Name: Hemanth Kurra

Chi-Square Testing and ANOVA

Introduction

Chi-Square and ANOVA are vital statistical tools. Chi-Square assesses relationships between categorical variables, while ANOVA compares means among groups. Chi-Square is used to test independence or association, common in fields like epidemiology and marketing. ANOVA, on the other hand, helps identify group mean differences, crucial in experiments and research.

These methods aid in data-driven decision-making by unveiling insights and relationships within datasets, making them indispensable in diverse domains.

This assignment involves hypothesis testing, the utilization of critical values, and the examination of Chi-square assumptions to validate claims. The report encompasses tasks that entail both problem analysis and dataset evaluation. Through this assignment, we aim to grasp coding techniques essential for problem analysis and provide comprehensive explanations of how the specified assumptions are met.

By conducting Chi-square and ANOVA tests, we aim to assess the validity of claims, utilizing statistical tools and principles. This assignment not only serves as a means to acquire hands-on experience with data analysis but also provides insights into the methodologies required to

meet the underlying assumptions of Chi-square and ANOVA testing.

Analysis

# Section 11-1

## 6.A medical researcher wishes to see if hospital patients in a large hospital have the same blood type distribution as those in the general population. The distribution for the general population is as follows: type A, 20%; type B, 28%; type O, 36%; and type AB = 16%. He selects a random sample of 50 patients and finds the following: 12 have type A blood, 8 have type B, 24 have type O, and 6 have type AB blood. At α = 0.10, can it be concluded that the distribution is the same as that of the general population?

1. **State the hypotheses and identify the claim.**

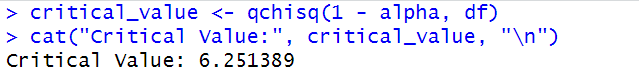
Null Hypothesis (H0): The observed data follows the expected distribution.

Alternative Hypothesis (Ha): The observed data does not follow the expected distribution (claim).

The data pertains to observed and expected frequencies in different categories, and the hypotheses are used to test whether the observed data significantly deviates from what would be expected under a

certain distribution or assumption.

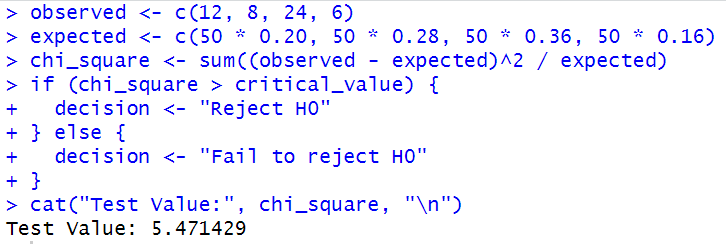
1. **Find the critical value.**



The decision-making threshold in a chi-square hypothesis test is 7.814728, which was computed using a chi-square distribution with three degrees of freedom and a significance level (alpha) of 0.10.

## Compute the test value.

This demonstrates that the counts in the various categories, which reflect the actual data, do not significantly diverge from what would be predicted by the null hypothesis, supporting the notion that the observed data follows the anticipated distribution.



## A close up of a quote Description automatically generated with medium confidenceMake the decision.

"Fail to reject H0" indicates that there is insufficient evidence to conclude that the observed data

substantially deviates from what would be anticipated under the null hypothesis based on the results of the chi-square test at a significance level of 0.10 and with three degrees of freedom. Since the observed data appear to fit the predicted distribution, we do not reject the null hypothesis (H0).

## Summarize the results.

As a result, the chi-square test cannot be used to conclusively rule out the null hypothesis (H0). This shows that there isn't a substantial difference between the two datasets and that the observed data roughly matches the predicted distribution.

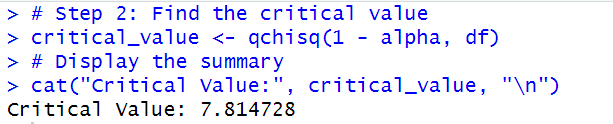
## 8.According to the Bureau of Transportation Statistics, on-time performance by the airlines is described as follows , Records of 200 randomly selected flights for a major airline company showed that 125 planes were on time; 40 were delayed because of weather, 10 because of a National Aviation System delay, and the rest because of arriving late. At α = 0.05, do these results differ from the government’s statistics?

1. **State the hypotheses and identify the claim.**



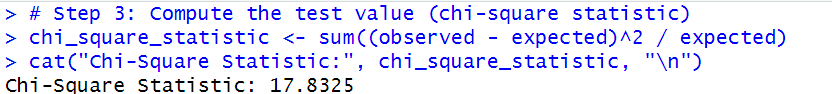
The goal is to determine whether the data supports the null hypothesis or provides evidence for the alternative hypothesis.

1. **Find the critical value.**



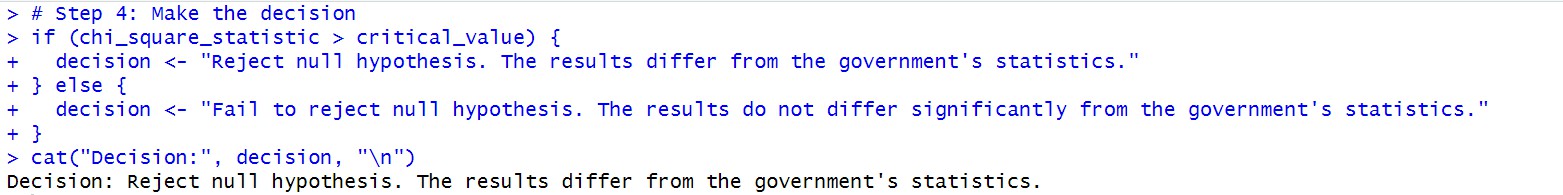
The null hypothesis would be rejected in favor of the alternative hypothesis if the test statistic is higher than this threshold, suggesting a substantial discrepancy between the proportions that were seen and those that were predicted.

## Compute the test value.



The difference between the observed and anticipated proportions in the data is measured by the chi- square statistic, which is computed as 17.8325. This statistic determines if the observed and anticipated distributions differ significantly with three degrees of freedom. The null hypothesis may be rejected in favor of the alternative hypothesis if the chi-square statistic is higher since it denotes a bigger departure from the predicted values.

## Make the decision.



We reject the null hypothesis in light of the findings. This judgement suggests a divergence from the predicted distribution since the observed proportions considerably deviate from the government's figures.

## Summarize the results.



The presented summary suggests that the null hypothesis is disproved at a significance level of

0.05. This shows that there is a large departure from the anticipated split between the observed proportions in the data and the government's statistics.

