

Exercise 1: Analyzing Correlation and Covariance for Employee Data Data Sample

An organization recorded the years of experience and monthly salaries of five employees:

Employee	Years of Experience	Monthly Salary (in \$)
1	2	3000
2	5	4500
3	7	5000
4	3	3500
5	6	4800

Tasks

- Calculate the Covariance:**
 - Find the mean of Years of Experience and Monthly Salary.
 - Use the covariance formula, calculating each value's deviation from the mean and finding the final sum.
- Calculate the Pearson Correlation Coefficient:**
 - First, calculate the standard deviations for Years of Experience and Monthly Salary.
 - Substitute the covariance and standard deviations into the correlation formula.

Interpretation:

- For **Covariance**, determine if the relationship is positive or negative based on the sign.
- For **Pearson Correlation**, interpret whether there's a strong relationship between years of experience and monthly salary.

Exercise 2: Using Chi-Square (χ^2) Measure for Categorical Data Data Sample

A retail store tracked the gender and purchase decision of ten customers:

Customer	Gender	Purchase Decision
1	Male	Purchased
2	Female	Purchased
3	Male	Did Not Purchase
4	Female	Purchased
5	Male	Purchased
6	Female	Did Not Purchase
7	Male	Did Not Purchase
8	Female	Purchased

9	Male	Purchased
10	Female	Did Not Purchase

Tasks

- Organize the Data:**
 - Count the frequency of each category combination (e.g., male/purchased, female/purchased).
- Calculate the Chi-Square Statistic:**
 - Use the observed and expected values to compute the Chi-square statistic using the formula:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

- where O_i is the observed count, and E_i is the expected count for each category.
- Draw a Conclusion:**
 - With a significance level of 0.05, decide if there's a significant association between gender and purchase decision.
 - If the Chi-square value is high, conclude that gender may influence purchasing behavior; if not, state that there is no significant association.

Exercise 3:

Given the next dataset with 5 rows for a line $y=ax+b$.

x	y
1	2
2	3.4
3	4.8
4	6.2
5	7.6

- Use the least squares method to find the line $y=ax+b$, then compute R^2 , and interpret the Results.
- Discuss the meaning of the slope, intercept, and R^2 in this context.

$$a = \frac{n \sum(xy) - \sum x \sum y}{n \sum x^2 - (\sum x)^2}, \quad b = \frac{\sum y - a \sum x}{n}, \quad R^2 = \frac{\sum (y_{\text{predicted}} - \bar{y})^2}{\sum (y - \bar{y})^2}$$