

Anekant Education Society’s

**Tuljaram Chaturchand College of Arts, Commerce and Science, Baramati - 413102**

Autonomous

A project report on

***“ Multivariate Analysis On Bank Data”***

SUBMITTED TO

**DEPARTMENT OF STATISTICS**

Msc.I. (Statistics)

**By**

Mr.Prasad Dattatray Beldar.

Under the Guidance of

Prof..Mrs.Jagtap.N.A



**Anekant Education Society’s**

**Tuljaram Chaturchand College of Arts, Commerce and Science, Baramati – 413102**

**DEPARTMENT OF STATISTICS**

## CERTIFICATE

This is to Certify that the project entitled is being submitted by Mr.Beldar Prasad Dattatray As partial fulfilment for the award of the degree of the Msc I to the Department Of Statistics.

This project has been conducted under my Supervision

and Guidence.

Place: Baramati Date:

Prof. Mrs.Jagtap.N.A Dr. Jagtap A.S.

**Project Guide Head of Department**

## TABLE OF CONTENT:

|  |  |  |
| --- | --- | --- |
| ***Sr. No*** | ***Contents*** | ***Page no.*** |
| 1 | Acknowledgement | **4** |
| 2 | Introduction | **5** |
| 3 | Abstract & Keywords | **6** |
| 4 | Objectives | **7** |
| 5 | Motivation | **8** |
| 6 | Methodology | **9** |
| 7 | Statistical Analysis | **10** |
| 8 | Conclusion | **14** |
| 9 | Limitations | **15** |
| 10 | References | **16** |

### **ACKNOWLEDGEMENT**

We Wish to express our sense of gratitude to Prof. Mrs.Jagtap.N.A for her inspiring guidance, immense motivation, constant encouragement & a critical approach at every stage, till project work complete successfully.

We are thankful to Dr. A. S. Jagtap (Head of Department of Statistics & vice principal of T. C. College, Baramati), for providing all necessary facilities & good co-operation.

We are also thankful all staff member of the Department of statistics for their valuable discussion and guidance for completion of our project.

We express our sincere thanks to my classmates for their help in the process of this work. Last but not least, we would like to thanks non-teaching staff of the department of the statistics.

**Introduction**

* PCA can be used for loan applicants as well. [For example, if you have many variables such as age, income, credit score, etc., PCA can help you identify which variables are most important for predicting loan approval](https://www.analyticsvidhya.com/blog/2022/05/loan-prediction-problem-from-scratch-to-end/)
* PCA is a statistical technique that can be used to reduce the number of variables in a dataset while retaining most of the information in the original set of variables.
* [It can be used to simplify large data tables, identify the most important variables, and find patterns in your data more easily**1**](https://towardsdatascience.com/classifying-loans-based-on-the-risk-of-defaulting-using-logistic-regression-9bd9c6b44640)
* [PCA can be particularly beneficial when you have many variables relative to the number of observations or when the variables are highly correlated**1**](https://towardsdatascience.com/classifying-loans-based-on-the-risk-of-defaulting-using-logistic-regression-9bd9c6b44640). [It helps resolve both problems by reducing the dataset to a smaller number of independent (i.e., uncorrelated) variables**1**](https://towardsdatascience.com/classifying-loans-based-on-the-risk-of-defaulting-using-logistic-regression-9bd9c6b44640).
* In general, PCA is just one step in an analytical process. [For example, you can use it before performing regression analysis, using a clustering algorithm, or creating a visualization**1**](https://towardsdatascience.com/classifying-loans-based-on-the-risk-of-defaulting-using-logistic-regression-9bd9c6b44640)

**SOFTWARE:**

* + Minitab
  + Microsoft Excel
  + Microsoft Word

### **ABSTRACT**

* [PCA can be used for loan applicants to reduce the number of variables and make the data easier to analyze1](https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/multivariate/how-to/principal-components/before-you-start/example/). [For example, a bank requires eight pieces of information from loan applicants: income, education level, age, length of time at current residence, length of time with current employer, savings, debt, and number of credit cards2](https://support.minitab.com/en-us/datasets/graphs-data-sets/loan-applicant-data/). [A bank administrator wants to analyze this data to determine the best way to group and report it2](https://support.minitab.com/en-us/datasets/graphs-data-sets/loan-applicant-data/). [The administrator performs a principal components analysis (PCA) to reduce the number of variables to make the data easier to analyze1](https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/multivariate/how-to/principal-components/before-you-start/example/). [The administrator wants enough components to explain 90% of the variation in the data](https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/multivariate/how-to/principal-components/before-you-start/example/)
  + - .

