Mariano Méndez-Suárez

DESCRIPTIVE AND INFERENTIAL STATISTICS

PROBLEMS AND SOLUTIONS

LIBROS PROFESIONALES DE EMPRESA

ESIC Editorial

Descriptive and Inferential Statistics

Problems and Solutions

Madrid, 2025

Mariano Méndez-Suárez

Descriptive and Inferential Statistics

Problems and Solutions

ESIC Editorial

April, 2025

Descriptive and Inferential Statistics: Problems and Solutions
Mariano Méndez-Suárez

Avda. de Valdenigriales, s/n
28223 Pozuelo de Alarcón (Madrid)
Tel.: 91 452 41 00
www.esic.edu/editorial
@EsicEditorial

ISBN: 978-84-1192-155-8
Legal Deposit: M-5076-2025

Cover: Zita Moreno Puig
Design: Santiago Díez Escribano
Printed by: Gráficas Dehon

A publication by

ESIC Editorial

Printed in Spain

This book has been printed with ecological ink and sustainable paper.

To Pilar

Contents

7 Regression Analysis

Regression analysis, also known as "supervised machine learning," is one of the primary statistical techniques used to understand relationships between variables, predict future outcomes, and make data-driven decisions. It helps identify trends and patterns, to understand how changes in one variable can affect another. Furthermore, regression allows analysts to predict future trends based on historical data, determine the impact of various factors on an outcome, and assess the strength of relationships.

In practice, regression methodology has a wide range of applications in a variety of business contexts. For example, it can be used to predict sales as a function of advertising spend, which helps companies allocate advertising spend more effectively. It also plays an important role in understanding how customer satisfaction affects repeat purchase rates, providing insight into customer loyalty and retention strategies. In addition, regression is critical for optimizing pricing strategies by analyzing the elasticity of demand, which helps determine the optimal price of products to maximize revenue. Another important application of regression is in the analysis of user engagement metrics to understand their effect on website conversion rates. This analysis is essential for improving the user experience and increasing sales. In addition, regression can fine-tune the targeting of promotional campaigns by predicting customer behavior through demographic data, making marketing efforts more efficient and effective.

7.1. Key Concepts

- **Dependent Variable (Y)**: The variable that you are trying to predict or explain. It is also known as the response or outcome variable..
- **Independent Variable (X)**: The variable that is used to predict the dependent variable. It is also known as the predictor or explanatory variable.
- **Linear Regression**: A type of regression analysis where the relationship between the dependent and independent variables is modeled as a straight line.

7.2. Simple Linear Regression

Equation

The equation for a simple linear regression line is:

$$Y = \alpha + \beta X + \varepsilon$$

- Y is the dependent variable.
- X is the independent variable.
- α is the y-intercept.
- β is the slope of the regression line.
- ε is the error term.

Interpretation

- **Y-Interpretation (**α): The value of Y when $X = 0$. It represents the starting point of the regression line.
- **Slope (**β): The change in Y for a one-unit change in X. It indicates the strength and direction of the relationship between X and Y. In mathematical terms, this can be interpreted as the derivative of Y with respect to X, denoted as $\frac{dY}{dX}$. The slope represents the rate of change of the dependent variable Y for each unit increase in the independent variable X.

Notice. Due to the large number of calculation steps and potential rounding errors in decimals, the calculated results may differ slightly.

7.3. Step-by-step exercises on Regression Calculation

Exercise 1: Predicting Monthly Sales Based on Advertising Spend

A marketing manager wants to predict monthly sales based on advertising spend. The following data is collected:

Advertising Spend (X)	Monthly Sales (Y)
8	20
12	25
18	30
22	32
30	35

Create a model and evaluate it based on R^2 and RMSE.

