**The Regression Analysis of Individual Financial Performance**



**Jusak Praveen Sikharam**

Table of Contents

[Project overview 3](#_Toc122123450)

[Metric used 3](#_Toc122123451)

[2. DATA ANALYSIS; 4](#_Toc122123452)

[3. Methodology 7](#_Toc122123453)

[Data preprocessing[4] 7](#_Toc122123454)

[Implementation 7](#_Toc122123455)

[4. Model validation and evaluation [1] 8](#_Toc122123456)

[Justification regression of model 8](#_Toc122123457)

[5. CONCLUSION AND IMPROVEMENT. 9](#_Toc122123458)

[REFERENCES. 10](#_Toc122123459)

# Project Overview

Regression analysis is a powerful method and is commonly used in the estimation of parameters.

This project will study earning patterns from different businesses set by individuals and try to check a company that has a positive or negative contribution to the total earnings of the owner.

Many individuals set up different businesses, and at a time, business collapses due to a lack of mathematical accountability per specific business.

This model will help such investors detect companies that may reduce earnings and do away with them.

The supermarket may also benefit from such models since they depend on different

sub-businesses to make a profit.

Data accompanying this study was collected from an investor who has set up three businesses:

a shop, a chemist, and rent from real estate investments.

The data is recorded daily.

# Metric Used

This study is based on an analysis of the regression line.

The regression technique will combine the three earnings sources and develop a reasonable model that will be relied on in the decision-making process.

The coefficient of the variables and their p values will be used to make statistical conclusions on whether variables have a significant or insignificant effect on earnings.

From that point, the final decision is made whether to close a business or adjust management for that business to impact total earnings positively.

# 2. DATA ANALYSIS

Data exploration and visualization [3].

|  |  |
| --- | --- |
|  | *shop1* |
|  |  |
| Mean | 1844.048634 |
| Standard Error | 253.8639611 |
| Median | 563 |
| Mode | 0 |
| Standard Deviation | 16033.67478 |
| Sample Variance | 257078727 |
| Kurtosis | 936.9422876 |
| Skewness | 29.53616739 |
| Range | 577875 |
| Minimum | 0 |
| Maximum | 577875 |
| Sum | 7355910 |
| Count | 3989 |

Let's explore the data and features before we do an analysis.

This step is significant because it helps us understand the data and assumptions accompanying any research.

Below are descriptive statistics for the data variables.

|  |  |
| --- | --- |
|  | *Earnings* |
|  |  |
| Mean | 6049.536475 |
| Standard Error | 630.0248963 |
| Median | 2382 |
| Mode | 0 |
| Standard Deviation | 39791.44675 |
| Sample Variance | 1583359235 |
| Kurtosis | 1242.43173 |
| Skewness | 32.65525331 |
| Range | 1781612 |
| Minimum | 0 |
| Maximum | 1781612 |
| Sum | 24131601 |
| Count | 3989 |

The above tables show descriptive statistics [3] for total earnings and earnings from shop1.

|  |  |
| --- | --- |
| *chemist* | |
|  |  |
| Mean | 1937.427 |
| Standard Error | 236.0006 |
| Median | 676 |
| Mode | 0 |
| Standard Deviation | 14905.45 |
| Sample Variance | 2.22E+08 |
| Kurtosis | 969.2038 |
| Skewness | 29.72791 |
| Range | 567801 |
| Minimum | 0 |
| Maximum | 567801 |
| Sum | 7728395 |
| Count | 3989 |

|  |  |
| --- | --- |
| *rent* | |
|  |  |
| Mean | 2268.061 |
| Standard Error | 183.462 |
| Median | 1051 |
| Mode | 0 |
| Standard Deviation | 11587.19 |
| Sample Variance | 1.34E+08 |
| Kurtosis | 2278.761 |
| Skewness | 43.47646 |
| Range | 635936 |
| Minimum | 0 |
| Maximum | 635936 |
| Sum | 9047296 |
| Count | 3989 |

The above tables show descriptive statistics for earnings from chemist and rent collection.

Graphs for distribution of different variables forming the model.

No more data preprocessing needed because the data was obtained from the investor was already in analytical format.

Model fitting

Graphical user interface, application, table, Excel

Description automatically generated

From the above model, we can conclude that accuracy is of the highest level. Referring to the summary output statistics, our model explains almost 100% of the total point (value of R-square) [2].

This means our model can be relied on in decision-making. Our final model would be obtained from an excel document to avoid rounding off points. Therefore, our model is as follows.