###### **KEYWORDS** :

* + Graphical Representation
  + **Correlation**
  + **Principal Component Analysis**.

### **OBJECTIVES:**

* To reduce the number of variables and make the data easier to analyze.
* wants enough components to explain 90% of the variation in the data.
* wants to analyze this data to determine the best way to group and report it
* Data visualization
* Outlier Detection

## MOTIVATION

* Multivariate data analysis is used to analyze data that has more than one variable. It is used to identify patterns and relationships between variables that may not be apparent when looking at each variable separately. Multivariate data analysis can help you understand how different variables are related to each other and how they affect each other.
* There are many reasons why you might want to use multivariate data analysis. For example, you might want to identify which variables are most important in predicting an outcome or you might want to identify which variables are most closely related to each other.
* Multivariate data analysis can also help you identify outliers or unusual observations in your data. This can be useful for identifying errors in your data or for identifying unusual patterns that may be of interest.
* Overall, multivariate data analysis can help you gain a deeper understanding of your data and can help you make more informed decisions based on your data
* In our curriculum, there is a course "Applied Multivariate Analysis" which is applicable more in Dimension Reduction . We have experience of handling different data sets in practical during our M.Sc.-I. During learning "Applied Multivariate Analysis" also get theoretical knowledge but I was much interested to knowing how the data is analyses in the actual sense. So, I decided to carry out PCA in my project.

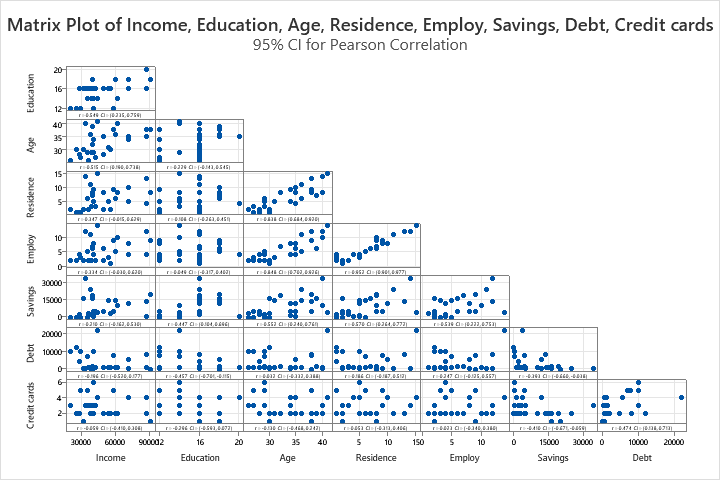
# **METHODOLOGY** :

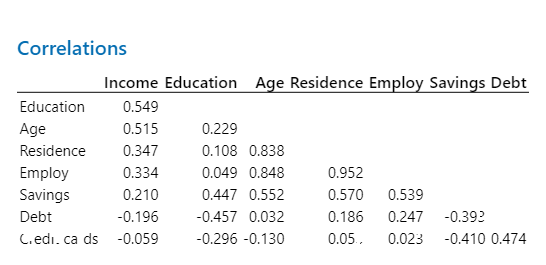
* As we decided to Study of Loan Applicants.We decided to Study Different factors like Income, Education, Age, Residency at current address, Years at current employer, Savings, Debt, and the number of credit cards. All information used in this project is collected from online source and various websites. Taking this data as a basis of research we decided to work on Study on Loan Applicants And we will Use Principal Component Analysis for the dimension Reduction Or variable Reduction
* In this project the data contains eight Variable so firstly we will plot Correlation Matrix Then Go for PCA and from this we will get important variable and interprete the result on this Basis as a we used some software like Minitab ,MS-Exel.

**STATISTICAL ANALYSIS**



**Correlation matrix:**



****

WORKSHEET 1

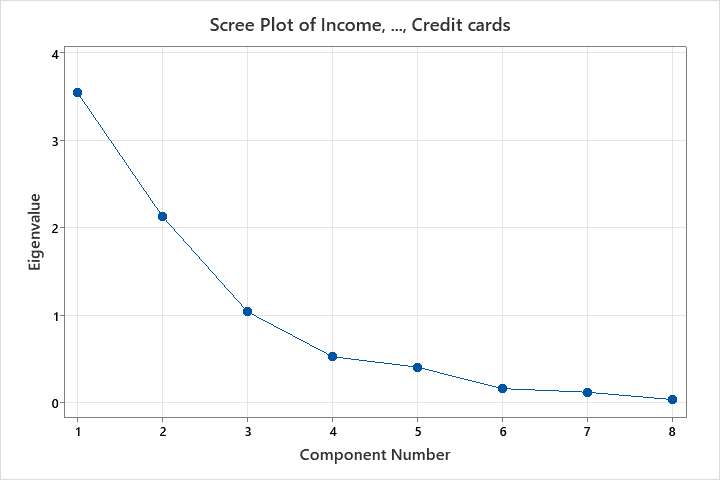
**Principal Component Analysis: Income, Education, Age, Residence, Employ, Savings, Debt, Credit cards**

