**Practical 10:**

**Objective**: Solving Linear and Multiple Regressions

**Part A: (Linear Regression – Using Microsoft Excel)**

A cloth manufacturer wants to determine the relationship between the thickness of a synthetic fiber and its tensile strength. Researchers took measurements at various pre-selected, known levels of fiber thickness, and the following data was collected.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fiber thickness, X | 40 | 31 | 34 | 44 | 49 | 36 | 41 | 50 | 39 | 45 |
| Tensile strength, Y | 83 | 74 | 72 | 70 | 75 | 73 | 70 | 76 | 79 | 72 |

1. Use Microsoft Excel to complete the following table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| 1 | 40 | 83 | 1600 | 6889 | 3320 |
| 2 | 31 | 74 | 961 | 5476 | 2294 |
| 3 | 34 | 72 | 1156 | 5184 | 2448 |
| 4 | 44 | 70 | 1936 | 4900 | 3080 |
| 5 | 49 | 75 | 2401 | 5625 | 3675 |
| 6 | 36 | 73 | 1296 | 5329 | 2628 |
| 7 | 41 | 70 | 1681 | 4900 | 2870 |
| 8 | 50 | 76 | 2500 | 5776 | 3800 |
| 9 | 39 | 79 | 1521 | 6241 | 3081 |
| 10 | 45 | 72 | 2025 | 5184 | 3240 |
| ∑ | 409 | 744 | 17077 | 55504 | 30436 |

1. Write the normal equations for the above dataset.

Normal equation : [10 409; 409 17077] \* [a0; a1] = [744; 30436]

1. Solve the above normal equations to find the linear regression model.

+MINVERSE(region of [10 409; 409 17077])

|  |  |
| --- | --- |
| 4.894526 | -0.11723 |
| -0.11723 | 0.002866 |
|  |  |

+MMULT(region of inversed matrix, region of [744; 30436])

|  |
| --- |
| 73.64976 |
| 0.018343 |

Y = a0 + a1\*x is

Y = 73.64976 + 0.018343x

1. If the fiber strength thickness was 45, what would be the predicted strength?

Predicted strength = 73.64976 + 0.018343\*45 = 74.47521

1. Compute sum of square of the residues (), standard error of estimate () and correlation coefficient ().

Sr = ∑(yi – (a0+a1\*xi))^2 = 173.2668123

Sy/x = (Sr / n-2)^(1/2) = 4.653853

r = (n\*∑(xi\*yi) - ∑xi \* ∑yi) / ((n\*∑xi^2 – (∑xi)^2)^(1/2) \* (n\*∑yi^2 – (∑yi)^2)^(1/2) = 0.027939

**Part B: (Linear Regression- Using Microsoft Excel’s built-in function)**

1. Fit the following data into a linear regression model.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 3 | 4 | 4 | 2 | 5 | 3 | 4 | 5 | 3 | 2 |
| y | 57 | 78 | 72 | 58 | 89 | 63 | 73 | 84 | 75 | 48 |

Method1.

- +SLOPE(y’s, x’s), +INTERCEPT(y’s, x’s)

- y = 31.53333 + 10.90746\*x

Method2.

- +LINEST(y’s, x’s)

- y = 31.53333 + 10.90746\*x

1. Predict the y values for and .

|  |  |
| --- | --- |
| New X | Predicted y |
| 15 | 195.1048 **≒** 195 |
| 20 | 249.6286 **≒** 250 |
| 30 | 358.6762 **≒** 359 |
| 40 | 467.7238 **≒** 468 |

1. Compute sum of square of the residues (), standard error of estimate () and correlation coefficient ().

* +LINEST(y’s,x’s,true,true), =CORREL(x’s,y’s)

Sr = 255.5048

Sy/x = 5.56138

r = 0.911113

**Part C: (Multiple Regression)**

In a study, 1000 adults aged 18+ (males and females) were recruited to study factors associated with BMI. Data collected are based on the following variables:

* BMI
* Sex (female=1)
* Calorie (calorie intake daily)
* Exercise (minutes of exercise per week)
* Income (monthly salary in dollars $)

1. Fit the dataset in BMI.xlsx to a multiple regression model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Coefficient** | **Std. err** | **t value** | **p value** |
| **Intercept** | 20.68253 | 0.166515 | 124.2084 | 0 |
| **Calorie** | 0.001886 | 4.84E-05 | 38.98907 | 1.2E-202 |
| **Exercise** | -0.0274 | 0.003245 | -8.44285 | 1.08E-16 |
| **Income** | 8.79E-05 | 4.79E-05 | 1.83607 | 0.066645 |

**Model : BMI = 20.68253 + 0.001886 \* calorie – 0.0274 \* exercise**

1. Predict the BMI value when given the calorie=2400 and exercise=21.

**Predicted BMI = 20.68253 + 0.001886\*2400 – 0.0274\*21 = 24.63353**

**Part D: (Programming Practice)**

Write a program to fit dataset into a linear regression model.