**Earnings = (shop1\*0.99999…) + (chemist\*1) + (rent\*1) – 6.3664629124105E-12**

All signs of regression coefficient value corresponding to each variable are positive.

This means that shop1, chemist, and rent collection positively increased earnings whenever a unit of each variable was introduced in our model. All P values are less than 0.05, indicating those variables had the effect to payments significantly

## 3. Methodology

## Data preprocessing [4]

Any type of processing done on raw data to prepare it for another data processing operation is referred to as data preprocessing, which is a part of data preparation. It has historically been a crucial first stage in the data mining process.

Data preparation methods have recently been modified to train AI and machine learning models and to run inferences against them.

Data preprocessing changes the data into a format that can be processed in data mining, machine learning, and other data science tasks more quickly and efficiently.

To ensure reliable findings, the techniques are typically applied at the very beginning of the machine learning and AI development pipeline.

Regarding this point, I did not perform more data preprocessing because I received complete and organized data ready for analysis.

## Implementation

A statistical technique used to determine the level effect of an independent variable (x) on a dependent variable is regression analysis (y). Estimating the influence of the independent variable or variables (X or collection of Xs) on the dependent variable is helpful (Y). The relationships between a group of independent variables and dependent variables are thus understood or described using it.

Regression analysis produces a mathematical equation that is frequently referred to as a regression equation.

Regression analysis can determine the relationships between the characteristics of a dataset that relate to a particular issue. We can use it, for instance, to calculate how much a car's curb weight affects its performance in terms of mileage. In a nutshell, it studies the interactions between various occurrences.

When attempting to estimate (predict) the value of a dependent variable using one or more predictors, regression is also helpful (independent variables). For instance, we will be able to forecast the use of electricity for that hour of the day based on the outdoor temperature, the time of day, and the family size.

## 4. Model validation and evaluation [1]

The model's performance is rarely a concern for most novices and professionals. The topic of the discussion is how to create a model that can be used across many domains.

A machine learning model cannot be 100% effective; otherwise, it would be considered biased, where the ideas of overfitting and underfitting are also included.

Comparing model coefficients with theory and gathering new data to validate model predictions are two methods to assess the validity of regression models.

Comparison of the outcomes to theoretical model calculations, as well as data splitting or cross-validation, uses some of the data to estimate the model coefficients and the remaining data to gauge the model's predictive power. An explanation of these techniques is provided. We find that when it is impractical to gather new data to test the model, data splitting is an efficient approach to model validation.

Justification regression of the model

From the statistics output of this study, a unit increase of shop1 causes earnings to increase by 0.09999… units. This indicates that shop1 is contributing a lot of payments to the investor.

The second business of a chemist has a positive effect on earnings. According to this study, the Introduction of chemist and rent collection units-imposed earnings to increase positively by 1.

We have a standard error which is very minimal, almost zero. This means our model was of high performance.

I decide to use the regression model over other models like moving averages.

Comparison between Moving average and regression analysis.

|  |  |
| --- | --- |
| REGRESSION MODEL | MOVING AVERAGE MODEL |
| Uses the whole sample to develop a model | Uses few samples to produce a model |
| Combines many variables at once | Develops a model from a single variable |
| Simple to use | Complex in application due to its nature |

### 5. CONCLUSION AND IMPROVEMENT

From the above analysis, I conclude the owner of the businesses under this study earns profitably.

If all other factors are kept constant, investors' earnings are contributed by three firms.

I can advise the investor to increase the number of businesses like the ones he operates.

With the help of the model, one can monitor business and avoid losses that may drag other enterprises behind.

Concerning this point, the owner of the companies under study was able to watch the business performance through regression analysis.

I advise investors to apply statistical theories and calculations to make decisions based on their previous data.

The regression model can be improved by transforming data to the required form if some assumptions are not met by the data used in the study.

### REFERENCES

1. Draper, Norman R., and Harry Smith. *Applied regression analysis*. Vol. 326. John Wiley & Sons, 1998.
2. Cameron, A. Colin, and Frank AG Windmeijer. "An R-squared measure of goodness of fit for some common nonlinear regression models." *Journal of econometrics* 77.2 (1997): 329-342.
3. Nick, Todd G. "Descriptive statistics." *Topics in biostatistics* (2007): 33-52
4. Yang, Hui. "Data preprocessing." *Pennsylvania State University: Citeseer* (2018).