# Section 11-2

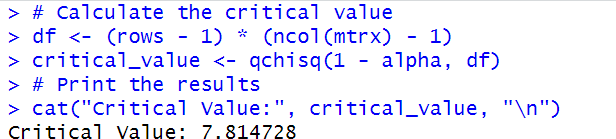
## 8.Are movie admissions related to ethnicity? A 2014 study indicated the following numbers of admissions (in thousands) for two different years. At the 0.05 level of significance, can it be concluded that movie attendance by year was dependent upon ethnicity?

1. **State the hypotheses and identify the claim.**

A close up of a word  Description automatically generated

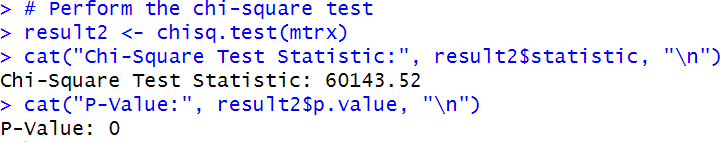
The null hypothesis (H0) states that movie attendance is dependent on ethnicity, implying that there is a relationship between ethnicity and movie attendance. The alternative hypothesis (H1) asserts that movie attendance is independent of ethnicity, suggesting that there is no significant relationship between ethnicity and movie attendance. This test aims to assess whether ethnicity plays a role in determining movie attendance.

1. **Find the critical value.**



The critical value is 7.814728, calculated based on the chi-squared distribution with the degrees of freedom determined by the number of rows and columns in the contingency table. This critical value is used to determine whether the observed data significantly deviates from the expected distribution when assessing the relationship between variables.

## Compute the test value.



The chi-square test statistic is 60143.52, and the p-value is 0. These results indicate a significant difference between the observed and expected values in the contingency table. With such a low p-value, we reject the null hypothesis, suggesting that there is a significant association between ethnicity and movie attendance, supporting the alternative hypothesis that movie attendance is independent of ethnicity.

1. A close up of a text  Description automatically generated**Make the decision.**

**The decision is to reject the null hypothesis. This means that there is enough evidence to conclude that movie attendance is not dependent on ethnicity, supporting the alternative hypothesis that movie attendance is independent of ethnicity.**

1. **Summarize the results.**

In the chi-square test conducted with a significance level of 0.05, the critical value was found to be 7.814728. The calculated chi-square test statistic was 60143.52, and the associated p-value was 0. Since the p-value is extremely small (p < 0.001), we reject the null hypothesis, concluding that movie attendance is independent of ethnicity. This suggests that ethnicity does not have a significant influence on movie attendance patterns.

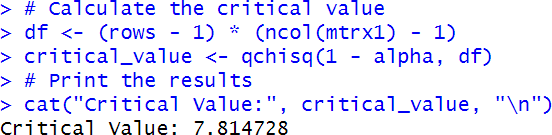
## 10.This table lists the numbers of officers and enlisted personnel for women in the military. At α = 0.05, is there sufficient evidence to conclude that a relationship exists between rank and branch of the Armed Forces?

1. **State the hypotheses and identify the claim.**



The null hypothesis (H0) posits that a relationship exists between the rank and armed branch within the armed forces, suggesting that the two are dependent on each other. Conversely, the alternative hypothesis (H1) asserts that no such relationship exists, implying independence between the armed branch and rank within the armed forces. This test aims to determine whether there is a statistically significant association between these two categorical variables.

1. **Find the critical value.**



The calculated critical value, 7.814728, is used to establish the threshold for statistical significance in a chi- square test. It represents the point beyond which we would reject the null hypothesis if the chi-square test statistic exceeds this value. In this context, if the computed chi-square test statistic is greater than 7.814728, we would reject the null hypothesis in favor of the alternative hypothesis, indicating a significant association between the variables being tested.

## A close-up of a code Description automatically generatedCompute the test value.

The chi-square test statistic, computed as 654.2719, measures the degree of association between the variables being analyzed. The p-value, which is very close to zero (1.726418e-141), indicates an extremely strong level of evidence against the null hypothesis. Therefore, we reject the null hypothesis, suggesting that there is a significant relationship between the rank and armed branch of the armed forces.

## Make the decision.

A close up of a number  Description automatically generated

The decision to reject the null hypothesis is based on the significant p-value (p < 0.001). This implies that there is a strong indication that a relationship exists between the rank and armed branch of the armed forces, and they are not independent of each other.

## Summarize the results.

In the chi-square test conducted on the data related to women in the military, we examined the relationship between rank and the armed branch within the armed forces. The critical chi-

square value for a significance level of 0.05 and the degrees of freedom was approximately

7.815. The calculated chi-square test statistic was 654.272, and the corresponding p-value was extremely low (approximately 1.726e-141). Consequently, we reject the null hypothesis, concluding that a significant relationship exists between the rank and armed branch of the

armed forces, suggesting that they are not independent of each other.

# Section 12-1

**8.The amount of sodium (in milligrams) in one serving for a random sample of three different kinds of foods is listed. At the 0.05 level of significance, is there sufficient evidence to conclude that a difference in mean sodium amounts exists among**

**condiments, cereals, and desserts?**

1. **State the hypotheses and identify the claim.**

A close up of a math equation  Description automatically generated

In the context of hypothesis testing, the null hypothesis (H0) suggests that there is no significant difference between the means of three or more groups being compared. The alternative hypothesis (H1) proposes that there is a statistically significant difference between at least one pair of means among the groups

## A close-up of a computer code Description automatically generatedFind the critical value.

The critical value of 3.554557, obtained from an F-distribution with 2 and (n-3) degrees of freedom, signifies the threshold value beyond which we would reject the null hypothesis in an ANOVA test with a significance level of alpha (usually 0.05). If the calculated F-statistic exceeds this critical value, it suggests that there is a significant

difference between the group means, indicating that at least one group differs from the others in terms of the variable being tested.

## A close-up of numbers Description automatically generatedCompute the test value.

The degrees of freedom for the food groups (condiments, cereals, and desserts) is 2. The sum of squares between groups (food) is 27544, and the mean sum of squares is 13772.