Solution

- Step 1: Calculate the Means of X and Y

$$\bar{X} = \frac{8 + 12 + 18 + 22 + 30}{5} = 18$$

$$\bar{Y} = \frac{20 + 25 + 30 + 32 + 35}{5} = 28.4$$

- Step 2: Calculate the Covariance of X and Y

$$\text{Cov}(X, Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
8	20	-10	-8.4	84.0
12	25	-6	-3.4	20.4
18	30	0	1.6	0.0
22	32	4	3.6	14.4
30	35	12	6.6	79.2

$$\text{Cov}(X, Y) = \frac{84 + 20.4 + 0 + 14.4 + 79.2}{5} = 39.6$$

- Step 3: Calculate the Variance of X

$$\sigma_X^2 = \frac{\sum(X_i - \bar{X})^2}{n}$$

X	$X - \bar{X}$	$(X - \bar{X})^2$
8	-10	100
12	-6	36
18	0	0
22	4	16
30	12	144

$$\sigma_X^2 = \frac{100 + 36 + 0 + 16 + 144}{5} = 59.2$$

- Step 4: Calculate the Slope and Intercept

$$\beta = \frac{\text{Cov}(X, Y)}{\sigma_X^2} = \frac{39.6}{59.2} = 0.6689$$

$$\alpha = \bar{Y} - \beta \cdot \bar{X} = 28.4 - (0.6689 \times 18) = 16.37$$

- Step 5: Regression Line Equation

$$Y = 16.37 + 0.6689X$$

- Step 6: Calculate the RMSE

$$e_i = Y_i - (\alpha + \beta X_i)$$

X	Y	$\hat{Y}$	Residual (e_i)	Residual Squared (e_i^2)
8	20	21.71	-1.71	2.92
12	25	24.39	0.61	0.37
18	30	28.37	1.63	2.65
22	32	31.12	0.88	0.77
30	35	36.44	-1.44	2.07

$$\text{RMSE} = \sqrt{\frac{2.92 + 0.37 + 2.65 + 0.77 + 2.07}{5}} = 1.58$$

- Step 7: Calculate the Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum\left(Y_i - \hat{Y}_i\right)^2}{\sum(Y_i - \bar{Y})^2}$$

$$R^2 = 1 - \frac{2.92 + 0.37 + 2.65 + 0.77 + 2.07}{(20 - 28.4)^2 + (25 - 28.4)^2 + (30 - 28.4)^2 + (32 - 28.4)^2 + (35 - 28.4)^2}$$

$$R^2 = 1 - \frac{8.78}{(70.56 + 11.56 + 2.56 + 12.96 + 43.56)} = 1 - \frac{8.78}{141.2} = 0.9378$$

- Step 8: Model Evaluation

  The model has an R^2 of 0.9378, meaning that 93.78% of the variability in monthly sales can be explained by advertising spend. The RMSE of 1.58 indicates the average magnitude of the prediction error.

Interpretation of Results

1. **Slope (**β**)**. The slope of 0.6689 suggests that for every unit increase in advertising spend, monthly sales increase by approximately 0.6689 units on average. This indicates a positive relationship between advertising spend and sales.
2. **Intercept (**α**)**. The intercept of 16.37 represents the estimated value of monthly sales when advertising spend is zero.
3. **Coefficient of Determination (**R^2**)**. An R^2 of 0.9378 indicates that approximately 93.78% of the variability in monthly sales can be explained by advertising spend.
4. **Root Mean Square Error (RMSE).** An RMSE of 1.58 suggests that, on average, the sales predictions are 1.58 units away from the actual sales.

Conclusions

The regression analysis indicates a significant positive relationship between advertising spend and monthly sales. Increasing advertising spend is likely to result in increased sales. However, for a more robust model, other factors influencing sales could be included.

Exercise 2: Marketing Budget Optimization

A company wants to optimize its marketing budget to increase product awareness. Data has been collected on marketing spending and the resulting number of leads generated. Using regression analysis, the goal is to determine how changes in marketing spend affect the number of leads generated and establish the optimal budget allocation.

