

Mariano Méndez-Suárez

# DESCRIPTIVE AND INFERENTIAL STATISTICS

PROBLEMS  
AND SOLUTIONS



# **Descriptive and Inferential Statistics**

Problems and Solutions

Madrid, 2025

Mariano Méndez-Suárez

# **Descriptive and Inferential Statistics**

Problems and Solutions

April, 2025

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

All rights reserved.

Any form of reproduction, distribution, public communication or transformation of this work may only be carried out with the authorization of its owners, with the exceptions provided for by law.

Please contact CEDRO (Spanish Centre for Reprographic Rights)  
if you need to photocopy or scan any part of this work ([www.cedro.org](http://www.cedro.org)).

© 2025, ESIC Editorial  
Avda. de Valdenigrales, s/n  
28223 Pozuelo de Alarcón (Madrid)  
Tel.: 91 452 41 00  
[www.esic.edu/editorial](http://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



Printed in Spain

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

To Pilar

# Contents

Chapter 1. <b>Introduction to Descriptive Statistics</b> .....	19
1.1. Key Concepts and Techniques for Frequency Tables .....	19
Frequency Table Overview .....	19
Key Calculations .....	20
Example Step-by-Step Frequency Table Calculation .....	20
Step-by-Step Solution .....	22
1.2. Exercises on Descriptive Statistics .....	24
Exercise 1: Daily Visits to a Website.....	24
Exercise 2: Number of Products Sold per Day (Marketing) .....	26
Exercise 3: Customer Reviews per Day (Marketing) .....	27
Exercise 4: Daily Transactions in a Business .....	28
Exercise 5: Units Produced per Day (Business) .....	30
1.3. Proposed Exercises .....	31
Exercise 1: Daily Visits to an E-commerce Website.....	31
Exercise 2: Number of Products Sold per Week (Marketing) .....	32
Exercise 3: Customer Support Tickets per Day (Digital Business) ....	32
Exercise 4: Monthly Subscriptions to a Digital Service .....	33
Exercise 5: Daily Transactions in an Online Store.....	33
Chapter 2. <b>Measures of Central Tendency</b> .....	35
2.1. Central tendency measures .....	35
2.2. Exercises on Mean, Median, Mode, and Percentiles .....	36
Exercise 1: Customer Daily Screen Time on a news webpage.....	36
Exercise 2: Customer Daily Screen Time on a sports App .....	37

Exercise 3: Customer Daily Screen Time on an App .....	38
Exercise 4: Number of Orders Processed per Day in an Online Store .....	39
Exercise 5: Daily User Sign-ups on a Digital Platform.....	41
2.3. Proposed Exercises on Mean, Median, Mode, and Percentiles.....	42
Exercise 1: Daily Visits to a Marketing Blog.....	42
Exercise 2: Weekly Sales of Products in an Online Store .....	43
Exercise 3: Monthly Subscriptions to a Digital Service .....	43
Exercise 4: Daily Revenue of an E-commerce Site (Digital Business) ..	44
Exercise 5: Weekly Active Users on a Social Media Platform (Digital Business) .....	44
2.4. Measures of Shape: Skewness and Kurtosis.....	45
Skewness .....	45
Kurtosis .....	46
Examples .....	47
Chapter 3. <b>Concentration Measures: Lorenz Curve and Gini Index</b> .....	49
3.1. Lorenz Curve .....	49
Key Points .....	50
3.2. Gini Index .....	50
Applications .....	50
Example 1: Income Distribution among a Population .....	50
Example 2: Daily Sales in a Retail Store.....	53
Example 3: Monthly Marketing Campaign Expenses.....	55
Example 4: Weekly Product Sales in an E-commerce Store .....	57
Example 5: Monthly Revenue from Digital Services .....	60
Example 6: Quarterly Profits of a Marketing Firm .....	62
Example 7: Yearly Sales Revenue of a Digital Product.....	64
Chapter 4. <b>Measures of Dispersion: Variance and Standard Deviation</b> ...	67
4.1. Calculating Variance and Standard Deviation .....	67
4.2. Variance Calculation when each Data Point has a Frequency of 1....	68
Exercise 1: Daily Revenue from a Marketing Campaign.....	68
Exercise 2: Weekly Product Sales in a Business.....	70
Exercise 3: Monthly Subscriptions to a Digital Service .....	71
Exercise 4: Daily Active Users on a Social Media Platform .....	72
Exercise 5: Weekly Revenue from an E-commerce Website .....	73
4.3. Exercises on Measures of Dispersion when each Data Point has a Frequency greater than 1 .....	75

