A STUDY ON WINE QUALITY DATASET

UG MAJOR ELECTIVE PROJECT

(19MT/ME/PR45)

submitted to

DEPARTMENT OF MATHEMATICS
STELLA MARIS COLLEGE (AUTONOMOUS)
Affiliated to the University of Madras
at the Major Level in the IV Semester
for the award of the degree of

BACHELOR OF SCIENCE
IN
MATHEMATICS

BY

ANGELINE A
(20 / UMTA / 018)

Under the guidance of

Dr. Josphine Lissie. A

DEPARTMENT OF MATHEMATICS

STELLA MARIS COLLEGE (AUTONOMOUS)

JUNE 2022

CERTIFICATE

This is to certify that the dissertation in the broad area of **STATISTICS** and **DATA ANALYTICS** titled ***A STUDY ON WINE QUALITY DATASET*** is Submitted to the Stella Maris College (Autonomous), affiliated to the University of Madras, by **ANGELINE A (20 / UMTA / 018)** at the Major Elective level for the degree of **BACHELOR OF SCIENCE IN MATHEMATICS** is a bonafide dissertation work done by her during the year 2021 – 2022 under my guidance and supervision.

Dr. Jude Annie Cynthia. V
The head
Department of Mathematics
Stella Maris College (Autonomous)
Chennai – 600086

Dr. Josphine Lissie. A
Supervisor
Dean of Academic affairs
Assistant Professor
Department of Mathematics
Stella Maris College (Autonomous)
Chennai – 600086

PLACE: Chennai
DATE:

DECLARATION

I hereby declare that the dissertation entitled **"A STUDY ON WINE QUALITY DATASET"** submitted to Stella Maris College (Autonomous), affiliated to the University of Madras, at the Major Elective level for the degree of **BACHELOR OF SCIENCE IN MATHEMATICS** is the record of genuine work done by me during 2021- 2022 under the supervision of Dr. Josphine Lissie A, Assistant professor, Department of Mathematics, Stella Maris College, Chennai – 600086.

PLACE: Chennai ANGELINE A

DATE: 20 / UMTA / 018

Dr. Josphine Lissie A

Supervisor

Dean of Academic Affairs

Assistant Professor

Department of Mathematics

Stella Maris College (Autonomous)

Chennai – 600086

ACKNOWLEDGEMENT

There are many people who have supported me during my academic journey and as I complete this project, I would like to acknowledge several individuals. First and foremost, I would like to thank God, my lord and savior, for providing me with guidance, strength, and wisdom throughout this process.

I wish to record my deep sense of gratitude and profound thanks to my research guide **Dr. Josphine Lissie A**, Dean of Academic affairs, Assistant Professor, Department of Mathematics, Stella Maris College, Chennai, for her keen interest, inspiring guidance, constant encouragement with my work during all stages, to bring this project into fruition.

I am extremely indebted to **Dr. Sr. Rosy Joseph**, Principal, Stella Maris College, Chennai and **Dr. Jude Annie Cynthia**, Head of the Department, Department of Mathematics, Stella Maris College, Chennai, for giving me this great opportunity and for their constant support.

I also thank the **faculty members** of the Department of Mathematics, Stella Maris College, Chennai, for their support and for helping me out to finish this project. I would like to thank the **Department of Mathematics** of Stella Maris College, Chennai for providing me this opportunity which has helped me to shine brighter in my life.

I would like to thank **my parents and my sister** for giving me the moral support for the completion of this project. I would like to thank each and every one who took their time to fill up my survey questions.

ANGELINE A

LIST OF SYMBOLS ABBREVIATIONS

X' or $\bar{x}$	Mean
σ	Standard Deviation
σ^2	Variance
R	Co-efficient of Correlation
Z	Test statistic
S.E	Standard error
r^2	Co-efficient of determination
OBS	Observations

TABLE OF CONTENTS

LIST OF TABLES

3.29	Average quality of different Brands	83
3.30	Average Price of different brands	83
3.31	Average quality of grapes used for the preparation of Red wines	84
3.32	Average quality of grapes used for the preparation of White wines	84

CHAPTER – 1
PRELIMINARIES

1.1: Statistics

What is statistics?

Statistics is **a branch of applied mathematics that involves the collection, description, analysis, and inference of conclusions from quantitative data**. The mathematical theories behind statistics rely heavily on differential and integral calculus, linear algebra, and probability theory.

TWO MAIN METHODS OF STATISTICS

- Descriptive Statistics - which summarize data from a sample using indexes such as the mean or standard deviation
- Inferential Statistics - which draw conclusions from data that are subject to random variation (e.g., observational errors, sampling variation).

What is Statistical problems?

The statistical problems in real life consist of sampling, inferential statistics, probability, estimating, enabling a team to develop effective projects in a problem-solving frame.

For instance, car manufacturers looking to paint the cars might include a wide range of people that include supervisors, painters, paint representatives, or the same professionals to collect the data, which is necessary for the whole process and make it successful.

Statistical problems contain four components,

- Put up the Question
- Collect the data and facts by-polls, studies, and reviews, etc.
- Analyze the data with different statistical techniques.
- Please find out the result and transmit it to the public.

FUNCTIONS OR USES OF STATISTICS

Statistics can be used **to improve data quality by developing specific experimental designs and survey samples**. Statistics also provides tools for prediction and forecasting. Statistics is applicable to a wide variety of academic disciplines, including natural and social sciences as well as government and business.

What is the use of Statistics in real life?

Statistics is used for the graphical representation of the collected data. Statistics can compare information through median, mean, and mode.

Therefore, statistics concepts can easily be applied to real life, such as for calculating the time to get ready for office, how much money is required to visit work in a month, gym, diet count of a weak, in education, and much more.

Besides this, statistics can be utilized for managing daily routines so that you can work efficiently.

BASIC USES OF STATISTICS IN DAILY LIFE

Government

Weather Forecasting

Political Campaigns

Sports

Research

Eduaction

Emergency prepadness

Prediction

Quality Testing

Insurance

Consumer goods

Financial market

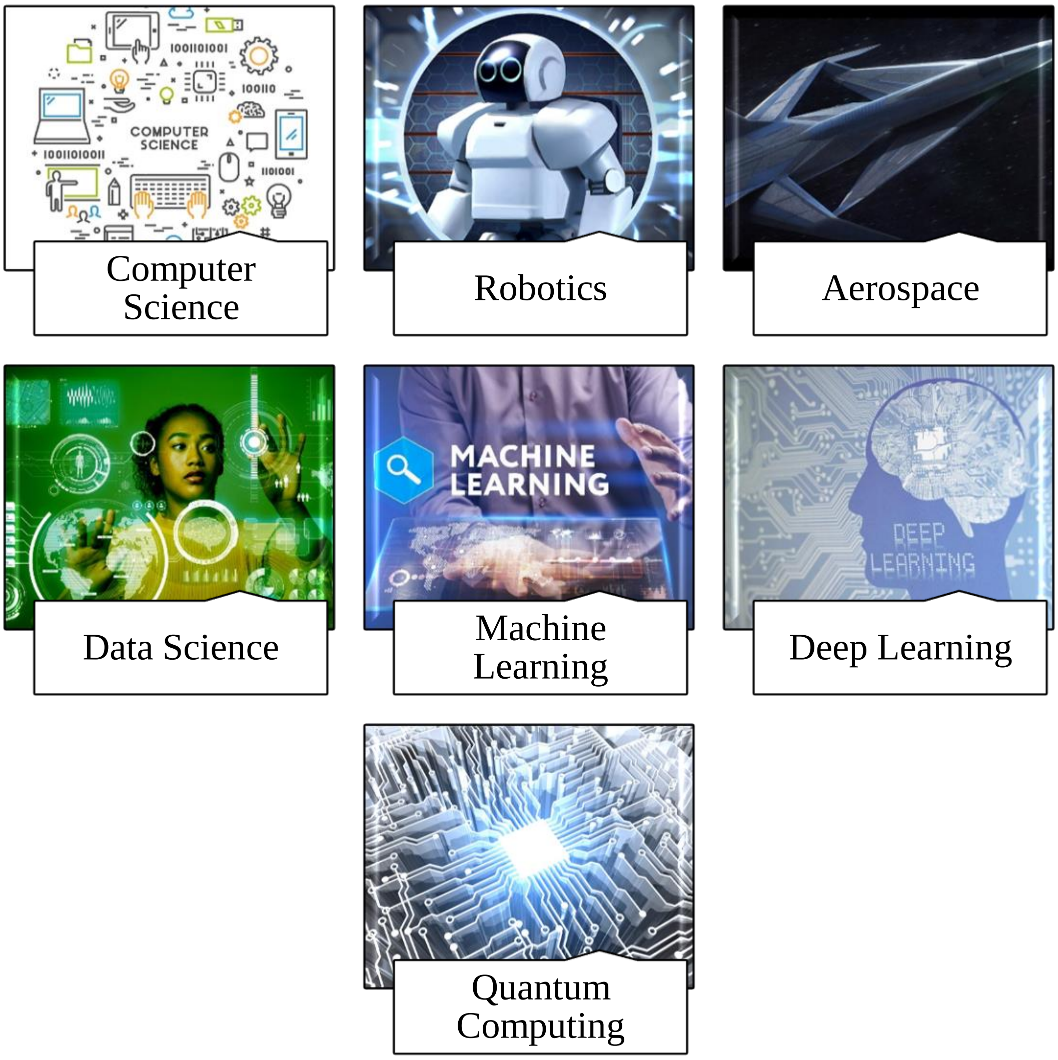
Business Statistics

Transportation

Cryptocurrency

Tourism

ADVANCED USES OF STATISTICS



1.1.1: NORMAL DISTRIBUTION

The normal distribution is a probably function that describes how the values of a variable are distributed. Normal distribution is probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence that data far from the mean.

The probability density of the normal distribution is

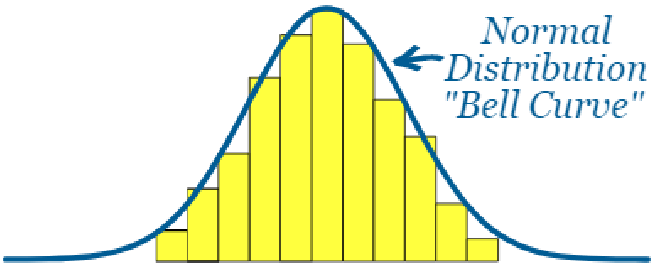
$$f\left(\frac{x}{\mu}, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where, μ – Mean

σ – Standard deviation

σ^2 – Variance

In graph form, normal distribution will appear as a bell curve.



1.1.2: CORRELATION

What is Correlation?

In statistics, correlation is **a statistic that establishes the relationship between two variables**. In other words, it is the measure of association of variables. Correlation is the numerical measurement showing the degree pf correlation between the two variables.

One variable may be called as "Subject" (Independent) and the other variable is called as "Relative" (Dependent). Relative variable is measured in terms of subject variable. Ex, Rainfall and production of agricultural products. Here Rainfall is independent and production of agricultural products is dependent on Rainfall.

Scatter Diagram

Step 1: To examine whether two variables x and y are correlated, we collect pairs of values of x and y. Let them be (x1,y1), (x_2,y_2),…….,(x_n,y_n)

Step 2: Then plot these points in a graph paper. The resulting figure is called the scatter diagram.

Step 3: From the scatter diagram, we can form the fairly good idea of correlation relationship between x and y.

If the points in the scatter diagram are dense or closely packed, it is an indication that x and y are correlated more.

On the other hand, if the points are widely scattered throughout the graph paper, we may conclude that x and y are either not correlated or poorly correlated.

Types of Correlation

1. Linear and Non-Linear Correlation
2. Positive and Negative Correlation
3. Simple and Multiple Correlation
4. Partial and Total Correlation

Linear and Non-Linear Correlation

Linear Correlation

If the points in the scatter diagram appear to be near a straight line, we may assume a linear correlation.

Non-Linear Correlation

If the points cluster round a well-defined curve other than a straight line, then it is assumed to have non-linear relationship.

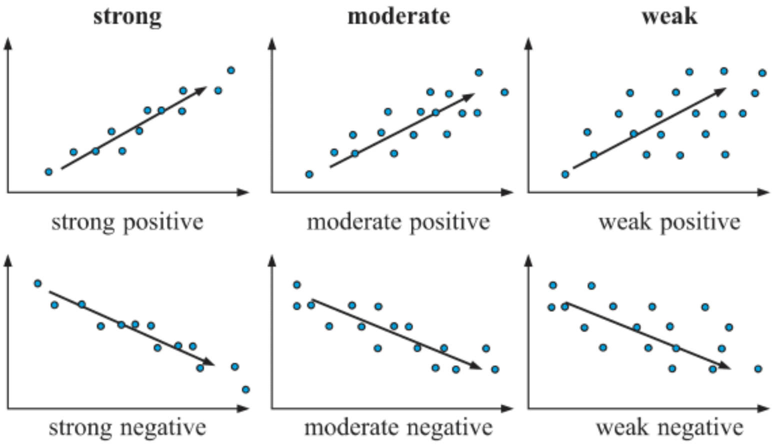
Positive and Negative Correlation

Positive Correlation

A positive correlation is the relationship between two variables in which both variables move in the same direction. Therefore, when one variable increases as the other variable decreases or one variable decreases while the other variable increases. Example: Height and Weight (i.e., Taller people tend to be heavier).

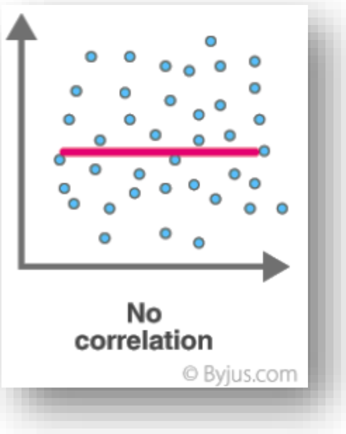
Negative Correlation

A negative correlation is the relationship between two variables in which an increase in one variable is associated with a decrease in the other. Example: Height above sea level and temperature [i.e., As you climb the mountain (increase in height), it f=gets colder (decrease in temperature).



Zero Correlation

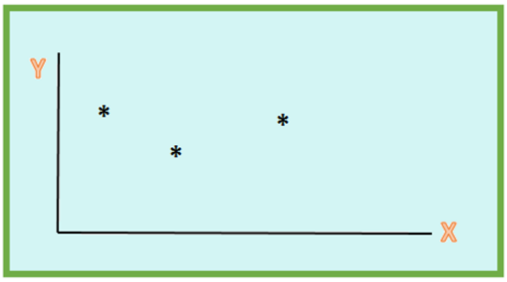
A zero correlation exist when there is no relationship between two variables. For example, there is no relationship between the amount of tea drunk and the level intelligence.



Simple and Multiple Correlation

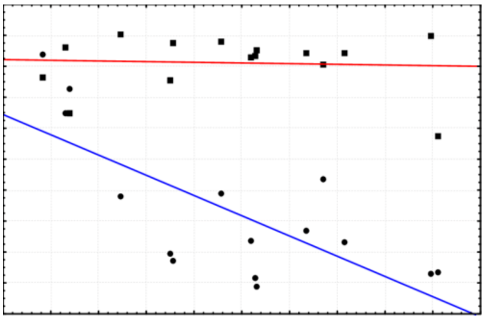
Simple Correlation

When We study the relationship between only two variables, the relationship is described as simple correlation. For example, quantity of money and price level, demand and price, etc.



Multiple Correlation

When we study the relationship between more than two variables simultaneously, the relationship is called as multiple correlation. For example, the relationship of price, demand and supply of a commodity.



Partial and Total Correlation

Partial Correlation

The study of two variables excluding some other variables is called partial correlation. For example, study price and demand, eliminating the supply .

Total Correlation

All the required facts are taken into account.

What is Correlation Coefficient?

The correlation coefficient is **the specific measure that quantifies the strength of the linear relationship between two variables in a correlation analysis**. The coefficient is what we symbolize with the r in a correlation report.

Properties of Coefficient of Correlation

Correlation coefficient is all about establishing relationships between two variables. Some properties of correlation coefficient are as follows:

1) Correlation coefficient remains in the same measurement as in which the two variables are.

2) The sign which correlations of coefficient have will always be the same as the variance.

3) The numerical value of correlation of coefficient will be in between -1 to + 1. It is known as real number value.

4) The negative value of coefficient suggests that the correlation is strong and negative. And if ‘r’ goes on approaching toward -1 then it means that the relationship is going towards the negative side.

When ‘r’ approaches to the side of + 1 then it means the relationship is strong and positive. By this we can say that if +1 is the result of the correlation then the relationship is in a positive state.

5) The weak correlation is signaled when the coefficient of correlation approaches to zero. When ‘r’ is near about zero then we can deduce that the relationship is weak.

6) Correlation coefficient can be very dicey because we cannot say that the participants are truthful or not.

The coefficient of correlation is not affected when we interchange the two variables.

7) Coefficient of correlation is a pure number without effect of any units on it. It also not get affected when we add the same number to all the values of one variable. We can multiply all the variables by the same positive number. It does not affect the correlation coefficient. As we discussed, ‘r ‘is not affected by any unit because ‘r’ is a scale invariant.

8) We use correlation for measuring the association but that does not mean we are talking about causation. By this, we simply mean that when we are correlating the two variables then it might be the possibility that the third variable may be influencing them.

KARL PEARSON’S CORRELATION CO-EFFICIENT FORMULA

The karl Pearson correlation co-efficient formula is the most widely used correlation statistics because it is easy to use and simple to understand.

The Pearson correlation co-efficient (r) is used to denote the linear relationship between two variables x and y whereby the Pearson’s value must be between -1 and 1.

If the Pearson’s is negative, then the relationship is also negative and if it is positive, the relationship is positive.

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}$$

Thus, for calculating the co-efficient of correlation between two variables x and y

(1)

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2}\sqrt{(Y-\overline{Y})^2}}$$

Where, $\overline{X}$ = mean of X variable

$\overline{Y}$ = mean of Y variable

This formula is used when deviations are measured from their means.

(2)

Assumed Mean Method:

$$d_x = X - A$$

$$d_y = Y - A$$

$$r = \frac{N\sum d_x d_y - (\sum d_x)(\sum d_y)}{\sqrt{N\sum d_x{}^2 - (\sum d_x)^2} - \sqrt{N\sum d_y{}^2 - (\sum d_y)^2}}$$

USES OF CORRELATION AND CORRELATION CO-EFFICIENT

1) It is used in deriving the degree and direction of relationship within the variables.
2) It is used in developing the concept of Regression.
3) It is used in research work.
4) It is used to study the strength of a relationship
5) In the field of science, it is used to make conclusions.
6) It is used to study the direction of a relationship.

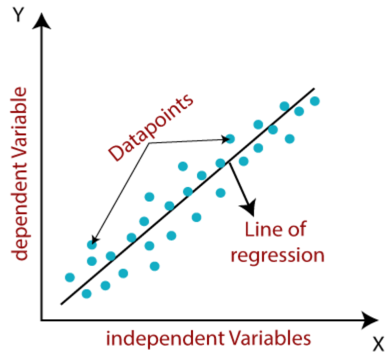
1.1.3: REGRESSION

What is Regression?

The statistical method which helps us to estimate the unknown value of one variable from the known value of the related variable is called regression.

What is line of Regression?

The line described in the average relationship between two variables is known as the line of regression.



USES OF REGRESSION ANALYSIS

1) Regression analysis is used in statistics in all those fields where two or more variables for having the tendency to go back to the average. It is widely used in social science like economics, natural and physical sciences.
2) Regression analysis predicts the value of the dependent variables from the values of independent variables.
3) It can calculate co-efficient of correlation(r), and co-efficient of determination(r^2) with the help of regression coefficients.

4) Regression analysis in statistical estimation of demand curves, supply curves, production function, cost function, consumption function, etc. can be predicted.

SIMPLE AND MULTIPLE REGRESSION

SIMPLE REGRESSION

The regression analysis confined to the study of only two variables at a time is termed as simple regression.

MULTIPLE REGRESSION

The regression analysis for studying more than two variables at a time is known as multiple regression.

PROPERTIES OF REGRESSION ANALYSIS

1) It indicates the cause-and-effect relationship between the variables and establishes a functional relationship.
2) If x is a random variable, then y is a fixed variable and vice versa.
3) Regression co-efficient is an absolute figure if we know the value of independent variable, we can find the value of the dependent variable.
4) The regression co-efficient explains that the degrees in one variable is associated with the increase in the other variable.

METHODS OF STUDYING REGRESSION

There are two methods, namely

1) Graphic method
2) Algebraic method

GRAPHIC METHOD

The points are plotted in a graph paper presenting pairs of values of concerned variables. It involves drawing a scatter diagram with independent variable on X-axis and dependent variable on Y-axis. After that a line is drawn in such a manner that it passes through most of the distribution, with remaining points distributed almost evenly on either side of the line.

A regression line is known as the line of best fit that summarizes the general moment of data. It shows the best mean values of the other. The regression line is based on the criteria that it is a straight line that minimizes the sum of squared deviations between the predicted and observed values of the dependent values.

ALGEBRAIC METHOD

It indicates the best probable mean value of one variable corresponding to the mean value of the other. Since a regression line is the line of best fit, it cannot be used conversely; therefore, there are always two regression lines constructed for the relationship between two variables x and y.

The algebraic method refers to various methods of solving a pair of linear equations including graphing, substitution and elimination. Algebraic method develops two regression equations of X on Y and Y on X.

REGRESSION EQUATION OF Y ON X

$$\mathbf{Y = a + bX}$$

REGRESSION EQUATION OF X ON Y

X = a + bY

These above regression equation is solved by the using normal equations for finding a and b.

For X and Y, the normal equations are,

$\sum x = Na + b\sum y$

$\sum xy = a \sum y + b \sum y^2$

For Y on X, the normal equations are,

$\sum y = Na + b \sum x$

$\sum xy = a \sum x + b \sum x^2$

The regression line gives the best average value of one variable for any given value of the other variable. The regression line of X on Y gives the most probable values of X for any given value of Y. In the same manner, the regression line of Y on X gives the most probable values of Y for any given value of X. Thus, there will be two regression lines in case of two variables.

When there is perfectly positive correlation (+1) or perfectly negative correlation (-1), the two regression lines will coincide with each other, i.e., there will be only one line. If the regression lines are nearer to each other, then there is higher degree of correlation. If the two lines are farther away from each other, then there is lesser degree of correlation. If r = 0, both variables are independent. Both will cut each other at right angle.

DEVIATIONS TAKEN FROM THE ARITHMETIC MEAN OF X AND Y

REGRESSION EQUATION OF X ON Y

$(X - X') = r \frac{\sigma x}{\sigma y} (Y - Y')$

Where, $b_{xy} = r \frac{\sigma x}{\sigma y}$ is the regression co-efficient of X on Y.

REGRESSION EQUATION OF Y ON X

$(Y - Y') = r \frac{\sigma y}{\sigma x} (X - X')$

Where, $b_{yx} = r \frac{\sigma y}{\sigma x}$ is the regression co-efficient of Y on X.

$r = \pm\sqrt{b_{xy} * b_{yx}}$

where r is the co-efficient of correlation.

FINDING REGRESSION EQUATIONS FROM DEVIATIONS TAKEN FROM THE ASSUMED MEAN

REGRESSION EQUATION OF X ON Y

$(X - X') = r\, b_{xy} (Y - Y')$

Where, $b_{xy} = \frac{N\sum dxdy - \sum dx \sum dy}{N\sum dy^2 - (\sum dy)^2}$

REGRESSION EQUATION OF Y ON X

$(Y - Y') = r\, b_{yx} (X - X')$

Where, $b_{yx} = \frac{N\sum dxdy - \sum dx \sum dy}{N\sum dx^2 - (\sum dx)^2}$

dx = x – A, where A is the assumed mean from X values.

dy = y – B, where B is the assumed mean from Y values.

MATHEMATICAL PROPERTIES OF REGRESSION CO-EFFICIENT

1. The geometric mean between regression co-efficient is the co-efficient of correlation.

   PROOF:

   We know that,

   $$b_{xy} = r\frac{\sigma x}{\sigma y} \text{ and } b_{yx} = r\frac{\sigma_y}{\sigma_x}$$

   $$b_{xy} \cdot b_{yx} = r\frac{\sigma_x}{\sigma_y} \cdot r\frac{\sigma_y}{\sigma_x} = r^2 \quad \Rightarrow \quad r = \sqrt{b_{xy} b_{yx}}$$

   where r is the coefficient of correlation

   Note- If b_{xy} is negative, b_{yx} will also be negative. Both regression coefficients will have the same algebraic sign. If b_{xy} and b_{yx} are positive, r will be positive and vice versa.

2. Arithmetic mean of b_{xy} and b_{yx} is equal to or greater that r, i.e.,

   $$\frac{b_{xy} + b_{yx}}{2} \geq r$$

3. Regression coefficients are independent of change of origin but not of scale.

1.1.4: Residual normality

Normality is the assumption that the underlying residuals are normally distributed, or approximately so. While a residual plot, or normal plot of the residuals can identify non-normality, you can formally test the hypothesis using the Shapiro-Wilk or similar test.

Why residual normality is important?

The basic assumption of regression model is normality of residual. If your residuals are not normal then there may be problem with the model fit, stability and reliability. In order to generalize a regression model beyond the sample, it is necessary to check some of the assumptions of regression residuals.

The null hypothesis states that the residuals are normally distributed, against the alternative hypothesis that they are not normally-distributed. If the test p-value is less than the predefined significance level, you can reject the null hypothesis and conclude the residuals are not from a normal distribution. If the p-value is greater than the predefined significance level, you cannot reject the null hypothesis.

1.1.5: Outliers

An outlier is an observation that lies an abnormal distance from other values in a random sample from a population. In a sense, this definition leaves it up to the analyst (or a consensus process) to decide what will be considered abnormal.

