Karnatak Law Society’s

GOGTE INSTITUTE OF TECHNOLOGY

Udyambag Belagavi -590008 Karnataka, India.



A Course Project Report on

**Exploring the Linear Relationship: Pearson's Correlation Analysis between Time Spent and Student Marks**

Submitted for the requirements of 5th semester B.E. in CSE for **Research Methodology and Intellectual Property Rights (21CS57)**

**Submitted by**

|  |  |
| --- | --- |
| **NAME** | **USN** |
| **1)Shreyas Huddar** | **2GI21IS050** |
| **2)Ayush anand** | **2GI21IS013** |
| **3) Anirudh Kulkarni** | **2GI21IS008** |
| **4) Raghupraveer Nippani** | **2GI21IS65** |

**Under the guidance of**

**Prof.Neelkanth Karekar**

**Dept. of ISE**

**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 **“Exploring the Linear Relationship: Pearson's Correlation Analysis between Time Spent and Student Marks”** carried out by **Students Ayush anand, Anirudh Kulkarni,Shreyas Huddar,Raghupraveer Nippani** bearing **USNs: 2GI21IS013 2GI21IS050,2GI21IS008,2GI21IS065** for **Research Methodology and Intellectual Property Rights (21IS57)** course is submitted in partial fulfilment of the requirements for 5th semester B.E. in **INFORMATION 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 | Prof.Neelkanth Karekar |

Department of ISE

KLS Gogte Institute of Technology, Belagavi

Karnatak Law Society’s

GOGTE INSTITUTE OF TECHNOLOGY

Udyambag Belagavi -590008

**Academic Year 2023-24 (Odd Semester)**

**Semester: V**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S**  **.**  **N**  **o** | **Project**  **Component** | **Ma**  **x.**  **Ma rks** | **Marks Earned** | | | |
| **2GI21IS013** | **2GI21IS008** | **2GI21IS050** | **2GI21IS065** |
|  |  |  |  |
| Ayush anand | Anirudh Kulkarni | Shreyas huddar | Raghupraveer Nippani |
|  |  |  |  |
| 1 | Relevance of the project and  its objectives | 01 |  |  |  |  |
| 2 | Tools/Fram ework used | 01 |  |  |  |  |
| 3 | Methodolog y / Design | 02 |  |  |  |  |
| 4 | Implementa tion and Results | 03 |  |  |  |  |
| 5 | Project Report | 03 |  |  |  |  |
|  | **Total** | **10** |  |  |  |  |

**Course: Research Methodology and Intellectual Property Rights**

**(21IS57)**Rubrics for evaluation of Course Project

**`ACKNOWLEDGEMENTS**

We extend our sincere gratitude to all those who contributed to the success of this project, making it a collaborative and enriching experience. This project, a culmination of efforts and insights, was undertaken as part of a course requirement. As a team of two, we would like to acknowledge the following:

Our heartfelt thanks to the instructors for their guidance, support, and invaluable feedback throughout the duration of the course. Their expertise and commitment fostered an environment conducive to learning and exploration.

A special acknowledgment to our teammate, whose collaboration and dedication played a pivotal role in the project's success. The synergy of our efforts enhanced the quality and depth of our analysis.

This project served as an invaluable learning experience, and the support from these individuals and entities greatly enriched our journey.

## ABSTRACT

This study explores the relationship between the time spent by students on a particular activity and the marks they scored. A simple regression analysis was conducted using data collected on these variables. The analysis revealed a strong positive correlation (r = 0.94255) between the time spent and the marks obtained. The regression equation, with a slope (β) of 5.68875 and a y-intercept (α) of 1.22385, suggests that an increase in time spent is associated with a significant increase in marks. However, it is important to acknowledge the potential influence of unobserved factors that might contribute to student performance. The abstract provides a concise overview of the study's objectives, methods, and key findings.