Data

Marketing Spend (X)	Leads (Y)
12	145
15	160
8	120
18	180
14	170
20	190

Solution

- Step 1: Calculate the Means of X and Y

$$\bar{X} = \frac{12 + 15 + 8 + 18 + 14 + 20}{6} = 14.5$$

$$\bar{Y} = \frac{145 + 160 + 120 + 180 + 170 + 190}{6} = 160.83$$

- Step 2: Calculate the Covariance of X and Y

$$\mathrm{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
12	145	-2.5	-15.83	39.58
15	160	0.5	-0.83	-0.42
8	120	-6.5	-40.83	265.42
18	180	3.5	19.17	67.10
14	170	-0.5	9.17	-4.58
20	190	5.5	29.17	160.42

$$\mathrm{Cov}(X, Y) = \frac{39.58 - 0.42 + 265.42 + 67.10 - 4.58 + 160.42}{6} = 87.25$$

- Step 3: Calculate the Variance of X

$$\sigma_X^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

X	$X - \bar{X}$	$(X - \bar{X})^2$
12	-2.5	6.25
15	0.5	0.25
8	-6.5	42.25
18	3.5	12.25
14	-0.5	0.25
20	5.5	30.25

$$\sigma_X^2 = \frac{6.25 + 0.25 + 42.25 + 12.25 + 0.25 + 30.25}{6} = 15.25$$

- Step 4: Calculate the Slope and Intercept

$$\beta = \frac{\text{Cov}(X, Y)}{\sigma_X^2} = \frac{87.25}{15.25} = 5.77$$

$$\alpha = \bar{Y} - \beta \cdot \bar{X} = 160.83 - (5.77 \times 14.5) = 77.24$$

- Step 5: Regression Line Equation

$$Y = 77.24 + 5.77X$$

- Step 6: Calculate the RMSE

$$e_i = Y_i - (\alpha + \beta X_i)$$

X	Y	$\hat{Y}$	Residual (e_i)	Residual Squared (e_i^2)
12	145	146.51	-1.51	2.28
15	160	163.84	-3.84	14.74
8	120	123.37	-3.37	11.37
18	180	181.12	-1.12	1.25
14	170	158.61	11.39	129.74
20	190	192.76	-2.76	7.61

$$\text{RMSE} = \sqrt{\frac{2.28 + 14.74 + 11.37 + 1.25 + 129.74 + 7.61}{6}} = 5.47$$

- Step 7: Calculate the Coefficient of Determination (R^2).

$$R^2 = \frac{\text{Variance Explained by the Model}}{\text{Total Variance of Y}}$$

$$R^2 = 0.94$$

- Step 8: Model Evaluation

  The model has an R^2 of 0.94, indicating that 94% of the variability in leads can be explained by the marketing spend. The RMSE of 5.47 shows that, on average, the predictions are 5.47 units away from actual values, suggesting a high level of prediction accuracy.

Conclusions

The analysis suggests that an increase in marketing spend positively impacts the number of leads generated. The R^2 value indicates that the model explains a significant portion of the variability in leads based on marketing spend. However, other factors such as market conditions, campaign effectiveness, and customer demographics may also play a role and should be considered for optimization.

Exercise 3: Predicting Employee Turnover Based on Job Satisfaction

A human resources manager wants to predict employee turnover based on job satisfaction scores. The following data is collected:

Job Satisfaction (X)	Employee Turnover (%) (Y)
2	25
5	22
7	20
9	15
11	10

Create a model and evaluate it based on R^2 and RMSE.

Solution

- Step 1: Calculate the Means of X and Y

$$\bar{X} = \frac{2 + 5 + 7 + 9 + 11}{5} = 6.8$$

$$\bar{Y} = \frac{25 + 22 + 20 + 15 + 10}{5} = 18.4$$

- Step 2: Calculate the Covariance of X and Y

$$\mathrm{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
2	25	-4.8	6.6	-31.68
5	22	-1.8	3.6	-6.48
7	20	0.2	1.6	0.32
9	15	2.2	-3.4	-7.48
11	10	4.2	-8.4	-35.28

Sum of $(X - \bar{X})(Y - \bar{Y})$:

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = -31.68 - 6.48 + 0.32 - 7.48 - 35.28 = -80.6$$