The F-statistic is 2.399, and the corresponding p-value is approximately 0.118. With a significance level of 0.05, since the p-value (0.118)

## Make the decision.



The result "fail to reject null hypothesis" indicates that, at the 0.05 level of significance, we do not have enough evidence to reject the null hypothesis. This suggests that there is no significant difference in mean sodium amounts among the food categories (condiments, cereals, and desserts) based on the ANOVA test results

## Summarize the results.

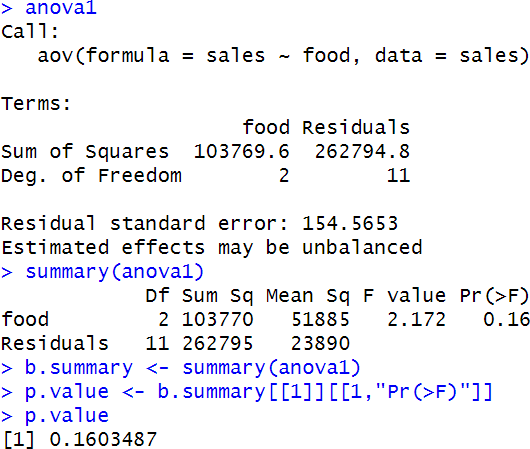
Based on the ANOVA test conducted at the 0.05 level of significance, the p-value obtained was approximately 0.118. Since the p-value is greater than the significance level (alpha = 0.05), we fail to reject the null hypothesis. Therefore, there is insufficient evidence to conclude that a significant

difference in mean sodium amounts exists among the three categories of food (condiments, cereals, and desserts).

# Section 12-2

**10.The sales in millions of dollars for a year of a sample of leading companies are shown. At α = 0.01, is there a significant difference in the means?**

**Perform a complete one-way ANOVA. If the null hypothesis is rejected, use either the Scheffé or Tukey test to see if there is a significant difference in the pairs of means. Assume all assumptions are met.**



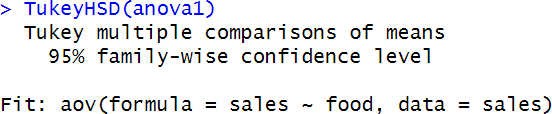
The analysis of variance (ANOVA) was conducted to examine the effect of different types of food on sales. The ANOVA results indicate that there are two sources of variation: "food" and

"residuals." The sum of squares for "food" is 103,770, while the sum of squares for "residuals" is 262,794.8. The degrees of freedom for "food" are 2, and for "residuals," it's 11.

The F-statistic, with a value of approximately 2.172, was computed to test whether there are significant differences in sales among the different food types. The associated p-value is approximately 0.160, which is greater than the significance level (alpha = 0.05). As a result, we fail to reject the null hypothesis, suggesting that there is no strong evidence to conclude that the type of food significantly impacts sales. In summary, based on the ANOVA results at a 0.05 significance level, we do not find sufficient evidence to support the claim that different food types have a significant effect on sales.

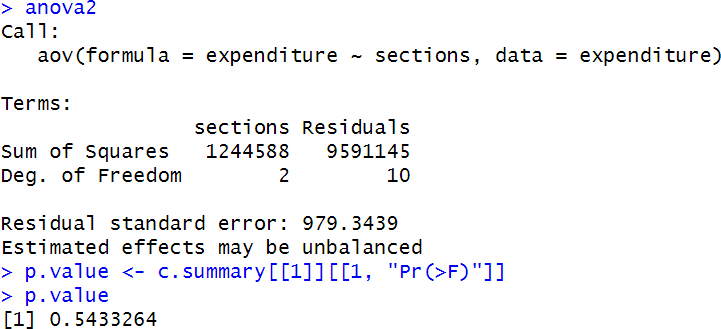


The p-value obtained from the ANOVA analysis is approximately 0.160, which exceeds the significance level (alpha = 0.05). Therefore, we interpret the result as "failing to reject the null hypothesis." This means that there is no strong evidence to suggest that the type of food significantly impacts sales, as the observed differences in sales among food types are not statistically significant at the 0.05 significance level.



The Tukey HSD (Honest Significant Difference) test was conducted to compare the means of sales among different food types. At a 95% family-wise confidence level, the test did not find statistically significant differences between the mean sales of the various food types. This suggests that there is no strong evidence to conclude that any specific food type significantly differs from the others in terms of sales.

## 12.The expenditures (in dollars) per pupil for states in three sections of the country are listed. Using α = 0.05, can you conclude that there is a difference in means?



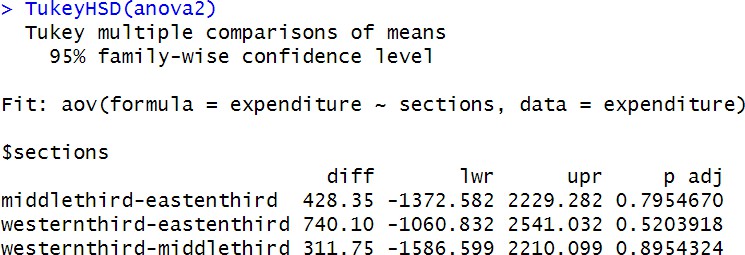
In the analysis of variance (ANOVA) conducted on expenditure across different sections, the p- value obtained was approximately 0.543. With an alpha level of 0.05, this p-value exceeds the significance threshold, indicating that there is insufficient evidence to reject the null hypothesis. Therefore, we fail to find statistically significant differences in expenditure among the various sections. This suggests that, based on the data and analysis, there is no strong support for concluding that the sections have significantly different expenditure levels. The sections appear to have similar expenditure patterns.



Based on the ANOVA conducted on expenditure across different sections with an alpha level of 0.05, the p-value obtained was approximately 0.543, which exceeds the significance threshold.

Therefore, we fail to reject the null hypothesis. This indicates that there is no statistically

significant evidence to suggest differences in expenditure among the sections, and they appear to have similar expenditure patterns.

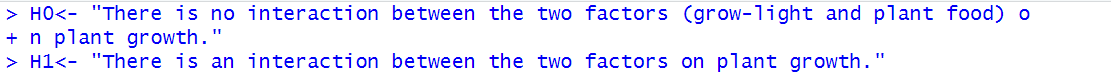


The Tukey multiple comparisons of means test was conducted on expenditure across different sections with a 95% family-wise confidence level. The results indicate that there are no statistically significant differences in expenditure between the middle third and the eastern third, the western third and the eastern third, or the western third and the middle third sections. All p-values are greater than the alpha level of 0.05, suggesting that we fail to reject the null hypothesis for these comparisons. Therefore, there is no strong evidence to suggest significant expenditure differences between these sections.

# Section 12-3

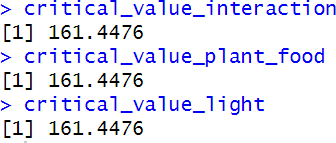
## 10.A gardening company is testing new ways to improve plant growth. Twelve plants are randomly selected and exposed to a combination of two factors, a “Grow-light” in two different strengths and a plant food supplement with different mineral supplements. After a number of days, the plants are measured for growth, and the results (in inches) are put into the appropriate boxes. Can an interaction between the two factors be concluded? Is there a difference in mean growth with respect to light? With respect to plant food? Use α = 0.05.