**Table of contents**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Content** | **Page No.** |
| 1 | Problem Statement | **7** |
| 2 | Introduction | **7** |
| 3 | Objectives | **8** |
| 4 | Methodology | **9** |
| 5 | Results | **10** |
| 6 | Conclusion | **11** |
| 7 | References | **12** |

### Problem Statement

In the realm of education, understanding the factors that contribute to student success is of paramount importance. This study seeks to investigate the relationship between the time spent by students on a specific activity and the corresponding marks they achieve. The overarching goal is to discern whether there exists a measurable and significant correlation, and if so, to quantify the strength and direction of this relationship.

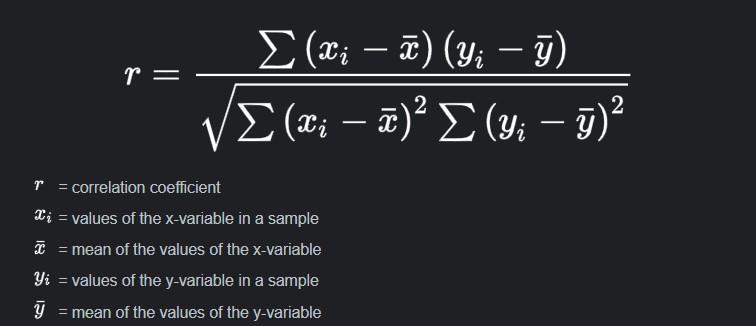
While the initial analysis points to a strong positive correlation between time spent and marks obtained, it is crucial to address the broader question: What other factors, perhaps unobserved or not included in our current analysis, might influence student performance? Recognizing and acknowledging the potential presence of these confounding variables is essential for a comprehensive understanding of the relationship between time investment and academic achievement.

### Introduction

This report delves into the analysis of the correlation between the time students allocate to a specific activity and the corresponding marks they achieve, employing Pearson's correlation coefficient (r) as the statistical measure. Pearson's correlation is a widely used method for quantifying the strength and direction of a linear relationship between two continuous variables.

Pearson's Correlation Coefficient (r)

Pearson's correlation is defined by the following formula:



Interpretation of Pearson's Correlation Coefficient

* ( r ) ranges between -1 and 1.
* A positive ( r ) indicates a positive linear relationship: as one variable increases, the other tends to increase.
* A negative ( r ) indicates a negative linear relationship: as one variable increases, the other tends to decrease.
* An ( r ) close to 1 or -1 suggests a strong linear relationship.
* An ( r ) close to 0 indicates a weak or no linear relationship.

This statistical method allows us to quantify the extent to which changes in one variable correspond to changes in another. In our analysis, ( r ) will serve as a numerical indicator of the linear association between the time spent on an activity and the marks achieved by students. As we progress, a deeper exploration of the Pearson correlation results will illuminate the nature of this relationship and provide valuable insights into the dynamics of student performance in relation to time allocation.

### Objectives

**The primary objectives of this study are as follows:**

1. **Examine the Relationship:** Investigate and quantify the relationship between the time spent by students on a specific activity and the marks they achieve. Utilize regression analysis to understand the nature and strength of this relationship.
2. **Understand Variable Contributions:** Explore the individual contributions of the time spent variable to student marks. Analyze the significance of the slope and y-intercept in the context of the regression model.

**Methodology/Design:**

#### Data Collection

The dataset for this analysis was obtained from Kaggle, a reputable platform for datasets related to various domains. The dataset includes two main variables: the time spent by students on a specific activity and the marks they achieved. The collection of data from Kaggle ensures a diverse and potentially well-curated set, providing a comprehensive basis for analysis.

#### Data Preprocessing

Prior to analysis, the dataset underwent standard preprocessing steps. This involved handling missing data, checking for outliers, and ensuring data consistency. Any necessary transformations were applied to prepare the dataset for regression analysis.

#### Regression Analysis

The primary analytical tool employed in this study is simple linear regression. The regression model was built using Excel, utilizing the least squares method to determine the coefficients (slope and y-intercept). The statistical output from the regression analysis, including the slope, y-intercept, and correlation coefficient, was then interpreted to understand the relationship between time spent and marks obtained.

