Karnatak Law Society's
GOGTE INSTITUTE OF TECHNOLOGY
Udyambag Belagavi -590008 Karnataka, India.

A Course Project Report on

"Simple Regression Method 3"

Submitted for the requirements of 5th semester B.E. in CSE for

Research Methodology & Intellectual property rights

(21CS57)

Submitted by

NAME	USN
1) Atharva Parulkar	2GI21CS105
2) Omkar Vishnu Patil	2GI21CS106
3) Omkar Patil	2GI21CS101
4) Prateet Gadavi	2GI21CS112

Under the guidance of

Dr. Sanjeev Sannakki

Dept. of CSE

Academic Year 2023-2024 (Odd semester)

Karnatak Law Society's

GOGTE INSTITUTE OF TECHNOLOGY

Udyambag Belagavi -590008 Karnataka, India.

Department of Computer Science and Engineering

Certificate

This is to certify that the Course Project work titled **"Simple Regression Method 3 "** carried out by **Students Atharva Parulkar, Omkar Vishnu Patil , Omkar Patil ,** Prateet Gadavi bearing USNs: 2GI21CS105, 2GI21CS106. 2GI21CS101, 2GI21CS112 for **Research Methodology & Intellectual property rights (21CS57)** course is submitted in partial fulfilment of the requirements for 5th semester B.E. in **COMPUTER SCIENCE AND ENGINEERING**i. It is certified that all corrections/ suggestions indicated have been incorporated in the report. The course project report has been approved as it satisfies the academic requirements prescribed for the said degree.

Date: Signature of Guide

Place: Belagavi Dr. Sanjeev Sannakki

Dept. of CSE

KLS Gogte Institute of Technology, Belagavi

ACKNOWLEDGEMENTS

I extend my sincere appreciation to my fellow team members for their collective dedication and collaborative spirit throughout the development of this project. Each team member played a crucial role in contributing their skills and expertise, making this endeavor a truly collaborative and rewarding experience.

I want to express my gratitude to the faculty and mentors who provided guidance and support, helping us navigate challenges and refine our approach. Their insights were invaluable in shaping the project and enhancing our understanding of the subject matter.

This project also benefited from the resources and facilities provided by GIT. The conducive academic environment and access to necessary infrastructure significantly facilitated the implementation and execution of our ideas.

Lastly, I want to acknowledge the collective effort of all those who, directly or indirectly, contributed to the successful completion of this project. Their support and encouragement were instrumental in turning our ideas into a tangible and meaningful outcome.

Thank you to everyone involved in making this college project a collaborative and enriching learning experience.

ABSTRACT

This project introduces research methodology refer to the specific, measurable goals that a researcher aims to achieve through a study. These objectives guide the research process and help define the purpose and scope of the investigation. Commonly, objectives articulate what the researcher intends to accomplish, such as testing a hypothesis, exploring relationships between variables, or describing a phenomenon.

For instance, in a study on the impact of a new teaching method on student performance, objectives might include assessing academic scores before and after implementation, comparing results between groups, and identifying any significant improvements.Clear and well-defined objectives contribute to the overall rigor and success of a research project by providing a roadmap for the research process.

In research methodology, simple regression is a statistical technique used to explore the relationship between two variables. It helps in understanding how changes in one variable are associated with changes in another. Simple regression is particularly useful when examining the influence of an independent variable on a dependent variable. Researchers analyze the data to estimate the parameters (slope and intercept) of the regression equation and assess the strength and significance of the relationship. This method provides valuable insights into the linear association between variables within a study.

.

Table of contents

Problem Statement

Simple Regression Method 3

Objectives

1. **Define the Research Problem :**
    - Clearly articulate the problem or issue that your research aims to address.
2. **Conduct a Literature Review :**
    - Review existing literature to identify gaps, theories, and relevant studies related to your research.
3. **Formulate Hypotheses or Research Questions :**
    - Develop specific hypotheses or research questions that your study seeks to answer.
4. **Design the Research Methodology:**
    - Define the research design, including the type of study (e.g., experimental, observational), sampling method, and data collection techniques.
5. **Collect Data :**
    - Implement the data collection process according to the defined methodology.
6. **Analyze Data :**
    - Conduct appropriate statistical or qualitative analyses on the collected data.
7. **Interpret Results:**
    - Interpret the findings in the context of your research questions or hypotheses.
8. **Draw Conclusions :**
    - Summarize the main conclusions drawn from your study.
9. **Discuss Implications :**

- Explore the implications of your findings for theory, practice, or future research.