Exercise 1: Analysis of Sales for a Digital Product.....	75
Exercise 2: Click-Through Rate (CTR) in Advertising Campaigns ....	76
Exercise 3: Social Media Engagement Rates .....	77
Exercise 4: Monthly Revenue from Subscriptions .....	78
Exercise 5: Customer Retention Rates .....	79
4.4. Proposed Exercises on Measures of Dispersion with absolute frequency equal to 1 .....	80
Exercise 1: Monthly Sales of a Product .....	80
Exercise 2: Daily Visitors to a Website .....	80
Exercise 3: Weekly Earnings from a Subscription Service .....	81
Exercise 4: Monthly Active Users on an App .....	81
Exercise 5: Quarterly Revenue of a Business .....	82
4.5. Proposed Exercises with absolute frequency greater than 1 .....	82
Exercise 1: Conversion Rate of a Campaign .....	82
Exercise 2: Customer Ratings for a New Product .....	83
Exercise 3: Ad Impressions .....	83
Exercise 4: Page Views .....	83
Exercise 5: Customer Purchase Frequency .....	84
Chapter 5. <b>Index Numbers</b> .....	85
5.1. Index numbers at a glance .....	85
5.2. Simple Indexes .....	85
5.3. Aggregate and Weighted Indexes .....	86
Aggregate Indexes .....	86
Weighted Indexes .....	86
5.4. Laspeyres Index $L_0^t$ .....	86
Definition.....	86
Practical use.....	87
Example Calculation.....	87
5.5. Paasche Index $P_0^t$ .....	87
Definition.....	87
Practical use.....	88
Example Calculation.....	88
5.6. Comparison of Laspeyres and Paasche Indexes .....	88
5.7. How the Consumer Price Index (CPI) is Calculated in Spain .....	89
Steps for Calculating CPI in Spain .....	89
Example of CPI Calculation .....	90
5.8. How the NASDAQ Index is Calculated .....	90



Type of Index Used.....	90
Steps for Calculating the NASDAQ Index .....	90
What is the Index Divisor? .....	90
Example Calculation.....	91
5.9. Exercises on Laspeyres and Paasche Indexes .....	91
Exercise 1: Consumer Goods Basket .....	91
Exercise 2: Stock Market Basket .....	92
Exercise 3: Energy Basket .....	94
Exercise 4: Technology Stock Basket.....	95
Exercise 5: Agricultural Goods Basket .....	96
Summary.....	97
5.10. Proposed Exercises for Laspeyres and Paasche calculation .....	97
Proposed Exercise 1: Technology Gadgets Basket .....	97
Proposed Exercise 2: Food Items Basket .....	98
Proposed Exercise 3: Energy Products Basket .....	98
Proposed Exercise 4: Stock Portfolio Basket.....	99
Proposed Exercise 5: Automobile Basket .....	99
<b>Chapter 6. Two-Dimensional Statistical Analysis .....</b>	<b>101</b>
6.1. Covariance .....	101
Definition.....	101
Formula .....	101
Calculation Steps .....	102
Example.....	102
6.2. Correlation .....	103
Definition.....	103
Formula .....	103
Calculation Steps .....	103
Example.....	103
Interpretation of Correlation Values.....	104
Examples of Correlation Values .....	104
Example.....	105
6.3. Contingency Table .....	106
Example.....	106
Key Concepts.....	106
Application in Business .....	106
6.4. Exercises .....	106
Exercise 1: Customer Satisfaction vs. Purchase Frequency .....	106
Exercise 2: Ad Clicks vs. Device Type.....	107