1.1.6: SAMPLING THEORY AND TEST OF SIGNIFICANCE

INTRODUCTION

1. The larger group from which the sample is drawn is called the universe or population.
2. From the population, a part which is representative of the whole is selected. This part is called a sample and the process of selection is known as sampling.
3. Statistical samples play a vital role in almost all statistical studies based on which decisions for future actions are to be made. Sometimes, the population is infinite. Even if the population is finite, study through samples is easier.
4. Sampling theory is a study of the relationship between a population and a sample drawn from it and the estimation of population parameters. Thus, statistical inference is divided into-
    - Estimation theory
    - Testing of hypothesis (relationship between the parameters of the population and the sample.)

TESTING OF HYPOTHESIS

Hypothesis is an assumption which may or may not be true about a population parameter.

LAYING DOWN OF HYPOTHESIS

To verify our assumption, we collect data and find out the difference between sample values and population. If no difference or if the difference is small, then the hypothesized value is correct. Generally, two hypotheses must be constructed if one of them is correct, the other one is rejected.

a. Null Hypothesis (H_0) - The hypothesis that we hence assumed is said to be a null hypothesis, i.e., it means that the true difference between the measure of sample and population is nil.
b. Alternative Hypothesis (H_1) - Rejection of H_0 leads to acceptance of alternative hypothesis H_1.
TABLE 1.1:

	DECISION FROM SAMPLES	
	ACCEPT H_0	**ACCEPT H_1**
H_0 TRUE	CORRECT	WRONG (TYPE I ERROR)
H_1 FALSE	WRONG (TYPE II ERROR)	CORRECT

1. What is type I error
   Rejection of H_0 when true is said to be a type I error.
2. What is type II error?
   Acceptance of H_1 when false is said to be a type II error.

Level of Significance (LOS)

The maximum probability of committing a type I error is known as level of significance. Generally, 5% LOS is fixed in statistical tests. This implies that we can have 95% confidence in accepting a hypothesis or it could be wrong 5%.

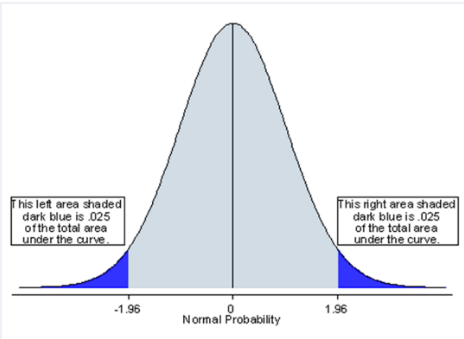
Critical Region

The range of variation has two regions namely acceptance region and critical region - region of rejection. If the sample statistic falls in the critical region, we have to reject the hypothesis as it leads to false decisions. So, we go for H_1 if the computed value of sample statistic falls in the rejecting region.

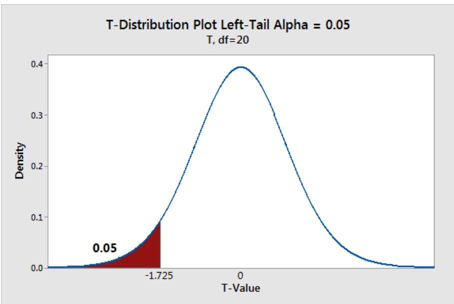
One Tailed and Two Tailed Tests

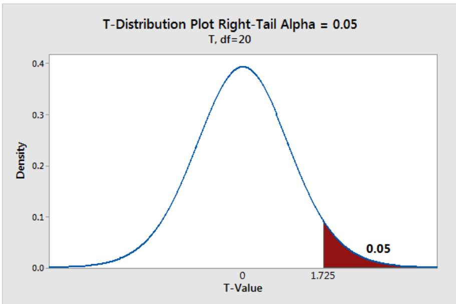
The critical region under a normal curve can be divided into two ways -

1. Two sides under a curve



2. One side under a curve and both are either at the right tail or the left tail.





Making a Decision or Conclusion

The decision will be taken on the basis of the computed value; whether it lies in the accepted or rejected region.

Standard Error

When the average amount of variability of observations of a population is computed, it is called standard deviation. However, when the average amount of variability of the observations of a sample distribution is computed, it is known as the standard error.

Utility of standard error

1. It is useful in testing the hypothesis. We may test the hypothesis at 5% LOS, which means if the difference between the observed and expected is more than

1.96 × S.E (standard error), the hypothesis is not accepted and the alternative hypothesis must be considered instead. The LOS can be 1%. Generally, the hypothesis is accepted if the difference is less than 3 S.E, 5% LOS is popular.
2. Reliability of a sample can be known.
3. The value of the parameters can be determined along with the limits.

Various Tests of Significance

Various tests of significance to be applied in various situations are discussed under the following,
1. Tests of significance for attributes
2. Tests of significance for variables (large samples)
3. Tests of significance for variables (small samples)

Tests of Significance for Large Samples

If n >30, then those samples may be recorded as large samples. The following are the assumptions made for large samples -
1. The random sampling distribution of statistics is approximately normal.
2. Sampling values are sufficiently close to the population values and can be used for the calculation of standard error of estimates.

STANDARD ERROR OF MEAN

$$S.E(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

i. When standard deviation of population is known, then $\sigma = \text{S.D of population}$
ii. When standard deviation of population is not known, then $\sigma = \text{S.D of sample}$

TESTING THE DIFFERENCE BETWEEN THE MEANS OF TWO SAMPLES

1. If two independent random samples with n_1 and n_2 respectively are drawn from the same population of standard deviation σ, the standard error of the difference between the sample means is given by -

$$S.E(\overline{X_1} - \overline{X_2}) = \sqrt{\sigma^2(\frac{1}{n_1} + \frac{1}{n_2})}$$

where, σ^2 – Standard deviation of the population.

2. If two random samples with $\overline{X_1}$, σ_1, n_1 and $\overline{X_2}$, σ_2, n_2 respectively are drawn from different populations, then the standard error of the difference between the means is given by,

$$S.E(\overline{X_1} - \overline{X_2}) = \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$$

3. If σ_1 and σ_2 are unknown, the standard error of the difference between the means is

$$S.E(\overline{X_1} - \overline{X_2}) = \sqrt{\frac{{S_1}^2}{n_1} + \frac{{S_2}^2}{n_2}}$$

Where S_1, S_2 represent the standard deviations of the two samples.

STANDARD ERROR OF THE DIFFERENCE BETWEEN TWO STANDARD DEVIATIONS

In case of two large random samples, each drawn from a normally distributed population, the standard error of the difference between the standard deviations is given by,

$$\text{S.E}(\sigma_1 - \sigma_2) = \sqrt{\frac{{\sigma_1}^2}{2n_1} + \frac{{\sigma_2}^2}{2n_2}}$$

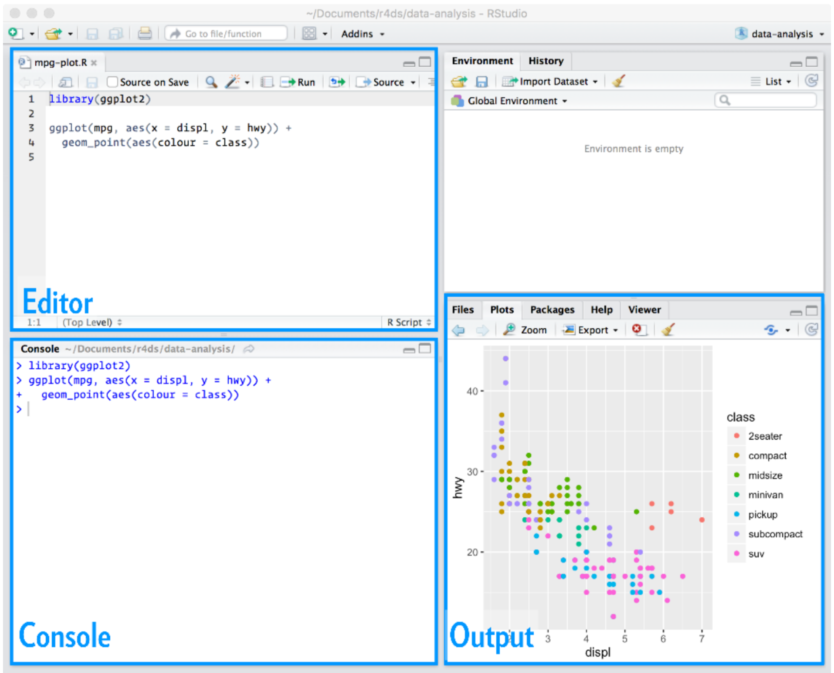
Where, σ_1 – standard deviation of the 1st sample
σ_2 – Standard deviation of the 2nd sample

1.2: R PROGRAMMING LANGUAGE

R is a software adapted by statisticians as a standard software package for data analysis. R has four features namely,

- **R-Console**: Using R console, analysts can write codes for running the data, and also view the output codes later, the codes can be written using R script.
- **R-Script**: R script is the interface where analysts can write codes, the process is quite simple, users just have to write the codes and run it.
- **R Environment**: R environment is the space to add external factors, this involves adding the actual data set, then adding variables, vectors and functions to run the data.
- **Graphical Output**: Once all the scripts and codes are added and data sets and variables are added to R, graphical output feature could be used to create graphs.

For working in R, First the file has been set as a working directory and then used for analysis purposes.



CHAPTER – 2

INTRODUCTION – WINE AND ITS PRODUCTION REGION

2.1: Introduction

My project is about determining the wine quality of the wine corresponding to their constituents present in the wine. The wine quality dataset comprises information about the wines that are cultivated in the specific region called vinho Verde in Portuguese. Interestingly, Vinho Verde (Portuguese for 'green wine') is neither a grape nor a style of wine. It is a Portuguese wine region that produces delicious whites, reds, and roses.

Vinho Verde wine has soared in popularity in the past decade due to its irresistible fruity flavor, festive fizziness, and freshness.

2.2: The Vinho Verde Wine Region

Vinho Verde is located in northwest Portugal and is the biggest Portuguese wine region. The Vinho Verde region starts at the Minho River, near the border with Spain, and ends at the Douro River in the south.

The Vinho Verde wine region has nine subregions, named after towns or rivers:

- Monção and Melgaço
- Lima
- Cávado
- Ave
- Basto
- Amarante
- Baião
- Sousa
- Paiva

The white, red, and rose wine blends made in Vinho Verde are released very young, just a few months after harvesting.

Some say that "Verde" refers to the green grapes which ripen late due to the cool weather.

2.3: A Brief History of Vinho Verde

When the Romans arrived in Portugal around 2000 years ago, people were already making fermented drinks. Wine in Portugal is part of daily life, and many families have a small plot of land where they grow grapes alongside vegetables and citrus trees. All over the Vinho Verde region, you'll see grapevines hung in the beautiful Pergola style, draped high above where the cool breeze protects them from moisture and mold, arranged in a square around a family or neighborhood garden.

In recent decades, the European Union has funded the modernization of Portugal's vineyards, so the commercial producers have upgraded their trellising systems. But the traditional style of grape-growing is in place in small towns all over Portugal. And here's a fun fact: while many producers do grow the white grapes that make most of the Vinho Verde we drink Stateside, it's red wine that most Portuguese drink at home. Traditionally, they would drink it in ceramic bowls, and it's considered an essential accompaniment to any meal. The red Vinho Verdes are difficult to find abroad, but if you travel to Portugal be sure to try them. They are often made from the grape Vinhão, which has low alcohol, medium tannins, and an inky texture with some sour flavors; you might love it or hate it, but it's worth a taste for sure.

2.4: Grapes Grown in Vinho Verde

There are many grape varieties used in the Vinho Verde blends, and some are used for single varietal wines:

2.4.1: White Grapes of Vinho Verde

These are some of the popular white grapes used in the Vinho Verde wines:

1. ***Alvarinho*** (called ***Albariño*** in Spain): This grape is found chiefly in the subregions of Monção and Melgaço, between Lima and the Spanish border in Northern Minho. The grape is planted in other subregions as well.
2. ***Arinto (Pedernã)***: Known for its high acidity, this grape variety is found in the warmer areas of *Basto* and *Amarante.*
3. ***Avesso***: Common in the subregions of *Amarante, Baião, Sousa, and Paiva*, these fragrant grapes are not as acidic as other grapes in Vinho Verde.
4. ***Azal Branco***: This is the second most planted grape variety in the Vinho Verde region. However, *Azal Branco* is not used in single-varietal wines and is often blended with *Alvarinho* and *Avesso* grapes. It's found in the areas of *Amarante, Basto, Baião,* and *Sousa*.
5. ***Batoca***: Found in the eastern region of *Basto*, this grape was traditionally grown to lend a distinct smoothness to the Vinho Verde wine blends.
6. ***Loureiro***: This grape variety is popular in the coastal areas of *Ave, Cavado, Lima*, and *Sousa*. The winemaker *Anselmo Mendes* creates outstanding single-varietal *Loureiro* wines from these areas.
7. ***Trajadura***: This grape is mainly planted in the Minho region. It has large berries that grow in clusters.

2.4.2: Red Grapes of Vinho Verde

Here we have the red grapes used to make red Vinho Verde wine. Most of these grapes are used to make Rose wines, too.

- ***Padeiro***: Also known as *Tinta Cao*, this rare grape is used in different red wine blends. Some winemakers use *Padeiro* to make single-variety Rose wines.
- ***Alvarelhão***: The popularity of the *Alvarelhao* grape sagged in the 20th century. But, in recent years, the grape experienced a resurgence and is now used in red Vinho Verde wines.
- ***Espadeiro***: The acidity of this red grape makes it ideal for both red and rose wines.
- ***Vinhao***: This is the most planted red grape in the Vinho Verde region. The *Vinhao* grape is included in many different blends due to its high acidity and firm tannins.
- Other red grape varieties are: *Amaral (Azal Tinto), Borraçal, Brancelho, Pedral*, and *Rabo de Ovelha.*

2.5: Vinho Verde Viticulture and Winemaking

The Vinho Verde region in Minho has 15,500 hectares of vineyards. The terroir consists of granitic soil and a cool, rainy climate perfect for vine growth. Wine producers in Vinho Verde grow their vines high above the ground (on trees, telephone poles, fences, wires, and more). They do this to allow better sun exposure, to prevent humidity and fungal diseases, and to free up space to grow other crops.

Northern Portugal is a very humid area, and some wineries have struggled to implement organic farming practices. However, some winemakers have successfully achieved this, and others have even become entirely biodynamic. Organic vineyards use natural fertilizers, while biodynamic vineyards have a more integrated farming system — for example, they use local herbs to nourish each vine.

2.5.1: The Vinho Verde Wine Making Process

The vinification process of Vinho Verde wines is relatively quick. The wine is bottled soon after fermentation. This allows malolactic fermentation to occur in the bottle. The released carbon dioxide is trapped in the bottle and dissolves in the wine, creating the distinct Vinho Verde fizzy mouthfeel. Some winemakers may also add carbon dioxide artificially, and others completely omit adding carbonation.

2.6: The story behind the spritz

Many of the Vinho Verde wines out there have a light fizz that can be extremely refreshing. Originally, this occurred when carbon dioxide, which is a natural byproduct of fermentation, was trapped alongside the freshly fermented wine during bottling. But these days, winemakers add it through a boost of carbon dioxide, because they know that drinkers abroad have come to associate Vinho Verde with light bubbles. Because of this, the wine has become known as a sort of “soda pop for adults.” Which is totally fine, if that’s what you’re after in a wine. There’s always a time and place for “cheap and cheerful.”

But many Vinho Verde wines don’t have any spritz, at all, and if you try them, you’ll find that the acidity and minerality of the wine shines through even more when those bubbles aren’t present. If you see Vinho Verde in a Burgundy-shaped bottle (the rounder bottom, as opposed to the thin Bordeaux-shaped one), this is an indicator that it won’t have spritz; as well, you can ask your retailer to recommend a small-production Vinho Verde and it also will not have spritzed, as it is mainly practiced by the large, cooperative bottlers who export en masse.

2.7: Single- Variety Vinho Verde is killer, and it is age-worthy

Most Vinho Verde is a blend of white grapes, all indigenous to Portugal, but there are two predominant grapes that winemakers are starting to see as more interesting than the others: Alvarinho and Loureiro. You may be familiar with Albarinho, a different spelling but essentially still the same grape, from northern Spain. In Spain, Albarinho tends to be somewhat rounder and softer than its Portuguese counterpart. Alvarinho displays tropical aromas and an overall lemony character and is high in acidity; Loureiro is more floral, and also acidic.

Many winemakers in the Vinho Verde region have begun making single-varietal Alvarinho and Loureiro, with very good results. It turns out that these wines age very well, and respond nicely to oak aging, developing complexity and character. But it’s very hard to find these wines on the market. Here’s why: the image of Vinho Verde as a young wine has caused importers to pressure Portuguese producers to deliver wines as soon as they are bottled, in early spring. They don’t really get the chance to hold their wines back. Perhaps this will change in coming years, but in the meantime, if you visit Portugal, be sure to look for aged Alvarinho and Loureiro wines.

2.8: Flavor Profiles of Vinho Verde Wines

Vinho Verde wines are light with 10-12% ABV, natural acidity, and a gentle effervescence. Many have likened drinking a Vinho Verde to sipping on a glass of sparkling wine. However, the Vinhos Verdes have lower carbon dioxide pressure compared to Prosecco or Champagne.

- The white wines have a lemon hue and a range of tropical fruit and floral notes, depending on the grape variety used.
- The red wine has a ruby hue and firm tannins, while the roses are refreshing and fruity.

2.9: The Best Food to Try with Vinho Verde Wines

Vinho Verde wine is a popular summer-time brunch drink in Portugal and around the world. It goes well with most types of food. Pair the wine with light dishes like salads and chicken or more filling meals like potato and pork dishes.

The white Vinho Verde also goes great with seafood dishes, which are very popular in Portuguese cuisine. It pairs great with cod, monkfish, and shrimp served with creamy sauces or rice on the side.

2.10: Compared to other wines, Vinho is a Bargain

The average bottle of Vinho Verde, from any one of the main producers, will be around $10 at your local retailer. It’s a wine for drinking anytime. And if you want to try one of the biodynamic or single-varietal Vinho Verde wines, it still won’t cost you much more than $18-20. Why so cheap? Portugal’s economy has suffered from the global downturn more than any other Western European nation, so land and labor prices are relatively low. But on a positive note, all of the recent investment in the wine industry, and the rise in Vinho Verde’s popularity abroad, could provide a boost to the nation, providing jobs and improving prospects for the future.

2.11: DATASET

We took the wine quality dataset representing the whole population of wines as Vinho Verde is one of the best of wines around the world. It’s because of it’s quality and cheap rate. The sample dataset that I have drawn from the huge population containing 1600 information about the wine quality of both red and white wine is included below with which we have performed certain statistical analyses like correlation, regression, testing of hypothesis, etc.

I collected my wine quality dataset from my cousin, a data analyst Mr. Christopherpraveen William who is currently working as a Senior data analyst in Irish Bank, Dublin, Ireland. It’s a secondary data in this case since I have collected it from a 2nd person.