10. Recommendations :

- Provide recommendations based on your study's outcomes.

11. Write and Present the Research :

- Communicate your research through a well-structured and coherent research paper or presentation.

12. Reflect on Limitations and Future Research :

- Identify and discuss the limitations of your study and suggest directions for future research.

Theory:

simple linear regression (SLR) is a linear regression model with a single explanatory variable. That is, it concerns two-dimensional sample points with one independent variable and one dependent variable (conventionally, the *x* and *y* coordinates in a Cartesian coordinate system) and finds a linear function (a non-vertical straight line) that, as accurately as possible, predicts the dependent variable values as a function of the independent variable. The adjective *simple* refers to the fact that the outcome variable is related to a single predictor.

It is common to make the additional stipulation that the ordinary least squares (OLS) method should be used: the accuracy of each predicted value is measured by its squared *residual* (vertical distance between the point of the data set and the fitted line), and the goal is to make the sum of these squared deviations as small as possible. In this case, the slope of the fitted line is equal to the correlation between *y* and *x* corrected by the ratio of standard deviations of these variables. The intercept of the fitted line is such that the line passes through the center of mass (*x*, *y*) of the data points.

Formulation and computation:

Consider the model function

$$y = \alpha + \beta x,$$

which describes a line with slope β and y-intercept α. In general such a relationship may not hold exactly for the largely unobserved population of values of the independent and dependent variables; we call the unobserved deviations from the above equation the errors. Suppose we observe n data pairs and call them $\{(x_i, y_i), i = 1, \ldots, n\}$. We can describe the underlying relationship between y_i and x_i involving this error term ε_i by

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

This relationship between the true (but unobserved) underlying parameters α and β and the data points is called a linear regression model.

The goal is to find estimated values and for the parameters α and β which would provide the "best" fit in some sense for the data points. As mentioned in the introduction, in this article the "best" fit will be understood as in the least-squares approach: a line that minimizes

$$\hat{\varepsilon}_i = y_i - \alpha - \beta x_i .$$

In other words, $\widehat{\alpha}$ and $\widehat{\beta}$ solve the following minimization problem:

$$(\hat{\alpha}, \hat{\beta}) = \operatorname{argmin}(Q(\alpha, \beta)),$$

where the objective function Q is:

$$Q(\alpha, \beta) = \sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2} = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 .$$

By expanding to get a quadratic expression in α and β, we can derive minimizing values of the function arguments, denoted $\widehat{\alpha}$ and $\widehat{\beta}$:[6]

$$\widehat{\alpha} = \bar{y} - (\widehat{\beta}\,\bar{x}),$$

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} \Delta x_i \Delta y_i}{\sum_{i=1}^{n} \Delta x_i^2}$$

Here we have introduced

- $\bar{x}$ and $\bar{y}$ as the average of the x_i and y_i, respectively
- Δx_i and Δy_i as the deviations in x_i and y_i with respect to their respective means.

Interpretation:

Relationship with the sample covariance matrix

The solution can be reformulated using elements of the covariance matrix:

$$\widehat{\beta} = \frac{s_{x,y}}{s_x^2} = r_{xy}\frac{s_y}{s_x}$$

where

- r_{xy} is the sample correlation coefficient between *x* and *y*
- s_x and s_y are the uncorrected sample standard deviations of *x* and *y*
- sample variance and sample covariance, respectively

$$\frac{\hat{y}-\bar{y}}{s_y} = r_{xy}\frac{x-\bar{x}}{s_x}.$$

This shows that r_{xy} is the slope of the regression line of the standardized data points (and that this line passes through the origin). Since then we get that if x is some measurement and y is a followup measurement from the same item, then we expect that y (on average) will be closer to the mean measurement than it was to the original value of x. This phenomenon is known as regressions toward the mean.