Covariance:

$$\text{Cov}(X, Y) = \frac{-80.6}{5} = -16.12$$

- Step 3: Calculate the Variance of X

$$\text{Var}(X) = \frac{\sum(X_i - \bar{X})^2}{n}$$

Sum of $(X - \bar{X})^2$:

$$\sum(X_i - \bar{X})^2 = 23.04 + 3.24 + 0.04 + 4.84 + 17.64 = 48.8$$

Variance:

$$\text{Var}(X) = \frac{48.8}{5} = 9.76$$

- Step 4: Calculate the Slope (β)

$$\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{-16.12}{9.76} \approx -1.65$$

- Step 5: Calculate the Intercept (α).

$$\alpha = \bar{Y} - \beta\bar{X}$$

$$\alpha = 18.4 - (-1.65) \times 6.8 \approx 29.62$$

- Step 6: Form the Regression Equation

$$Y = 29.62 - 1.65X$$

- *Step 7: Calculate R^2.*

First, calculate the correlation coefficient (ρ) using the formula:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Where σ_X and σ_Y are the standard deviations of X and Y.

Calculate σ_X:

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{9.76} \approx 3.12$$

Calculate the variance and standard deviation of Y:

$$\text{Var}(Y) = \frac{\sum(Y_i - \bar{Y})^2}{n}$$

Sum of $(Y - \bar{Y})^2$:

$$\sum(Y_i - \bar{Y})^2 = 42.56 + 12.96 + 2.56 + 11.56 + 70.56 = 140.2$$

Variance:

$$\text{Var}(Y) = \frac{140.2}{5} = 28.04$$

Standard deviation:

$$\sigma_Y = \sqrt{28.04} \approx 5.30$$

Calculate the correlation coefficient:

$$\rho = \frac{-16.12}{3.12 \times 5.30} \approx -0.98$$

Calculate R^2:

$$R^2 = \rho^2 = (-0.98)^2 = 0.96$$

- Step 8: Calculate RMSE

1. Calculate the residuals (errors):

$$e_i = Y_i - \hat{Y}_t$$

X	Y	$\hat{Y}_i$	Residual (e_i)	e_i^2
2	25	29.62 - 1.65(2)	25 - 26.32	1.74
5	22	29.62 - 1.65(5)	22 - 21.37	0.40
7	20	29.62 - 1.65(7)	20 - 18.07	3.72
9	15	29.62 - 1.65(9)	15 - 14.77	0.05
11	10	29.62 - 1.65(11)	10 - 11.47	2.16

2. Sum the squared residuals:

$$\sum e_i^2 = 1.74 + 0.40 + 3.72 + 0.05 + 2.16 = 8.07$$

3. Calculate the Mean Squared Error (MSE):

$$MSE = \frac{\sum e_i^2}{n} = \frac{8.07}{5} = 1.61$$

4. Calculate the Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{1.61} \approx 1.27$$

Summary of Results

- Intercept (α): 29.62
- Slope (β): -1.65
- R^2: 0.96
- RMSE: 1.27

Interpretation

The regression equation $Y = 29.62 - 1.65X$ suggests that for every additional unit of job satisfaction, employee turnover decreases by approximately 1.65%. The model explains 96% of the variability in employee turnover, with an average prediction error of 1.27%.

Exercise 4: Predicting Product Sales Based on Customer Ratings

A product manager wants to predict product sales based on customer ratings. The following data is collected.

Customer Rating (X)	Product Sales (Y)
1	10
3	15
4	20
6	25
8	28

Create a model and evaluate it based on R^2 and RMSE.

