**SR University warangal**

**Advance Data Science**

**October 21, 2025**

**Submitted by**: Kassim Sesay

Enrol #: 2503B05101

1. (a) Derive the regression equation=a+bx using the least square method and calculate a (intercept) and b (slope). Also compute the value of **(ΣX, ΣY, ΣXY, ΣX2),**

|  |  |  |  |
| --- | --- | --- | --- |
| Temperature (X) | Power Consumption (Y) | ΣXY | ΣX2 |
| 10 | 300 | 3000.00 | 100.00 |
| 12 | 310 | 3720.00 | 144.00 |
| 14 | 320 | 4480.00 | 196.00 |
| 16 | 330 | 5280.00 | 256.00 |
| 18 | 345 | 6210.00 | 324.00 |
| 20 | 360 | 7200.00 | 400.00 |
| 22 | 370 | 8140.00 | 484.00 |
| 24 | 390 | 9360.00 | 576.00 |
| 26 | 420 | 10920.00 | 676.00 |
| 28 | 450 | 12600.00 | 784.00 |
| ΣX 190.00 | ΣY 3595.00 | ΣXY 70910.00 | ΣX2 3940.00 |

Number of observations, n = 10

ΣX = 190.00

ΣY = 3595.00

ΣXY = 70910.00

ΣX^2 = 3940.00

Mean of X = 19.0000

Mean of Y = 359.5000

Regression equation: Ŷ = 209.515152 + (7.893939) X

Slope (b) formula: b = (n\*ΣXY - ΣX\*ΣY) / (n\*ΣX2 - (ΣX) 2)

Compute numerator: n\*ΣXY - ΣX\*ΣY = 10\*70910.00 - 190.00\*3595.00 = 26050.0000

Compute denominator: n\*ΣX2 - (ΣX) 2 = 10\*3940.00 - (190.00)2 = 3300.0000

Slope (b) = numerator / denominator = 26050.0000 / 3300.0000 = 7.893939

Intercept (a) formula: a = mean(Y) - b \* mean(X)

Intercept (a) = 359.500000 - (7.893939)\*19.000000 = 209.515152

Predicted values (Ŷ) and R²

|  |  |  |  |
| --- | --- | --- | --- |
| X | Y (actual) | Ŷ (predicted) | (Y - Ŷ)2 |
| 10 | 300.00 | 288.4545 | 133.2975 |
| 12 | 310.00 | 304.2424 | 33.1497 |
| 14 | 320.00 | 320.0303 | 0.0009 |
| 16 | 330.00 | 335.8182 | 33.8512 |
| 18 | 345.00 | 351.6061 | 43.6400 |
| 20 | 360.00 | 367.3939 | 54.6703 |
| 22 | 370.00 | 383.1818 | 173.7603 |
| 24 | 390.00 | 398.9697 | 80.4555 |
| 26 | 420.00 | 414.7576 | 27.4830 |
| 28 | 450.00 | 430.5455 | 378.4793 |

Residual sum of squares (SS\_res) = 958.787879

Total sum of squares (SS\_tot) = 21522.500000

R² = 1 - SS\_res / SS\_tot = 0.955452

Regression equation (final)

Estimated regression line: Ŷ = 209.515152 + (7.893939) X

Interpretation: The slope is positive, which means power consumption increases with temperature in this dataset

1. **Statsmodels OLS (for comparison)**

Below is the key output from statsmodels OLS fit:

OLS Regression Results   
==============================================================================  
Dep. Variable: y R-squared: 0.955  
Model: OLS Adj. R-squared: 0.950  
Method: Least Squares F-statistic: 171.6  
Date: Tue, 21 Oct 2025 Prob (F-statistic): 1.10e-06  
Time: 15:56:35 Log-Likelihood: -37.005  
No. Observations: 10 AIC: 78.01  
Df Residuals: 8 BIC: 78.61  
Df Model: 1   
Covariance Type: nonrobust   
==============================================================================  
 coef std err t P>|

This is my short findings:

* + - The fitted linear model has a positive slope, indicating that higher temperatures are associated with higher power consumption.
      * The regression equation is Ŷ = 209.5152 + 7.893939 X.
      * R² = 0.9555, which indicates the proportion of variance in Y explained by X.

This completes the step-by-step least squares derivation and Python verification.