Generalizing the notation, we can write a horizontal bar over an expression to indicate the average value of that expression over the set of samples. For example:

$$\overline{xy} = \frac{1}{n}\sum_{i-1}^{n} x_i y_i.$$

This notation allows us a concise formula for r_{xy}:

$$r_{xy} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{\left(\overline{x^2} - \bar{x}^2\right)\left(\overline{y^2} - \bar{y}^2\right)}}.$$

The coefficient of determination ("R squared") is equal to when the model is linear with a single independent variable. See sample correlation coefficient for additional details.

Interpretation about the slope:

By multiplying all members of the summation in the numerator by :

$\frac{(x_i - \bar{x})}{(x_i - \bar{x})} = 1$ (thereby not changing it):

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \frac{(y_i - \bar{y})}{(x_i - \bar{x})}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2} \frac{(y_i - \bar{y})}{(x_i - \bar{x})}$$

We can see that the slope (tangent of angle) of the regression line is the weighted average of

$\frac{(y_i - \bar{y})}{(x_i - \bar{x})}$ that is the slope (tangent of angle) of the line that connects the i-th point to the average of all points, weighted by $(x_i - \bar{x})^2$ because the further the point is the more "important" it is, since small errors in its position will affect the slope connecting it to the center point more.

Interpretation about the intercept

$$\widehat{\alpha} = \bar{y} - \widehat{\beta}\,\bar{x},$$

Given $\widehat{\beta} = \tan(\theta) = dy/dx \to dy = dx \times \widehat{\beta}$ with θ the angle the line makes with the positive x axis, we have $y_{\text{intersection}} = \bar{y} - dx \times \widehat{\beta} = \bar{y} - dy$

Applications of the Project:

Simple regression analysis is a statistical method that can be applied in various fields to explore relationships between two variables. Here are three common applications of simple regression:

1. **Economics and Finance:**
   - *Application:* In economics and finance, simple regression is frequently used to analyze the relationship between two economic variables. For example, researchers might explore how changes in interest rates (independent variable) affect consumer spending (dependent variable). By applying simple regression, analysts can estimate the impact of interest rate changes on spending, providing valuable insights for economic forecasting and policy-making.

2. **Health Sciences:**
   - *Application:* In health sciences, researchers often use simple regression to examine the association between independent and dependent variables related to health outcomes. For instance, a study might investigate the relationship between the number of hours of physical activity per week (independent variable) and cardiovascular health indicators such as blood pressure (dependent variable). Simple regression can help quantify the extent to which changes in physical activity levels are associated with changes in blood pressure.

3. **Marketing and Sales:**
   - *Application:* Simple regression is useful in marketing to understand the impact of various marketing strategies on sales performance. For example, a company might analyze the relationship between

advertising expenditure (independent variable) and monthly sales revenue (dependent variable). By employing simple regression, marketers can estimate the effectiveness of their advertising campaigns and make data-driven decisions to optimize marketing budgets.

Conclusion

In conclusion the simple regression method serves as a valuable statistical tool with diverse applications across various disciplines. Through the exploration of relationships between two variables, this method offers insights into patterns, trends, and potential cause-and-effect associations. In this brief overview, we highlighted three specific applications to illustrate the method's versatility.

In the field of economics and finance, simple regression enables researchers to assess the impact of changes in one variable on another, contributing to informed economic forecasting and policy decisions. In health sciences, the method allows for the examination of relationships between lifestyle factors and health outcomes, aiding in the development of targeted interventions. Additionally, in marketing and sales, simple regression empowers professionals to quantify the effectiveness of marketing strategies on sales performance, guiding strategic decision-making.

References :

- Panneer Selvam, Research Methodology, PHI Learning Pvt. Ltd., 2007.
- Dr. B.L. Wadhera -Intellectual Property Rights, Universal Law Publishing Co. Ltd.. 2002
- William G Zikmund, Business Research Methods, Indian edition, South western Publishers, 8th Indian Reprint – 2009
- https://drive.google.com/file/d/1Qh_E3BmdhcX3ICGC8QsD72NrVs7tXNTl/view?usp=drivesdk