VARIABLES AFFECTING QUALITIES OF BANANAS

**Summary.** This paper investigates the relationship between numeric characteristics of different attributes of bananas and the qualities of bananas. The various strategy used to achieve the goal includes finding measures of spreads and locations of numerical data of different attributes. Graphical representation of data for example normal QQ plot and side-by-side boxplots between attributes are also included in this paper. It is found that ripeness of bananas is the most relevant attributes associated to the quality of bananas.

1. INTRODUCTION

The results from this paper is vital as B.Pathare (2023, Volume 134) had mentioned that consumer interest in high fruit quality had increased due to the international food trade. The quality of banana is predicted and classified by optical properties such as absorption, suggested by Hashim (2016, Volume 212). On the other hand, the ripeness of banana can be identified by new feature such as ripening factor as proposed by Mazen (2019). This paper will first analyse the numerical characteristics of a banana, this includes various attributes, including Size, Weight, Sweetness, Softness, Ripeness, Acidity, and Harvest Time. We will then discuss the association between numerical characteristic and the quality of banana.

1. METHOD

The data collected contain negative entries as the data are demeaned measures, which means it is already subtracted with mean of each variable from its original value. We find the summary statistics for all variables to summarise the data. A matrix of normal QQ plot and scatterplots are also included, the scatterplot is used to show the pairwise relationship between variables. We also have 2 histograms of Harvest Time, one for “Good” and one for “Bad” bananas. 6 pair of side-by-side boxplots corresponding to the variables are also included.

1. A black text and numbers

   Description automatically generatedA close up of numbers

   Description automatically generatedA black text on a white background

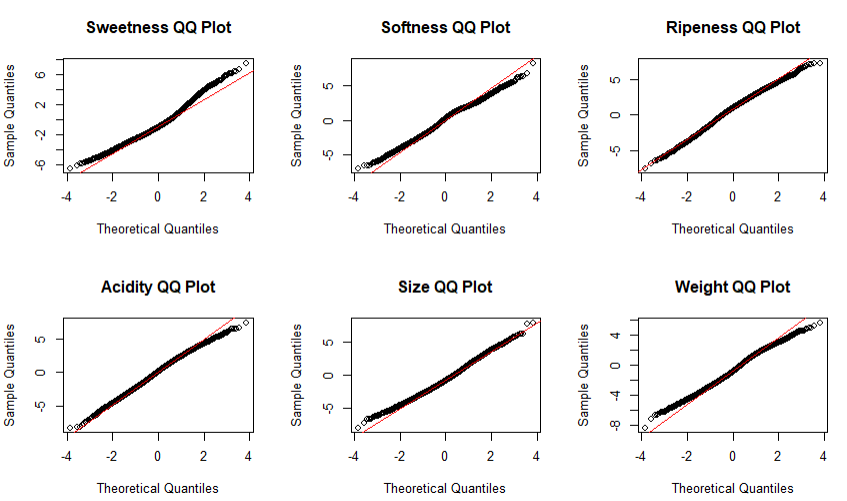
   Description automatically generatedA black text on a white background

   Description automatically generatedA close up of a number

   Description automatically generatedA close up of numbers

   Description automatically generatedA number and text on a white background

   Description automatically generatedRESULTS AND DISCUSSION

From the data we can see that the value of median and mean of all 7 attributes are almost the same, this shows that the data may have symmetric distribution around the centre. Even if the distribution is not perfectly symmetric, it still suggests that the skewness of the data is relatively mild.

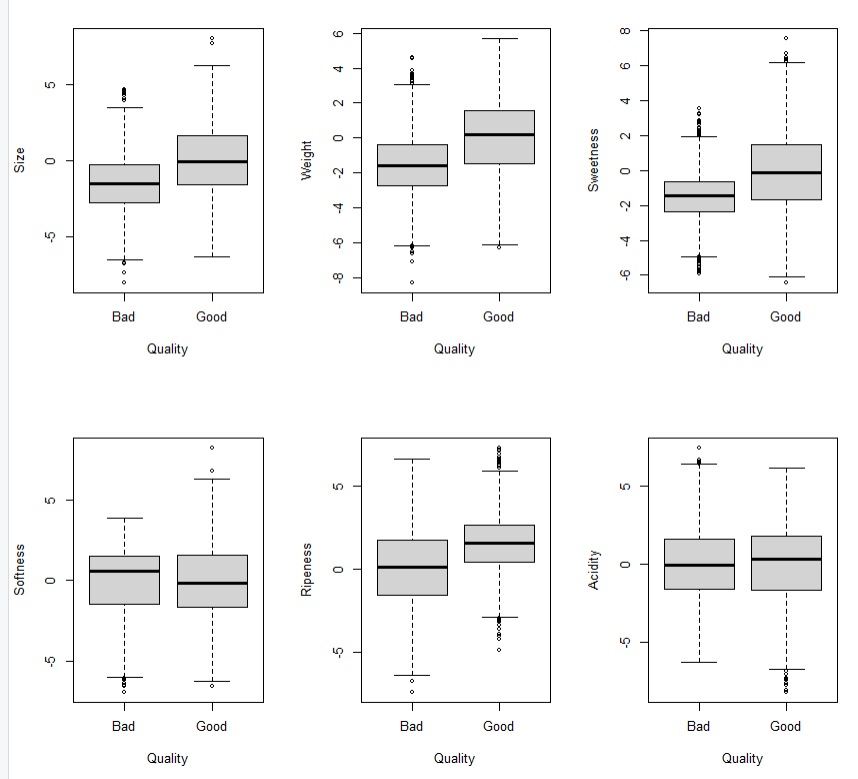
We can observe from the normal QQ plot that the points for sweetness, softness and weight slightly deviate from the diagonal reference line, especially in both tails of the distribution. This indicates that the distribution may be skewed or non-normal. For points in ripeness, acidity, and size QQ plot, the points follow more closely to the reference line, suggesting the distribution to be more likely symmetrically distributed around the mean. One interesting common feature we can observe is that all 6 normal QQ plots have a few outliers on both tails.