TABULAR COLUMN OF WINE QUALITY DATASET

Type	Name	Year of Manufacturing	Price	Fixed Acidity	Volatile Acidity	Citric Acid	Residual Sugar	Chlorides	Free Sulfur dioxide	Total Sulfur dioxide	Density	pH	Sulphates	Alcohol	Quality	Name and Year of Manufacturing
Red	Quinta do Vale Meao Tinto	2000	₹ 33,521.00	7.4	0.6	0.26	7.3	0.07	36	121	0.9982	3.37	0.49	9.4	7	Quinta do Vale Meao Tinto - 2000
Red	Zafirah Vinha da Rocha Tinto	2019	₹ 2,859.83	10.4	0.44	0.73	6.55	0.074	38	76	0.999	3.17	0.85	12	7	Zafirah Vinha da Rocha Tinto - 2019
Red	Aphros 'Yakkos' Grande Reserva Espumante Tinto	2006	₹ 2,283.75	10.5	0.51	0.64	2.4	0.107	6	15	0.9973	3.09	0.66	11.8	7	Aphros 'Yakkos' Grande Reserva Espumante Tinto - 2006
Red	Anselmo Mendes Pardusco Private Tinto	2018	₹ 2,129.05	12.9	0.35	0.49	5.8	0.066	5	35	1.0014	3.2	0.66	12	7	Anselmo Mendes Pardusco Private Tinto - 2018
Red	Antonio Pereira Tinto Bom	2016	₹ 1,799.50	7.7	0.915	0.12	2.2	0.143	7	23	0.9964	3.35	0.65	10.2	7	Antonio Pereira Tinto Bom - 2016
Red	Zafirah Graciosidade Graciousness Tinto	2018	₹ 1,539.52	8.4	0.25	0.39	2	0.041	4	10	0.99386	3.27	0.71	12.5	7	Zafirah Graciosidade Graciousness Tinto - 2018
Red	Marcio Lopes Pequenos Rebentos Tinto Atlantico	2020	₹ 1,386.01	6.7	0.28	0.28	2.4	0.012	36	100	0.99064	3.26	0.39	11.7	6	Marcio Lopes Pequenos Rebentos Tinto Atlantico - 2020
Red	Via Latina Espumante Tinto Bruto	2012	₹ 1,217.57	6.8	0.59	0.06	6	0.06	11	18	0.9962	3.41	0.59	10.8	6	Via Latina Espumante Tinto Bruto - 2012
Red	Adega de Moncao Tinto	2020	₹ 1,029.74	8.2	0.31	0.4	2.2	0.058	6	10	0.99536	3.31	0.68	11.2	6	Adega de Moncao Tinto - 2020
Red	Adega Ponte da Barca 'Este' Tinto	2016	₹ 909.96	7.7	0.58	0.01	1.8	0.088	12	18	0.99568	3.32	0.56	10.5	6	Adega Ponte da Barca 'Este' Tinto - 2016
Red	Casa da Tojeira Espumante Tinto Reserva Bruto	2020	₹ 905.10	7.1	0.59	0	2.1	0.091	9	14	0.99488	3.42	0.55	11.5	6	Casa da Tojeira Espumante Tinto Reserva Bruto - 2020
Red	Adega Cooperativa Ponte de Lima Tinto	2015	₹ 762.40	9.1	0.3	0.41	2	0.068	10	24	0.99523	3.27	0.85	11.7	6	Adega Cooperativa Ponte de Lima Tinto - 2015
Red	Casal Gracia Tinto	2018	₹ 705.00	5.1	0.51	0.18	2.1	0.042	16	101	0.9924	3.46	0.87	12.9	6	Casal Gracia Tinto - 2018
Red	Turra Tinto	2018	₹ 702.00	8.2	0.33	0.39	2.5	0.074	29	48	0.99528	3.32	0.88	12.4	5	Turra Tinto - 2018
Red	Casa Santa Eulalia 'Plainas' Tinto	2018	₹ 702.00	7.1	0.66	0	2.4	0.052	6	11	0.99318	3.35	0.66	12.7	5	Casa Santa Eulalia 'Plainas' Tinto - 2018
Red	Quinta de Santa Cristina Tinto	2018	₹ 601.00	5.1	0.42	0	1.8	0.044	18	88	0.99157	3.68	0.73	13.6	5	Quinta de Santa Cristina Tinto - 2018
Red	Quinta do Outeiro de Baixo Tinto Reserva	2019	₹ 568.73	6.1	0.4	0.16	1.8	0.069	11	25	0.9955	3.42	0.74	10.1	5	Quinta do Outeiro de Baixo Tinto Reserva - 2019
Red	Terras de Felgueiras Espumante Tinto Bruto	2015	₹ 503.00	6.6	0.56	0.14	2.4	0.064	13	29	0.99397	3.42	0.62	11.7	5	Terras de Felgueiras Espumante Tinto Bruto - 2015
Red	Casa Santa Eulalia 'Plainas' Tinto	2020	₹ 445.51	9.4	0.3	0.56	2.8	0.08	6	17	0.9964	3.15	0.92	11.7	5	Casa Santa Eulalia 'Plainas' Tinto - 2020
Red	Adega Ponte da Barca Tinto	2020	₹ 403.15	7.2	0.33	0.33	1.7	0.061	3	13	0.996	3.23	1.1	10	5	Adega Ponte da Barca Tinto - 2020
Red	Casa da Tojeira 'Portal da Tojeira' Tinto	2017	₹ 347.25	7.3	0.65	0	1.2	0.065	15	21	0.9946	3.39	0.47	10	5	Casa da Tojeira 'Portal da Tojeira' Tinto - 2017
Red	Monte Baixo Vinho Verde Tinto	2020	₹ 335.12	5.4	0.835	0.08	1.2	0.046	13	93	0.9924	3.57	0.85	13	5	Monte Baixo Vinho Verde Tinto - 2020
Red	Caves Campelo Tinto	2010	₹ 233.41	11	0.3	0.58	2.1	0.054	7	19	0.998	3.31	0.88	10.5	5	Caves Campelo Tinto - 2010
Red	Tr3s Castas Tinto	2020	₹ 214.71	15	0.21	0.44	2.2	0.075	10	24	1.00005	3.07	0.84	9.2	5	Tr3s Castas Tinto - 2020
Red	Casal de Ventozela 'Vento'Z' Espadeiro Rose	2020	₹ 583.71	6.4	0.29	0.57	1	0.06	15	120	0.9924	3.06	0.41	9.5	6	Casal de Ventozela 'Vento'Z' Espadeiro Rose - 2020
Red	Casa da Bouca Espadeiro Rose	2020	₹ 1,011.00	6.4	0.57	0.12	2.3	0.12	25	36	0.99519	3.47	0.71	11.3	5	Casa da Bouca Espadeiro Rose - 2020
Red	Terras de Felgueiras Espadeiro Rose	2020	₹ 503.10	7.3	0.48	0.32	2.1	0.062	31	54	0.99728	3.3	0.65	10	7	Terras de Felgueiras Espadeiro Rose - 2020
Red	Espadeiro Vinha Ingleses	2020	₹ 503.00	5.3	0.47	0.11	2.2	0.048	16	89	0.99182	3.54	0.88	13.6	5	Espadeiro Vinha Ingleses - 2020
Red	Aveleda Parcela do Roseiral	2018	₹ 1,883.88	7.3	0.24	0.39	17.95	0.057	45	149	0.9999	3.21	0.36	8.6	7	Aveleda Parcela do Roseiral - 2018
Red	Quinta da Calcada Edicao Rose	2020	₹ 1,147.85	6.1	0.27	0.44	6.7	0.041	61	230	0.99505	3.12	0.4	8.9	7	Quinta da Calcada Edicao Rose - 2020
Red	Casa de Vila Verde 'Pluma' Rose	2021	₹ 2,982.65	8.9	0.4	0.51	2.6	0.052	13	27	0.995	3.32	0.9	13.4	6	Casa de Vila Verde 'Pluma' Rose - 2021
Red	Casa Ferreirinha Quinta da Leda	2019	₹ 2,760.40	10.5	0.24	0.42	1.8	0.077	6	22	0.9976	3.21	1.05	10.8	6	Casa Ferreirinha Quinta da Leda - 2019
Red	Quinta do Crasto Vinhas Velhas - Old Vines Reserva	2012	₹ 2,712.21	10.2	0.49	0.63	2.9	0.072	10	26	0.9968	3.16	0.78	12.5	6	Quinta do Crasto Vinhas Velhas - Old Vines Reserva - 2012
Red	Quinta de Soalheiro Bruto Rose	2020	₹ 1,525.56	8.2	0.28	0.4	2.4	0.052	4	10	0.99356	3.33	0.7	12.8	6	Quinta de Soalheiro Bruto Rose - 2020
Red	Adega de Moncao 'Muralhas de Moncao' Rose	2017	₹ 1,470.11	4.9	0.42	0	2.1	0.048	16	42	0.99154	3.71	0.74	14	6	Adega de Moncao 'Muralhas de Moncao' Rose - 2017
Red	Quinta de Santiago Rose	2018	₹ 1,251.44	10.7	0.52	0.38	2.6	0.066	29	56	0.99577	3.15	0.79	12.1	5	Quinta de Santiago Rose - 2018
Red	Padeiro Bardado	2020	₹ 1,217.84	7.4	0.635	0.1	2.4	0.08	16	33	0.99736	3.58	0.69	10.8	5	Padeiro Bardado - 2020
Red	Casal de Ventozela Ve - Ve Rose	2020	₹ 1,071.28	8.3	0.28	0.48	2.1	0.093	6	12	0.99408	3.26	0.62	12.4	5	Casal de Ventozela Ve - Ve Rose - 2020
Red	Adega Ponte da Barca 'Estreia' Rose	2021	₹ 995.47	8.8	0.33	0.41	5.9	0.073	7	13	0.99658	3.3	0.62	12.1	5	Adega Ponte da Barca 'Estreia' Rose - 2021
Red	Adega Cooperativa Ponte de Lima Rose	2019	₹ 915.03	8.9	0.38	0.4	2.2	0.068	12	28	0.99486	3.27	0.75	12.6	5	Adega Cooperativa Ponte de Lima Rose - 2019
Red	Casal Gracia Rose	2019	₹ 831.00	8.9	0.48	0.53	4	0.101	3	10	0.99586	3.21	0.59	12.1	5	Casal Gracia Rose - 2019
Red	Sol Real Rose	2019	₹ 776.13	8.5	0.34	0.4	4.7	0.055	3	9	0.99738	3.38	0.66	11.6	5	Sol Real Rose - 2019
Red	Casa Vila Verde 'Tiroliro' Rose	2019	₹ 764.98	9.2	0.31	0.36	2.2	0.079	11	31	0.99615	3.33	0.86	12	5	Casa Vila Verde 'Tiroliro' Rose - 2019
Red	Seastone Rose	2018	₹ 731.91	8.5	0.28	0.35	1.7	0.061	6	15	0.99524	3.3	0.74	11.8	5	Seastone Rose - 2018
Red	Seaside Cellars Rose	2018	₹ 650.92	7.2	0.36	0.46	2.1	0.074	24	44	0.99534	3.4	0.85	11	5	Seaside Cellars Rose - 2018
Red	Aveleda Fonte Rose	2019	₹ 604.00	6.2	0.39	0.43	2	0.071	14	24	0.99428	3.45	0.87	11.2	5	Aveleda Fonte Rose - 2019
Red	Orlana Rose	2019	₹ 566.93	7.2	0.37	0.32	2	0.062	15	28	0.9947	3.23	0.73	11.3	5	Orlana Rose - 2019
Red	Adega Cooperativa de Guimaraes Rose	2020	₹ 535.02	7.8	0.32	0.44	2.7	0.104	8	17	0.99732	3.33	0.78	11	5	Adega Cooperativa de Guimaraes Rose - 2020
Red	Quinta da Lixa 'Anjos de Portugal' Rose	2017	₹ 474.54	7.9	0.35	0.46	3.6	0.078	15	37	0.9973	3.35	0.86	12.8	5	Quinta da Lixa 'Anjos de Portugal' Rose - 2017
Red	Quinta de Curvos 'Afectus' Rose	2017	₹ 411.17	5.5	0.49	0.03	1.8	0.044	28	87	0.9908	3.5	0.82	14	5	Quinta de Curvos 'Afectus' Rose - 2017
Red	Espiral Rose	2013	₹ 340.08	8	0.59	0.16	1.8	0.065	3	16	0.9962	3.42	0.92	10.5	5	Espiral Rose - 2013
Red	Quinta da Raza Dom Diago Padeiro	2017	₹ 466.14	10.3	0.32	0.45	6.4	0.073	5	13	0.9976	3.23	0.82	12.6	5	Quinta da Raza Dom Diago Padeiro - 2017
Red	Tr3s Castas Rose	2020	₹ 230.97	5.2	0.48	0.04	1.6	0.054	19	106	0.9927	3.54	0.62	12.2	4	Tr3s Castas Rose - 2020
Red	Escudo Real Rose	2020	₹ 188.56	8.9	0.4	0.32	5.6	0.087	10	47	0.9991	3.38	0.77	10.5	4	Escudo Real Rose - 2020
Red	Quinta da Raza Dom Diago Padeiro	2021	₹ 205.77	10	0.31	0.47	2.6	0.085	14	33	0.99965	3.36	0.8	10.5	4	Quinta da Raza Dom Diago Padeiro - 2021
Red	Broadbent Rose	2018	₹ 135.22	7.1	0.875	0.05	5.7	0.082	3	14	0.99808	3.4	0.52	10.2	3	Broadbent Rose - 2018
White	Anselmo Mendes Curtimenta Alvarinho	2011	₹ 6,160.05	10.3	0.17	0.47	1.4	0.037	5	33	0.9939	2.89	0.28	9.6	9	Anselmo Mendes Curtimenta Alvarinho - 2011
White	Adega Ponte da Barca 'Estreia' Alvarinho Grande Escolha	2014	₹ 5,417.10	8.6	0.55	0.35	15.55	0.057	35.5	366.5	1.0001	3.04	0.63	11	8	Adega Ponte da Barca 'Estreia' Alvarinho Grande Escolha - 2014
White	Quinta de Soalheiro Premeiras Vinhas Alvarinho	2011	₹ 4,506.00	6.1	0.2	0.34	9.5	0.041	38	201	0.995	3.14	0.44	10.1	8	Quinta de Soalheiro Premeiras Vinhas Alvarinho - 2011
White	Anselmo Mendes Private 5 Barricas Alvarinho	2018	₹ 3,661.96	7.2	0.32	0.47	5.1	0.044	19	65	0.991	3.03	0.41	12.6	8	Anselmo Mendes Private 5 Barricas Alvarinho - 2018
White	Quinta de Santiago 'Santiago na Amphora do Rocim' Alvarinho	2020	₹ 3,374.45	5.8	0.36	0.38	0.9	0.037	3	75	0.9904	3.28	0.34	11.4	8	Quinta de Santiago 'Santiago na Amphora do Rocim' Alvarinho - 2020
White	Adega de Moncao Deu-La-Deu Alvarinho Reserva	2018	₹ 3,236.66	7.2	0.31	0.46	5	0.04	3	29	0.9906	3.04	0.53	12.5	8	Adega de Moncao Deu-La-Deu Alvarinho Reserva - 2018
White	Casa Santa Eulalia Terroir Velho Mundo Alvarinho	2016	₹ 3,007.73	7.4	0.17	0.4	5.5	0.037	34	161	0.9935	3.05	0.62	11.5	8	Casa Santa Eulalia Terroir Velho Mundo Alvarinho - 2016
White	Adega de Moncao Deu-La-Deu Terracos Alvarinho	2016	₹ 2,977.27	8.4	0.19	0.42	1.6	0.047	9	101	0.994	3.06	0.65	11.1	8	Adega de Moncao Deu-La-Deu Terracos Alvarinho - 2016
White	Anselmo Mendes 'Parcela Unica' Alvarinho	2014	₹ 2,813.85	6.5	0.52	0.17	1.4	0.047	5	26	0.9932	3.26	0.32	10	8	Anselmo Mendes 'Parcela Unica' Alvarinho - 2014
White	Quinta de Santiago e Mira Do O 'Sou' Alvarinho	2018	₹ 2,755.17	9.2	0.35	0.39	0.9	0.042	15	61	0.9924	2.96	0.28	10.4	8	Quinta de Santiago e Mira Do O 'Sou' Alvarinho - 2018
White	Quinta da Pedra Alvarinho	2017	₹ 2,734.69	9.2	0.28	0.49	11.8	0.042	29	137	0.998	3.1	0.34	10.1	8	Quinta da Pedra Alvarinho - 2017
White	Anselmo Mendes Alvarinho	2009	₹ 2,389.70	6.4	0.43	0.27	1.1	0.054	5	110	0.9939	3.24	0.52	9.1	8	Anselmo Mendes Alvarinho - 2009
White	Quintas de Melgaco QM Alvarinho	2014	₹ 2,328.32	9.8	0.93	0.45	8.6	0.052	34	187	0.9994	3.12	0.59	10.2	8	Quintas de Melgaco QM Alvarinho - 2014
White	Quinta Edmun do Val Alvarinho Grande Reserva	2008	₹ 2,263.51	6	0.33	0.18	3	0.036	5	85	0.99125	3.28	0.4	11.5	8	Quinta Edmun do Val Alvarinho Grande Reserva - 2008
White	Quinta de Soalheiro Alvarinho	2010	₹ 2,205.00	9.2	0.19	0.42	2	0.047	16	104	0.99517	3.09	0.66	10	8	Quinta de Soalheiro Alvarinho - 2010
White	Quinta de Soalheiro Premeiras Vinhas Alvarinho	2014	₹ 2,203.00	6.3	0.38	0.17	8.8	0.08	50	212	0.99803	3.47	0.66	9.4	8	Quinta de Soalheiro Premeiras Vinhas Alvarinho - 2014
White	Quintas de Melgaco QM Alvarinho Nature	2016	₹ 2,108.23	6.7	0.54	0.27	7.1	0.049	8	178	0.99502	3.16	0.38	9.4	8	Quintas de Melgaco QM Alvarinho Nature - 2016
White	Quinta do Regueiro Alvarinho Reserva	2016	₹ 2,103.87	6.7	0.54	0.27	7.1	0.049	8	178	0.99502	3.16	0.38	9.4	8	Quinta do Regueiro Alvarinho Reserva - 2016

White	Quinta do Regueiro 'Regueiro' Barricas Alvarinho	2019	₹	2,098.00	7.2	0.615	0.1	1.4	0.068	25	154	0.99499	3.2	0.48	9.7	8	Quinta do Regueiro 'Regueiro' Barricas Alvarinho - 2019
White	Adega Ponte da Barca 'Estreia' Alvarinho Reserva	2019	₹	2,053.50	7.1	0.27	0.28	1.25	0.023	3	89	0.98993	2.95	0.3	11.4	8	Adega Ponte da Barca 'Estreia' Alvarinho Reserva - 2019
White	Quinta de Covela 'Edicao Nacional' Alvarinho	2026	₹	2,023.46	7.3	0.28	0.42	1.2	0.033	29	142	0.99205	3.17	0.43	10.7	8	Quinta de Covela 'Edicao Nacional' Alvarinho - 2026
White	Quinta de Pacos 'Casa do Capitao-Mor' Sobre Lias Alvarinho	2016	₹	1,953.59	8.1	0.27	0.41	1.45	0.033	11	63	0.9908	2.99	0.56	12	7	Quinta de Pacos 'Casa do Capitao-Mor' Sobre Lias Alvarinho - 2016
White	Anselmo Mendes Muros de Melgaco Alvarinho	2011	₹	1,923.91	8.6	0.23	0.4	4.2	0.035	17	109	0.9947	3.14	0.53	9.7	7	Anselmo Mendes Muros de Melgaco Alvarinho - 2011
White	Quinta do Feital Dorado Alvarinho	2010	₹	1,913.00	7.6	0.67	0.14	1.5	0.074	25	168	0.9937	3.05	0.51	9.3	7	Quinta do Feital Dorado Alvarinho - 2010
White	Quinta de Santiago Alvarinho Reserva	2017	₹	1,906.55	7.3	0.28	0.43	1.7	0.08	21	123	0.9905	3.19	0.42	12.8	7	Quinta de Santiago Alvarinho Reserva - 2017
White	Adega Cooperative Ponte de Braca Estreia Reserva Alvarinho Vinho Verde	2017	₹	1,900.00	6.5	0.39	0.23	5.4	0.051	25	149	0.9934	3.24	0.35	10	7	Adega Cooperative Ponte de Braca Estreia Reserva Alvarinho Vinho Verde - 2017
White	Quinta de Pedra Alvarinho	2011	₹	1,825.20	6	0.21	0.24	12.1	0.05	55	164	0.997	3.34	0.39	9.4	7	Quinta de Pedra Alvarinho - 2011
White	Quinta de Soalheiro Alvarinho Reserva	2020	₹	1,804.00	7.2	0.27	0.46	18.75	0.052	45	255	1	3.04	0.52	8.9	7	Quinta de Soalheiro Alvarinho Reserva - 2020
White	Quinta de Alderiz Alvarinho Espumante	2017	₹	1,748.31	6.6	0.24	0.27	15.8	0.035	46	188	0.9982	3.24	0.51	9.2	7	Quinta de Alderiz Alvarinho Espumante - 2017
White	Anselmo Mendes 'Autentico' Alvarinho	2018	₹	1,635.27	6.9	0.54	0.32	13.2	0.05	53	236	0.9973	3.2	0.5	9.6	7	Anselmo Mendes 'Autentico' Alvarinho - 2018
White	Quinta de Pacos 'Casa do Capitao-Mor' Alvarinho	2019	₹	1,564.36	6.9	0.29	0.4	19.45	0.043	36	156	0.9996	2.93	0.47	8.9	7	Quinta de Pacos 'Casa do Capitao-Mor' Alvarinho - 2019
White	Quinta de Soalheiro Alvarinho Bruto Barrica Espumante	2014	₹	1,544.56	7.7	0.26	0.4	1.1	0.042	9	60	0.9915	2.89	0.5	10.6	7	Quinta de Soalheiro Alvarinho Bruto Barrica Espumante - 2014
White	Quinta do Regueiro 'Foral de Melgaco' Alvarinho	2020	₹	1,530.73	6.6	0.24	0.35	7.7	0.031	36	135	0.9938	3.19	0.37	10.5	7	Quinta do Regueiro 'Foral de Melgaco' Alvarinho - 2020
White	Quinta do Tamariz Alvarinho Reserva	2017	₹	1,510.00	7.1	0.39	0.35	12.5	0.044	26	72	0.9941	3.17	0.29	11.6	7	Quinta do Tamariz Alvarinho Reserva - 2017
White	Quinta D'amares Vinesa Alvarinho Barrica	2018	₹	1,457.21	7.2	0.6	0.2	9.9	0.07	21	174	0.9971	3.03	0.54	9.1	7	Quinta D'amares Vinesa Alvarinho Barrica - 2018
White	Casa de Canhotos Alvarinho Espumante Bruto	2018	₹	1,410.08	6.5	0.3	0.29	2.25	0.037	8	210	0.9937	3.19	0.62	9.9	7	Casa de Canhotos Alvarinho Espumante Bruto - 2018
White	Casa Vila Nova Alvarinho	2018	₹	1,354.23	6.7	0.34	0.54	16.3	0.047	44	181	0.9987	3.04	0.56	8.8	7	Casa Vila Nova Alvarinho - 2018
White	Quinta do Regueiro 'Primitivo' Alvarinho	2017	₹	1,348.09	5.7	0.18	0.22	4.2	0.042	25	111	0.994	3.35	0.39	9.4	7	Quinta do Regueiro 'Primitivo' Alvarinho - 2017
White	Quinta de Gomariz Alvarinho Bruto	2018	₹	1,321.99	7.5	0.23	0.68	11	0.047	37	133	0.9978	2.99	0.38	8.8	7	Quinta de Gomariz Alvarinho Bruto - 2018
White	Quinta de Soalheiro 'Granit' Alvarinho	2017	₹	1,302.00	6.3	0.26	0.25	7.8	0.058	44	166	0.9961	3.24	0.41	9	7	Quinta de Soalheiro 'Granit' Alvarinho - 2017
White	Anselmo Mendes Muros Antigos Alvarinho	2020	₹	1,289.00	6.2	0.36	0.26	13.2	0.051	54	201	0.9976	3.25	0.46	9	7	Anselmo Mendes Muros Antigos Alvarinho - 2020
White	Quinta do Regueiro Alvarinho Espumante Metodo Classico Bruto	2017	₹	1,274.96	6.3	0.34	0.28	14.7	0.047	49	198	0.9977	3.23	0.46	9.5	7	Quinta do Regueiro Alvarinho Espumante Metodo Classico Bruto - 2017
White	Quinta da Pedra 'Longos Vales' Alvarinho	2017	₹	1,255.64	6.4	0.31	0.39	7.5	0.04	57	213	0.99475	3.32	0.43	10	7	Quinta da Pedra 'Longos Vales' Alvarinho - 2017
White	Quinta do Regueiro Alvarinho	2015	₹	1,237.72	7	0.21	0.34	8	0.057	19	101	0.9954	2.99	0.59	9.4	7	Quinta do Regueiro Alvarinho - 2015
White	Quintas de Melgaco QM Alvarinho Vinhas Velhas	2020	₹	1,208.00	6.2	0.3	0.17	2.8	0.04	24	125	0.9939	3.01	0.46	9	7	Quintas de Melgaco QM Alvarinho Vinhas Velhas - 2020
White	Anselmo Mendes 'Momento Ousado' Alvarinho	2016	₹	1,175.85	6.1	0.28	0.35	12.8	0.048	63	229	0.9975	3.08	0.4	8.9	7	Anselmo Mendes 'Momento Ousado' Alvarinho - 2016
White	Casal de Ventozela Alvarinho	2017	₹	1,086.82	6.4	0.5	0.16	12.9	0.042	26	138	0.9974	3.28	0.33	9	7	Casal de Ventozela Alvarinho - 2017
White	Casa de Canhotos Alvarinho	2018	₹	1,086.82	9.8	0.36	0.45	1.6	0.042	11	124	0.9944	2.93	0.46	10.8	7	Casa de Canhotos Alvarinho - 2018
White	Quinta de Soalheiro Docil Alvarinho	2018	₹	1,048.19	7.6	0.26	0.47	1.6	0.068	5	55	0.9944	3.1	0.45	9.6	7	Quinta de Soalheiro Docil Alvarinho - 2018
White	Casa de Vila Nova 'Filipa de Lencastre' Alvarinho	2019	₹	738.27	6	0.23	0.15	9.7	0.048	101	207	0.99571	3.05	0.3	9.1	6	Casa de Vila Nova 'Filipa de Lencastre' Alvarinho - 2019
White	Adega Ponte da Barca Alvarinho Reserva	2020	₹	634.12	7.9	0.21	0.39	2	0.057	21	138	0.99176	3.05	0.52	10.9	6	Adega Ponte da Barca Alvarinho Reserva - 2020
White	Casa de Vila Boa Alvarinho	2019	₹	608.92	6.2	0.28	0.51	7.9	0.056	49	206	0.9956	3.18	0.52	9.4	6	Casa de Vila Boa Alvarinho - 2019
White	Quinta de Covela 'Edicao Nacional' Avesso Reserva	2018	₹	1,913.00	6.5	0.31	0.14	7.5	0.044	34	133	0.9955	3.22	0.5	9.5	7	Quinta de Covela 'Edicao Nacional' Avesso Reserva - 2018
White	Singellus Private Avesso	2016	₹	1,826.77	7.4	0.25	0.37	13.5	0.06	52	192	0.9975	3	0.44	9.1	7	Singellus Private Avesso - 2016
White	Quinta da Raza Avesso	2018	₹	1,609.69	5.6	0.245	0.25	9.7	0.032	12	68	0.994	3.31	0.34	10.5	7	Quinta da Raza Avesso - 2018
White	Quinta de Santa Teresa Avesso	2020	₹	1,601.40	5.8	0.26	0.24	9.2	0.044	55	152	0.9961	3.31	0.38	9.4	7	Quinta de Santa Teresa Avesso - 2020
White	Casa de Vilacetinho Avesso - Alvarinho	2019	₹	1,509.02	6.3	0.39	0.35	5.9	0.04	82.5	260	0.9941	3.12	0.66	10.1	7	Casa de Vilacetinho Avesso - Alvarinho - 2019
White	A&D Wines 'Monologo' Avesso P69	2019	₹	1,458.69	5.8	0.28	0.27	2.6	0.054	30	156	0.9914	3.53	0.42	12.4	7	A&D Wines 'Monologo' Avesso P69 - 2019
White	Quinta de Guimaraes 'Cazas Novas' Avesso	2018	₹	1,415.00	5.8	0.25	0.24	13.3	0.044	41	137	0.9972	3.34	0.42	9.5	7	Quinta de Guimaraes 'Cazas Novas' Avesso - 2018
White	Quinta de Covela 'Edicao Nacional' Avesso	2016	₹	1,220.39	6.1	0.27	0.3	16.7	0.039	49	172	0.99985	3.4	0.45	9.4	7	Quinta de Covela 'Edicao Nacional' Avesso - 2016
White	Casa de Vilacetinho Avesso Superior	2017	₹	1,049.87	8.2	0.61	0.45	5.4	0.03	15	118	0.9954	3.14	0.34	9.6	7	Casa de Vilacetinho Avesso Superior - 2017
White	A&D Wines 'Monologo' Avesso P70	2017	₹	948.24	6.6	0.25	0.51	8	0.047	61	189	0.99604	3.22	0.49	9.2	6	A&D Wines 'Monologo' Avesso P70 - 2017
White	A&D Wines 'Monologo' Avesso P68	2020	₹	919.68	6.6	0.45	0.43	7.2	0.064	31	186	0.9954	3.12	0.44	9.4	6	A&D Wines 'Monologo' Avesso P68 - 2020
White	A&D Wines 'Monologo' Avesso P71	2016	₹	910.44	6.5	0.15	0.55	5.9	0.045	75	162	0.99482	2.97	0.4	9.3	6	A&D Wines 'Monologo' Avesso P71 - 2016
White	Quinta da Raza Avesso - Alvarinho	2017	₹	910.00	7.9	0.25	0.35	6.7	0.039	22	64	0.99362	2.93	0.49	10.7	6	Quinta da Raza Avesso - Alvarinho - 2017
White	A&D Wines 'Monologo' Avesso P67	2021	₹	835.67	7.4	0.23	0.38	8.6	0.052	41	150	0.99534	3.06	0.46	10.3	6	A&D Wines 'Monologo' Avesso P67 - 2021
White	Anselmo Mendes 'Muros Antigos' Avesso	2021	₹	751.70	6.9	0.29	0.41	7.8	0.046	52	171	0.99537	3.12	0.51	9.6	6	Anselmo Mendes 'Muros Antigos' Avesso - 2021
White	Casa de Vilacetinho Avesso - Azal	2017	₹	751.70	6.6	0.24	0.27	10.3	0.047	54	219	0.99742	3.04	0.45	8.8	6	Casa de Vilacetinho Avesso - Azal - 2017
White	Casal de Ventozela Avesso	2017	₹	718.11	6.7	0.35	0.48	8.8	0.056	35	167	0.99628	3.04	0.47	9.4	6	Casal de Ventozela Avesso - 2017
White	Quinta da Raza Avesso - Alvarinho	2019	₹	631.92	7.4	0.34	0.28	12.1	0.049	31	149	0.99677	3.22	0.49	10.3	6	Quinta da Raza Avesso - Alvarinho - 2019
White	Casa de Vila Nova 'Vila Nova' Avesso	2017	₹	561.05	6.4	0.41	0.01	6.1	0.048	20	70	0.99362	3.19	0.42	10	6	Casa de Vila Nova 'Vila Nova' Avesso - 2017
White	Quintas de Caiz 'Encostas de Caiz' Avesso	2019	₹	415.75	6	0.26	0.42	5.2	0.027	70	178	0.9914	3.4	0.4	12.3	4	Quintas de Caiz 'Encostas de Caiz' Avesso - 2019
White	Aveleda Manoel Pedro Guedes Branco	2018	₹	4,875.59	7.5	0.32	0.24	4.6	0.053	8	134	0.9958	3.14	0.5	9.1	8	Aveleda Manoel Pedro Guedes Branco - 2018
White	Casa do Valle Branco	2006	₹	3,285.98	7.2	0.4	0.62	10.8	0.041	70	189	0.9976	3.08	0.49	8.6	8	Casa do Valle Branco - 2006
White	Quinta do Tamariz Grande Reserva Branco	2017	₹	1,711.00	6.4	0.22	0.56	14.5	0.055	27	159	0.998	2.98	0.4	9.1	7	Quinta do Tamariz Grande Reserva Branco - 2017
White	Adega Cooperativa de Guimaraes Branco	2012	₹	1,664.34	7.1	0.32	0.24	13.1	0.05	52	204	0.998	3.1	0.49	8.8	7	Adega Cooperativa de Guimaraes Branco - 2012
White	Quinta de Azevedo Branco Reserva	2019	₹	1,611.86	6.8	0.57	0.29	2.2	0.04	15	77	0.9938	3.32	0.74	10.2	7	Quinta de Azevedo Branco Reserva - 2019
White	Adega de Moncao Branco	2017	₹	1,530.17	6.6	0.24	0.35	7.7	0.031	36	135	0.9938	3.19	0.37	10.5	7	Adega de Moncao Branco - 2017
White	Casal de Ventozela Prime Selection Branco	2017	₹	1,482.41	6.8	0.26	0.22	7.7	0.047	57	210	0.9959	3.1	0.47	9	7	Casal de Ventozela Prime Selection Branco - 2017
White	Quinta da Calcada 'Lago' Cerqueira Branco	2019	₹	1,415.00	6.7	0.47	0.34	8.9	0.043	31	172	0.9964	3.22	0.6	9.2	7	Quinta da Calcada 'Lago' Cerqueira Branco - 2019
White	Casa de Cello Quinta de San Joanne Terroir Mineral Branco	2015	₹	1,210.28	8.4	0.18	0.42	5.1	0.036	7	77	0.9939	3.16	0.52	11.7	7	Casa de Cello Quinta de San Joanne Terroir Mineral Branco - 2015
White	Quinta do Portal Trevo Branco	2020	₹	1,177.31	7.7	0.15	0.29	1.3	0.029	10	64	0.9932	3.35	0.39	10.1	7	Quinta do Portal Trevo Branco - 2020
White	Quinta do Tamariz Superior Branco	2019	₹	1,175.85	5.9	0.26	0.3	1	0.036	38	114	0.9928	3.58	0.48	9.4	7	Quinta do Tamariz Superior Branco - 2019
White	Quinta de Gomariz QG Premium Branco	2017	₹	1,121.76	5.1	0.52	0.06	2.7	0.052	30	79	0.9932	3.32	0.43	9.3	7	Quinta de Gomariz QG Premium Branco - 2017
White	Quinta de Covela Trellis Branco	2018	₹	1,107.00	7	0.31	0.52	1.7	0.029	5	61	0.9918	3.07	0.43	10.4	7	Quinta de Covela Trellis Branco - 2018
White	Casal Gracia Branco	2012	₹	1,103.00	7.3	0.24	0.34	7.5	0.048	29	152	0.9962	3.1	0.54	9	7	Casal Gracia Branco - 2012
White	Quinta da Lixa Vinha Real Branco	2021	₹	1,091.02	6.3	0.16	0.4	1.6	0.033	59	148	0.9914	3.44	0.53	11.4	7	Quinta da Lixa Vinha Real Branco - 2021
White	Casa Santos Lima 'Lab' Branco	2019	₹	1,068.22	6.7	0.25	0.23	7.2	0.038	61	220	0.9952	3.14	0.35	9.5	7	Casa Santos Lima 'Lab' Branco - 2019
White	Casa de Vila Verde Branco	2020	₹	1,015.00	8	0.61	0.38	12.1	0.301	24	220	0.9993	2.94	0.48	9.2	7	Casa de Vila Verde Branco - 2020
White	Aveleda Branco	2020	₹	994.70	8.3	0.4	0.41	8.2	0.05	15	122	0.9979	3.39	0.49	9.3	6	Aveleda Branco - 2020
White	Casa da Bouca Branco	2020	₹	839.00	6.3	0.28	0.3	3.1	0.039	24	115	0.9942	3.05	0.43	8.6	6	Casa da Bouca Branco - 2020
White	Casal Garcia Sweet Branco	2019	₹	838.35	7.1	0.28	0.35	3.5	0.028	35	91	0.99022	2.96	0.33	12.1	6	Casal Garcia Sweet Branco - 2019
White	Quinta do Barco Grande Escolha Branco	2020	₹	797.70	6.1	0.22	0.31	1.4	0.039	40	129	0.99193	3.45	0.59	10.9	6	Quinta do Barco Grande Escolha Branco - 2020