#### Scatter Plot Visualization

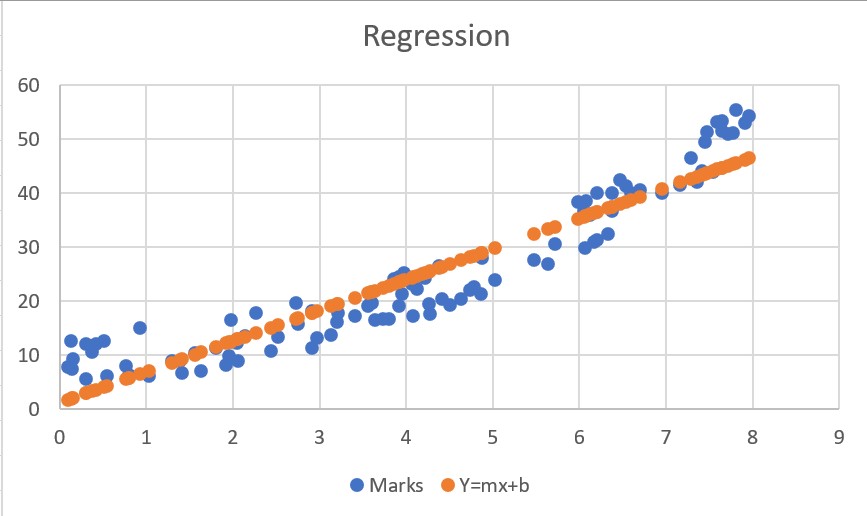
To complement the numerical analysis, a scatter plot was created to visually represent the distribution of data points and the regression line. This visualization aids in understanding the pattern and strength of the relationship between the variables.

### Results

The regression analysis revealed the following key findings:

1. **Slope (β):** The slope of the regression line is 5.68875, indicating that, on average, for every additional unit of time spent by a student on the activity, their marks are expected to increase by approximately 5.68875 units.
2. **Y-Intercept (α):** The y-intercept is 1.22385, representing the predicted marks when the time spent is zero. However, interpreting this value in the absence of a meaningful zero point for the time variable requires caution.
3. **Correlation Coefficient (r):** The correlation coefficient is 0.94255, signifying a strong positive correlation between the time spent on the activity and the marks obtained. This suggests that as the time spent increases, there is a notable tendency for marks to increase as well.

#### Scatter Plot Visualization



The scatter plot visually reinforces the observed correlation, with data points clustering around the regression line. The upward trend in the plot aligns with the positive correlation coefficient, providing a clear representation of the relationship between time spent and student marks.

#### Predictive Capability

The regression model exhibits a high predictive capability, as indicated by the strong correlation and the well-defined regression line. This suggests that the time spent variable is a significant predictor of the marks obtained by students in the given context.

#### Consideration of External Factors

While the analysis provides valuable insights into the observed relationship, it is crucial to acknowledge the potential influence of external factors not included in the dataset. Other variables such as study habits, socioeconomic factors, or individual learning styles might contribute to student performance and were not explicitly examined in this analysis.

### Conclusion

In summary, the Pearson's correlation analysis conducted on the dataset examining the relationship between time spent on a specific activity and student marks reveals a strong positive correlation ( r = 0.94255 ). This indicates a robust linear association, suggesting that as the time invested increases, there is a notable tendency for marks to increase as well. The statistical method employed provides a quantifiable measure of this relationship, contributing valuable insights into the interplay between time management and academic performance.

#### References

* [**https://sites.utexas.edu/sos/guided/inferential/numeric/bivariate/cor/**](https://sites.utexas.edu/sos/guided/inferential/numeric/bivariate/cor/)
* [**https://www.kaggle.com/datasets/yasserh/student-marks-dataset**](https://www.kaggle.com/datasets/yasserh/student-marks-dataset)
* [**https://drive.google.com/file/d/1Qh\_E3BmdhcX3ICGC8QsD72NrVs7tXNTl/view**](https://drive.google.com/file/d/1Qh_E3BmdhcX3ICGC8QsD72NrVs7tXNTl/view)