Exercise 3: Marketing Channel vs. Lead Generation .....	108
Exercise 4: Customer Age vs. Preferred Product Category .....	108
Exercise 5: Email Campaign Type vs. Click-Through Rate .....	109
<b>Chapter 7. Regression Analysis</b> .....	111
7.1. Key Concepts .....	111
7.2. Simple Linear Regression .....	112
Equation .....	112
Interpretation .....	112
7.3. Step-by-step exercises on Regression Calculation .....	112
Exercise 1: Predicting Monthly Sales Based on Advertising Spend ...	112
Exercise 2: Marketing Budget Optimization.....	115
Exercise 3: Predicting Employee Turnover Based on Job Satisfaction	117
Exercise 4: Predicting Product Sales Based on Customer Ratings ....	120
Exercise 5: Predicting Revenue Based on Number of Products Sold ..	122
Exercise 6: Predicting Conversion Rates Based on Email Campaign Quality .....	125
Exercise 7: Predicting Website Traffic Based on Marketing Spend....	128
Exercise 8: Predicting Customer Retention Based on Service Quality	131
Exercise 9: Predicting Customer Churn Based on Call Duration .....	134
Exercise 10: Predicting Employee Performance Based on Training Hours .....	137
7.4. Proposed Exercises on Regression .....	139
Exercise 1: Predicting Monthly Sales Based on Advertising Spend....	139
Exercise 2: Predicting Customer Satisfaction Based on Service Time	140
Exercise 3: Estimating Product Demand Based on Price .....	140
Exercise 4: Analyzing Employee Performance Based on Training Hours .....	140
Exercise 5: Forecasting Revenue Based on Social Media Mentions...	140
<b>Chapter 8. Probability</b> .....	141
8.1. Basic Concepts .....	141
8.2. Conditional Probability .....	141
Definition.....	142
Example.....	142
Interpretation .....	142
8.3. Bayes' Theorem.....	143
Definition.....	143
Example.....	143
Interpretation .....	143
8.4. Applications of Probability and Bayes' Theorem .....	144

Customer Segmentation Using Probability.....	144
Application .....	144
Using Baye's Theorem.....	144
Interpretation .....	144
8.5. Predicting Customer Churn with Conditional Probability.....	144
Context.....	144
Application .....	145
Using Baye's Theorem.....	145
Interpretation .....	145
8.6. Optimizing Marketing Campaigns with Multi- Channel Attribution ..	145
Context.....	145
Application .....	145
Solution .....	146
Interpretation .....	146
8.7. Solved Exercises.....	146
Exercise 1: Marketing Campaign Effectiveness.....	146
Exercise 2: Customer Referral Program .....	147
Exercise 3: Product Recall in Manufacturing .....	147
Exercise 4: Email Campaign Success Rate .....	148
Exercise 5: Customer Satisfaction Survey.....	148
8.8. Proposed Exercises .....	149
Exercise 1: Customer Referral Program .....	149
Exercise 2: Defective Product in Manufacturing.....	149
Exercise 3: Ad Click Conversion Rate .....	150
Exercise 4: Machine Learning Model Success .....	151
Exercise 5: Software Bug Detection .....	151
Chapter 9. <b>Random Variables</b> .....	153
9.1. Discrete Variable .....	153
Binomial Distribution.....	153
Applications in Business and Marketing.....	153
Example: Overbooking in Airline Reservations .....	154
Conclusion .....	155
Poisson Distribution .....	155
Applications in Digital Business and Marketing.....	156
9.2. Binomial Distribution Exercises .....	156
Exercise 1: Email Marketing Campaign.....	156
Exercise 2: Online Ad Click-Through .....	156
Exercise 3: Subscription Service.....	157
Exercise 4: Product Returns .....	157

Exercise 5: Customer Feedback .....	157
9.3. Poisson Distribution Exercises.....	158
Exercise 1: Website Traffic .....	158
Exercise 2: Email Sign-Ups.....	158
Exercise 3: Customer Complaints.....	158
Exercise 4: App Downloads.....	159
Exercise 5: Social Media Mentions .....	159
9.4. Binomial Distribution Proposed Exercises.....	159
Exercise 1: Website Conversion Rate .....	159
Exercise 2: Product Returns .....	160
Exercise 3: Customer Sign-Ups.....	160
Exercise 4: Click-Through Rate .....	160
Exercise 5: Quality Control.....	161
9.5. Poisson Distribution Proposed Exercises .....	161
Exercise 1: Website Visits .....	161
Exercise 2: Customer Service Calls .....	161
Exercise 3: App Installs.....	162
Exercise 4: Online Orders.....	162
Exercise 5: Software Bug Reports.....	162
Chapter 10. <b>Normal Distribution</b> .....	163
10.1. Key Characteristics.....	163
Formula of the Normal Distribution .....	163
10.2. Practical Applications.....	163
10.3. The Standard Normal Distribution .....	164
10.4. Standard normal distribution characteristics .....	164
10.5. Standardizing Any Normal Distribution .....	165
Example 1: Quality Control in Manufacturing.....	165
Example 2: Employee Performance Evaluation .....	165
10.6. Interpreting Z-Scores.....	166
10.7. Using the Standard Normal Distribution Table (Two-Tailed) .....	166
Example 1: Probability of a Value Less Than a Given Value .....	166
Example 2: Probability Between Two Values .....	167
Example 3: Probability Greater Than a Given Value.....	167
10.8. Exercises on Normal Distribution.....	167
Exercise 1: Customer Satisfaction Scores .....	167
Exercise 2: Website Load Times .....	168
Exercise 3: Email Open Rates.....	169
Exercise 4: Product Defect Rates .....	170

