A STARBUCKS ANALYSIS USING CORRELATION, REGRESSION AND AVERAGE

Ananda Reddy Y, M.Sc(Maths); M.Sc.(Stats) *1, K. Shekar, M.Sc (Stats) *2, Dhana Lakshmi, M.Sc.(Stats) *3

[*1]HOD, Dept. of statistics, Siva Sivani Degree college,Kompally,Sec'Bad-100

[*2]Lecturer in statistics, Siva Sivani Degree college,Kompally,Sec'Bad-100

[*3]Lecturer in statistics, Siva Sivani Degree college,Kompally,Sec'Bad-100

ABSTRACT

This summary provides a comprehensive overview of the coffee industry, highlighting the different kinds that contribute to the diverse global coffee culture. This research delves into the origins, unique characteristics, and cultural significance of several coffee bean varieties, offering a thorough understanding of the global coffee industry. The cultural relevance of different coffee kinds is also discussed, emphasizing how, in many countries, coffee has evolved from a beverage to a symbol of identity, rituals, and social interaction. The study takes into account the historical, social, and economic factors that have influenced coffee production and consumption across a range of civilizations, from customary coffee ceremonies to modern cafe trends. Additionally, the study discusses the most recent developments and trends in the coffee sector, such as the popularity of single-origin and specialty coffees, the effect of sustainable practices, and the effect of changing consumer tastes on the world coffee market.

INTRODUCTION

Correlation:

A statistical tool used to quantify the degree of relationship or movement between two variables is correlation. It offers information on the direction and intensity of the link between the variables. Ranges from -1 to 1. The correlation coefficient, commonly represented by "r," A direct relationship, where one variable tends to increase as the other does, is shown by a positive correlation. On the other hand, an inverse relationship—where one variable tends to decrease as the other increases—is implied by a negative correlation. Two variables may not always be related to one another, even when there is a correlation between them. This is known as correlation but not causation. Simply put, correlation quantifies the relationship between variables.

Regression Analysis:

A statistical method for simulating the relationship between a dependent variable and one or more independent variables is regression analysis. Its goal is to forecast the dependent variable's value by using the independent variables' values as a guide. The outcome is a regression equation that can be utilized for forecasting, comprehending the connections between variables, and determining the importance and strength of these

connections.While there are other forms of regression analysis as well, linear regression is the most widely used. The regression equation in linear regression assumes that the connection between the variables is linear and is expressed as a straight line.

Correlation Coefficient (r): A measure of the strength and direction of the linear relationship between two variables. It ranges from -1 to 1.

Scatterplots: Graphical representations of data points in a two-variable space, often used to visually assess the correlation between variables.

Regression Equation: The mathematical formula derived from regression analysis that expresses the relationship between the dependent and independent variables.

MEAN

A measure of central tendency, or the average of a collection of data, is referred to as a "mean". A dataset's mean is found by adding up each value and dividing the result by the total number of values in the dataset. It is a basic idea that is used to characterize the average value or center of gravity within a group of data.The arithmetic mean is the most widely used kind of mean among the several varieties.

The formula for the arithmetic mean (often simply called "the mean") is:

Mean= Sum of all values/ Number of values

The mean provides a single representative value that is useful for understanding the general trend or average behavior of a dataset. It is sensitive to extreme values, also known as outliers, which can significantly impact the mean.

OBJECTIVE

Although they have diverse uses, correlation and regression are two closely related statistical methods that are frequently combined. Measuring the magnitude and direction of a linear relationship between two variables is the main goal of

correlation. The degree to which the variables are associated numerically is indicated by the correlation coefficient.When two variables move in different directions or have no obvious linear link, correlation can be used to determine whether the relationship is positive (both variables travel in the same way) or negative (they move in opposite directions). Outliers or extreme data points that might have a disproportionate influence on the relationship between the variables can be found with the use of correlation analysis.

Making a prediction model that explains the relationship between a dependent variable and one or more independent variables is the main goal of regression analysis. It seeks to identify the curve or line that best fits the data and minimizes the discrepancy between the actual and anticipated values. When other variables are held constant, the regression coefficients in an equation show how much a change in one variable is related to a change in the dependent variable. Finding out whether a certain independent variable has a statistically significant effect on the dependent variable is made easier by regression. By evaluating the degree to which the model fits the observed data, regression models can be used to verify theories or hypotheses on the relationships between variables.

CAUSES OF STUDYING

Studying many coffee varieties explores cultural, economic, agricultural, and environmental factors in addition to flavor perception. Coffee has a wide range of flavor profiles that are determined by processing techniques, growth environments, and bean varietal. A person's taste and tasting abilities are improved by studying various coffee varieties, which promotes an awareness for the subtleties in flavor and scent. Studying many coffee varieties provides a comprehensive understanding of this beloved beverage around the world, taking into account aspects related to culture, economy, agriculture, and the environment. It serves a broad spectrum of interests, including those of

economics, environmentalists, medical professionals, and connoisseurs of culture.