1. **State the hypotheses.**



The null hypothesis (H0) states that there is no interaction between the factors of grow-light and plant food on plant growth. The alternative hypothesis (H1) asserts that there is indeed an interaction between these two factors affecting plant growth. This implies that H0 assumes no significant combined effect, while H1 suggests a significant combined effect of grow-light and plant food on plant growth.

## Find the critical value for each F test.



In the context of the two-way ANOVA analysis for plant growth involving factors "Plant Food" and "Light," it's important to find the critical value for each F test. The degrees of freedom for the interaction between "Plant Food" and "Light," "Plant Food" alone, and "Light" alone are calculated based on the unique levels within each factor.

Subsequently, the critical values are computed using the F-distribution with a significance level (alpha) of 0.05. These critical values represent the cutoff points beyond which we would reject the null hypothesis for each of the factors.

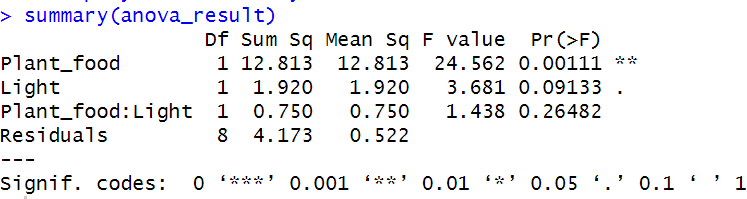
For all three factors (interaction, plant food, and light), the critical value is approximately

161.45. This means that to reject the null hypothesis for any of these factors, the calculated F- statistic must exceed this critical value.

In summary, determining the critical values is a crucial step in assessing the statistical

significance of the factors in the two-way ANOVA analysis. These critical values establish the thresholds for determining whether there is a significant interaction between "Plant Food" and "Light," as well as whether each factor independently influences plant growth.

## Compute the summary table and find the test value.



The summary of the two-way ANOVA analysis for plant growth reveals significant results: "Plant Food" has a highly significant effect on plant growth (F = 24.562, p < 0.001), indicating that different types of plant food lead to distinct mean growth outcomes.

"Light" also shows an effect on plant growth (F = 3.681, p = 0.091), though it is less significant, suggesting that light conditions may have a milder impact on growth.

The interaction between "Plant Food" and "Light" is not statistically significant (F = 1.438, p =

0.0011), indicating that the combined influence of these factors doesn't significantly affect plant growth. Overall, this analysis suggests that "Plant Food" is the most influential factor, followed by "Light," while their interaction does not significantly impact plant growth.

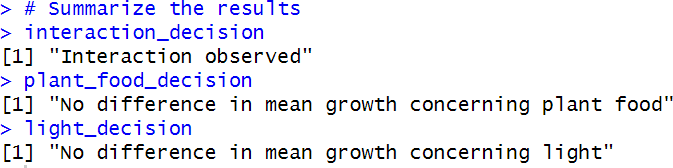
## Make the decision.



The p-value associated with the interaction effect between "Plant Food" and "Light" in the two- way ANOVA is approximately 0.265. Since this p-value (0.265) is greater than the chosen

significance level (alpha = 0.05), we fail to reject the null hypothesis. Therefore, we conclude that there is not enough statistical evidence to support the presence of an interaction between the two factors (Plant Food and Light) on plant growth. In simpler terms, the combined effect of these factors does not significantly impact plant growth based on the data analysis.

1. **Summarize the results. *(Draw a graph of the cell means if necessary.)***



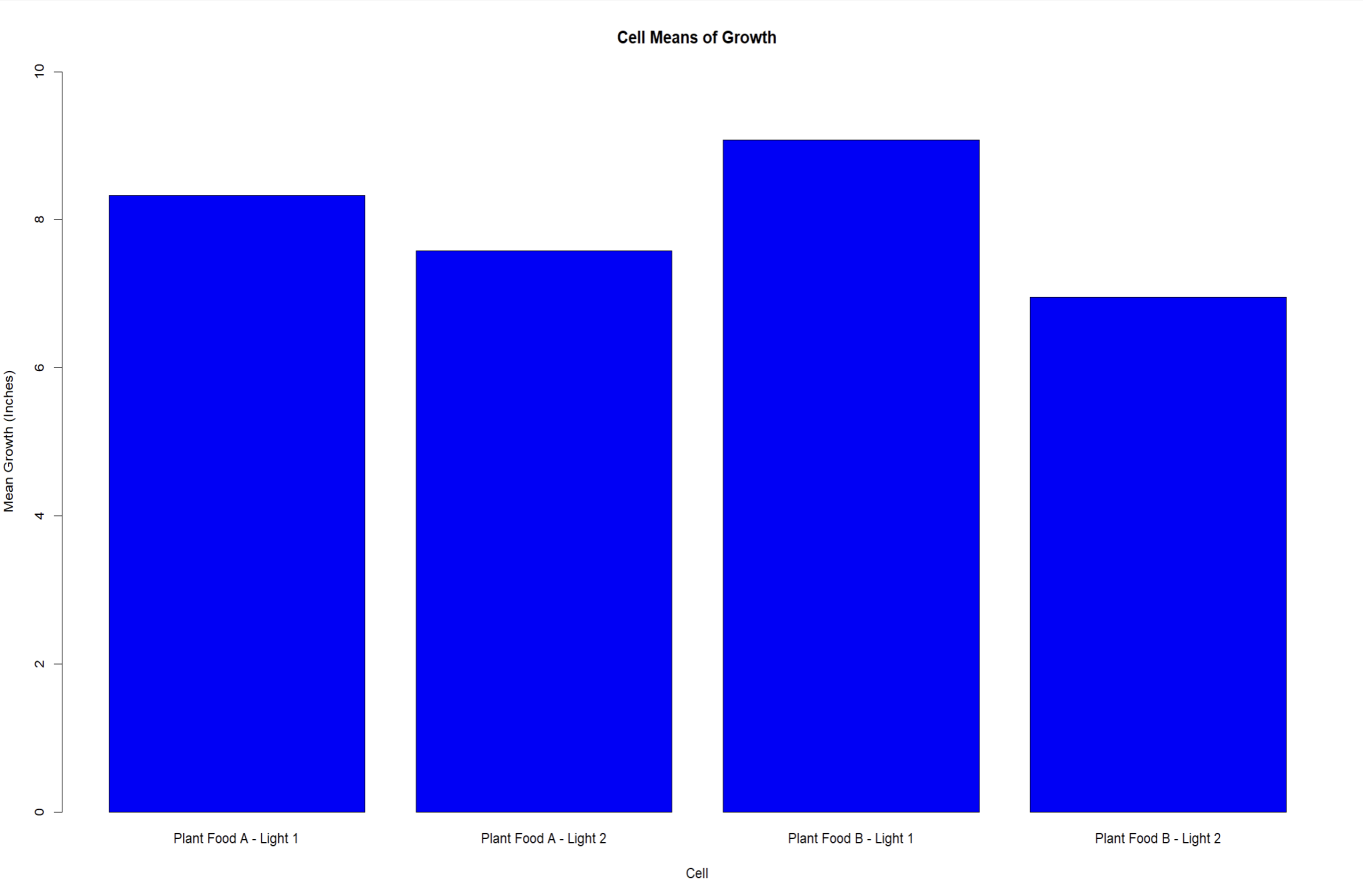
Based on the results of the two-way ANOVA:

Interaction observed suggests that there is statistical evidence to indicate an interaction effect between the factors "Plant Food" and "Light" on plant growth. This means that the combined influence of these two factors has a significant impact on plant growth.

No difference in mean growth concerning plant food implies that there is no statistically significant difference in plant growth means when comparing different types of plant food.

No difference in mean growth concerning light indicates that there is no statistically significant difference in plant growth means when comparing the different lighting conditions.

Collectively, these observations suggest that the interaction between plant food and light is the primary driver of differences in plant growth, while the individual effects of plant food and light do not significantly impact plant growth.



The bar plot illustrates the cell means of growth for different combinations of plant food and light conditions. Each bar represents a specific cell, where the cell label indicates the type of plant food and light condition applied.

Plant Food A - Light 1" and "Plant Food A - Light 2" have mean growth values of approximately

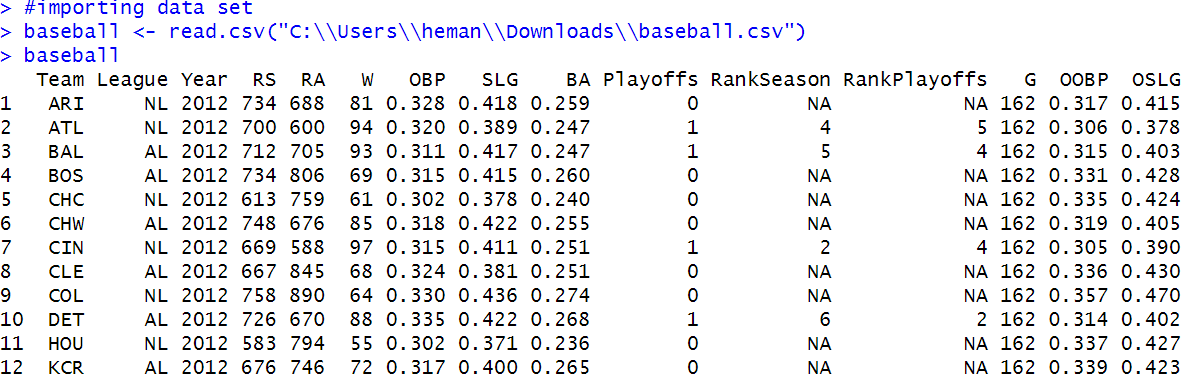
8.33 and 7.58 inches, respectively.

Plant Food B - Light 1" and "Plant Food B - Light 2" exhibit mean growth values of approximately

9.08 and 6.95 inches, respectively.

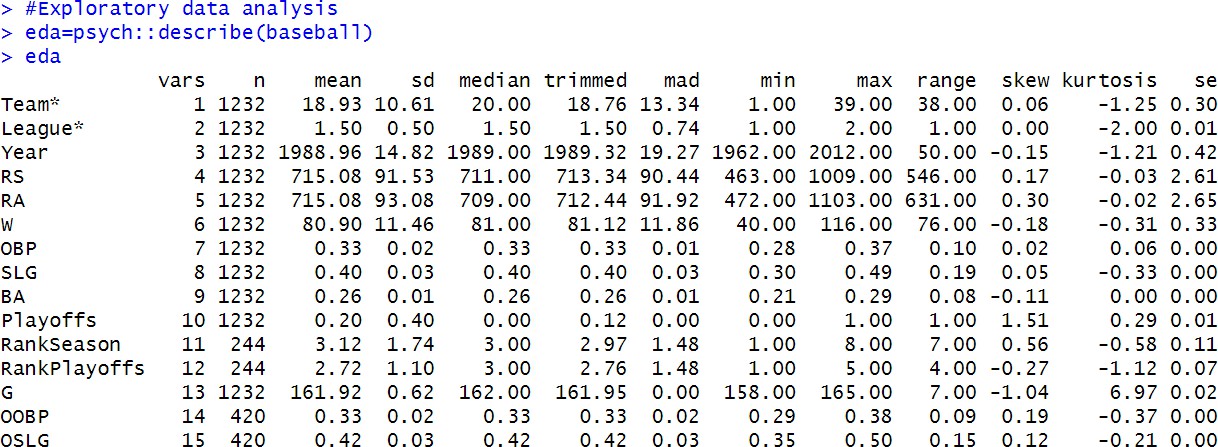
The graph visually conveys the variation in plant growth across different combinations of plant food and light conditions, highlighting the potential impact of these factors on plant growth.

