PABNA UNIVERSITY OF SCIENCE AND TECHNOLOGY

INFORMATION AND COMMUNICATION ENGINEERING

Course Name: Engineering Statistics

STAT-2201

ASSIGNMENT

SUBMITTED BY

JUBAYER AHMMED

ROLL : 220638

SUBMITTED TO:

Dr. Md. Sarwar Hosain

Associate-Professor

Department of Information and Communication Engineering

Pabna University of Science and Technology

Test for Correlation and Regression Coefficients

Introduction

In the field of statistics, analyzing the relationship between two or more variables is essential for understanding patterns, making predictions, and guiding decisions. Two foundational concepts in this regard are correlation and regression analysis. These tools not only describe relationships but also provide a framework to test whether these relationships are statistically significant or simply due to chance.

While correlation quantifies the degree to which two variables are related, regression analysis goes further to model this relationship, allowing us to predict the value of one variable based on another. To assess the validity of these findings, we employ hypothesis testing for correlation and regression coefficients. This assignment will explore these tests in depth, covering theoretical background, hypothesis formulation, test statistics, interpretation, and real-world applications.

1. Test for Correlation Coefficient

What is a Correlation Coefficient?

Correlation is a statistical measure that expresses the extent to which two variables are linearly related. It answers questions like: “Do students who study more score higher?” or “Does temperature affect ice cream sales?”

The most widely used measure is the Pearson correlation coefficient, denoted by r. It ranges from -1 to +1:

- $r = +1$: Perfect positive linear relationship.
- $r = -1$: Perfect negative linear relationship.
- $r = 0$: No linear relationship.

Types of Correlation

Positive Correlation: As one variable increases, the other tends to increase (e.g., height and weight).
Negative Correlation: As one variable increases, the other tends to decrease (e.g., exercise time and body fat percentage).
Zero Correlation: No consistent relationship (e.g., shoe size and intelligence).

Testing the Significance of the Correlation Coefficient

Why Test?

Just observing a correlation is not enough. We must test whether the observed correlation in the sample reflects a true correlation in the population or if it's due to random chance.

Hypothesis Testing for Correlation

To determine if the observed correlation is statistically significant, we perform a hypothesis test:
- Null Hypothesis (H_0): $\rho = 0$ (There is no correlation in the population.)
- Alternative Hypothesis (H_1): $\rho \neq 0$ (There is a significant correlation.)

Test Statistic

The test statistic for the correlation coefficient is calculated as:

$$t = \frac{r\sqrt{(n-2)}}{\sqrt{(1-r^2)}}$$

Where:
- r is the sample correlation coefficient,
- n is the number of data points.

This test statistic follows a t-distribution with (n - 2) degrees of freedom.

Decision Rule

- Compute the p-value using the t-distribution.
- If p-value < α (usually 0.05), reject the null hypothesis. There is a significant correlation.

2. Test for Regression Coefficient

What is Regression?

Regression analysis models the relationship between a dependent variable (Y) and one or more independent variables (X). In simple linear regression, the relationship is expressed as:

What is a Regression Coefficient?

In simple linear regression, the equation is:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Where:
- β_0 is the intercept,
- β_1 is the regression coefficient (slope),
- ε is the error term.

The regression coefficient β_1 measures the change in Y for a one-unit change in X.

Multiple Linear Regression

When there are multiple predictors (X_1, X_2, ..., X?), the model becomes:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ \ ..... + \beta_k X_k + \varepsilon$$

Each β_i represents the partial effect of X_i on Y, controlling for other variables.

Hypothesis Testing for Regression Coefficient

We test whether the predictor variable X has a significant impact on the response variable Y:
- Null Hypothesis (H_0): $\beta_1 = 0$ (X has no effect on Y)
- Alternative Hypothesis (H_1): $\beta_1 \neq 0$ (X significantly affects Y)

Test Statistic

The test statistic for β_1 is calculated as:

$$t = \frac{(b_1 - \beta_1)}{SE(b_1)}$$

Where:
- b_1 is the estimated regression coefficient from the sample,
- $SE(b_1)$ is the standard error of the coefficient.

We often assume $\beta_1 = 0$ under the null hypothesis, so the formula becomes:

$$t = \frac{b_1}{SE(b_1)}$$

This also follows a t-distribution with (n - 2) degrees of freedom.

Decision Rule

- Calculate the t-statistic and corresponding p-value.
- If p-value < α (commonly 0.05), reject the null hypothesis. It indicates that the predictor X significantly contributes to the model.

Interpretation

If both correlation and regression tests are significant, it means there is a linear relationship between the two variables, and changes in the independent variable are associated with predictable changes in the dependent variable. However, it is important to note that correlation does not imply causation. Regression gives more insight into the nature and magnitude of the relationship.

Real-Life Applications

Business & Marketing: Analyzing the effect of advertising budget on sales.

Medicine: Studying the correlation between drug dosage and patient recovery rates.

Education: Measuring the relationship between study hours and GPA.

Economics: Forecasting inflation based on interest rates and GDP.

Social Sciences: Exploring the effect of income on happiness levels.

Assumptions and Limitations

Assumptions for Pearson Correlation

- Linearity
- Interval or ratio-level data
- Normality
- No significant outliers

Assumptions for Regression

- Linearity
- Independence of errors
- Homoscedasticity (equal variance)
- Normal distribution of residuals

Limitations

- Correlation does not imply causation.
- Regression assumes linear relationships.
- Outliers can greatly affect results.

Conclusion

Testing for correlation and regression coefficients plays a crucial role in statistical analysis. These tests help us understand not only whether variables are related, but also whether that relationship is strong and significant enough to be useful in making predictions. Whether in scientific research, economics, or social sciences, these tools form the backbone of data-driven decision-making.