Solution

- Step 1: Calculate the Means of X and Y

$$\bar{X} = \frac{1 + 3 + 4 + 6 + 8}{5} = 4.4$$

$$\bar{Y} = \frac{10 + 15 + 20 + 25 + 28}{5} = 19.6$$

- Step 2: Calculate the Covariance of X and Y

$$\mathrm{Cov}(X, Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
1	10	-3.4	-9.6	32.64
3	15	-1.4	-4.6	6.44
4	20	-0.4	0.4	-0.16
6	25	1.6	5.4	8.64
8	28	3.6	8.4	30.24

Sum of $(X - \bar{X})(Y - \bar{Y})$:

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 32.64 + 6.44 - 0.16 + 8.64 + 30.24 = 77.8$$

Covariance:

$$\mathrm{Cov}(X, Y) = \frac{77.8}{5} = 15.56$$

- Step 3: Calculate the Variance of X

$$\text{Var}(X) = \frac{\sum(X_i - \bar{X})^2}{n}$$

Sum of $(X - \bar{X})^2$:

$$\sum(X_i - \bar{X})^2 = 11.56 + 1.96 + 0.16 + 2.56 + 12.96 = 29.2$$

Variance:

$$\text{Var}(X) = \frac{29.2}{5} = 5.84$$

- Step 4: Calculate the Slope (β)

$$\beta = \frac{\text{Cov}(X,Y)}{\text{Var}(X)} = \frac{15.56}{5.84} \approx 2.66$$

- Step 5: Calculate the Intercept (α)

$$\alpha = \bar{Y} - \beta\bar{X}$$

$$\alpha = 19.6 - 2.66 \times 4.4 \approx 7.86$$

- Step 6: Form the Regression Equation

$$Y = 7.86 + 2.66X$$

- Step 7: Calculate R^2. First, calculate the correlation coefficient (ρ) using the formula:

$$\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Where σ_X and σ_Y are the standard deviations of X and Y. Calculate σ_X:

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{5.84} \approx 2.42$$

Calculate the variance and standard deviation of Y:

$$\text{Var}(Y) = \frac{\sum(Y_i - \bar{Y})^2}{n}$$

Sum of $(Y - \bar{Y})^2$:

$$\sum(Y_i - \bar{Y})^2 = 92.16 + 21.16 + 0.16 + 29.16 + 69.16 = 211.8$$

Variance:

$$\text{Var}(Y) = \frac{211.8}{5} = 42.36$$

Standard deviation:

$$\sigma_Y = \sqrt{42.36} \approx 6.51$$

Calculate the correlation coefficient:

$$\rho = \frac{15.56}{2.42 \times 6.51} \approx 0.98$$

Calcular R^2:

$$R^2 = \rho^2 = 0.98^2 = 0.96$$

- Step 8: Calculate RMSE

1. Calculate the residuals (errors):

$$e_i = Y_i - \hat{Y}_t$$

X	Y	$\hat{Y}_i$	Residual (e_i)	e_i^2
1	10	7.86 + 2.66(1)	10 - 10.52	0.27
3	15	7.86 + 2.66(3)	15 - 15.84	0.71
4	20	7.86 + 2.66(4)	20 - 18.5	2.25
6	25	7.86 + 2.66(6)	25 - 23.82	1.39
8	28	7.86 + 2.66(8)	28 - 29.14	1.30

2. Sum the squared residuals:

$$\sum e_i^2 = 0.27 + 0.71 + 2.25 + 1.39 + 1.30 = 5.92$$

3. Calculate the Mean Squared Error (MSE):

$$MSE = \frac{\sum e_i^2}{n} = \frac{5.92}{5} = 1.18$$

4. Calculate the Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{1.18} \approx 1.09$$

Summary of Results

- Intercept (α): 7.86.
- Slope (β): 2.66.
- R^2: 0.96.
- RMSE: 1.09.

Interpretation

The regression equation $Y = 7.86 + 2.66X$ suggests additional unit increase in customer rating, product sales increase by approximately 2.66 units. The model explains 96% of the variability in product sales, with an average prediction error of 1.09 units.

Exercise 5: Predicting Revenue Based on Number of Products Sold

A business analyst wants to predict revenue based on the number of products sold. The following data is collected:

Number of Products Sold (X)	Revenue (Y)
5	50
12	70
19	85
25	90
30	95

Create a model and evaluate it based on R^2 and RMSE.