Exercise 5: Digital Ad Clicks .....	170
Exercise 6: Sales Performance.....	171
Exercise 7: App Usage Time.....	172
Exercise 8: Return on Investment (ROI).....	172
10.9. Proposed Exercises on Normal Distribution with Solution .....	173
Exercise 1: Email Campaign Click-Through Rate .....	173
Exercise 2: Customer Retention .....	174
Exercise 3: Social Media Engagement .....	174
Exercise 4: Brand Awareness Survey .....	174
Exercise 5: Advertising Recall.....	175
Exercise 6: Product Return Rate.....	175
Exercise 7: Inventory Turnover .....	175
Exercise 8: Quarterly Sales.....	176
Exercise 9: Employee Productivity .....	176
Exercise 10: Manufacturing Defect Rate .....	176
Exercise 11: Website Bounce Rate .....	177
Exercise 12: App Download Rate .....	177
Exercise 13: Server Downtime .....	177
Exercise 14: Online Transaction Volume .....	178
Exercise 15: Conversion Rate .....	178
Chapter 11. <b>Sampling</b> .....	179
11.1. Central Limit Theorem (CLT) .....	179
Key Points of the Central Limit Theorem .....	179
11.2. Sampling the Mean .....	180
Properties of the Sample Mean.....	180
Practical Example .....	180
11.3. Sampling Proportions .....	180
Properties of the Sample Proportion .....	181
Practical Example .....	181
11.4. Confidence Intervals.....	181
Understanding confidence intervals .....	181
Key Concepts.....	182
When Are Confidence Intervals Used in Real Life? .....	182
Example of Confidence Intervals for the Mean. Estimating Average Monthly Sales .....	183
11.5. Confidence intervals for proportions.....	183
Estimating Customer Satisfaction .....	184
11.6. Proposed Exercises on estimation of confidence intervals with mean sampling .....	184

Exercise 1: Estimating Average Delivery Time.....	184
Exercise 2: Estimating Average Marketing Spend per Campaign....	185
Exercise 3: Estimating Average Page Load Time.....	186
Exercise 4: Estimating Average Customer Lifetime Value (CLV) .....	187
Exercise 5: Estimating Average Monthly web Visits .....	188
Exercise 6: Estimating Average Click-Through Rate .....	188
Exercise 7: Estimating the Average Weight of Products .....	189
Exercise 8: Estimating the Average Time Spent on a Website .....	190
Exercise 9: Estimating the Average Monthly Rent .....	191
Exercise 10: Estimating the Average Daily Sales .....	191
Exercise 11: Estimating the Average Commute Time .....	192
Exercise 12: Estimating the Average Lifespan of Batteries.....	193
11.7. Proposed Exercises on estimation of confidence intervals with proportion sampling .....	194
Exercise 1: Estimating Proportion of Satisfied Customers.....	194
Exercise 2: Estimating Proportion of Clicked Ads .....	194
Exercise 3: Estimating Proportion of Repeat Purchasers .....	195
Exercise 4: Estimating Proportion of Defective Products .....	196
Exercise 5: Estimating the Proportion of Returning Customers .....	197
Exercise 6: Estimating the Proportion of Successful Campaigns ....	198
Exercise 7: Estimating the Proportion of High-Engagement Users...	198
Exercise 8: Estimating the Proportion of Products Returned .....	199
Exercise 9: Estimating the Proportion of Positive Reviews .....	200
Exercise 10: Estimating the Proportion of Email Open Rates.....	201
Exercise 11: Estimating the Proportion of Mobile App Conversions...	202
Exercise 12: Estimating the Proportion of Subscribers to a Newsletter	202
Chapter 12. <b>Hypothesis Testing</b> .....	205
12.1. P-Value for Hypothesis Testing .....	206
How to Interpret the p-value? .....	207
12.2. Hypothesis Testing: Examples for Three Scenarios .....	208
Two-Tailed Test Example: Evaluating a New Marketing Strategy ....	208
One-Tailed Test (Upper Tail) Example: Testing a New Feature in a Digital Product .....	209
One-Tailed Test (Lower Tail) Example: Evaluating Cost Reduction in a Business Process .....	210
12.3. Exercises on Hypothesis Testing for the Mean.....	211
Exercise 1: Average Monthly Sales .....	211
Exercise 2: Average Conversion Rate Improvement .....	211
Exercise 3: Average Time on App .....	212
Exercise 4: Reducing Manufacturing Costs.....	213
Exercise 5: Testing Average Customer Satisfaction Score .....	214