White	Casa de Vila Nova Marques de Lara Branco	2019	₹	789.50	5.1	0.21	0.28	1.4	0.047	48	148	0.99168	3.5	0.49	10.4	6	Casa de Vila Nova Marques de Lara Branco - 2019
White	Adega Ponte da Barca 'Estreia' Branco Vinho Verde	2016	₹	764.68	6.5	0.25	0.45	7.8	0.048	52	188	0.99576	3.2	0.53	9.1	6	Adega Ponte da Barca 'Estreia' Branco Vinho Verde - 2016
White	Quinta das Arcas 'Bicudo' Branco	2020	₹	762.40	7.1	0.18	0.39	14.5	0.051	48	156	0.99947	3.35	0.78	9.1	6	Quinta das Arcas 'Bicudo' Branco - 2020
White	Quinta de Azevedo Branco	2019	₹	726.35	7	0.48	0.12	4.5	0.05	23	86	0.99398	2.86	0.35	9	6	Quinta de Azevedo Branco - 2019
White	Adega Ponte da Barca Las Lilas Branco	2018	₹	600.00	5.7	0.14	0.3	5.4	0.045	26	105	0.99469	3.32	0.45	9.3	6	Adega Ponte da Barca Las Lilas Branco - 2018
White	Quinta da Lixa QL-RE 'O tal vinho da Lixa' Branco	2019	₹	586.24	7.2	0.4	0.24	8.5	0.055	45	151	0.99626	3.2	0.52	9.2	6	Quinta da Lixa QL-RE 'O tal vinho da Lixa' Branco - 2019
White	Aveleda Branco	2018	₹	535.26	6.7	0.31	0.34	6.8	0.059	51	215	0.99538	3.33	0.56	10.3	6	Aveleda Branco - 2018
White	Casal de Ventozela Branco	2018	₹	499.74	7.2	0.17	0.34	6.4	0.042	16	111	0.99278	2.99	0.4	10.8	5	Casal de Ventozela Branco - 2018
White	Quinta de Curvos 'Afectus' Branco	2017	₹	491.40	5.1	0.35	0.26	6.8	0.034	36	120	0.99188	3.38	0.4	11.5	5	Quinta de Curvos 'Afectus' Branco - 2017
White	Adega de Moncao Branco	2021	₹	463.41	6.1	0.31	0.58	5	0.039	36	114	0.9909	3.3	0.6	12.3	4	Adega de Moncao Branco - 2021
White	Quinta da Lixa Monsenhor Branco	2020	₹	457.49	6	0.25	0.28	2.2	0.026	54	126	0.9898	3.43	0.65	12.9	4	Quinta da Lixa Monsenhor Branco - 2020
White	Adega cooperativa Ponte de Lima Branco	2020	₹	412.53	8.2	0.37	0.36	1	0.034	17	93	0.9906	3.04	0.32	11.7	4	Adega cooperativa Ponte de Lima Branco - 2020
White	Quinta & Casa das Hortas 'Portal das Hortas' Branco	2018	₹	382.15	7.1	0.28	0.49	6.5	0.041	28	111	0.9926	3.41	0.58	12.2	4	Quinta & Casa das Hortas 'Portal das Hortas' Branco - 2018
White	Adega Ponte da Barca Branco	2020	₹	299.11	7.1	0.47	0.29	14.8	0.024	22	142	0.99518	3.12	0.48	12	4	Adega Ponte da Barca Branco - 2020
White	Quintas de Melgaco 'Leira do Canhoto' Branco	2020	₹	256.17	7	0.17	0.36	6.4	0.055	42	123	0.99318	3.11	0.5	11	4	Quintas de Melgaco 'Leira do Canhoto' Branco - 2020
White	Quinta & Casa das Hortas Saiote Branco	2014	₹	239.37	7.1	0.26	0.37	5.5	0.025	31	105	0.99082	3.06	0.33	12.6	4	Quinta & Casa das Hortas Saiote Branco - 2014
White	Casa de Vila Boa Branco	2016	₹	206.56	5.3	0.3	0.16	4.2	0.029	37	100	0.9905	3.3	0.36	11.8	4	Casa de Vila Boa Branco - 2016
White	Casal da Seara Branco	2020	₹	205.77	6.5	0.43	0.31	3.6	0.046	19	143	0.99022	3.15	0.34	12	4	Casal da Seara Branco - 2020
White	Quinta D'amares Loureiro - Arinto	2019	₹	6,343.96	6.9	0.39	0.4	4.6	0.022	5	19	0.9915	3.31	0.37	12.6	9	Quinta D'amares Loureiro - Arinto - 2019
White	Quinta de Covela 'Edicao Nacional' Arinto	2018	₹	2,023.46	6.2	0.255	0.24	1.7	0.039	138.5	272	0.99452	3.53	0.53	9.6	8	Quinta de Covela 'Edicao Nacional' Arinto - 2018
White	A&D Wines 'Casa do Arrabalde' Avesso - Alvarinho - Arinto	2018	₹	1,456.11	6.1	0.2	0.25	1.2	0.038	34	128	0.9921	3.24	0.44	10.1	7	A&D Wines 'Casa do Arrabalde' Avesso - Alvarinho - Arinto - 2018
White	Agri Roncao Quinta de Linhares Arinto	2017	₹	1,112.69	7.6	0.39	0.22	2.8	0.036	19	113	0.9926	3.03	0.29	10.2	7	Agri Roncao Quinta de Linhares Arinto - 2017
White	Quinta da Raza Dom Diago Arinto	2020	₹	1,086.90	5.8	0.415	0.13	1.4	0.04	11	64	0.9922	3.29	0.52	10.5	7	Quinta da Raza Dom Diago Arinto - 2020
White	A&D Wines 'Monologo' Arinto P28	2017	₹	948.28	6.8	0.35	0.53	10.1	0.053	37	151	0.9963	3.07	0.4	9.4	6	A&D Wines 'Monologo' Arinto P28 - 2017
White	A&D Wines 'Monologo' Arinto P25	2020	₹	919.68	6.6	0.22	0.34	11.6	0.05	59	140	0.99526	3.22	0.4	10.8	6	A&D Wines 'Monologo' Arinto P25 - 2020
White	A&D Wines 'Monologo' Arinto P26	2019	₹	902.00	7	0.31	0.39	7.5	0.055	42	218	0.99652	3.37	0.54	10.3	6	A&D Wines 'Monologo' Arinto P26 - 2019
White	A&D Wines 'Monologo' Arinto P24	2021	₹	886.09	5.2	0.17	0.27	0.7	0.03	11	68	0.99218	3.3	0.41	9.8	6	A&D Wines 'Monologo' Arinto P24 - 2021
White	A&D Wines 'Monologo' Arinto P27	2018	₹	709.71	7.3	0.31	0.25	6.65	0.032	30	138	0.99244	2.9	0.37	11.1	6	A&D Wines 'Monologo' Arinto P27 - 2018
White	A&D Wines 'Monologo' Arinto P29	2016	₹	668.55	6.4	0.27	0.45	8.3	0.05	52	196	0.9955	3.18	0.48	9.5	6	A&D Wines 'Monologo' Arinto P29 - 2016
White	Casal de Ventozela 'Vento'Z' Arinto	2018	₹	520.73	6.3	0.3	0.48	7.4	0.053	34	149	0.99472	3.18	0.53	9.8	6	Casal de Ventozela 'Vento'Z' Arinto - 2018
White	Casal de Ventozela Arinto	2020	₹	482.94	5.1	0.35	0.26	6.8	0.034	36	120	0.99188	3.38	0.4	11.5	5	Casal de Ventozela Arinto - 2020
White	Quinta D'amares Loureiro - Arinto	2020	₹	432.00	7.3	0.25	0.36	2.1	0.034	30	177	0.99085	3.25	0.4	11.9	4	Quinta D'amares Loureiro - Arinto - 2020
White	Quinta da Raza Dom Diago Arinto	2017	₹	358.95	6.9	0.15	0.28	4.4	0.029	14	107	0.99347	3.24	0.46	10.4	4	Quinta da Raza Dom Diago Arinto - 2017
White	Quinta da Raza Grande Escolha Alvarinho - Trajadura	2021	₹	1,336.96	6.9	0.41	0.33	10.1	0.043	28	152	0.9968	3.2	0.52	9.4	7	Quinta da Raza Grande Escolha Alvarinho - Trajadura - 2021
White	Quinta da Raza Blanco Grande Escolha Alvarinho - Trajadura	2017	₹	1,291.83	6.4	0.27	0.32	4.5	0.24	61	174	0.9948	3.12	0.48	9.4	7	Quinta da Raza Blanco Grande Escolha Alvarinho - Trajadura - 2017
White	Quinta do Regueiro Alvarinho - Trajadura	2020	₹	1,068.43	8.9	0.34	0.32	1.3	0.041	12	188	0.9953	3.17	0.49	9.5	7	Quinta do Regueiro Alvarinho - Trajadura - 2020
White	Quinta da Lixa 'Aromas das Castas' Alvarinho - Trajadura	2020	₹	994.70	7.1	0.31	0.17	1	0.042	21	144	0.99304	3.13	0.4	9.6	6	Quinta da Lixa 'Aromas das Castas' Alvarinho - Trajadura - 2020
White	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura	2016	₹	842.56	6.2	0.25	0.38	7.9	0.045	54	208	0.99572	3.17	0.46	9.1	6	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura - 2016
White	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura	2015	₹	800.43	7.6	0.27	0.29	2.5	0.059	37	115	0.99328	3.09	0.37	9.8	6	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura - 2015
White	Casa Santa Eulalia Alvarinho - Trajadura	2018	₹	746.81	6.7	0.29	0.45	14.3	0.054	30	181	0.99869	3.14	0.57	9.1	6	Casa Santa Eulalia Alvarinho - Trajadura - 2018
White	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura	2019	₹	608.77	7.3	0.36	0.34	14.8	0.057	46	173	0.99751	3.14	0.57	10.2	6	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura - 2019
White	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura	2020	₹	606.00	8.6	0.22	0.33	1.2	0.031	38	95	0.99239	2.83	0.31	10.3	6	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura - 2020
White	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura	2021	₹	543.45	6.9	0.37	0.23	9.5	0.057	54	166	0.99568	3.23	0.42	10	6	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura - 2021
White	Casa Santa Eulalia Alvarinho - Trajadura	2021	₹	478.14	7.5	0.26	0.31	1.6	0.032	36	109	0.99044	2.97	0.43	11.9	5	Casa Santa Eulalia Alvarinho - Trajadura - 2021
White	Quintas de Melgaco Torre de Menagem Alvarinho - Trajadura	2018	₹	379.15	6.4	0.15	0.36	1.8	0.034	43	150	0.9922	3.42	0.69	11	4	Quintas de Melgaco Torre de Menagem Alvarinho - Trajadura - 2018
White	Quinta da Lixa Trajadura	2020	₹	373.75	7.6	0.2	0.3	14.2	0.056	53	212.5	0.999	3.14	0.46	8.9	4	Quinta da Lixa Trajadura - 2020
White	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura	2018	₹	335.34	5.8	0.315	0.27	1.55	0.026	15	70	0.98994	3.37	0.4	11.9	4	Quinta da Lixa Escolha Alvarinho - Loureiro - Trajadura - 2018
White	Quinta D'amares Loureiro - Arinto	2019	₹	6,343.96	6.9	0.39	0.4	4.6	0.022	5	19	0.9915	3.31	0.37	12.6	9	Quinta D'amares Loureiro - Arinto - 2019
White	Quinta do Ameal Escolha Loureiro	2017	₹	4,681.14	7.1	0.49	0.22	2	0.047	146.5	307.5	0.9924	3.24	0.37	11	8	Quinta do Ameal Escolha Loureiro - 2017
White	Anselmo Mendes Private Loureiro	2019	₹	3,649.52	6.5	0.28	0.28	8.5	0.047	54	210	0.9962	3.09	0.54	8.9	8	Anselmo Mendes Private Loureiro - 2019
White	Quinta do Ameal Reserva Loureiro	2019	₹	3,129.66	6.8	0.29	0.16	1.4	0.038	122.5	234.5	0.9922	3.15	0.47	10	8	Quinta do Ameal Reserva Loureiro - 2019
White	Quinta de Ameal Loureiro	2015	₹	2,690.00	6.9	0.25	0.24	3.6	0.057	13	85	0.9942	2.99	0.48	9.5	8	Quinta de Ameal Loureiro - 2015
White	Quinta da Palmirinha Loureiro	2019	₹	2,276.54	7.7	0.28	0.63	11.1	0.039	58	179	0.9979	3.08	0.44	8.8	8	Quinta da Palmirinha Loureiro - 2019
White	Quintas do Homem 'Vale do Homem' Loureiro	2016	₹	1,560.00	6.9	0.29	0.4	19.45	0.043	36	156	0.9996	2.93	0.47	8.9	7	Quintas do Homem 'Vale do Homem' Loureiro - 2016
White	Anselmo Mendes 'Muros Antigos' Loureiro Escolha	2019	₹	1,301.55	6	0.32	0.12	5.9	0.041	34	190	0.9944	3.16	0.72	10	7	Anselmo Mendes 'Muros Antigos' Loureiro Escolha - 2019
White	Quinta de Santiago Alvarinho - Loureiro	2017	₹	1,297.62	6.4	0.32	0.5	10.7	0.047	57	206	0.9968	3.08	0.6	9.4	7	Quinta de Santiago Alvarinho - Loureiro - 2017
White	Quintas de Melgaco QM Loureiro - Alvarinho	2020	₹	1,263.61	8.4	0.4	0.7	13.1	0.042	29	197	0.998	3.06	0.64	9.7	7	Quintas de Melgaco QM Loureiro - Alvarinho - 2020
White	Casal de Ventozela 'Vento'Z' Loureiro	2020	₹	1,234.25	6.4	0.16	0.31	5.3	0.043	42	157	0.99455	3.35	0.47	10.5	7	Casal de Ventozela 'Vento'Z' Loureiro - 2020
White	Quinta de Curvos 'Curvos' Loureiro	2017	₹	1,085.35	9.6	0.23	0.4	1.5	0.044	19	135	0.9937	2.96	0.49	10.9	7	Quinta de Curvos 'Curvos' Loureiro - 2017
White	Quinta da Pousada Areal Superior Loureiro - Alvarinho	2020	₹	1,072.42	7.4	0.41	0.34	4.7	0.042	19	127	0.9953	3.25	0.42	10.4	7	Quinta da Pousada Areal Superior Loureiro - Alvarinho - 2020
White	Anselmo Mendes 'Passaros' Loureiro	2020	₹	1,068.43	9.1	0.27	0.32	1.1	0.031	15	151	0.9936	3.03	0.41	10.6	7	Anselmo Mendes 'Passaros' Loureiro - 2020
White	Casa de Vila Nova 'Vila Nova' Loureiro	2019	₹	1,006.68	7	0.16	0.73	1	0.138	58	150	0.9936	3.08	0.3	9.2	7	Casa de Vila Nova 'Vila Nova' Loureiro - 2019
White	Quinta D'amares Loureiro	2019	₹	1,005.81	7.2	0.14	0.35	1.2	0.036	15	73	0.9938	3.46	0.39	9.9	7	Quinta D'amares Loureiro - 2019
White	Casa da Bouca Loureiro	2020	₹	1,004.51	7.4	0.31	0.74	10.7	0.039	51	147	0.9977	3.02	0.43	8.7	7	Casa da Bouca Loureiro - 2020
White	Quinta de Curvos Loureiro Vinho Verde	2017	₹	987.87	7.8	0.25	0.34	13.7	0.044	66	184	0.99976	3.22	0.75	8.9	6	Quinta de Curvos Loureiro Vinho Verde - 2017
White	Quinta de Soalheiro 'Allo' Alvarinho - Loureiro	2019	₹	911.00	6.4	0.22	0.31	13.9	0.04	57	135	0.99672	3.21	0.38	10.7	6	Quinta de Soalheiro 'Allo' Alvarinho - Loureiro - 2019
White	Quinta da Calcada Loureiro - Alvarinho	2015	₹	890.29	5.3	0.21	0.29	0.7	0.028	11	66	0.99215	3.3	0.4	9.8	6	Quinta da Calcada Loureiro - Alvarinho - 2015
White	Quinta do Gradil 'Ruelas' Superior Alvarinho - Loureiro	2019	₹	838.71	6.2	0.25	0.38	7.9	0.045	54	208	0.99572	3.17	0.46	9.1	6	Quinta do Gradil 'Ruelas' Superior Alvarinho - Loureiro - 2019
White	Quinta da Lixa Loureiro	2018	₹	806.34	6.4	0.21	0.34	16.05	0.04	56	142	0.99678	3.11	0.38	10.6	6	Quinta da Lixa Loureiro - 2018
White	Casa Santa Eulalia 'Plainas' Loureiro	2018	₹	710.55	6.4	0.37	0.2	5.6	0.117	61	183	0.99459	3.24	0.43	9.5	6	Casa Santa Eulalia 'Plainas' Loureiro - 2018
White	Loureiro Quinta do Barco	2020	₹	587.79	6.4	0.25	0.57	1	0.062	21	122	0.99238	3	0.4	9.5	6	Loureiro Quinta do Barco - 2020
White	Adega Cooperativa Ponte de Lima Loureiro Colheita Seleccionada	2017	₹	508.13	7	0.44	0.24	12.1	0.056	68	210	0.99718	3.05	0.5	9.5	6	Adega Cooperativa Ponte de Lima Loureiro Colheita Seleccionada - 2017
White	Quinta D'amares Alvarinho - Loureiro	2019	₹	482.94	7	0.15	0.29	16.4	0.058	45	110	0.9978	3.15	0.37	9.7	5	Quinta D'amares Alvarinho - Loureiro - 2019
White	Casal de Ventozela Loureiro	2020	₹	482.94	8	0.27	0.33	1.2	0.05	41	103	0.99002	3	0.45	12.4	5	Casal de Ventozela Loureiro - 2020
White	Quinta das Arcas Conde Villar Loureiro	2017	₹	480.42	6.7	0.37	0.25	2.5	0.028	24	84	0.9909	3.14	0.36	11.7	5	Quinta das Arcas Conde Villar Loureiro - 2017

White	Quinta de Soalheiro 'Allo' Alvarinho - Loureiro	2017	₹	478.00	6.2	0.41	0.22	1.9	0.023	5	56	0.98928	3.04	0.79	13	4	Quinta de Soalheiro 'Allo' Alvarinho - Loureiro - 2017
White	Quinta da Lixa 'Flowers' Loureiro	2020	₹	455.61	6.5	0.36	0.28	3.2	0.037	29	119	0.9908	3.25	0.65	12.4	4	Quinta da Lixa 'Flowers' Loureiro - 2020
White	Quinta de Gomariz QG Loureiro Colheita Seleccionada	2021	₹	419.59	7.9	0.28	0.41	2	0.044	50	152	0.9934	3.45	0.49	10.7	4	Quinta de Gomariz QG Loureiro Colheita Seleccionada - 2021
White	Quinta do Tamariz Loureiro	2019	₹	407.53	7.8	0.21	0.39	1.8	0.034	62	180	0.991	3.09	0.75	12.6	4	Quinta do Tamariz Loureiro - 2019
White	Aveleda Loureiro - Alvarinho Vinho Verde	2020	₹	401.23	6.8	0.35	0.32	2.4	0.048	35	103	0.9911	3.28	0.46	12	4	Aveleda Loureiro - Alvarinho Vinho Verde - 2020
White	Adega cooperativa Ponte de Lima Loureiro	2020	₹	391.84	8	0.44	0.49	9.1	0.031	46	151	0.9926	3.16	0.27	12.7	4	Adega cooperativa Ponte de Lima Loureiro - 2020
White	Adega Ponte da Barca Colheita Seleccionada Alvarinho - Loureiro	2019	₹	384.67	6.2	0.22	0.49	6	0.029	31	128	0.9928	3.41	0.36	11.3	4	Adega Ponte da Barca Colheita Seleccionada Alvarinho - Loureiro - 2019
White	Adega Ponte da Barca 'Estreia' Loureiro Grande Escolha	2019	₹	332.60	7	0.36	0.32	10.05	0.045	37	131	0.99352	3.09	0.33	11.7	4	Adega Ponte da Barca 'Estreia' Loureiro Grande Escolha - 2019

CHAPTER – 3

STATISTICAL ANALYSIS

3.1: TESTING OF HYPOTHESIS (t-TEST BETWEEN THE TWO SAMPLES DRAWN FROM SAME POPULATION)

OBJECTIVE:

Hypothesis is an assumption which may or may not be true about a population parameter. To verify our assumption, we collect data and find out the difference between sample values and population. If no difference or if the difference is small, then the hypothesized value is correct. Generally, two hypotheses must be constructed if one of them is correct, the other one is rejected.

a. Null Hypothesis (H_0) - The hypothesis that we hence assumed is said to be a null hypothesis, i.e., it means that the true difference between the measure of sample and population is nil.

b. Alternative Hypothesis (H_1) - Rejection of H_0 leads to acceptance of alternative hypothesis H_1.

By using the t-test, we can find the difference between the means of two samples drawn from the same population. In this case, we can come up with the conclusion whether both the samples have same mean or not.

Two samples containing 213 and 230 data are drawn from the same population.