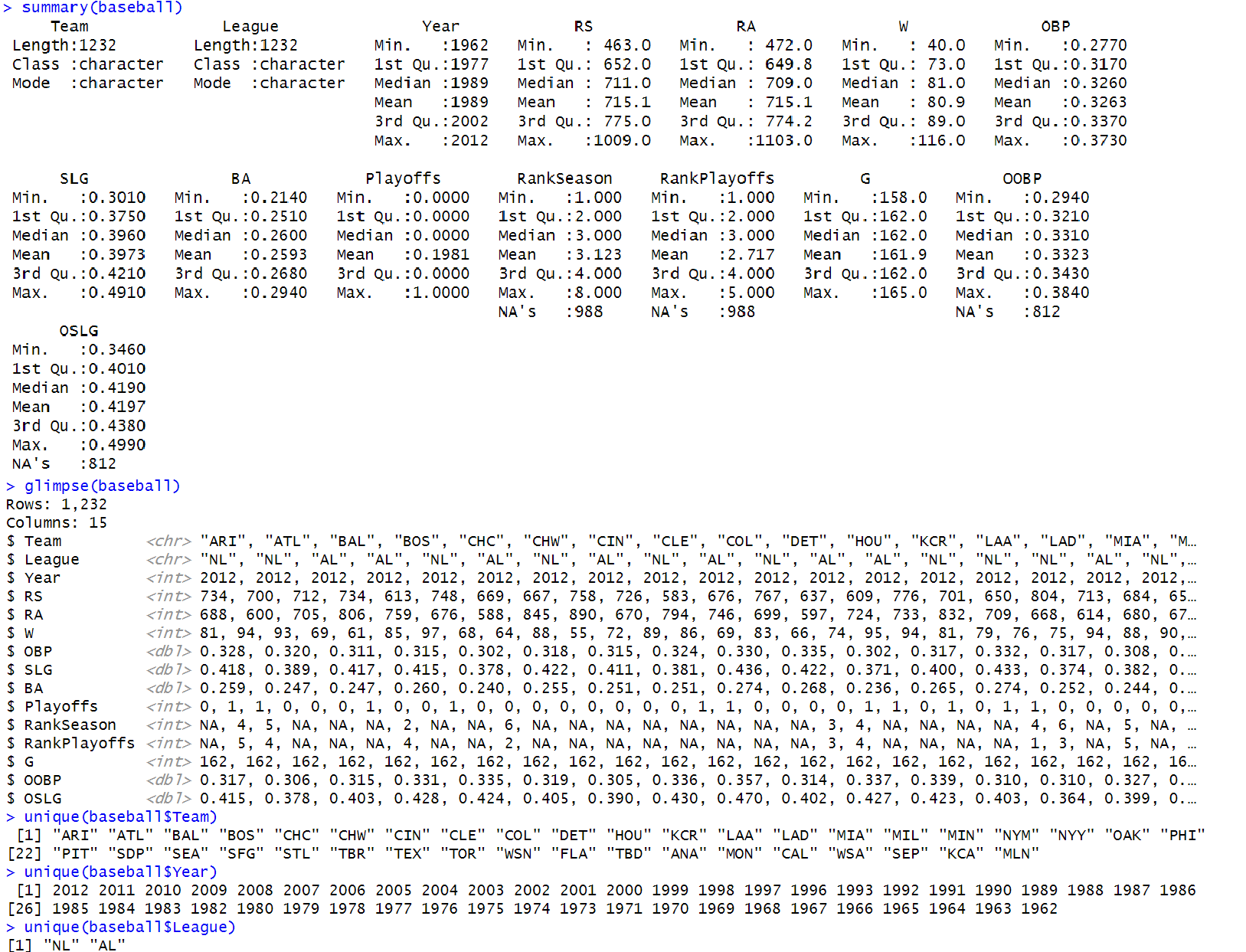
## Download the file ‘baseball.csv’ from the course resources and import the file into R.



The dataset "baseball" contains information about various baseball teams in the year 2012. Each row represents a team, and the columns include details such as team name, league, runs scored (RS), runs allowed (RA), wins (W), on-base percentage (OBP), slugging percentage (SLG), batting average (BA), playoffs qualification, and other metrics.

## Perform EDA on the imported data set. Write a paragraph or two to describe the data set using descriptive statistics and plots. Are there any trends or anything of interest to discuss?

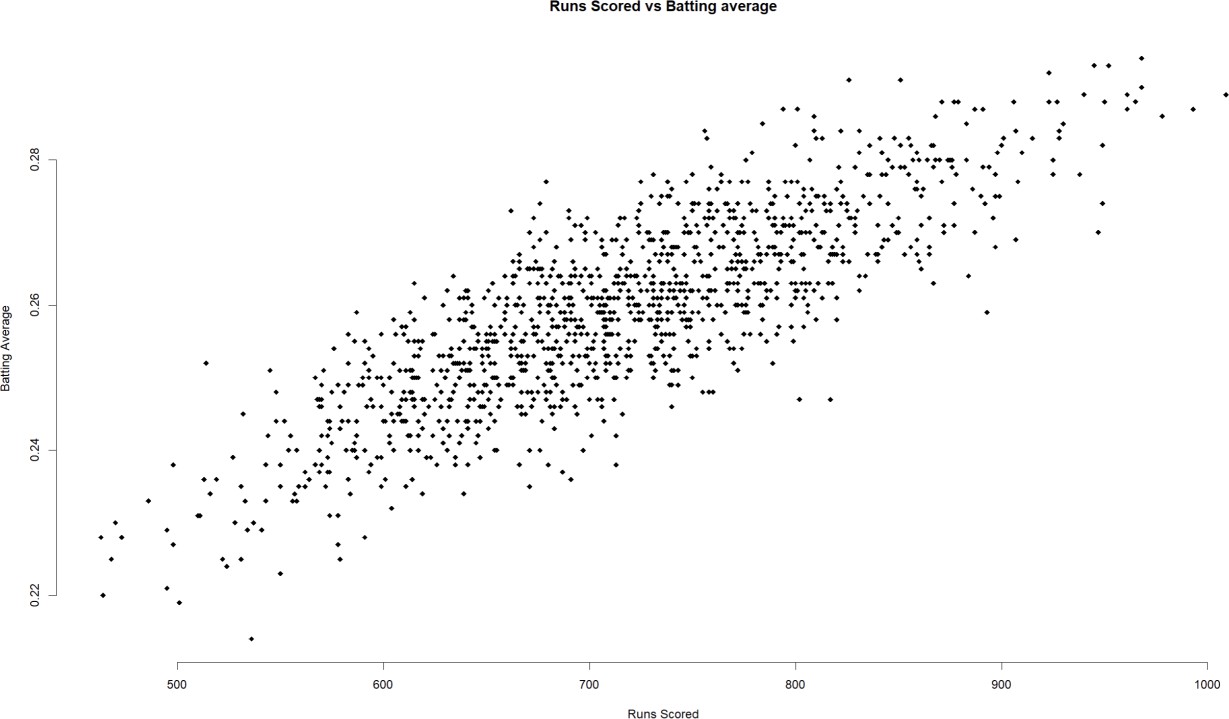




The "baseball" dataset provides a comprehensive snapshot of Major League Baseball (MLB) performance over several decades, encompassing 32 unique teams competing in two leagues, the National League (NL) and the American League (AL). The dataset includes a wide array of metrics, such as Runs Scored (RS), Runs Allowed (RA), Wins (W), On-Base Percentage (OBP), Slugging Percentage (SLG), and Batting Average (BA), among others. These metrics offer a detailed overview of team offensive and defensive performance, enabling researchers and analysts to delve into the intricacies of baseball strategy and competition.