A graph of different sizes of bananas

Description automatically generatedA chart of different sizes of objects

Description automatically generated with medium confidenceMost of the pairwise scatterplots did not show any clear trend, indicating no correlation between the attributes. Exceptionally, the scatterplot between weight and acidity shows a moderate positive correlation whereas scatterplot between softness and acidity shows evidence of weak negative correlation.

There is a slight difference in the harvest time of bananas as shown above, the harvest time of good bananas tends to be earlier.

A clear difference between good and bad bananas can be seen in attributes of size, weight, sweetness, and ripeness. The interquartile range of these attributes tend to be higher for good bananas. As for the aspect of softness and acidity, they have similar interquartile range, though there is a slight difference in the median. One interesting feature found is that the range of data for specific attributes appeared to have large difference, for example range of sweetness for bad bananas is much smaller than the range of good bananas. Outliers can also clearly be seen from the diagram above, which refers to data points that fall beyond the whiskers of the boxplot. From the boxplot, we can conclude that good bananas have bigger size, heavier weight, higher rating of sweetness and higher rating of ripeness.

A screenshot of a computer

Description automatically generatedSummary of Good Bananas

Summary of Bad Bananas

A screenshot of a computer

Description automatically generated

The summary supports the statement that “good bananas have bigger size, heavier weight, higher rating of sweetness and higher rating of ripeness.” For example, good bananas have a mean size of 0.005001 whereas bad bananas only have a mean size of -1.5029. The difference is quite significant.

The study has also further investigated and found that when the ripeness ratings are above the mean, there is a higher occurrence of good qualities bananas. Among the 4006 good qualities bananas, 2712 bananas have their ripeness rating above the mean.

1. CONCLUSION

We can conclude that the ripeness rating has a direct association with the qualities of the bananas. However, the data is limited as there may be other factors which affects the qualities of the bananas, for example geographical location and weather conditions. In order to improve the accuracy of this study, comparison of data collected from different regions can be made and analyse how other factors may impact bananas qualities across different regions.

(806 words)

REFERENCES

Al-Dairi, M., Pathare, P. B., Al-Yahyai, R., Jayasuriya, H. P. W., & Al‐Attabi, Z. (2023). Postharvest quality, technologies, and strategies to reduce losses along the supply chain of banana: A review. *Trends in Food Science and Technology*, *134*, 177–191. https://doi.org/10.1016/j.tifs.2023.03.003

Mazen, F. M. A., & Nashat, A. A. (2019). Ripeness classification of bananas using an artificial neural network. *Arabian Journal for Science and Engineering*, *44*(8), 6901–6910. <https://doi.org/10.1007/s13369-018-03695-5>

Adebayo, S. E., Hashim, N., Abdan, K., Hanafi, M., & Mollazade, K. (2016). Prediction of quality attributes and ripeness classification of bananas using optical properties. *Scientia Horticulturae*, *212*, 171–182. https://doi.org/10.1016/j.scienta.2016.09.045

APPENDIX

banana.data = read.csv("banana.csv", header = TRUE)

attach(banana.data)

summary(Size)

summary(Weight)

summary(Sweetness)

summary(Softness)

summary(Ripeness)

summary(Acidity)

summary(HarvestTime)

par(mfrow=c(2,3))

qqnorm(banana.data$Sweetness, main="Sweetness QQ Plot")

qqline(banana.data$Sweetness, col="red")

qqnorm(banana.data$Softness, main="Softness QQ Plot")

qqline(banana.data$Softness, col="red")

qqnorm(banana.data$Ripeness, main="Ripeness QQ Plot")

qqline(banana.data$Ripeness, col="red")

qqnorm(banana.data$Acidity, main="Acidity QQ Plot")

qqline(banana.data$Acidity, col="red")

qqnorm(banana.data$Size, main="Size QQ Plot")

qqline(banana.data$Size, col="red")

qqnorm(banana.data$Weight, main="Weight QQ Plot")

qqline(banana.data$Weight, col="red")

c = data.frame(Size, Weight, Sweetness, Softness,Ripeness, Acidity)

pairs(c)

good = subset(banana.data,Quality == "Good")

hist(good$HarvestTime,main="Harvest Time of Good Bananas",xlab ="Harvest Time",ylab = “Number of Bananas” )

bad = subset(banana.data,Quality == "Bad")

hist(bad$HarvestTime,main="Harvest Time of Bad Bananas",xlab ="Harvest Time",ylab = “Number of Bananas” )

par(mfrow=c(2,3))

boxplot(Size~Quality,data = banana.data)

boxplot(Weight~Quality,data = banana.data)

boxplot(Sweetness~Quality,data = banana.data)

boxplot(Softness~Quality,data = banana.data)

boxplot(Ripeness~Quality,data = banana.data)

boxplot(Acidity~Quality,data = banana.data)

summary(good)

summary(bad)

x = subset(banana.data,Quality == "Good" & Ripeness >= 0.7811)