The two independent samples from the same population $n_1$and n_2 is given by
$n_1 = 233 \; and \; n_2 = 230$

FORMULA:

The standard error of the difference between the sample means is given by

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Since we have sample from same population, we take the above the formula.

where, $\bar{x}_1$ – Mean of the 1st sample

$\bar{x}_2$ – Mean of the 2nd sample

σ^2 – Square of the standard deviation of the population

n_1 – No of observations in 1st sample

n_2 - No of observations in 2nd sample

3.1.1: t-TEST FOR FIXED ACIDITY OF THE WINE

$$H_0: There\ is\ no\ significant\ difference\ between\ the\ means\ of\ the\ fixed\ acidity.$$

$$H_1: There\ is\ a\ significant\ difference\ between\ the\ means\ of\ the\ fixed\ acidity.$$

OBSERVATIONS FROM EXCEL

TABLE 3.1:

$\bar{x}_1$	7.1939
$\bar{x}_2$	7.2877
n_1	233
n_2	230
σ	1.4823

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(1.4823)^2\left(\frac{1}{233} + \frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.1374$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|7.1939 - 7.2877|}{0.1374}$$

$Z = \mathbf{0.6827}$ = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is less than Critical value so we accept our Null hypothesis H_0.

INFERENCE:

Therefore, we conclude that **“There is no significant difference between the means of the fixed acidity”.**

3.1.2: t-TEST FOR VOLATILE ACIDITY OF THE WINE

$$H_0: There\ is\ no\ significant\ difference\ between\ the\ means\ of\ the\ Volatile\ acidity.$$

$$H_1: There\ is\ a\ significant\ difference\ between\ the\ means\ of\ the\ Volatile\ acidity.$$

OBSERVATIONS FROM EXCEL

TABLE 3.2:

$\bar{x}_1$	0.3404
$\bar{x}_2$	0.3748
n_1	233
n_2	230
σ	0.1569

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(0.1569)^2\left(\frac{1}{233} + \frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.0145$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|0.3404 - 0.3748|}{0.0145}$$

$Z = \mathbf{2.3724}$ = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is greater than Critical value so we reject our Null hypothesis H_0 and accepts our alternative hypothesis H_1 .

INFERENCE:

Therefore, we conclude that **"There is a significant difference between the means of the Volatile acidity".**

3.1.3: t-TEST FOR CITRIC ACID OF THE WINE

H_0: *There is no significant difference between the means of the Citric acid.*

H_1: *There is a significant difference between the means of the Citric acid.*

OBSERVATIONS FROM EXCEL

TABLE 3.3:

$\bar{x}_1$	0.3345

$\bar{x}_2$	0.3283
n_1	233
n_2	230
σ	0.1581

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(0.1581)^2\left(\frac{1}{233}+\frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.0147$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|0.3345 - 0.3283|}{0.0147}$$

$Z = \mathbf{0.4217}$ = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is less than Critical value so we accept our Null hypothesis H_0.

INFERENCE:

Therefore, we conclude that **"There is no significant difference between the means of the Citric acid".**

3.1.4: t-TEST FOR RESIDUAL SUGAR OF THE WINE

H_0: *There is no significant difference between the means of the Residual sugar.*

H_1: *There is a significant difference between the means of the Residual sugar.*

OBSERVATIONS FROM EXCEL

TABLE 3.4:

$\bar{x}_1$	5.7864
$\bar{x}_2$	5.7275
n_1	233
n_2	230

σ	4.6930

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(4.6930)^2 \left(\frac{1}{233} + \frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.4362$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|5.7864 - 5.7275|}{0.4362}$$

$Z = \mathbf{0.1350}$ = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is less than Critical value so we accept our Null hypothesis H_0.

INFERENCE:

Therefore, we conclude that **"There is no significant difference between the means of the Residual sugar".**

3.1.5: t-TEST FOR CHLORIDES OF THE WINE

$$H_0: There\ is\ no\ significant\ difference\ between\ the\ means\ of\ the\ Chlorides.$$

$$H_1: There\ is\ a\ significant\ difference\ between\ the\ means\ of\ the\ chlorides.$$

OBSERVATIONS FROM EXCEL

TABLE 3.5:

$\bar{x}_1$	0.0525
$\bar{x}_2$	0.0652
n_1	233
n_2	230
σ	0.0371

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(0.0371)^2\left(\frac{1}{233} + \frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.0034$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|0.0525 - 0.0652|}{0.0034}$$

$Z = \mathbf{3.7353}$ = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is greater than Critical value so we reject our Null hypothesis H_0.

INFERENCE:

Therefore, we conclude that **"There is a significant difference between the means of the Chlorides".**

3.1.6: t-TEST FOR DENSITY OF THE WINE

$$H_0: There\ is\ no\ significant\ difference\ between\ the\ means\ of\ the\ Density.$$

$$H_1: There\ is\ a\ significant\ difference\ between\ the\ means\ of\ the\ Density.$$

OBSERVATIONS FROM EXCEL

TABLE 3.6:

$\bar{x}_1$	0.9948
$\bar{x}_2$	0.9955
n_1	233
n_2	230
σ	0.0027

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(0.0027)^2\left(\frac{1}{233} + \frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.00025$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|0.9948 - 0.9955|}{0.00025}$$

Z = **2.8** = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is greater than Critical value so we reject our Null hypothesis H_0.

INFERENCE:

Therefore, we conclude that **"There is a significant difference between the means of the density".**

3.1.7: t-TEST FOR pH OF THE WINE

$$H_0: There\ is\ no\ significant\ difference\ between\ the\ means\ of\ the\ pH.$$

$$H_1: There\ is\ a\ significant\ difference\ between\ the\ means\ of\ the\ pH.$$

OBSERVATIONS FROM EXCEL

TABLE 3.7:

$\bar{x}_1$	3.2079
$\bar{x}_2$	3.2195
n_1	233
n_2	230
σ	0.1565

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(0.1565)^2 \left(\frac{1}{233} + \frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.0145$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|3.2079 - 3.2195|}{0.0145}$$

Z = **0.8** = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is less than Critical value so we accept our Null hypothesis H_0.

INFERENCE:

Therefore, we conclude that **"There is no significant difference between the means of the pH".**

3.1.8: t-TEST FOR SULPHITES OF THE WINE

$$H_0: There\ is\ no\ significant\ difference\ between\ the\ means\ of\ the\ sulphites.$$

$$H_1: There\ is\ a\ significant\ difference\ between\ the\ means\ of\ the\ sulphites.$$

OBSERVATIONS FROM EXCEL

TABLE 3.8:

$\bar{x}_1$	0.5261
$\bar{x}_2$	0.5196
n_1	233
n_2	230
σ	0.1699

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(0.1699)^2\left(\frac{1}{233} + \frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.0158$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|0.5261 - 0.5196|}{0.0158}$$

$Z = \mathbf{0.4114}$ = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is less than Critical value so we accept our Null hypothesis H_0.

INFERENCE:

Therefore, we conclude that **"There is no significant difference between the means of the Sulphites".**

3.1.9: t-TEST FOR ALCOHOL OF THE WINE

$$H_0: There\ is\ no\ significant\ difference\ between\ the\ means\ of\ the\ alcohol.$$

$$H_1: There\ is\ a\ significant\ difference\ between\ the\ means\ of\ the\ alcohol.$$

OBSERVATIONS FROM EXCEL

TABLE 3.9:

$\bar{x}_1$	10.5293
$\bar{x}_2$	9.9895
n_1	233
n_2	230
σ	1.1532

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(1.1532)^2\left(\frac{1}{233} + \frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.1072$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|10.5293 - 9.9895|}{0.1072}$$

$Z = \mathbf{5.0354}$ = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is greater than Critical value so we reject our Null hypothesis H_0.

INFERENCE:

Therefore, we conclude that **"There is a significant difference between the means of the Alcohol".**

3.1.10: t-TEST FOR QUALITY OF THE WINE

$$H_0: There\ is\ no\ significant\ difference\ between\ the\ means\ of\ the\ Quality.$$

$$H_1: There\ is\ a\ significant\ difference\ between\ the\ means\ of\ the\ Quality.$$

OBSERVATIONS FROM EXCEL

TABLE 3.10:

$\bar{x}_1$	6.2112
$\bar{x}_2$	6.6943
n_1	233
n_2	230
σ	1.0889

Now, substituting all the values in the standard error we get,

$$S.E(\bar{x}_1 - \bar{x}_2) = \sqrt{(1.0889)^2\left(\frac{1}{233} + \frac{1}{230}\right)}$$

$$S.E(\bar{x}_1 - \bar{x}_2) = 0.1012$$

Test statistic is given by,

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{S.E(\bar{x}_1 - \bar{x}_2)}$$

$$Z = \frac{|6.2112 - 6.6943|}{0.1012}$$

Z = **4.774** = Calculated value

Critical value at 5% LOS, the table value = 1.96

CONCLUSION:

Since Test statistic is greater than Critical value so we reject our Null hypothesis H_0.

INFERENCE:

Therefore, we conclude that **"There is a significant difference between the means of Quality".**

CONCLUSION:

From the above cases, we can conclude that some of the constituents of the samples drawn from the same population has no significant difference between the constituents of the samples and some has significant difference. So, we can represent the whole population by the sample that we have taken.

3.2: CORRELATION

OBJECTIVE:

In statistics, correlation is **a statistic that establishes the relationship between two variables**. In other words, it is the measure of association of variables. Correlation is the numerical measurement showing the degree of correlation between the two variables.

In this section, we will discuss about the relationship between the constituents of the wine and the quality of the wine. By finding the relationship between them, we can talk about contribution of the constituents to the wine to the quality of the wine.

CORRELATION FORMULA USED IN THE CALCULATION

$$r = \frac{N\sum dxdy - \sum dx\sum dy}{\sqrt{N\sum dx^2 - (\sum dx)^2} \times \sqrt{N}\sum dy^2 - (\sum dy)^2}$$

Where **N** – No of observations

$\boldsymbol{\sum dxdy}$ – Summation of $dxdy$

$\boldsymbol{\sum dx}$ – Summation of dx

$\boldsymbol{\sum dy}$ – Summation of dy

$\boldsymbol{\sum dx^2}$ – Summation of the squared value of dx

$\boldsymbol{\sum dy^2}$ – Summation of the squared value of dy

3.2.1: CORRELATION BETWEEN THE CONSTITUENTS OF WINE AND THEIR QUALITY

3.2.1.1: RED WINE

CORRELATION BETWEEN THE **FIXED ACIDITY** OF THE **RED WINE** AND THE **QUALITY** OF THE **RED WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.11:

No of observations **(N)**	56
Mean value of X ***(X')***	8.007142857 ≈ 8
Mean value of Y ***(Y')***	5.428571429 ≈ 5
$\boldsymbol{\sum dxdy}$	10.9

$\sum dx$	0.4
$\sum dy$	24
$\sum dx^2$	213.06
$\sum dy^2$	52

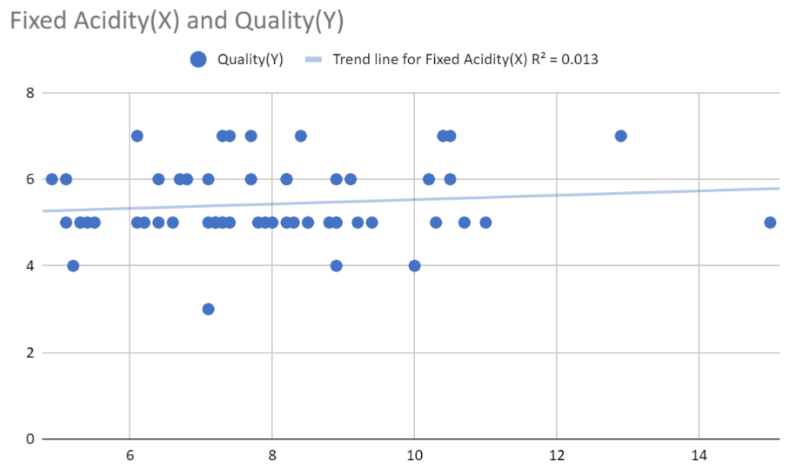
CALCULATION:

$$r = \frac{(56)(10.9)-(0.4)(24)}{\sqrt{(56)(213.06)-(0.4)^2} \times \sqrt{(56)(52)-(24)^2}} = \mathbf{0.1139}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Fixed acidity	Quality	0.1139	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week direct relationship between the fixed acidity of the red wine and the quality of the red wine. They are also approximately 11% weekly positively correlated which means increase in the fixed acidity level in the red wine may or may not increase the quality of the wine.

CORRELATION BETWEEN THE **VOLATILE ACIDITY** OF THE **RED WINE** AND THE **QUALITY** OF THE **RED WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.12:

No of observations **(N)**	56
Mean value of X **(X')**	0.4317857143 ≈ 1
Mean value of Y **(Y')**	5.428571429 ≈ 5
$\sum \boldsymbol{dxdy}$	-14.51
$\sum \boldsymbol{dx}$	-31.82
$\sum \boldsymbol{dy}$	24
$\sum \boldsymbol{dx^2}$	19.4957
$\sum \boldsymbol{dy^2}$	52

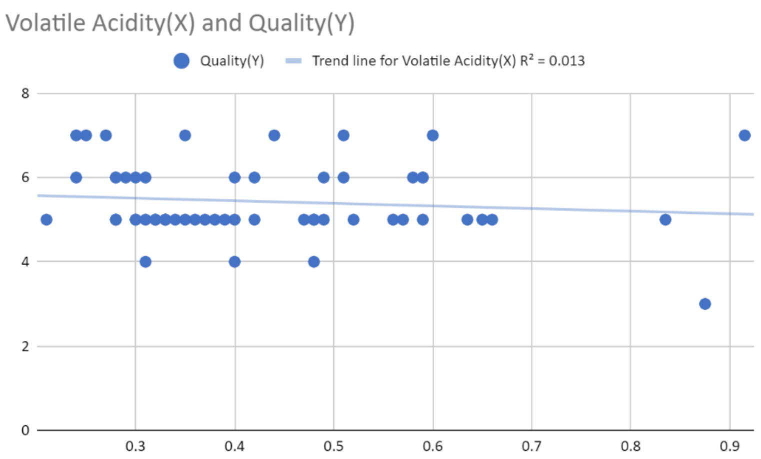
CALCULATION:

$$r = \frac{(56)(-14.51)-(-31.82)(24)}{\sqrt{(56)(19.4957)-(-31.82)^2} \times \sqrt{(56)(52)-(24)^2}} = \mathbf{-0.1136}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Volatile acidity	Quality	-0.1136	Negative

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week inverse relationship between the volatile acidity of the red wine and the quality of the red wine. They are also approximately 11% weekly negatively correlated which means decrease in the volatile acidity level in the red wine may or may not increase the quality of the wine.

CORRELATION BETWEEN THE **CITRIC ACID** OF THE **RED WINE** AND THE **QUALITY** OF THE **RED WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.13:

No of observations **(N)**	56
Mean value of X **(*X'*)**	0.3142857143 ≈ 1
Mean value of Y **(*Y'*)**	5.428571429 ≈ 5
$\sum dxdy$	-14.14
$\sum dx$	-38.4
$\sum dy$	24
$\sum dx^2$	28.4748
$\sum dy^2$	52

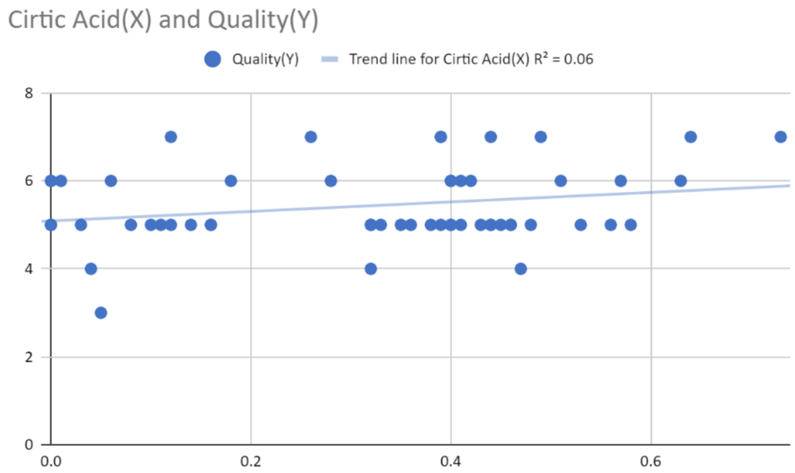
CALCULATION:

$$r = \frac{(56)(-14.14)-(-38.4)(24)}{\sqrt{(56)(28.4748)-(-38.4)^2} \times \sqrt{(56)(52)-(24)^2}} = \mathbf{0.2451}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Citric acid	Quality	0.2451	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a week direct relationship between the citric acid of the red wine and the quality of the red wine. They are also approximately 24% weekly positively correlated which means increase in the citric acid level in the red wine can increase the quality of the wine but to certain extinct as I have already mentioned in introduction that everything needs to be certain limit.

CORRELATION BETWEEN THE **RESIDUAL SUGAR** OF THE **RED WINE** AND THE **QUALITY** OF THE **RED WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.14:

No of observations **(N)**	56
Mean value of X **(*X'*)**	3.155357143 ≈ 3
Mean value of Y **(*Y'*)**	5.428571429 ≈ 5
$\sum \boldsymbol{dxdy}$	40
$\sum \boldsymbol{dx}$	8.7
$\sum \boldsymbol{dy}$	24
$\sum \boldsymbol{dx^2}$	364.885
$\sum \boldsymbol{dy^2}$	52

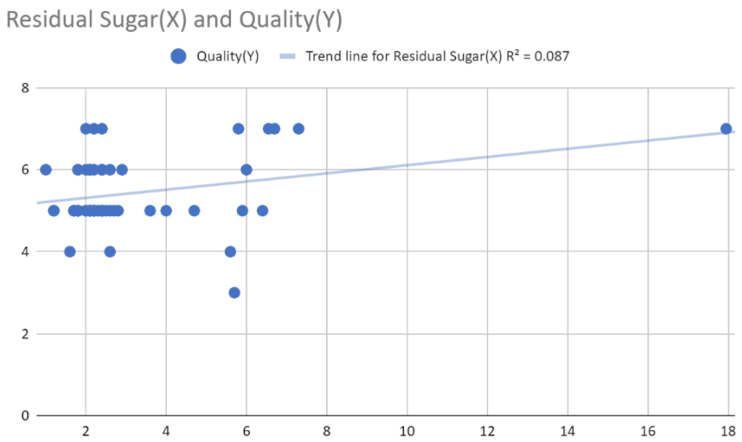
CALCULATION:

r = $\frac{(56)(40)-(8.7)(24)}{\sqrt{(56)(364.885)-(8.7)^2} \times \sqrt{(56)(52)-(24)^2}}$ = **0.2945**

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Residual sugar	Quality	0.2945	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a week direct relationship between the residual sugar of the red wine and the quality of the red wine. They are also approximately 29% weekly positively correlated which means increase in the residual sugar level in the red wine can increase the quality of the wine.

CORRELATION BETWEEN THE **CHLORIDES** OF THE **RED WINE** AND THE **QUALITY** OF THE **RED WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.15:

No of observations **(N)**	56
Mean value of X **(*X'*)**	0.06871428571 ≈ 0.05
Mean value of Y **(*Y'*)**	5.428571429 ≈ 5
$\sum \boldsymbol{dxdy}$	0.388
$\sum \boldsymbol{dx}$	1.048
$\sum \boldsymbol{dy}$	24
$\sum \boldsymbol{dx^2}$	0.044642
$\sum \boldsymbol{dy^2}$	52

CALCULATION:

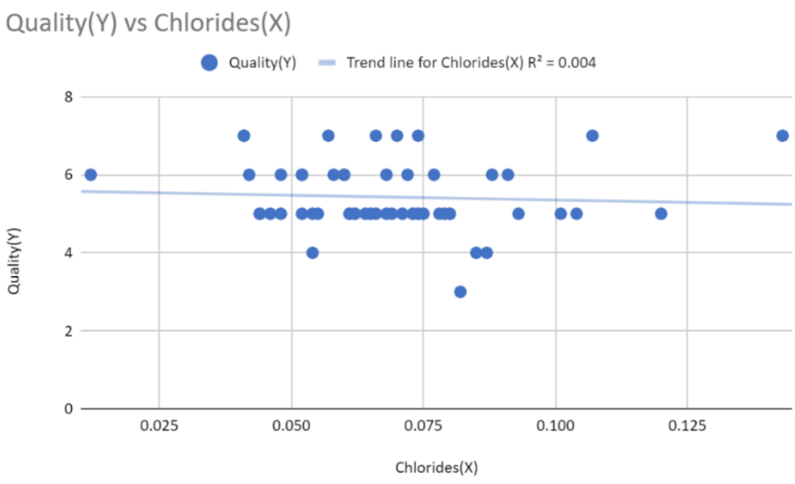
$$r = \frac{(56)(0.388)-(1.048)(24)}{\sqrt{(56)(0.044642)-(1.048)^2} \times \sqrt{(56)(52)-(24)^2}} = \mathbf{-0.05984}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE

Chlorides	Quality	-0.05984	Negative

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week inverse relationship between the chlorides of the red wine and the quality of the red wine. They are also approximately 5% weekly negatively correlated which means decrease in the chlorides (salt level) in the red wine doesn't make any change in the quality of the wine. Here by, I can say that they are independent of each other.

CORRELATION BETWEEN THE **DENSITY** OF THE **RED WINE** AND THE **QUALITY** OF THE **RED WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.16:

No of observations **(N)**	56
Mean value of X ***(X')***	0.9956426786 ≈ 1
Mean value of Y ***(Y')***	5.428571429 ≈ 5
$\sum \boldsymbol{dxdy}$	-0.0981
$\sum \boldsymbol{dx}$	-0.24401
$\sum \boldsymbol{dy}$	24
$\sum \boldsymbol{dx^2}$	0.0013824865
$\sum \boldsymbol{dy^2}$	52

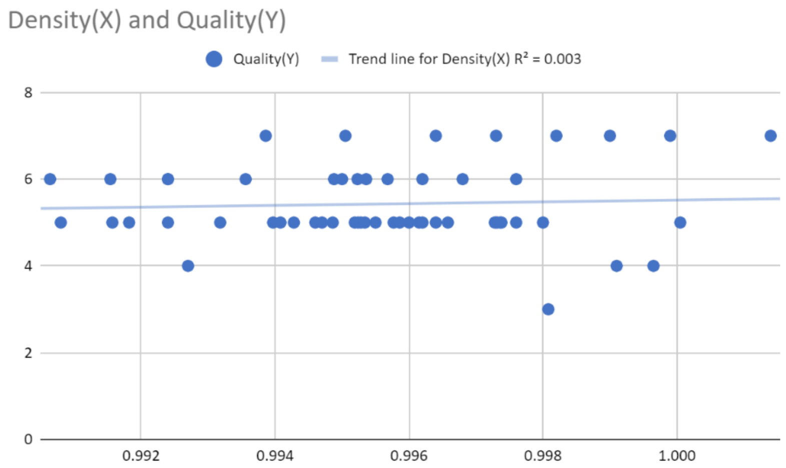
CALCULATION:

$$r = \frac{(56)(-0.0981)-(-0.24401)(24)}{\sqrt{(56)(0.0013824865)}-(-0.24401)^2 \times \sqrt{(56)(52)}-(24)^2} = \mathbf{0.05611}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Density	Quality	0.05611	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week direct relationship between the density of the red wine and the quality of the red wine. They are also approximately 5% weekly positively correlated which means increase in the density of the red wine doesn't make any change in the quality of the wine. Moreover, they are independent of each other.

CORRELATION BETWEEN THE **pH** OF THE **RED WINE** AND THE **QUALITY** OF THE **RED WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.17:

No of observations **(N)**	56
Mean value of X **(*X'*)**	3.332321429 ≈ 3
Mean value of Y **(*Y'*)**	5.428571429 ≈ 5
$\sum dxdy$	5.72
$\sum dx$	18.61
$\sum dy$	24
$\sum dx^2$	7.2761
$\sum dy^2$	52

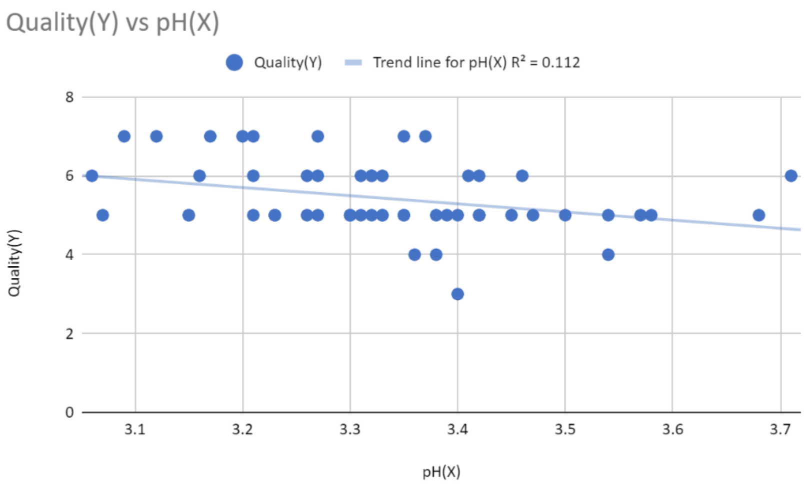
CALCULATION:

$$r = \frac{(56)(5.72)-(18.61)(24)}{\sqrt{(56)(7.2761)-(18.61)^2} \times \sqrt{(56)(52)-(24)^2}} = \mathbf{-0.3343}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
pH	Quality	-0.3343	Negative

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a week inverse relationship between the pH level of the red wine and the quality of the red wine. They are also approximately 33% weekly negatively correlated which means decrease in the pH level of red wine can make changes in the quality of the wine.