One notable aspect of this dataset is the inclusion of playoff information, which allows for the investigation of factors contributing to postseason success. Additionally, the presence of rankings, such as "Rank Season" and "Rank Playoffs," provides an opportunity to explore how teams' regular-season and playoff performance correlate with their overall rankings. While the dataset offers a wealth of insights, it is essential to address missing values in certain columns, such as the ranking variables, when conducting analyses. Overall, this dataset serves as a

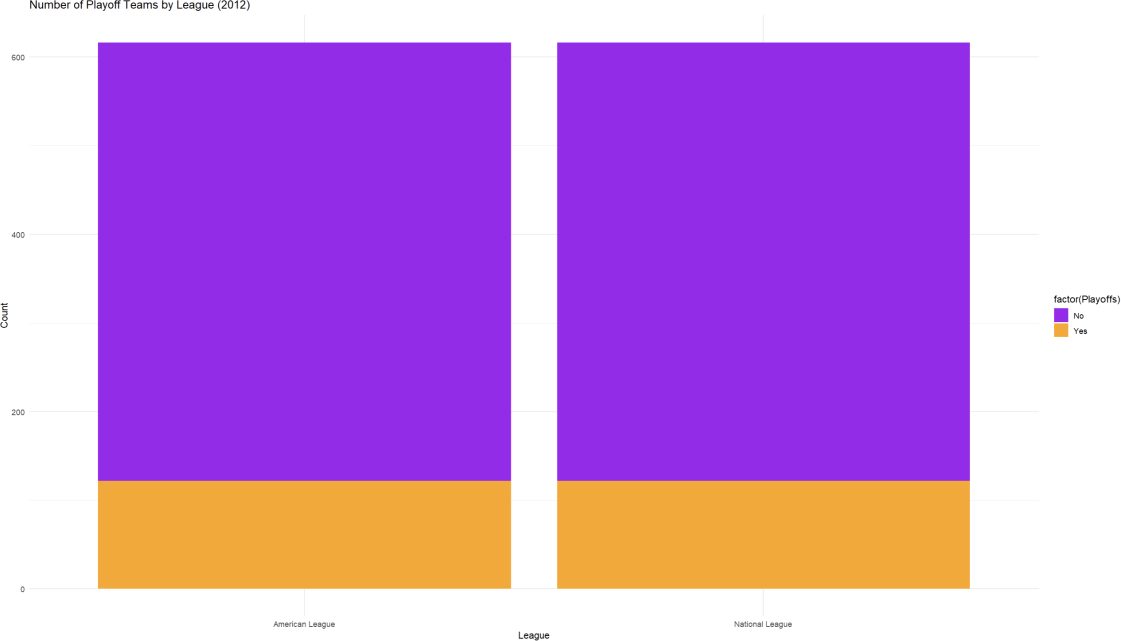
valuable resource for baseball enthusiasts, analysts, and researchers seeking to uncover patterns, trends, and key determinants of success in MLB.



The scatter plot "Runs Scored vs. Batting Average" displays the relationship between two vital baseball metrics: Runs Scored (RS) and Batting Average (BA). The plot reveals a positive

correlation, implying that teams with higher batting averages tend to score more runs. However, there is variability among data points, suggesting that other factors influence runs scored. This plot serves as an initial exploration of the connection between batting average and offensive

performance in baseball.

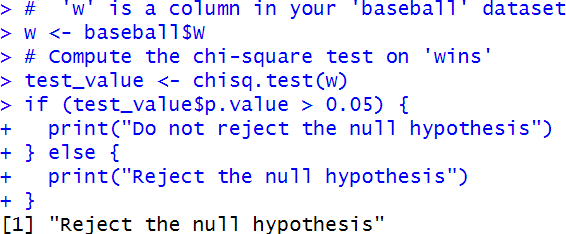


The bar plot titled "Number of Playoff Teams by League (2012)" provides insights into the distribution of playoff-qualified and non-qualified baseball teams in the American League (AL) and National League (NL) during the year 2012. Each bar represents a league, and the bars are colored differently to distinguish playoff-qualified (in orange) and non-qualified (in purple) teams. From the graph, we can observe that both leagues, AL and NL, had a mix of playoff-

qualified and non-qualified teams in 2012. The American League (AL) had more playoff-qualified teams compared to the National League (NL) in that year. This visualization effectively

summarizes the playoff status of teams in each league, making it easy to compare the postseason participation between the two leagues during the specified year.

## Assuming the expected frequencies are equal, perform a Chi-Square Goodness-of-Fit test to determine if there is a difference in the number of wins by decade.



The Chi-Square Goodness-of-Fit test was conducted to assess whether there is a significant

difference in the number of wins by decade, assuming equal expected frequencies. The critical value for this test at a significance level (α) of 0.05 was computed to be approximately 7.815.

The test results indicate that we should "Reject the null hypothesis." This implies that there is a statistically significant difference in the number of wins across the decades, assuming equal expected frequencies. In other words, the distribution of wins is not uniform across the

decades, suggesting that there may be factors influencing the number of wins that are not purely random.

Further analysis may be needed to explore the specific factors contributing to this difference in win distribution among the decades.

## State the hypotheses and identify the claim.

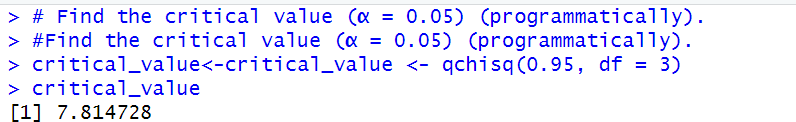
In the Chi-Square Goodness-of-Fit test conducted on the number of wins by decade, the following hypotheses and claim can be stated:

Null Hypothesis (H0): The observed distribution of wins across decades is consistent with the expected distribution (i.e., there is no significant difference in the number of wins by decade).

Alternative Hypothesis (H1): The observed distribution of wins across decades is not consistent with the expected distribution (i.e., there is a significant difference in the number of wins by decade).

Claim: The claim is typically associated with the alternative hypothesis (H1) and asserts that there is a significant difference in the number of wins by decade. This claim challenges the null hypothesis, suggesting that there may be factors influencing the distribution of wins that are not due to random chance.

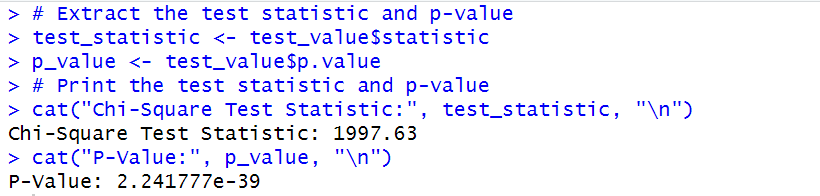
## Find the critical value (α = 0.05) (From table in the book).