**Eigenanalysis of the Correlation Matrix**

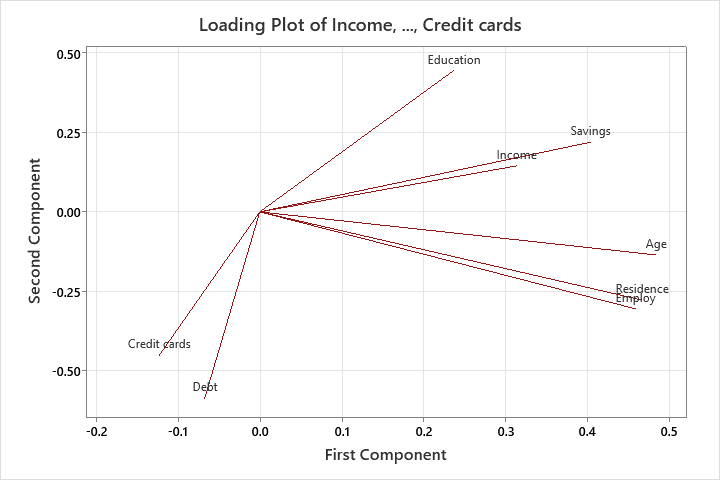
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Eigenvalue | 3.5476 | 2.1320 | 1.0447 | 0.5315 | 0.4112 | 0.1665 | 0.1254 | 0.0411 |
| Proportion | 0.443 | 0.266 | 0.131 | 0.066 | 0.051 | 0.021 | 0.016 | 0.005 |
| Cumulative | 0.443 | 0.710 | 0.841 | 0.907 | 0.958 | 0.979 | 0.995 | 1.000 |

**Eigenvectors**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **PC1** | **PC2** | **PC3** | **PC4** | **PC5** | **PC6** | **PC7** | **PC8** |
| Income | 0.314 | 0.145 | -0.676 | -0.347 | -0.241 | 0.494 | 0.018 | -0.030 |
| Education | 0.237 | 0.444 | -0.401 | 0.240 | 0.622 | -0.357 | 0.103 | 0.057 |
| Age | 0.484 | -0.135 | -0.004 | -0.212 | -0.175 | -0.487 | -0.657 | -0.052 |
| Residence | 0.466 | -0.277 | 0.091 | 0.116 | -0.035 | -0.085 | 0.487 | -0.662 |
| Employ | 0.459 | -0.304 | 0.122 | -0.017 | -0.014 | -0.023 | 0.368 | 0.739 |
| Savings | 0.404 | 0.219 | 0.366 | 0.436 | 0.143 | 0.568 | -0.348 | -0.017 |
| Debt | -0.067 | -0.585 | -0.078 | -0.281 | 0.681 | 0.245 | -0.196 | -0.075 |
| Credit cards | -0.123 | -0.452 | -0.468 | 0.703 | -0.195 | -0.022 | -0.158 | 0.058 |



This scree plot shows that the flat portion of the curve begins at component 4, indicating we should use the first three principal components



The lower section of the output shows the [correlations](https://statisticsbyjim.com/basics/correlations/), or loadings, between the variables and the components. Use the loadings to identify the features that most strongly correlate with each component.

For example, the first principal component has high loadings on age (0.484), residence (0.466), employ (0.459), and savings (0.404). Hence PC1 is strongly associated with these variables.

Below, the loading plot displays the correlations graphically using the first and second components for the X and Y axis

**CONCLUSION**

* So, PC1 measures long-term stability—older, more time at current employer, and living longer in the current residence.
* Conversely, credit cards and debt have negative correlations with PC2. It’s a measure of debt usage.

**LIMITATION:**

* **We have study on 30 samples so we can increase sample size Interpret again.**
* **Lack of previous research studies on this topic**
* **Unable to apply more statistical technique**

# REFERENCES:

##### Web Search-

[**Principal Component Analysis Guide & Example - Statistics By Jim**](https://statisticsbyjim.com/basics/principal-component-analysis/)**.**