CORRELATION BETWEEN THE **SULPHITES** OF THE **RED WINE** AND THE **QUALITY** OF THE **RED WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.18:

No of observations **(N)**	56
Mean value of X **(X')**	0.7225 ≈ 0.7
Mean value of Y **(Y')**	5.428571429 ≈ 5
$\sum dxdy$	-1.4

$\sum dx$	1.26
$\sum dy$	24
$\sum dx^2$	1.428
$\sum dy^2$	52

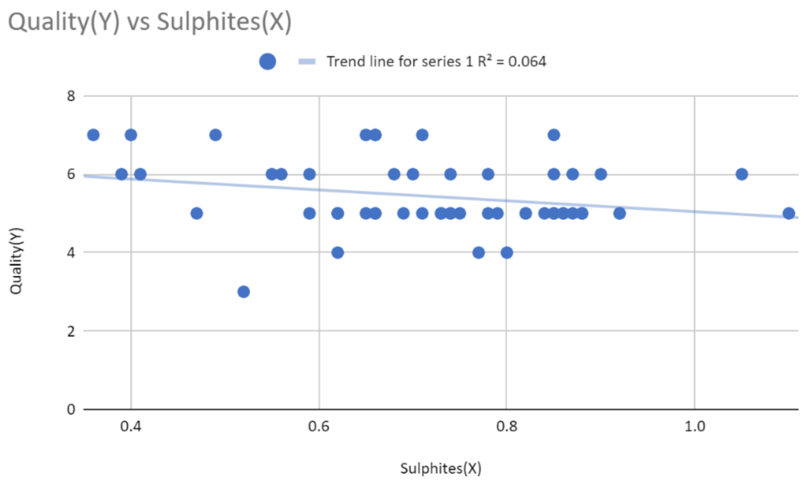
CALCULATION:

$$r = \frac{(56)(-1.4)-(1.26)(24)}{\sqrt{(56)(1.428)-(1.26)^2} \times \sqrt{(56)(52)-(24)^2}} = \mathbf{-0.2539}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Sulphites	Quality	-0.2539	Negative

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a week inverse relationship between the sulphites content of the red wine and the quality of the red wine. They are also approximately 25% weekly negatively correlated which means decrease in sulphites in the red wine can make changes in the quality of the red wine.

CORRELATION BETWEEN THE **ALCOHOL** OF THE **RED WINE** AND THE **QUALITY** OF THE **RED WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.19:

No of observations **(N)**	56
Mean value of X **(*X'*)**	11.49642857 ≈ 11

Mean value of Y (Y')	$5.428571429 \approx 5$
$\sum dxdy$	17.5
$\sum dx$	-28.2
$\sum dy$	24
$\sum dx^2$	105.06
$\sum dy^2$	52

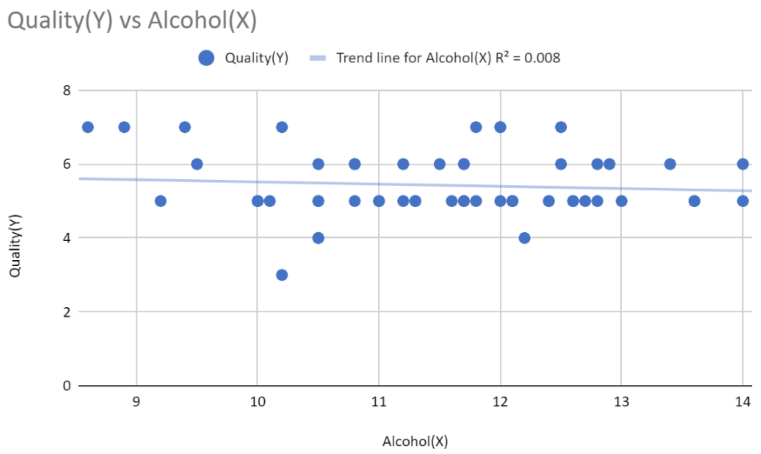
CALCULATION:

$$r = \frac{(56)(17.5)-(-28.2)(24)}{\sqrt{(56)(105.06)}-(-28.2)^2 \times \sqrt{(56)(52)}-(24)^2} = \mathbf{-0.0879}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Alcohol	Quality	-0.0879	Negative

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week inverse relationship between alcohol content of the red wine and the quality of the red wine. They are also approximately 8% weekly negatively correlated which means decrease in alcohol content in the red wine doesn't make changes in the quality of the red wine.

CONCLUSION:

By looking through the about results on correlation coefficients between the constituents of the red wine and the quality of the red wine, we can say that among all the constituents the pH level present in the red wine contributes much to the quality of the red wine. i.e., decrease in pH level in the red wine increases the quality of the red wine. Simply, we can say that less the pH level in the red wine, more the quality of the red wine.

3.2.1.2: WHITE WINE

CORRELATION BETWEEN THE **FIXED ACIDITY** OF THE **WHITE WINE** AND THE **QUALITY** OF THE **WHITE WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.20:

No of observations **(N)**	176
Mean value of X **(X')**	6.944886364 ≈ 7
Mean value of Y **(Y')**	6.460227273 ≈ 7
$\sum dxdy$	45.4
$\sum dx$	-9.7
$\sum dy$	-95
$\sum dx^2$	163.81
$\sum dy^2$	321

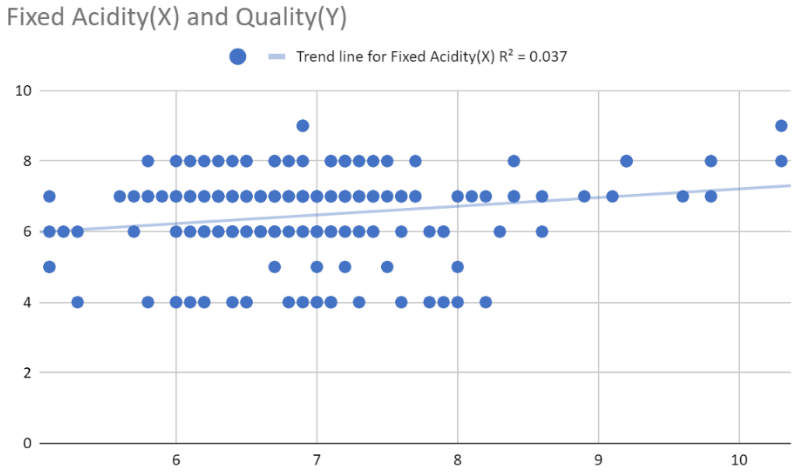
CALCULATION:

$$r = \frac{(176)(45.4)-(-9.7)(-95)}{\sqrt{(176)(163.81)-(-9.7)^2} \times \sqrt{(176)(321)-(-95)^2}} = \mathbf{0.1838}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Fixed acidity	Quality	0.1838	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week direct relationship between fixed acidity of the white wine and the quality of the white wine. They are also approximately 18% weekly positively correlated which means increase in fixed acidity of the white wine may or may not make changes in the quality of the white wine.

CORRELATION BETWEEN THE **VOLATILE ACIDITY** OF THE **WHITE WINE** AND THE **QUALITY** OF THE **WHITE WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.21:

No of observations **(N)**	176
Mean value of X **(X')**	0.3112784091 ≈ 0.3
Mean value of Y **(Y')**	6.460227273 ≈ 7
$\sum dxdy$	3.465
$\sum dx$	1.985
$\sum dy$	-95
$\sum dx^2$	2.412725
$\sum dy^2$	321

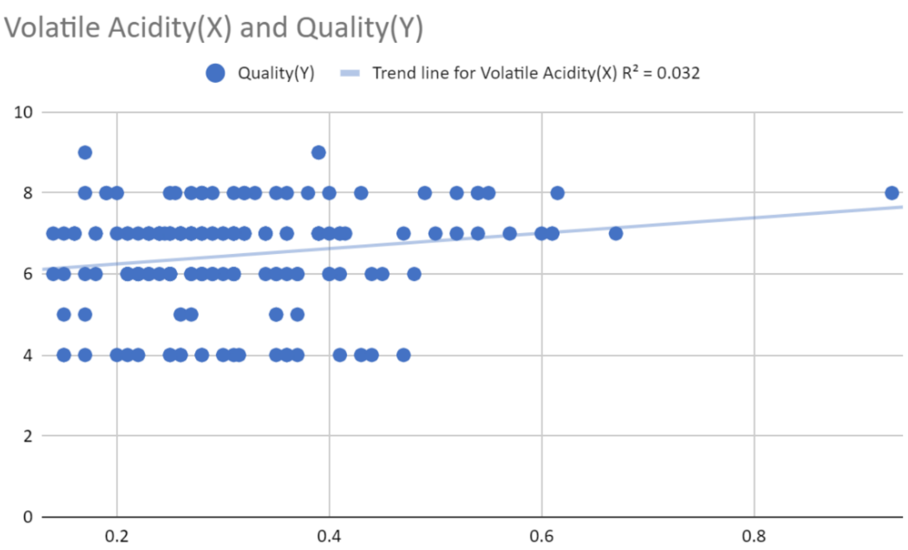
CALCULATION:

$$r = \frac{(176)(3.465)-(1.985)(-95)}{\sqrt{(176)(2.412725)-(1.985)^2} \times \sqrt{(176)(321)-(-95)^2}} = \mathbf{0.1787}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Volatile acidity	Quality	0.1787	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week direct relationship between volatile acidity of the white wine and the quality of the white wine. They are also approximately 17% weekly positively correlated which means increase in volatile acidity of the white wine may or may not make changes in the quality of the white wine.

CORRELATION BETWEEN THE **CITRIC ACID** OF THE **WHITE WINE** AND THE **QUALITY** OF THE **WHITE WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.22:

No of observations **(N)**	176
Mean value of X **(*X'*)**	0.3409659091 ≈ 0.3
Mean value of Y **(*Y'*)**	6.460227273 ≈ 7
$\sum dxdy$	-3.82
$\sum dx$	7.21
$\sum dy$	-95
$\sum dx^2$	2.9569
$\sum dy^2$	321

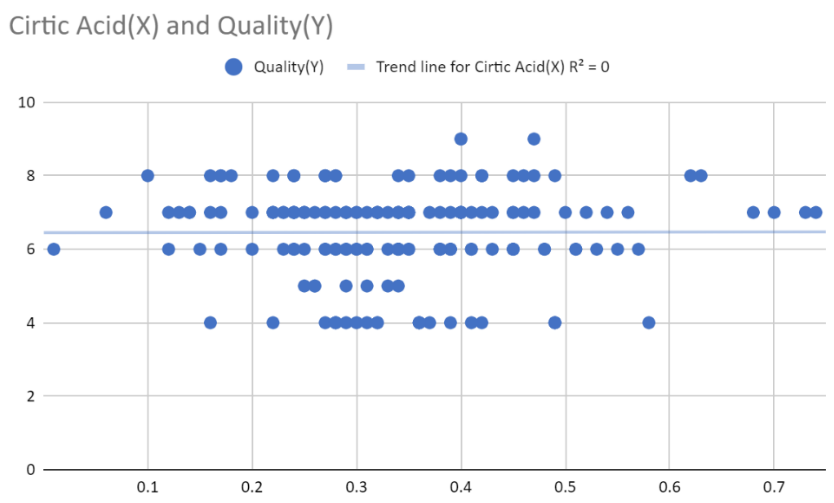
CALCULATION:

$$r = \frac{(176)(-3.82)-(7.21)(-95)}{\sqrt{(176)(2.9569)-(7.21)^2} \times \sqrt{(176)(321)-(-95)^2}} = \mathbf{0.0027}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Citric acid	Quality	0.0027	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week direct relationship between citric acid of the white wine and the quality of the white wine. They are also approximately 0.2% weekly positively correlated which means increase in citric acid present in the white wine doesn't affect the quality of the white wine. Simply, they are independent of each other.

CORRELATION BETWEEN THE **RESIDUAL SUGAR** OF THE **WHITE WINE** AND THE **QUALITY** OF THE **WHITE WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.23:

No of observations **(N)**	176
Mean value of X ***(X')***	6.623579545 ≈ 6
Mean value of Y ***(Y')***	6.460227273 ≈ 7
$\sum dxdy$	60

$\sum dx$	-66.25
$\sum dy$	-95
$\sum dx^2$	3982.6375
$\sum dy^2$	321

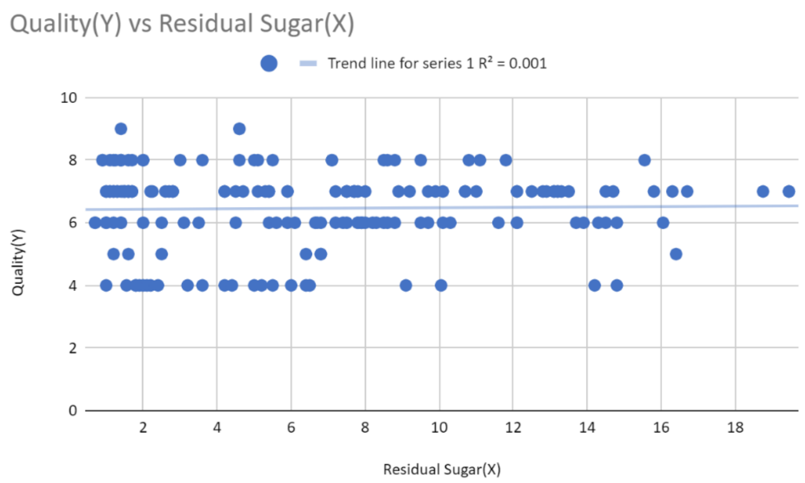
CALCULATION:

$$r = \frac{(176)(60)-(-66.25)(-95)}{\sqrt{(176)(3982.6375)-(-66.25)^2} \times \sqrt{(176)(321)-(-95)^2}} = \mathbf{0.0235}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Residual sugar	Quality	0.0235	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week direct relationship between residual sugar of the white wine and the quality of the white wine. They are also approximately 2% weekly positively correlated which means increase in residual sugar present in the white wine doesn't make any changes in the quality of the white wine. Simply, they are independent of each other since the correlation between them is almost equal to zero.

CORRELATION BETWEEN THE **CHLORIDES** OF THE **WHITE WINE** AND THE **QUALITY** OF THE **WHITE WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.24:

No of observations **(N)**	176
Mean value of X **(*X'*)**	6.623579545 ≈ 6
Mean value of Y **(*Y'*)**	6.460227273 ≈ 7
$\sum dxdy$	0.929
$\sum dx$	-0.479
$\sum dy$	-95
$\sum dx^2$	0.138119
$\sum dy^2$	321

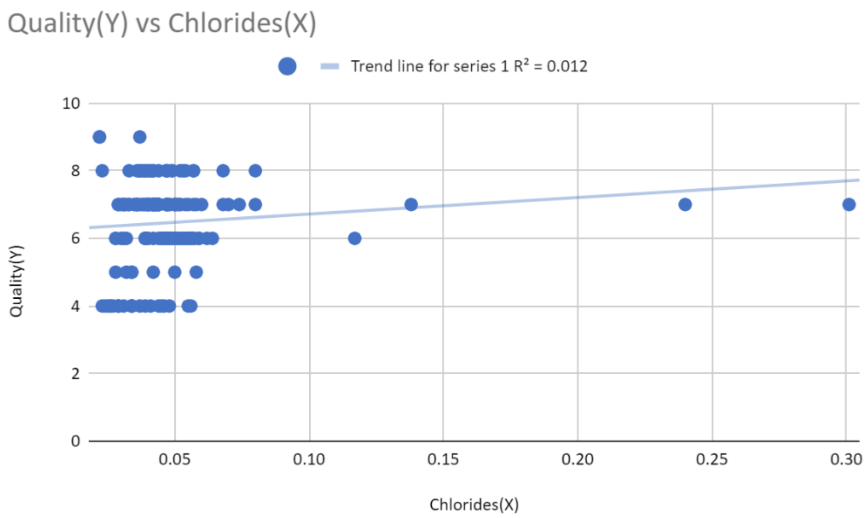
CALCULATION:

$$r = \frac{(176)(0.929)-(-0.479)(-95)}{\sqrt{(176)(0.138119)-(-0.479)^2} \times \sqrt{(176)(321)-(-95)^2}} = \mathbf{0.1104}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Chlorides	Quality	0.1104	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week direct relationship between chlorides of the white wine and the quality of the white wine. They are also approximately 11% weekly positively correlated which means increase in chlorides (salt content) in white wine doesn't may or may not make increase in the quality of the white wine.

CORRELATION BETWEEN THE **DENSITY** OF THE **WHITE WINE** AND THE **QUALITY** OF THE **WHITE WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.25:

No of observations **(N)**	176
Mean value of X **(*X'*)**	0.9944836364 ≈ 1
Mean value of Y **(*Y'*)**	6.460227273 ≈ 7
$\sum \boldsymbol{dxdy}$	0.68697
$\sum \boldsymbol{dx}$	-0.97088
$\sum \boldsymbol{dy}$	-95
$\sum \boldsymbol{dx^2}$	0.0065469902
$\sum \boldsymbol{dy^2}$	321

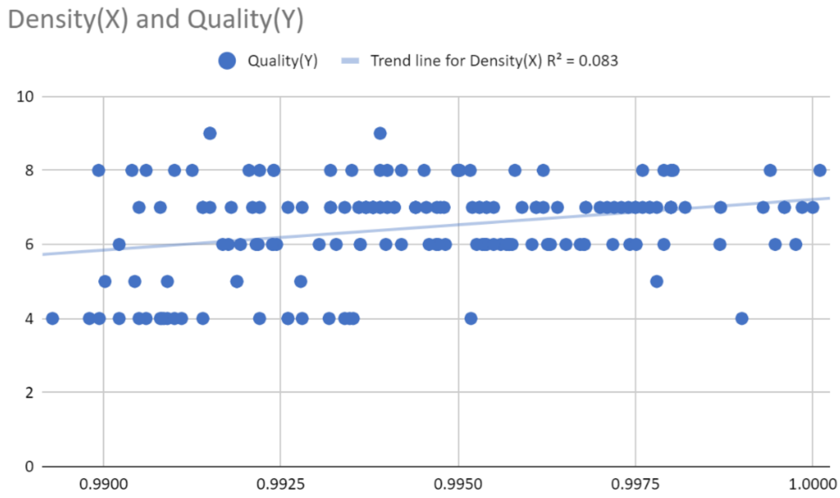
CALCULATION:

$$r = \frac{(176)(0.68697)-(-0.97088)(-95)}{\sqrt{(176)(0.0065469902)-(-0.97088)^2} \times \sqrt{(176)(321)-(-95)^2}} = \mathbf{0.2874}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Density	Quality	0.2874	Positive

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a week direct relationship between density of the white wine and the quality of the white wine. They are also approximately 28% weekly positively correlated which means increase in density of the white wine might increase in the quality of the white wine.

CORRELATION BETWEEN THE **pH** OF THE **WHITE WINE** AND THE **QUALITY** OF THE **WHITE WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.26:

No of observations **(N)**	176
Mean value of X (X')	3.168352273 ≈ 3
Mean value of Y (Y')	6.460227273 ≈ 7
$\sum dxdy$	-20.48
$\sum dx$	29.63
$\sum dy$	-95
$\sum dx^2$	8.9281
$\sum dy^2$	321

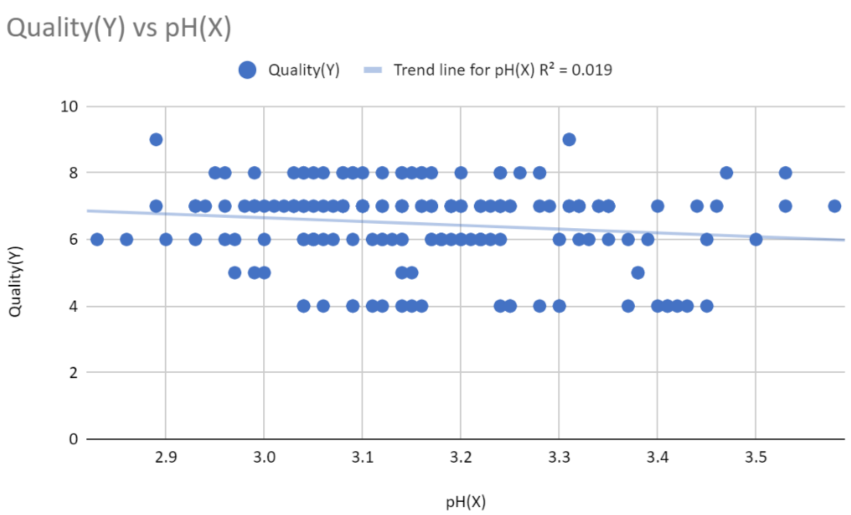
CALCULATION:

$$r = \frac{(176)(-20.48)-(29.63)(-95)}{\sqrt{(176)(8.9281)-(29.63)^2} \times \sqrt{(176)(321)-(-95)^2}} = \mathbf{-0.1376}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
pH	Quality	-0.1376	Negative

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a very week inverse relationship between pH level in white wine and the quality of the white wine. They are also approximately 13% weekly negatively correlated which means decrease in pH level in white wine may or may not increase the quality of the white wine since they are almost independent to each other.

CORRELATION BETWEEN THE **SULPHITES** OF THE **WHITE WINE** AND THE **QUALITY** OF THE **WHITE WINE**

OBSERVATIONS FROM EXCEL

TABLE 3.27:

No of observations **(N)**	176
Mean value of X **(X')**	0.4636363636 ≈ 0.5
Mean value of Y **(Y')**	6.460227273 ≈ 7
$\sum dxdy$	2.48
$\sum dx$	-6.4
$\sum dy$	-95
$\sum dx^2$	2.1502
$\sum dy^2$	321

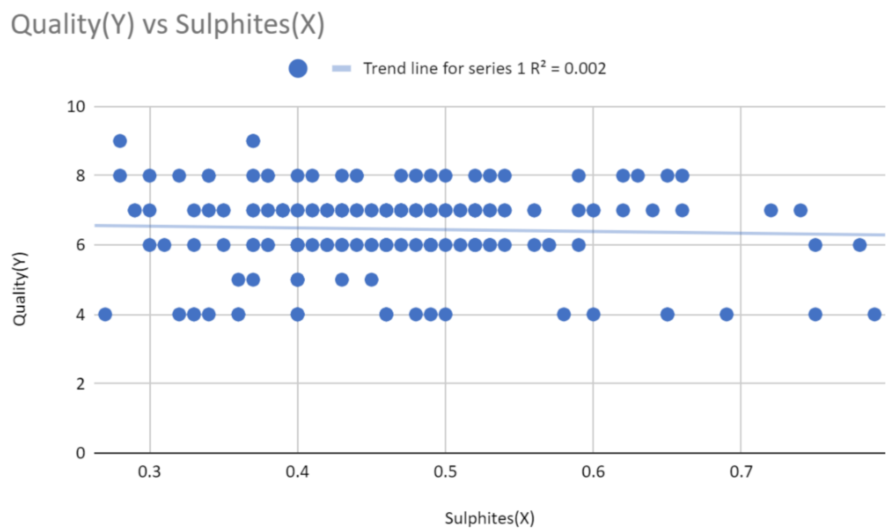
CALCULATION:

$$r = \frac{(176)(2.48)-(-6.4)(-95)}{\sqrt{(176)(2.1502)-(-6.4)^2} \times \sqrt{(176)(321)-(-95)^2}} = \mathbf{-0.04285}$$

<u>CORRELATION TABLE:</u>

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Sulphites	Quality	-0.04285	Negative

SCATTER PLOT:



<u>INFERENCE:</u>

From the above observations we can say that there exists a very week inverse relationship between sulphites present in the white wine and the quality of the white wine. They are also approximately 4% weekly negatively correlated which means decrease in sulphites in white wine doesn't make changes in the quality of the white wine since they are almost independent to each other.

CORRELATION BETWEEN THE **ALCOHOL** OF THE **WHITE WINE** AND THE **QUALITY** OF THE **WHITE WINE**

<u>OBSERVATIONS FROM EXCEL</u>

TABLE 3.28:

No of observations **(N)**	176
Mean value of X ***(X')***	10.22159091 ≈ 10
Mean value of Y ***(Y')***	6.460227273 ≈ 7

$\sum dxdy$	-116.4
$\sum dx$	39
$\sum dy$	-95
$\sum dx^2$	240.38
$\sum dy^2$	321

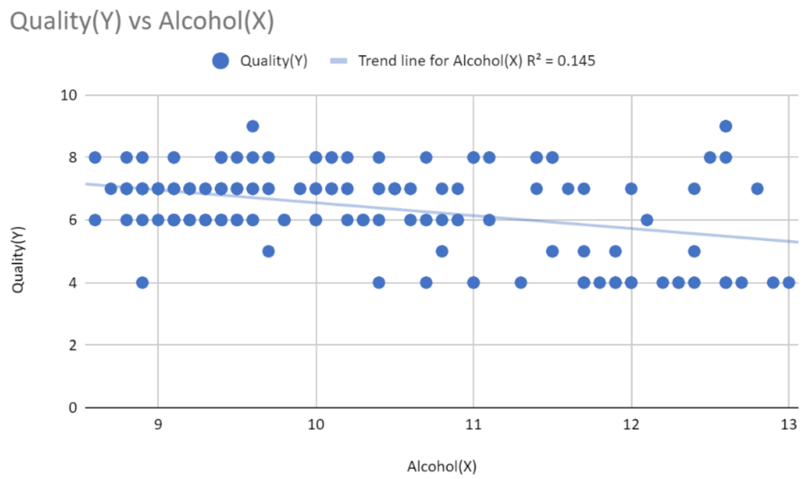
CALCULATION:

$$r = \frac{(176)(-116.4)-(39)(-95)}{\sqrt{(176)(240.38)-(39)^2} \times \sqrt{(176)(321)-(-95)^2}} = \mathbf{-0.3814}$$

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Alcohol	Quality	-0.3814	Negative

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a week inverse relationship between alcohol content present in the white wine and the quality of the white wine. They are also approximately 38% weekly negatively correlated which means decrease in alcohol in white wine can increase the quality of the white wine.

