Practice Exercises 4

Andres Gregori

May 24, 2023

Disclosure

1 Z-tests, T-tests, and Linear Regression

1. A researcher wants to determine if there is a significant difference in the mean scores of two groups, Group A and Group B. The researcher collects the following data:

   Group A: 9, 7, 6, 10, 12 Group B: 8, 5, 11, 14, 10

   Perform a test to determine if there is a significant difference between the means of Group A and Group B. Use a significance level of 0.05.

2. A company is developing two different versions of a website to see which one leads to higher user engagement. They randomly select 20 users and divide them into two groups: Group X and Group Y. After one month, they measure the number of pages visited by each user. The data is as follows:

   Group X: 10, 8, 12, 9, 7, 11, 13, 6, 9, 10 Group Y: 8, 6, 11, 7, 5, 9, 10, 7, 8, 6

   Perform a test to determine if there is a significant difference in user engagement between Group X and Group Y. Use a significance level of 0.05.

3. A manufacturer of light bulbs claims that the average lifespan of their bulbs is 1200 hours. To test this claim, a sample of 50 bulbs is selected, and their lifespans are recorded. The sample mean is found to be 1175 hours, with a standard deviation of 100 hours.

   Perform a test to determine if there is a significant difference between the sample mean and the claimed population mean of 1200 hours. Use a significance level of 0.05.

4. A car rental company claims that the average fuel efficiency of their fleet is 35 miles per gallon (mpg). To investigate this claim, a random sample of 100 cars is selected, and their fuel efficiencies are measured. The sample mean is calculated to be 34.2 mpg, with a standard deviation of 2.5 mpg.

   Perform a test to determine if there is a significant difference between the sample mean and the claimed population mean of 35 mpg. Use a significance level of 0.05.

5. A researcher wants to investigate the relationship between the number of hours studied and the exam scores of students. The researcher collects data from 20 students, recording the number of hours studied (independent variable) and their corresponding exam scores (dependent variable). The data is as follows:

   Hours studied: 5, 7, 4, 6, 8, 9, 3, 5, 7, 6, 2, 4, 6, 8, 9, 3, 5, 7, 6, 4 Exam scores: 65, 75, 60, 70, 80, 85, 55, 65, 75, 70, 50, 60, 70, 80, 85, 55, 65, 75, 70, 60

   Perform a simple linear regression to determine the equation of the regression line and assess the strength of the relationship between hours studied and exam scores.

6. A company is testing the effectiveness of two different training methods (Method A and Method B) for improving employee productivity. They randomly assign 30 employees to each method and measure their productivity scores after completing the training. The data is as follows:

   Method A: 78, 82, 85, 73, 80, 88, 76, 79, 81, 84, 77, 75, 79, 83, 80, 87, 82, 75, 78, 81, 84, 79, 76, 82, 80, 83, 85, 77, 81, 79 Method B: 76, 72, 85, 70, 79, 81, 73, 75, 79, 82, 75, 71, 80, 78, 84, 82, 69, 74, 77, 81, 79, 76, 73, 80, 77, 81, 83, 75, 78, 80

   Perform a test to determine if there is a significant difference in productivity scores between employees trained with Method A and Method B. Use a significance level of 0.05.

7. A company claims that their product has an average defect rate of 2%. To test this claim, a sample of 400 products is inspected, and it is found that 15 of them are defective.

   Perform a test to determine if there is a significant difference between the sample defect rate and the claimed population defect rate of 2%. Use a significance level of 0.05.

8. A real estate agent wants to determine if there is a relationship between the size of a house (in square feet) and its selling price (in thousands of dollars). The agent collects data from 25 recently sold houses and records their sizes and selling prices. The data is as follows:

   House Size: 1500, 1800, 1200, 1350, 2000, 1600, 1900, 1750, 1300, 1400, 1650, 1550, 1450, 1700, 1550, 1650, 1800, 1400, 1850, 1600, 1500, 1750,

1300, 1450, 1550 Selling Price: 250, 280, 220, 240, 320, 270, 300, 290, 230, 235, 280, 260, 250, 300, 260, 280, 290, 240, 310, 270, 255, 290, 235, 250, 270

Perform a simple linear regression to determine the equation of the regression line and predict the selling price for a house with a size of 1700 square feet.

9. A company claims that the average response time of their customer service hotline is 30 seconds. To test this claim, a sample of 50 randomly selected calls is analyzed, and the response times (in seconds) are recorded. The sample mean is found to be 32 seconds, with a standard deviation of 5 seconds.

   Perform a test to determine if there is a significant difference between the sample mean response time and the claimed population mean of 30 seconds. Use a significance level of 0.05.

10. A marketing analyst wants to explore the relationship between advertising expenditure (in thousands of dollars) and sales revenue (in thousands of dollars) for a specific product. The analyst collects data from 12 different advertising campaigns and records the advertising expenditure and corresponding sales revenue. The data is as follows:

    Advertising Expenditure: 8, 10, 12, 6, 14, 16, 4, 9, 11, 7, 13, 15 Sales Revenue: 12, 15, 18, 10, 21, 23, 8, 14, 16, 11, 20, 22

    Perform a simple linear regression to determine the equation of the regression line and predict the sales revenue for an advertising expenditure of $13,000.

2 Solutions

2.1 Z-tests, T-tests, and Linear Regression

1. The t-statistic is calculated as -0.163 and the p-value is 0.877. There is no significant difference between the means of Group A and Group B.

2. The t-statistic is calculated as 0.267 and the p-value is 0.793. There is no significant difference in user engagement between Group X and Group Y.

3. The z-statistic is calculated as -2.5 and the p-value is 0.006. There is a significant difference between the sample mean and the claimed population mean.

4. The z-statistic is calculated as -2.26 and the p-value is 0.024. There is a significant difference between the sample mean and the claimed population mean.

5. The equation of the regression line is: Exam scores = 53.68 + 6.04 * Hours studied. The relationship between hours studied and exam scores is moderately strong, with a positive slope indicating that as the number of hours studied increases, the exam scores tend to increase.

6. The t-statistic is calculated as 0.659 and the p-value is 0.514. There is no significant difference in productivity scores between employees trained with Method A and Method B.

7. The z-statistic is calculated as -1.732 and the p-value is 0.083. There is no significant difference between the sample defect rate and the claimed population defect rate.