Solution

- Step 1: Calculate the Means of X and Y

$$\bar{X} = \frac{5 + 12 + 19 + 25 + 30}{5} = 18.2$$

$$\bar{Y} = \frac{50 + 70 + 85 + 90 + 95}{5} = 78$$

- Step 2: Calculate the Covariance of X and Y

$$\text{Cov}(X, Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
5	50	-13.2	-28	369.6
12	70	-6.2	-8	49.6
19	85	0.8	7	5.6
25	90	6.8	12	81.6
30	95	11.8	17	200.6

Sum of $(X - \bar{X})(Y - \bar{Y})$:

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 369.6 + 49.6 + 5.6 + 81.6 + 200.6 = 707$$

Covariance:

$$\text{Cov}(X, Y) = \frac{707}{5} = 141.4$$

- Step 3: Calculate the Variance of X.

$$\text{Var}(X) = \frac{\sum(X_i - \bar{X})^2}{n}$$

Sum of $(X - \bar{X})^2$:

$$\sum(X_i - \bar{X})^2 = 174.24 + 38.44 + 0.64 + 46.24 + 139.24 = 398.8$$

Variance:

$$\text{Var}(X) = \frac{398.8}{5} = 79.76$$

- Step 4: Calculate the Slope (β).

$$\beta = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{141.4}{79.76} \approx 1.77$$

- Step 5: Calculate the Intercept (α).

$$\alpha = \bar{Y} - \beta \bar{X}$$

$$\alpha = 78 - 1.77 \times 18.2 \approx 46.76$$

- Step 6: Form the Regression Equation

$$Y = 46.76 + 1.77X$$

- Step 7: Calculate R^2.

  First, calculate the correlation coefficient (ρ) using the formula:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

  Where σ_X and σ_Y are the standard deviations of X and Y.

  Calculate σ_X:

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{79.76} \approx 8.93$$

  Calculate the variance and standard deviation of Y:

$$\text{Var}(Y) = \frac{\sum (Y_i - \bar{Y})^2}{n}$$

  Sum of $(Y - \bar{Y})^2$:

$$\sum (Y_i - \bar{Y})^2 = 784 + 64 + 49 + 144 + 289 = 1\,330$$

  Variance:

$$\text{Var}(Y) = \frac{1\,330}{5} = 266$$

  Standard deviation:

$$\sigma_Y = \sqrt{266} \approx 16.31$$

  Calculate the correlation coefficient:

$$\rho = \frac{141.4}{8.93 \times 16.31} \approx 0.98$$

  Calculate R^2:

$$R^2 = \rho^2 = 0.98^2 = 0.96$$

- Step 8: Calculate RMSE

1. Calculate the residuals (errors):

$$e_i = Y_i - \hat{Y}_t$$

X	Y	$\hat{Y}_i$	Residual (e_i)	e_i^2
5	50	46.76 + 1.77(5)	50 - 55.61	31.47
12	70	46.76 + 1.77(12)	70 - 68.96	1.08
19	85	46.76 + 1.77(19)	85 - 80.39	21.24
25	90	46.76 + 1.77(25)	90 - 91.01	1.02
30	95	46.76 + 1.77(30)	95 - 99.86	23.62

2. Sum the squared residuals:

$$\sum e_i^2 = 31.47 + 1.08 + 21.24 + 1.02 + 23.62 = 78.43$$

3. Calculate the Mean Squared Error (MSE):

$$MSE = \frac{\sum e_i^2}{n} = \frac{78.43}{5} = 15.69$$

4. Calculate the Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{MSE} = \sqrt{15.69} \approx 3.96$$

Summary of Results

- Intercept (α): 46.76.
- Slope (β): 1.77.
- R^2: 0.96.
- RMSE: 3.96.

Interpretation

The regression equation $Y = 46.76 + 1.77X$ suggests that for every additional product sold, revenue increases by approximately \$1.77. The model explains 96% of the variability in revenue, with an average prediction error of \$3.96.

Exercise 6: Predicting Conversion Rates Based on Email Campaign Quality

A digital marketing manager wants to predict conversion rates based on email campaign quality scores. The following data is collected:

Email Campaign Quality (X)	Conversion Rate (Y)
3	5
6	10
9	12
12	18
15	20