CONCLUSION:

By looking through the about results on correlation coefficients between the constituents of the white wine and the quality of the white wine, we can say that among all the constituents, the alcohol content present in the white wine contributes much to the quality of the red wine. i.e., decrease in alcohol content in the white wine increases the quality of the white wine. Simply, we can say that less the alcohol content in the white wine, more the quality of the red wine.

3.2.2: CORRELATION BETWEEN THE MANUFACTURING YEAR AND THE QUALITY

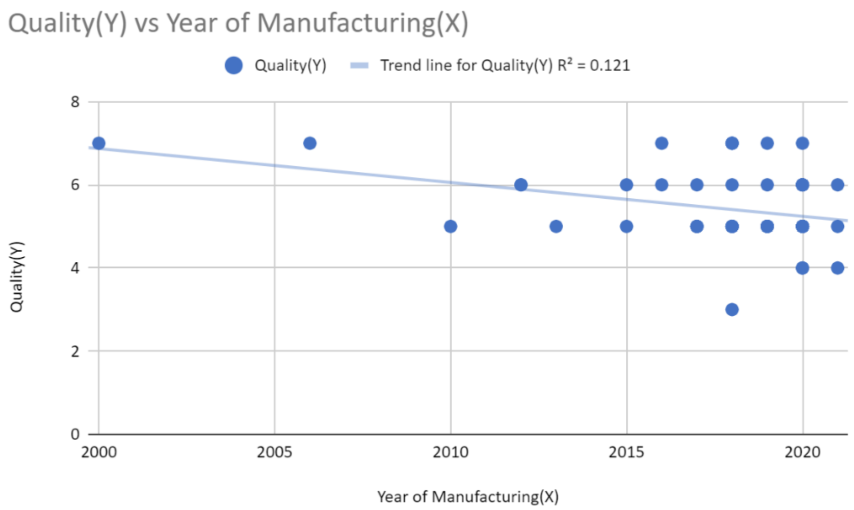
3.2.2.1: RED WINE

OBSERVATION FROM R STUDIO

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Year (Red wine)	Quality	-0.3482	Negative

SCATTER PLOT:



INFERENCE:

From the above observations we can say that there exists a week inverse relationship between year of the red wine and the quality of the red wine. They are also approximately 34% weekly negatively correlated which means older the red wine higher the quality of the red wine.

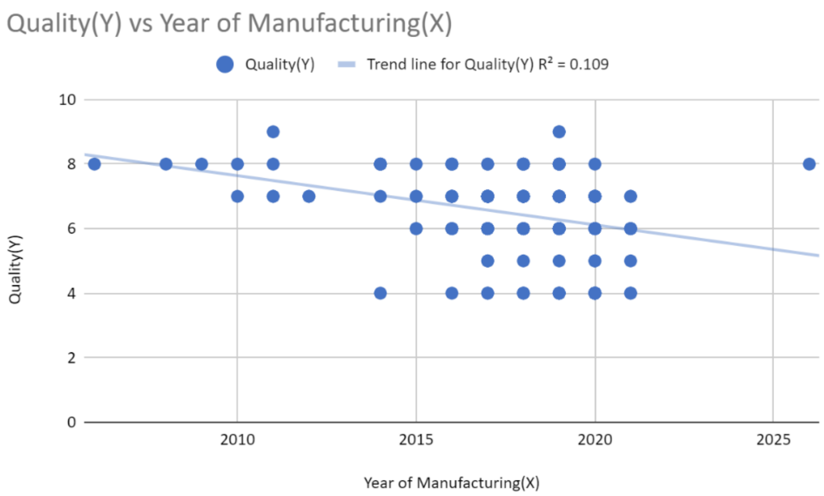
3.2.2.2: WHITE WINE

OBSERVATION FROM R STUDIO

CORRELATION TABLE:

VARIABLE 1	VARIABLE 2	CO-EFFICIENT OF CORRELATION	POSITIVE/NEGATIVE
Year (White wine)	Quality	-0.33	Negative

SCATTER PLOT:



<u>INFERENCE:</u>

From the above observations we can say that there exists a week inverse relationship between year of the white wine and the quality of the red wine. They are also approximately 33% weekly negatively correlated which means older the white wine higher the quality of the white wine.

CONCLUSION:

From the calculations, we can conclude that addition or taking back of the constituents in the wine may increase or decrease the quality of the wine and the range of r is from -0.4 to 0.4 and among the constituents of the red wine, only pH level of the red wine contributes to the quality of the wine, i.e., there exist an inverse relationship between the pH level and the quality of the wine, whereas in the white wine, the alcohol level contributes more to the quality, i.e., decrease in the level of alcohol can increase the quality of the wine a bit. finding the correlation between the year the quality, we can conclude how the quality varies according to the tear of manufacturing.

3.2.3: R coding for calculating the co-efficient of correlation

```
rm(list = ls())
x = c(values of x (i.e., constituents of the wine))
y = c(values of y(i.e., quality of the wine))
cor.test(x, y,alternative = "two.sided", method = "pearson", exact=FALSE )
```

3.3: REGRESSION ANALYSIS

OBJECTIVE:

The statistical method which helps us to estimate the unknown value of one variable from the known value of the related variable is called regression. The line described in the average relationship between two variables is known as the line of regression.

From the regression analysis, we will be able to find the distribution of the constituents, i.e., whether they are normally distributed or not. We can also find how the quality varies when there is a change in the level of constituents.

3.3.1: REGRESSION ANALYSIS BETWEEN THE CONSTITUENTS OF THE WINE AND THE QUALITY OF THE WINE

3.3.1.1: RED WINE

REGRESSION ANALYSIS BETWEEN THE **FIXED ACIDITY** AND THE **QUALITY** OF THE **RED WINE**

VARIABLE TABLE:

Independent variable	***x***	*Fixed acidity*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*5.0254*
b	*0.05036*

The linear regression equation is as follows,

$$Y = 5.0254 + 0.05036X$$

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.01295. It means that 1.3% of the variability of Y is explained by X.

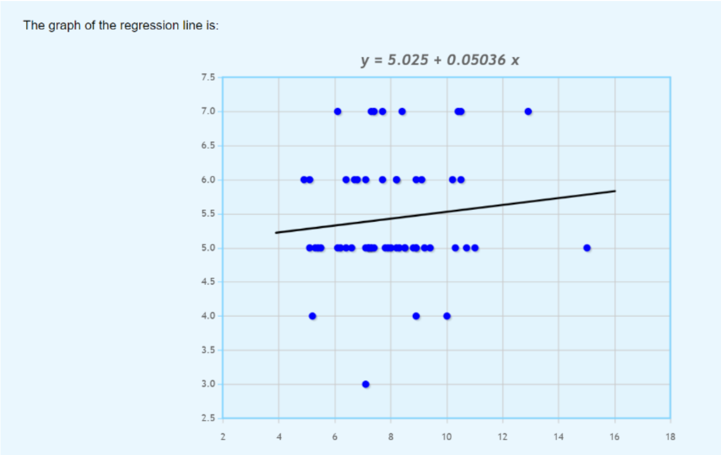
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.0001847. Since the null hypothesis is accepted for the fixed acidity in the testing of hypothesis test that we have done before, we can say that the data are normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of fixed acidity and quality is distributed slightly away from the regression line of Y on X, i.e., they are weekly positively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **VOLATILE ACIDITY** AND THE **QUALITY** OF THE **RED WINE**

VARIABLE TABLE:

Independent variable	***x***	*Volatile acidity*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*5.6949*
b	*0.6168*

The linear regression equation is as follows,

$\boxed{Y = 5.6949 + 0.6168X}$

<u>RESULTS OBTAINED FROM R PROGRAMMING (R-studio)</u>

- r Square (r^2) equals 0.01291. It means that 1.3% of the variability of Y is explained by X.

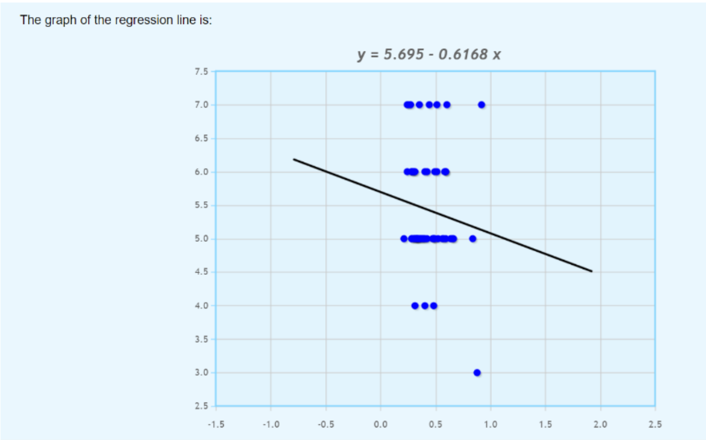
<u>Residual Normality:</u>

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.0001216. Since the null hypothesis is rejected for the volatile acidity in the t-test that we have done before, we can say that the data aren't normally distributed.

<u>Outliers:</u>

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



<u>INFERENCE:</u>

From the above graph of the regression line of Y on X, I can say that the plots of volatile acidity and quality is distributed slightly away from the regression line of Y on X, i.e., they are weekly negatively correlated to each other.

<u>REGRESSION ANALYSIS BETWEEN THE **CITRIC ACID** AND THE **QUALITY** OF THE **RED WINE**</u>

<u>VARIABLE TABLE:</u>

Independent variable	***x***	*Citric acid*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*5.0888*

b	*1.0811*

The linear regression equation is as follows,

$$\boxed{Y = 5.0888 + 1.0811X}$$

<u>RESULTS OBTAINED FROM R PROGRAMMING (R-studio)</u>

- r Square (r^2) equals 0.06005. It means that 6% of the variability of Y is explained by X.

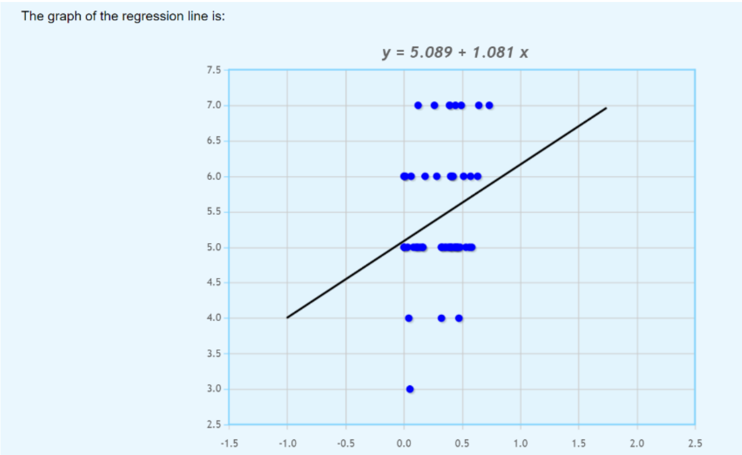
<u>Residual Normality:</u>

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.01437. Since the null hypothesis is accepted for the citric acid in the t-test that we have done before, we can say that the data are normally distributed.

<u>Outliers:</u>

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



<u>INFERENCE:</u>

From the above graph of the regression line of Y on X, I can say that the plots of citric acid and quality is distributed somewhat closer to the regression line of Y on X, i.e., they are weekly positively correlated to each other.

<u>REGRESSION ANALYSIS BETWEEN THE **RESIDUAL SUGAR** AND THE **QUALITY** OF THE **RED WINE**</u>

<u>VARIABLE TABLE:</u>

Independent variable	***x***	*Residual sugar*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*5.1137*
b	*0.09977*

The linear regression equation is as follows,

$$\boxed{Y = 5.1137 + 0.09977X}$$

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.08676. It means that 8.6% of the variability of Y is explained by X.

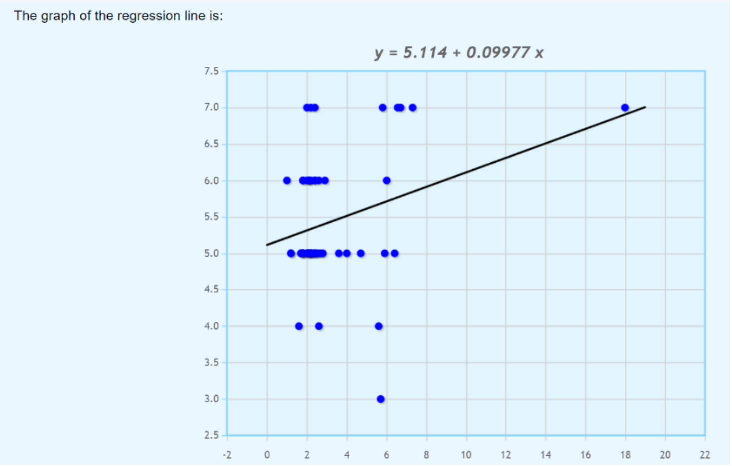
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.0003485. Since the null hypothesis is rejected for the residual sugar in the t-test that we have done before, we can say that the data aren't normally distributed.

Outliers:

The data contains outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of residual sugar and quality is distributed somewhat closer to the regression line of Y on X, i.e., they are weekly positively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **CHLORIDES** AND THE **QUALITY** OF THE **RED WINE**

VARIABLE TABLE:

Independent variable	***x***	*Chlorides*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*5.5964*
b	*2.4428*

The linear regression equation is as follows,

$$\boxed{Y = 5.5964 + 2.4428X}$$

<u>RESULTS OBTAINED FROM R PROGRAMMING (R-studio)</u>

- r Square (r^2) equals 0.003581. It means that 0.4% of the variability of Y is explained by X.

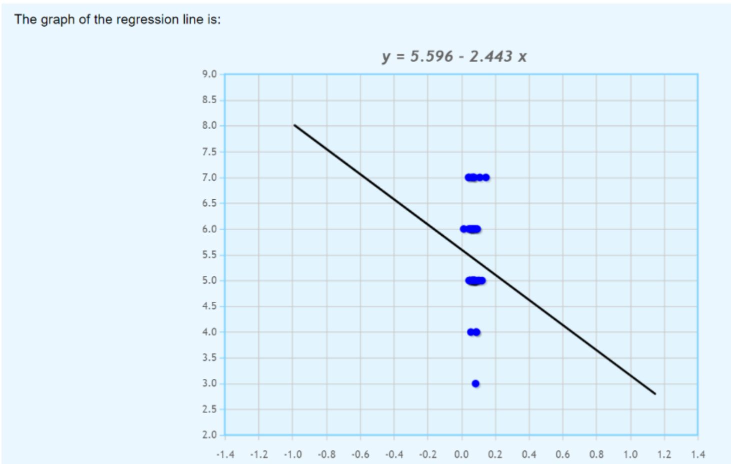
<u>Residual Normality:</u>

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.00001566. Since the null hypothesis is rejected for the chlorides in the t-test that we have done before, we can say that the data aren't normally distributed.

<u>Outliers:</u>

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



<u>INFERENCE:</u>

From the above graph of the regression line of Y on X, I can say that the plots of chlorides and quality is distributed somewhat closer to the regression line of Y on X, i.e., they are weekly negatively correlated to each other.

<u>REGRESSION ANALYSIS BETWEEN THE **DENSITY** AND THE **QUALITY** OF THE **RED WINE**</u>

<u>VARIABLE TABLE:</u>

Independent variable	***x***	*Density*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*-14.7668*
b	*20.2837*

The linear regression equation is as follows,

$$\boxed{Y = -14.7668 + 20.2837X}$$

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.003149. It means that 0.3% of the variability of Y is explained by X.

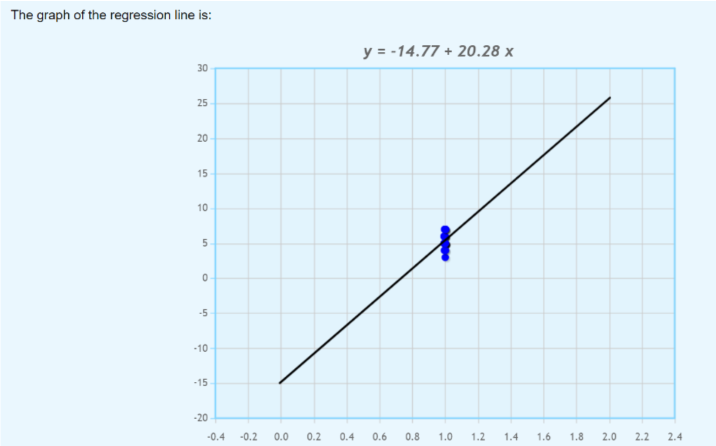
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.00001508. Since the null hypothesis is rejected for the density in the t-test that we have done before, we can say that the data aren't normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of density and quality is distributed very closer to the regression line of Y on X, i.e., they are positively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **pH** AND THE **QUALITY** OF THE **RED WINE**

VARIABLE TABLE:

Independent variable	***x***	*pH*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*12.3146*
b	*2.0664*

The linear regression equation is as follows,

$$\boxed{Y = 12.3146 + 2.0664X}$$

<u>RESULTS OBTAINED FROM R PROGRAMMING (R-studio)</u>

- r Square (r^2) equals 0.1117. It means that 11.2% of the variability of Y is explained by X.

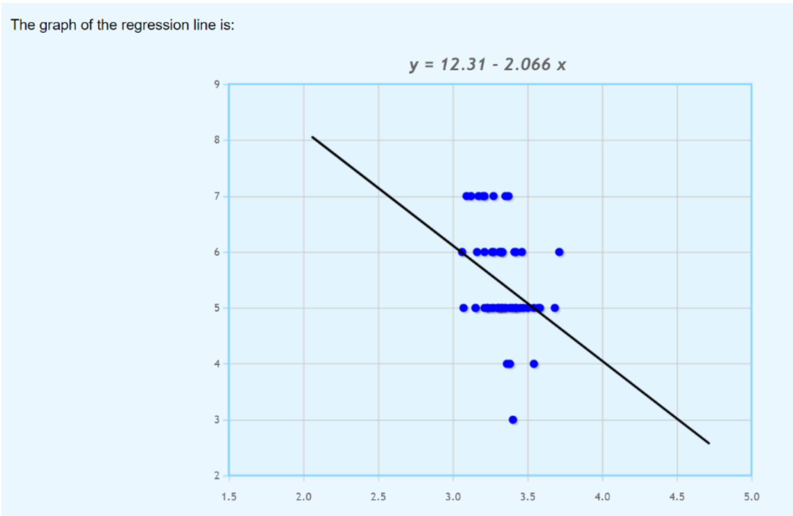
<u>Residual Normality:</u>

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.1029. Since the null hypothesis is accepted for the pH level in the t-test that we have done before, we can say that the data are normally distributed.

<u>Outliers:</u>

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



<u>INFERENCE:</u>

From the above graph of the regression line of Y on X, I can say that the plots of pH level and quality is distributed very closer to the regression line of Y on X, i.e., they are strongly negatively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **SULPHITES** AND THE **QUALITY** OF THE **RED WINE**

VARIABLE TABLE:

Independent variable	***x***	*Sulphites*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*6.43*
b	*1.3861*

The linear regression equation is as follows,

Y = 6.43 + 1.3861X

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.06446. It means that 6.4% of the variability of Y is explained by X.

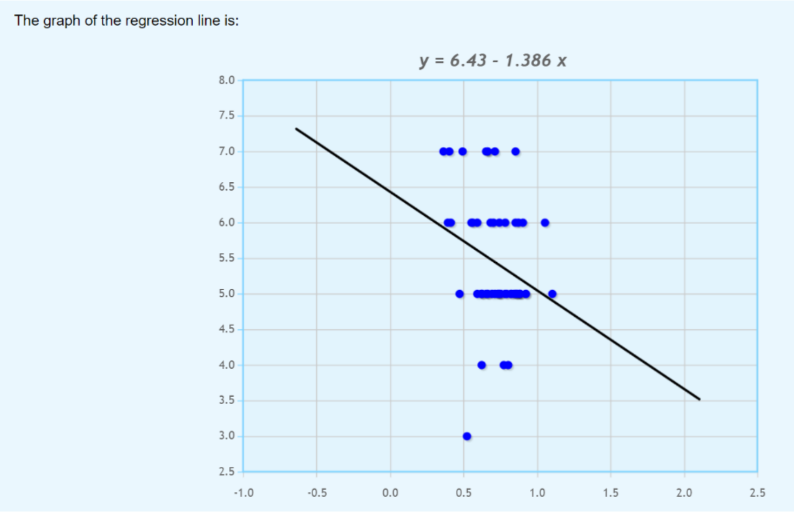
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.005312. Since the null hypothesis is accepted for the sulphites in the t-test that we have done before, we can say that the data are normally distributed.

Outliers:

The data contains outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of sulphites level and quality is distributed closer to the regression line of Y on X, i.e., they are negatively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **ALCOHOL** AND THE **QUALITY** OF THE **RED WINE**

VARIABLE TABLE:

Independent variable	***x***	*Alcohol*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*6.1136*
b	*0.05959*

The linear regression equation is as follows,

$$\boxed{Y = 6.1136 + 0.05959X}$$

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.007734. It means that 0.8% of the variability of Y is explained by X.

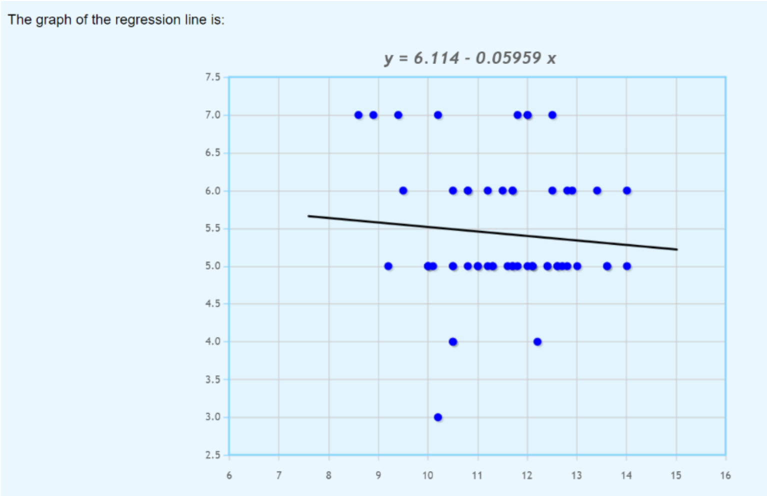
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.0000658. Since the null hypothesis is rejected for the alcohol in the t-test that we have done before, we can say that the data aren't normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of alcohol content and quality is distributed moderately around the regression line of Y on X, i.e., they are negatively correlated to each other.

3.3.1.2: WHITE WINE

REGRESSION ANALYSIS BETWEEN THE **FIXED ACIDITY** AND THE **QUALITY** OF THE **WHITE WINE**

VARIABLE TABLE:

Independent variable	***x***	*Fixed acidity*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*4.7775*
b	*0.2426*

The linear regression equation is as follows,

Y = 4.7775 + 0.2426X

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.03378. It means that 3.4% of the variability of Y is explained by X.

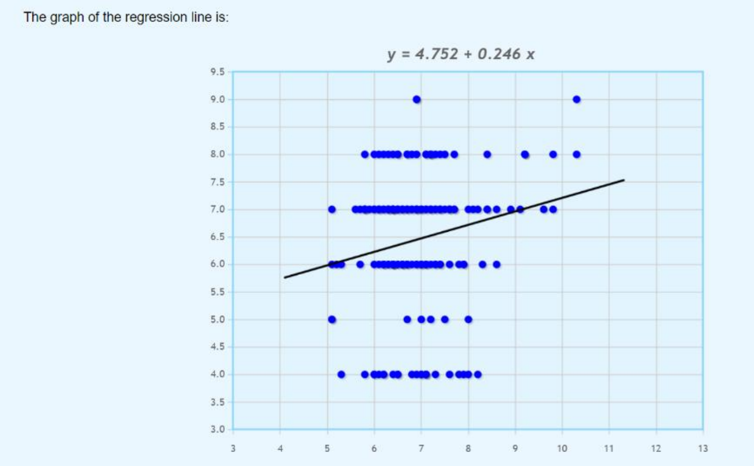
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals $1.396e^{-7}$. Since the null hypothesis is accepted for the fixed acidity in the testing of hypothesis test that we have done before, we can say that the data are normally distributed.

<u>Outliers:</u>

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



<u>INFERENCE:</u>

From the above graph of the regression line of Y on X, I can say that the plots of fixed acidity and quality is distributed slightly away from the regression line of Y on X, i.e., they are weekly positively correlated to each other.

<u>REGRESSION ANALYSIS BETWEEN THE **VOLATILE ACIDITY** AND THE **QUALITY** OF THE **WHITE WINE**</u>

<u>VARIABLE TABLE:</u>

Independent variable	***x***	*Volatile acidity*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*5.8695*
b	*1.8978*

The linear regression equation is as follows,

Y = 5.8695 + 1.8978X

<u>RESULTS OBTAINED FROM R PROGRAMMING (R-studio)</u>

- r Square (r^2) equals 0.03192. It means that 3.2% of the variability of Y is explained by X.

