two variables vary together. A positive covariance means that when one variable increases, the other tends to increase too.

* Formula:  
  sXY=∑(x⋅y)Ns\_{XY} = \frac{\sum (x \cdot y)}{N}sXY​=N∑(x⋅y)​

10. Coefficient of Rank Correlation (Spearman’s rrankr\_{\text{rank}}rrank​)

This measures the relationship between two sets of rankings.

* Formula:  
  rrank=1−6∑d2n(n2−1)r\_{\text{rank}} = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}rrank​=1−n(n2−1)6∑d2​  
  Where:
  + ddd is the difference between the ranks of corresponding values.
  + nnn is the number of pairs.