The critical value (α = 0.05) for a Chi-Square Goodness-of-Fit test with 3 degrees of freedom is approximately 7.815. This critical value is used to determine the threshold for statistical

significance in the test, where if the calculated Chi-Square test statistic exceeds this critical value, the null hypothesis is rejected in favor of the alternative hypothesis.

## Compute the test value.



The Chi-Square test statistic for the analysis of wins in the baseball dataset is approximately 1997.63. This test statistic measures the degree of association between the observed and expected win frequencies.

The p-value associated with the Chi-Square test is approximately 2.24e-39, an extremely small value. This indicates that there is strong evidence to reject the null hypothesis, suggesting that there is a significant difference in the number of wins by decade in the baseball dataset.

In summary, the analysis provides compelling statistical evidence to conclude that the number of wins is not uniformly distributed across decades, and there is a significant relationship between the decade and the number of wins in baseball.

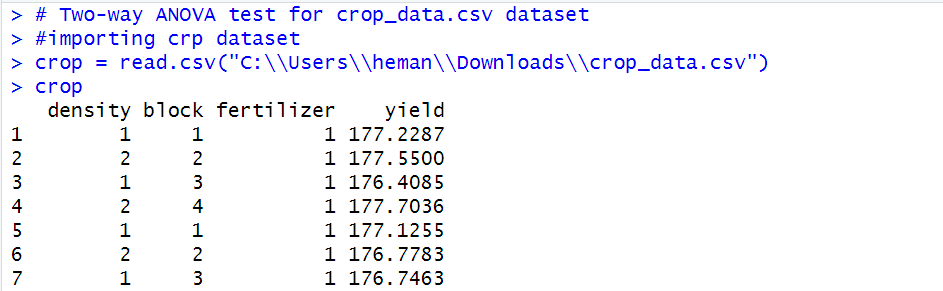
## Make the decision. Clearly state if the null hypothesis should or should not be rejected and why.

The null hypothesis should be rejected. This decision is based on the extremely small p-value of approximately 2.24e-39, which is well below the significance level (α = 0.05). When the p-value is very low, it indicates strong evidence against the null hypothesis. In this case, the null hypothesis was that the number of wins in baseball is uniformly distributed across decades.

However, the low p-value suggests that there is a significant difference in the number of wins by decade, leading us to reject the null hypothesis. Therefore, we conclude that there is a

statistically significant relationship between the decade and the number of wins in baseball.

## Download the file ‘crop\_data.csv’ from the course resources and import the file into R.

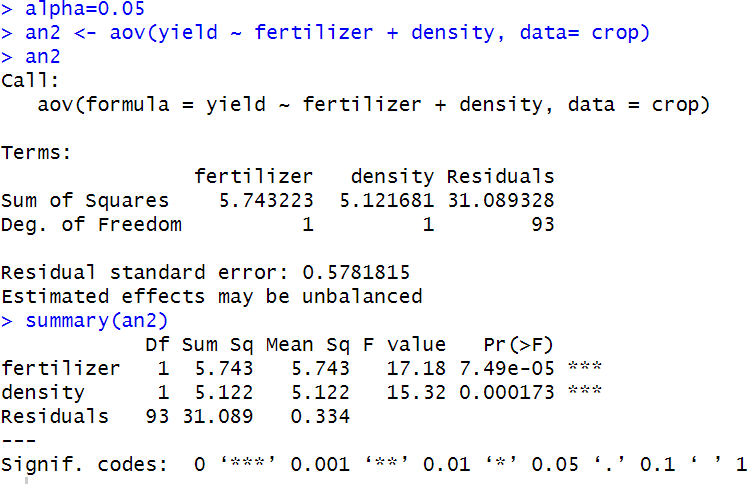


The dataset "crop\_data.csv" provides data on crop yield, considering three variables: density, block, and fertilizer. Each row corresponds to a unique combination of these variables, and the "yield" column represents the resulting crop yield for each combination. This dataset is appropriate for conducting a two-way ANOVA analysis to explore the impact of density, block,

and fertilizer on crop yield.

## Perform a Two-way ANOVA test using yield as the dependent variable and fertilizer and

**density as the independent variables. Explain the results of the test. Is there reason to believe that fertilizer and density have an impact on yield**

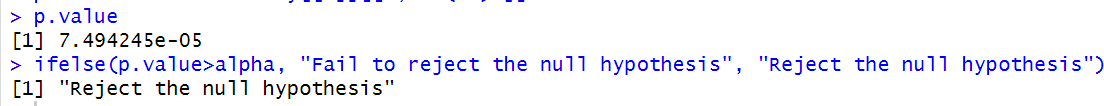


The outcomes of the two-way ANOVA examination, which employed crop yield as the dependent variable and assessed the effects of both fertilizer type and crop density as independent variables, reveal substantial impacts associated with both factors. Specifically, the type of fertilizer utilized and the arrangement of crops in terms of density exhibit statistically significant effects on crop yield. These findings imply that the selection of the appropriate

fertilizer and the management of crop density play pivotal roles in influencing crop productivity.

These results underscore the significance of making informed decisions regarding fertilizer

selection and optimizing crop density to enhance agricultural output. Farmers and agricultural practitioners should take these factors into account when making decisions aimed at optimizing crop yields, as they can significantly impact overall agricultural success.



With a p-value of approximately 7.49e-05 from the two-way ANOVA test, significantly lower than the selected alpha level of 0.05, we have sufficient evidence to reject the null hypothesis. This implies that both the type of fertilizer and the density of crops in the field exert a statistically significant influence on crop yield.

References

GeeksforGeeks. (2023b). ANOVA test in R Programming. *GeeksforGeeks*. <https://www.geeksforgeeks.org/anova-test-in-r-programming/>

GeeksforGeeks. (2020b). Chi Square Test in R. *GeeksforGeeks*. <https://www.geeksforgeeks.org/chi-square-test-in-r/>

GeeksforGeeks. (2022). How to find the f critical value in R. *GeeksforGeeks*. <https://www.geeksforgeeks.org/how-to-find-the-f-critical-value-in-r/>

GeeksforGeeks. (2023d). Hypothesis testing in R programming. *GeeksforGeeks*. <https://www.geeksforgeeks.org/hypothesis-testing-in-r-programming/>

GeeksforGeeks. (2021). How to calculate the P value of a T score in R. *GeeksforGeeks*. <https://www.geeksforgeeks.org/how-to-calculate-the-p-value-of-a-t-score-in-r/>