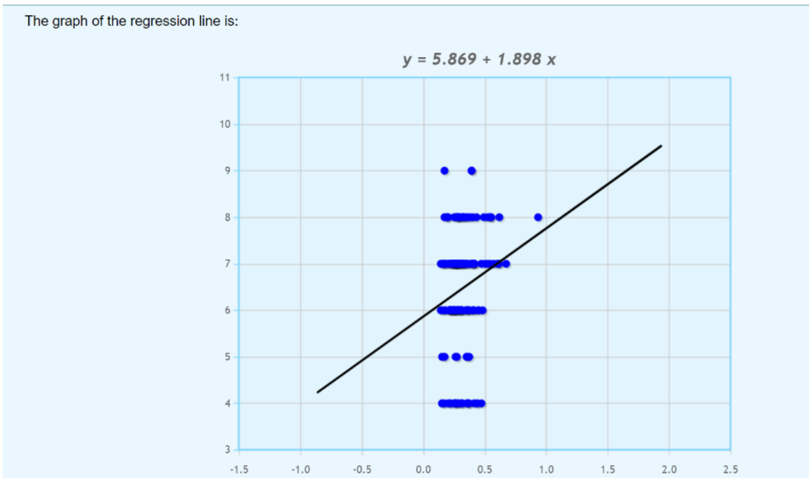
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals $2.257e^{-7}$. Since the null hypothesis is rejected for the volatile acidity in the t-test that we have done before, we can say that the data aren't normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of volatile acidity and quality is distributed slightly away from the regression line of Y on X, i.e., they are weekly positively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **CITRIC ACID** AND THE **QUALITY** OF THE **WHITE WINE**

VARIABLE TABLE:

Independent variable	***x***	*Citric acid*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*6.451*
b	*0.02696*

The linear regression equation is as follows,

Y = 6.451 + 0.02696X

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.000007174. It means that 0.0007% of the variability of Y is explained by X.

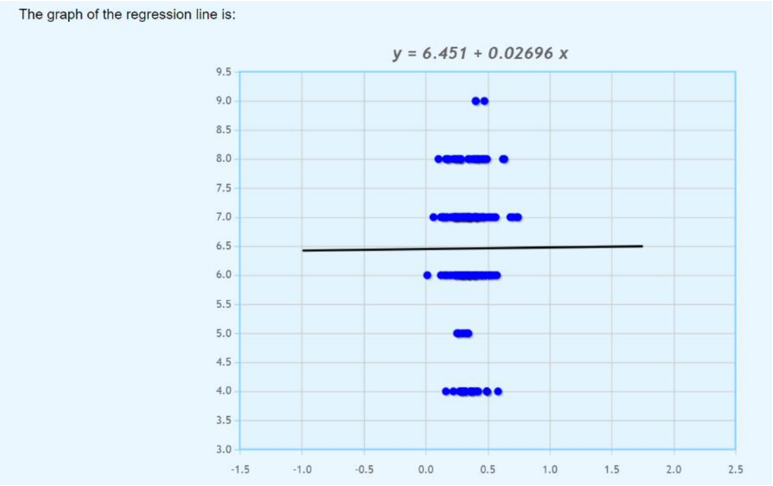
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals $1.057e^{-10}$. Since the null hypothesis is accepted for the citric acid in the t-test that we have done before, we can say that the data are normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of citric acid and quality is distributed away from the regression line of Y on X, i.e., they are very weekly positively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **RESIDUAL SUGAR** AND THE **QUALITY** OF THE **WHITE WINE**

VARIABLE TABLE:

Independent variable	*x*	*Residual sugar*
Dependent variable	*y*	*Quality*

The coefficients are as follows

Intercept(a)	*6.4197*
b	*0.006125*

The linear regression equation is as follows,

$Y = 6.4197 + 0.006125X$

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.0005504. It means that 0.06% of the variability of Y is explained by X.

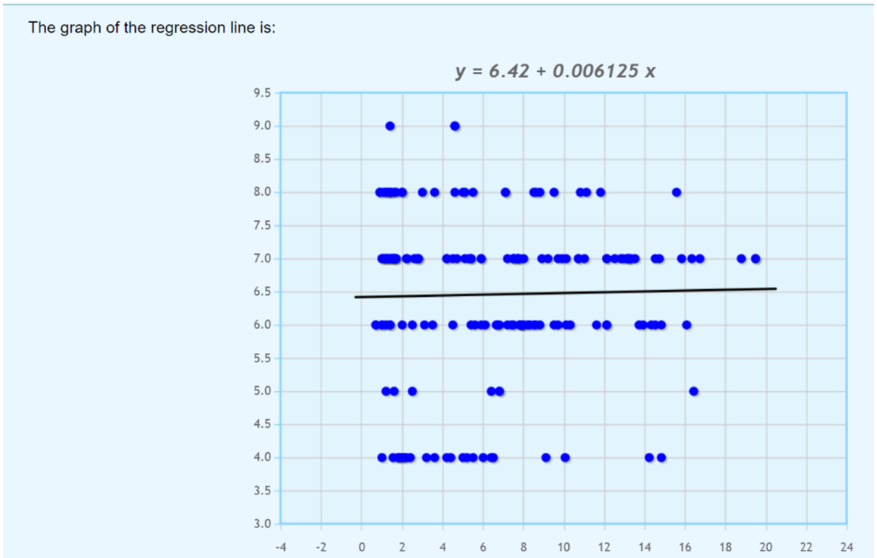
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals $4.711e^{-10}$. Since the null hypothesis is rejected for the residual sugar in the t-test that we have done before, we can say that the data aren't normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of residual sugar and quality is distributed slightly away from the regression line of Y on X, i.e., they are weekly positively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **CHLORIDES** AND THE **QUALITY** OF THE **WHITE WINE**

VARIABLE TABLE:

Independent variable	***x***	*Chlorides*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*6.2285*
b	*4.9004*

The linear regression equation is as follows,

Y = 6.2285 + 4.9004X

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.01218. It means that 1.2% of the variability of Y is explained by X.

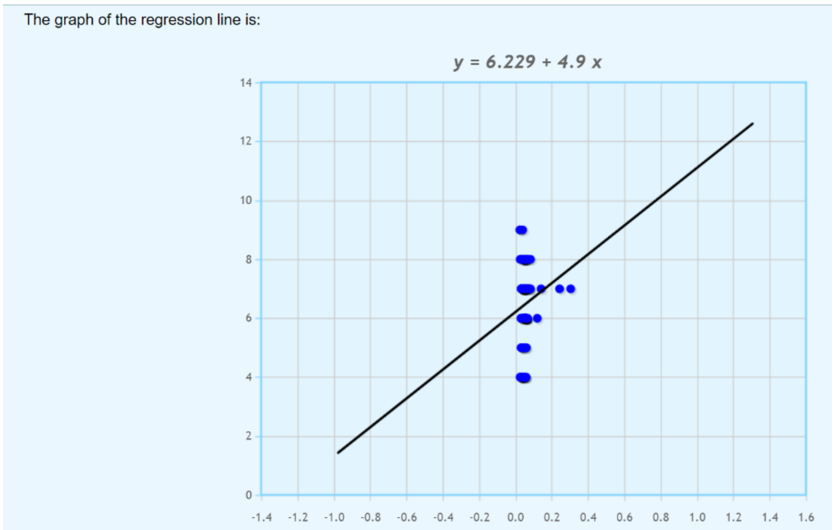
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals $4.329e^{-9}$. Since the null hypothesis is rejected for the chlorides in the t-test that we have done before, we can say that the data aren't normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of chlorides and quality is distributed somewhat closer to the regression line of Y on X, i.e., they are weekly positively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **DENSITY** AND THE **QUALITY** OF THE **WHITE WINE**

VARIABLE TABLE:

Independent variable	***x***	*Density*

Dependent variable	*y*	*Quality*

The coefficients are as follows

Intercept(a)	*-129.5439*
b	*136.7586*

The linear regression equation is as follows,

$\boxed{Y = -129.5439 + 136.7586X}$

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.0826. It means that 8.2% of the variability of Y is explained by X.

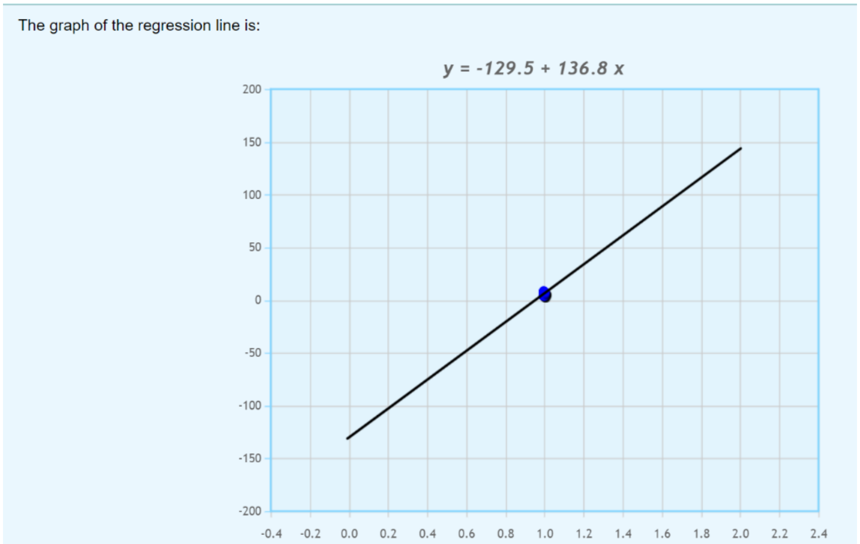
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.03308. Since the null hypothesis is rejected for the density in the t-test that we have done before, we can say that the data aren't normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of density and quality is distributed very closer to the regression line of Y on X, i.e., they are positively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **pH** AND THE **QUALITY** OF THE **WHITE WINE**

VARIABLE TABLE:

Independent variable	***x***	*pH*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*10.0682*
b	*1.1388*

The linear regression equation is as follows,

Y = 10.0682 + 1.1388X

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.01894. It means that 1.8% of the variability of Y is explained by X.

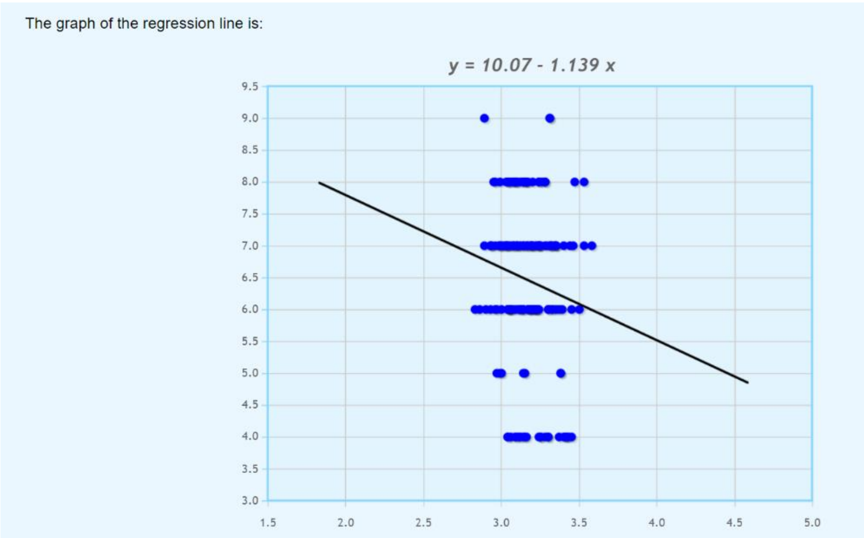
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals $5.678e^{-7}$. Since the null hypothesis is accepted for the pH level in the t-test that we have done before, we can say that the data are normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of pH level and quality is distributed slightly away from the regression line of Y on X, i.e., they are weekly negatively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **SULPHITES** AND THE **QUALITY** OF THE **WHITE WINE**

VARIABLE TABLE:

Independent variable	***x***	*Sulphites*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*6.6959*
b	*0.5082*

The linear regression equation is as follows,

$$\boxed{Y = 6.6959 + 0.5082X}$$

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.001836. It means that 0.2% of the variability of Y is explained by X.

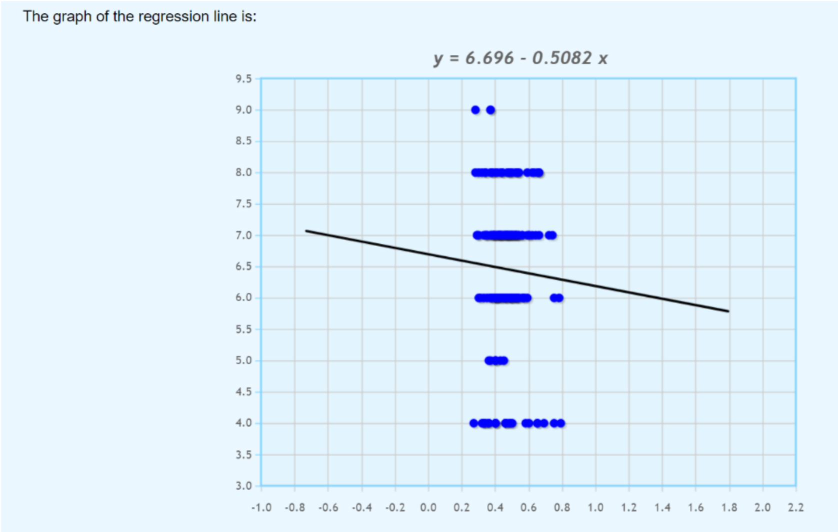
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals $1.156e^{-9}$. Since the null hypothesis is accepted for the sulphites in the t-test that we have done before, we can say that the data are normally distributed.

Outliers:

The data does not contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of sulphites level and quality is distributed closer to the regression line of Y on X, i.e., they are negatively correlated to each other.

REGRESSION ANALYSIS BETWEEN THE **ALCOHOL** AND THE **QUALITY** OF THE **WHITE WINE**

VARIABLE TABLE:

Independent variable	***x***	*Alcohol*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*10.6659*
b	*0.4115*

The linear regression equation is as follows,

$$\boxed{Y = 10.6659 + 0.4115X}$$

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.1455. It means that 14.5% of the variability of Y is explained by X.

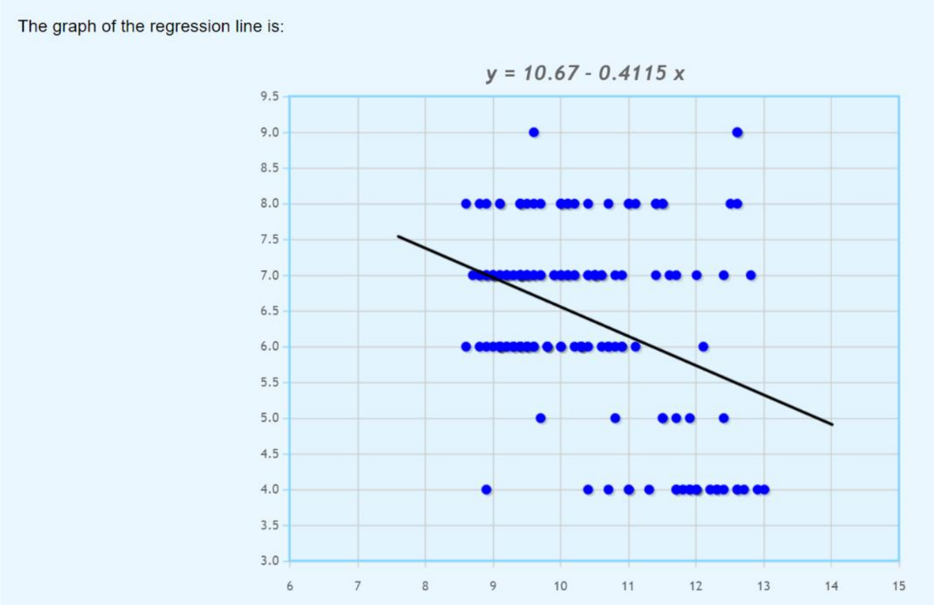
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.2373. Since the null hypothesis is rejected for the alcohol in the t-test that we have done before, we can say that the data aren't normally distributed.

Outliers:

The data contains outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of alcohol content and quality is distributed closely around the regression line of Y on X, i.e., they are negatively correlated to each other.

3.3.2: **REGRESSION ANALYSIS** BETWEEN **MANUFACTURING YEAR** THE WINE AND **THE QUALITY** OF THE WINE

3.3.2.1: RED WINE

REGRESSION ANALYSIS BETWEEN THE **MANUFACTURING YEAR** AND THE **QUALITY** OF THE **RED WINE**

VARIABLE TABLE:

Independent variable	***x***	*Year (Red wine)*
Dependent variable	***y***	*Quality*

The coefficients are as follows

Intercept(a)	*169.6699*
b	*0.0814*

The linear regression equation is as follows,

$$Y = 169.6699 + 0.0814X$$

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.1213. It means that 12.1% of the variability of Y is explained by X.

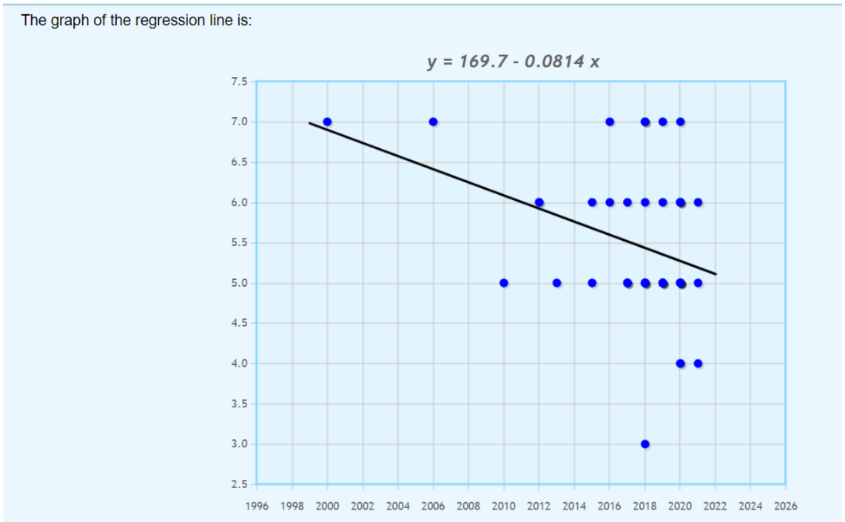
<u>Residual Normality:</u>

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.001393. The data are normally distributed.

<u>Outliers:</u>

The data doesn't contain any outliers.

REGRESSION LINE IN SCATTER PLOT



<u>INFERENCE:</u>

From the above graph of the regression line of Y on X, I can say that the plots of manufacturing year of the red wines and quality is distributed around the regression line of Y on X, i.e., they are negatively correlated to each other.

<u>3.3.2.2: WHITE WINE</u>

<u>REGRESSION ANALYSIS BETWEEN THE **MANUFACTURING YEAR** AND THE **QUALITY** OF THE **WHITE WINE**</u>

<u>VARIABLE TABLE:</u>

Independent variable	*x*	*Year (White wine)*
Dependent variable	*y*	*Quality*

The coefficients are as follows

Intercept(a)	*313.9079*
b	*0.1524*

The linear regression equation is as follows,

Y = 313.9079+ 0.1524X

RESULTS OBTAINED FROM R PROGRAMMING (R-studio)

- r Square (r^2) equals 0.1089. It means that 10.9% of the variability of Y is explained by X.

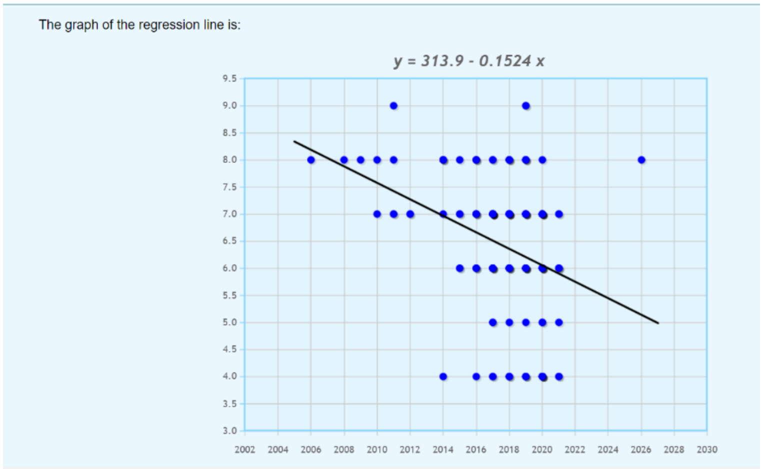
Residual Normality:

The linear regression model assumes normality for residual errors. Shapiro will p-value equals 0.00001123. The data are normally distributed.

Outliers:

The data doesn’t contain any outliers.

REGRESSION LINE IN SCATTER PLOT



INFERENCE:

From the above graph of the regression line of Y on X, I can say that the plots of manufacturing year of the white wines and quality is distributed around the regression line of Y on X, i.e., they are negatively correlated to each other.

3.4: BEST BRAND AND GRAPES

3.4.1: BEST BRAND

3.4.1.1: BEST BRAND FROM QUALITY

OBSERVATIONS FROM EXCEL:

TABLE 3.29: AVERAGE QUALITY OF DIFFERENT BRANDS

BRAND NAME	AVERAGE QUALITY
Quintas'	6.4273
Anselmo Mendes	7.3846
Aveleda	6
Adega	5.8
Casa/Casal	6.0204
A&D Wines	6.1667

INFERENCE:

From the above table, we can infer that the wine produced by the **Anselmo Mendes** has the best quality of all the brands that have been taken into account.

3.4.1.2: BEST AND CHEAPEST BRAND FROM PRICE

OBSERVATIONS FROM EXCEL:

TABLE 3.30: AVERAGE PRICE OF DIFFERENT BRANDS

BRAND NAME	AVERAGE PRICE
Quintas'	₹ 1,689.39
Anselmo Mendes	₹ 2,303.83
Aveleda	₹ 1,549.11
Adega	₹ 1,223.64
Casa/Casal	₹ 1,027.41
A&D Wines	₹ 963.60

INFERENCE:

From the above table, I can conclude that the cheapest brand is A&D wines but also all the other wines brands are affordable nowadays. When it comes to the best brand, we have to take quality also into the account. By comparing the quality and the price, we can conclude that A&D wines are best in both the ways. It is cheaper as well has good quality

3.4.2: BEST GRAPE TYPE

3.4.2.1: GRAPES USED FOR THE PREPARATION OF RED WINE

OBSERVATIONS FROM EXCEL:

TABLE 3.31: AVERAGE QUALITY OF GRAPES USED FOR PREPARATION OF RED WINE

NAME OF THE GRAPE	AVERAGE QUALITY
Tinto	5.7917
Espadeiro	5.25
Padeiro	5.1429

INFERENCE:

From the above table, we can infer that the red wine prepared from the grape type Tinto has the better quality when compared to all the other grape types taken into account.

3.4.2.2: GRAPES USED FOR THE PREPARATION OF WHITE WINE

OBSERVATIONS FROM EXCEL:

TABLE 3.32: AVERAGE QUALITY OF GRAPES USED FOR PREPARATION OF WHITE WINE

NAME OF THE GRAPE	AVERAGE QUALITY
Alvarinho (Albarinho)	7.3654
Avesso	6.35
Branco	5.9744
Arinto	6.2
Trajadura	5.7142
Loureiro	6.1389

INFERENCE:

From the above table, we can infer that the white wine prepared from the grape type *Alvarinho* has the best quality when compared to all the other grape types used for the preparation of white wine

CONCLUSION:

If we observe the data clearly, most of the wines produced by *Anselmo Mendes* brand is prepared from the grape type *Alvarinho*. Thus, we can conclude that the *Anselmo Mendes* uses the best quality grapes for their white wine production. And in the same way, A&D brand wines are manufactured from the white grape types *Arinto* and *Avesso* which has the best quality next to *Alvarinho.*

CHAPTER – 4

CONCLUSION

I hereby conclude my project on the topic **"A STUDY ON WINE QUALITY DATASET".** This analysis has made me get a deeper insight into the topics of correlation, regression and testing of hypothesis, which I've used to determine which constituent of a particular wine contributes the most towards its overall quality. On doing this project, I've not only learnt in detail how to apply statistical methods in the solving the problems in inferential statistics but also the preparation of wine, its various constituents, cultivation regions of various specific brands and about the various popular brands of wine.

By the testing of hypothesis method, I could see the similarities between the sample statistics and the population parameter by using the difference between the means.

By the correlation and regression analysis between the constituents of the wine and the quality, I can infer that all the correlation co-efficient lies between the range -0.4 to 0.4 which implies only minute changes will be introduced to the wine when a particular constituent is added. And by the same way, calculating correlation and regression analysis between the manufacturing year and the quality, I came to know that the manufacturing year and quality has inverse relationship between them. Since the taste of the wine becomes more and more tastier as it gets older.

From the mean quality of the brand and price, we came to know that the best brand is *Alvarinho* by quality and the best brand is *A&D wines* by price. And from the mean quality of the specific grapes of both red and white, I can infer *Tinto* and *Alvarinho* as the best quality grapes used for the preparation of both red and white wine

It has been quite the enjoyment learning both the academic and non-academic aspects of this paper. Wine has a lot of health benefits that also vary with the type, for instance, red wine provides antioxidants. India too has a rich history of wines and therein lies my motivation towards this paper.

REFERENCES

1. Vinovest - https://www.vinovest.co/blog/vinho-verde
2. Red wine github - https://kchaaa.github.io/red-wine/
3. Vinepair - https://vinepair.com/wine-blog/7-things-you-need-to-know-about-vinho-verde/
4. Wine searcher - https://www.wine-searcher.com/regions-vinho+verde
5. WineEnthusiast - https://www.winemag.com/2018/09/12/white-wines-vinho-ve
6. Vivino - https://www.vivino.com/toplists/top-25-portuguese-vinho-verde-white-wines-united-states-righ
7. Tasteatlas - https://www.tasteatlas.com/most-popular-wines-varieties-and-appellations-in-portugal
8. Advanced Mixology - https://advancedmixology.com/blogs/art-of-mixology/best-portuguese-red-wines
9. Wines Portugal - https://www.winesofportugal.com/en/portuguese-wines/grape-varieties/