METHODOLOGY

Original data that is gathered straight from the source for a particular study or analytical goal is referred to as "primary data". No one else has ever before gathered, handled, or evaluated this data. It is first-hand knowledge that is acquired using a variety of techniques, including surveys, experiments, interviews, direct measurements, and observations. Primary data collection can take a lot of time and money, particularly in terms of labor, technology, and resources. To guarantee the quality of their data, researchers must properly organize and carry out data collecting.

Data that has already been gathered, processed, and analyzed by another party for a purpose different from your current investigation or study is referred to as "secondary data." This kind of information is gathered from already-existing sources rather than being directly gathered by the researcher. A range of sources, such as scholarly journals, government publications, market research reports, and other pre-existing databases, can provide secondary data.
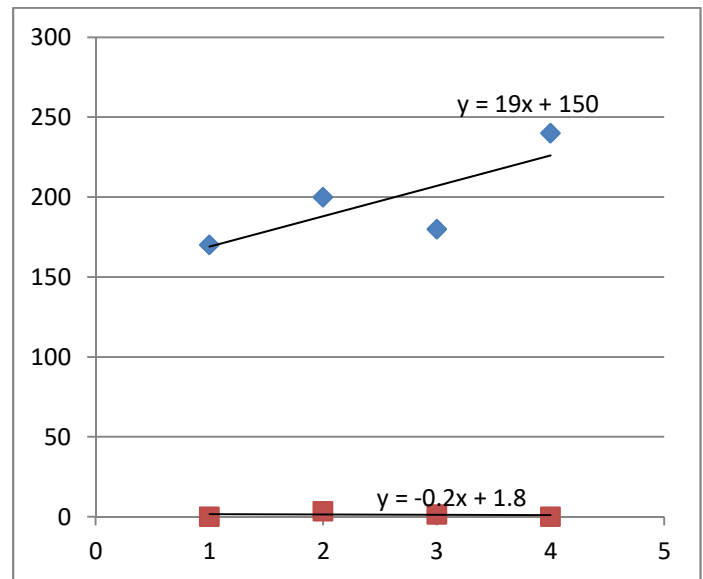
When utilizing secondary data, researchers must evaluate its quality attentively. This entails assessing the reliability of the source, the original data gathering methods, and the data's applicability to the goals of the current study.

ANALYSIS:

Scatter plot on Trans fat and Saturated Fat.
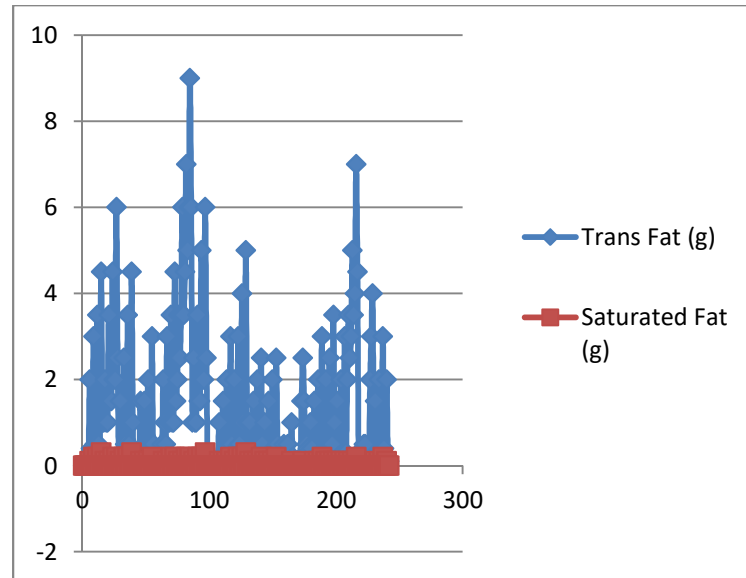
Regression line of Total Fat and Trans Fat

Regression line of Trans Fat is y= 19x+150

Regression line of Saturated Fat is y = -0.2x + 1.8

Correlation between Total Fat and Trans Fat is 0.890992

Mean of Total Carbohydrates in Classic Espresso is 104.6724

Mean of Total Carbohydrates in Signature Espresso is 138.25

Mean of Total Carbohydrates in TazoÂ® Tea is 89.90385

CONCLUSION

From the above correlation value we can conclude that there is a positive correlation between total fat and trans fat and by comparing means we can conclude that total carbohydrates are more in Signature Espresso. We can estimate Saturatd fat and Trans fat by using regression lines.

REFERENCES

"Statistics" by Robert S. Witte and John S. Witte, "Regression Analysis by Example" by Samprit Chatterjee and Ali S. Hadi

DATA SOURCE: The dataset is publically available on the kaggle website