Exercise 6: Average Product Price .....	215
Exercise 7: Average Website Load Time .....	215
Exercise 8: Average Daily Sales .....	216
Exercise 9: Average Customer Satisfaction Rating.....	217
Exercise 10: Average Marketing Spend Per Campaign.....	218
12.4. Exercises on Hypothesis Testing for Proportions .....	218
Exercise 1: Proportion of Satisfied Customers .....	218
Exercise 2: Proportion of Returning Customers .....	219
Exercise 3: Proportion of Clicked Ads .....	220
Exercise 4: Proportion of Email Opens .....	221
Exercise 5: Proportion of Positive Reviews .....	222
Exercise 6: Proportion of Successful Campaigns .....	222
Exercise 7: Proportion of Customer Complaints .....	223
Exercise 8: Proportion of Products Returned .....	224
Exercise 9: Proportion of Engaged Users.....	225
Exercise 10: Proportion of Defective Products .....	225
12.5. Proposed Exercises on Hypothesis Testing .....	226
Exercise 1: Average Delivery Time .....	226
Exercise 2: Average Time Spent on Platform .....	226
Exercise 3: Average Number of Items Sold Per Day .....	227
Exercise 4: Average Production Time.....	227
Exercise 5: Email Open Rate.....	227
Exercise 6: Customer Satisfaction .....	228
Exercise 7: User Engagement with New Feature.....	228
Exercise 8: Returned Products .....	228
Exercise 9: Bug Fix Time .....	229
Exercise 10: Consumer preference .....	229
Appendix. <b>Standard Normal Distribution Table</b> .....	230

7.1. Key Concepts. | 7.2. Simple Linear Regression. | 7.3. Step-by-step exercises on Regression Calculation. | 7.4. Proposed Exercises on Regression.

**R**egression 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.
- $\alpha$  is the y-intercept.
- $\beta$  is the slope of the regression line.
- $\varepsilon$  is the error term.

### Interpretation

- **Y-Interpretation ( $\alpha$ ):** The value of  $Y$  when  $X = 0$ . It represents the starting point of the regression line.
- **Slope ( $\beta$ ):** 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 ( $R^2$ )

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

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

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

- 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 ( $\beta$ ).** 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 ( $\alpha$ ).** 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

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

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
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

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

- 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

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

X	Y	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
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 ( $\beta$ )

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

- Step 5: Calculate the Intercept ( $\alpha$ ).

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

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

- Step 6: Form the Regression Equation

$$Y = 29.62 - 1.65X$$

- Step 7: Calculate  $R^2$ .

First, calculate the correlation coefficient ( $\rho$ ) using the formula:

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

Where  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of X and Y.

Calculate  $\sigma_X$ :

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

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):

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

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 ( $\alpha$ ): 29.62
- Slope ( $\beta$ ): -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

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

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

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

- 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 ( $\beta$ )

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

- Step 5: Calculate the Intercept ( $\alpha$ )

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

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

- Step 6: Form the Regression Equation

$$Y = 7.86 + 2.66X$$

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

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

Where  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of  $X$  and  $Y$ . Calculate  $\sigma_X$ :

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

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):

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

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 ( $\alpha$ ): 7.86.
- Slope ( $\beta$ ): 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 ( $\beta$ ).

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

- Step 5: Calculate the Intercept ( $\alpha$ ).

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

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

- Step 6: Form the Regression Equation

$$Y = 46.76 + 1.77X$$

- Step 7: Calculate  $R^2$ .

First, calculate the correlation coefficient ( $\rho$ ) using the formula:

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

Where  $\sigma_X$  and  $\sigma_Y$  are the standard deviations of  $X$  and  $Y$ .

Calculate  $\sigma_X$ :

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

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):

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

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 ( $\alpha$ ): 46.76.
- Slope ( $\beta$ ): 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