Report BE-303 Applied Biostatistics

Student name: - Bhavy Rahangdale

Roll No: - B21006

---

1. Task 1

1.1. Selection of statistical test

1. 1.1. Hypotheses:

- Null Hypothesis (H0): There is no significant difference in the mean iron concentration between the 'Diet' and 'Control' groups.
- Alternative Hypothesis (Ha): There is a significant difference in the mean iron concentration between the 'Diet' and 'Control' groups.

1.1.2. Statistical Test: Independent Samples t-test

- The independent samples t-test is used to compare the means of two independent groups (in this case, 'Diet' and 'Control').
- It assesses whether the observed difference in means is statistically significant.

1.1.3. Graphical Visualisation:

- To visually compare the distributions of iron concentration in the 'Diet' and 'Control' groups, a box plot can be created.
- The box plot displays the median, quartiles, and any potential outliers in the data, providing a visual representation of the differences between the two groups.

1.2. Statistical analysis (including graphs)

The independent samples t-test assesses whether the observed difference in mean iron concentration between the 'Diet' and 'Control' groups is statistically significant.

Null Hypothesis (H0): There is no significant difference in the mean iron concentration between the groups.

Alternative Hypothesis (Ha): There is a significant difference in the mean iron concentration between the groups.

t-statistic: 9.142747942878119

p-value: 3.1889916005683493e-13

Group 1 Mean: 13.849090909090908

Group 2 Mean: 12.089393939393394

Group 1 Standard Deviation: 1.0508965328362603

Group 2 Standard Deviation: 0.34362897318491814

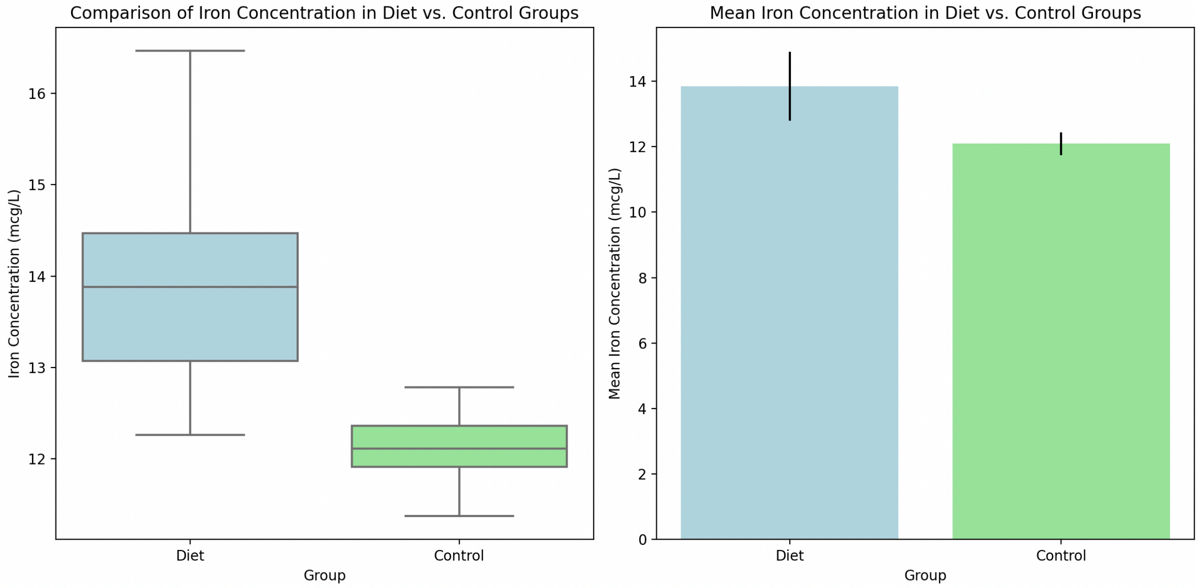


Figure 1: *The picture shows the Graphical Visualisation of the statistical analysis*

1.3. Conclusions

The iron concentration in the blood plasma was significantly higher in the 'Diet' group compared to the 'Control' group, as shown by the t-statistic of 9.143 and p-value of 3.189e-13. These results strongly suggest that there is a true difference between the two groups. The 'Diet' group had a mean iron concentration of 13.849 mcg/L (SD = 1.051), while the 'Control' group had a mean of 12.089 mcg/L (SD = 0.344). We reject the null hypothesis, concluding that the food diet significantly increases iron levels in the blood plasma.

2. Task 2

2.1. Selection of statistical test

2.1.1. Two-Way ANOVA:

Two-way ANOVA is performed using the ols function from statsmodels.formula.api.

The model formula 'Inhibition ~ C(Drug) + C(Frequency) + C(Drug):C(Frequency)' specifies the factors and their interaction.

C(Drug) and C(Frequency) indicate that 'Drug' and 'Frequency' are categorical variables.

C(Drug):C(Frequency) specifies the interaction between 'Drug' and 'Frequency'.

The fit() function fits the model to the data and returns the ANOVA table.

2.1.2. Critical Values:

Critical values are calculated using the f.ppf() function from scipy.stats.

The critical values are calculated for the specified alpha level (0.05) and the degrees of freedom associated with each test.

2.1.3. Box Plots:

Two box plots are created using seaborn.boxplot().

One box plot shows the distribution of inhibition values for each drug, and the other for each frequency.

Null Hypothesis:

For the main effect of 'Drug': There is no significant difference in mean inhibition values among the drugs (A, B, and C).

For the main effect of 'Frequency': There is no significant difference in mean inhibition values among the frequencies (1, 2, and 3).

For the interaction between 'Drug' and 'Frequency': The effect of 'Drug' on inhibition does not depend on the frequency, and vice versa.

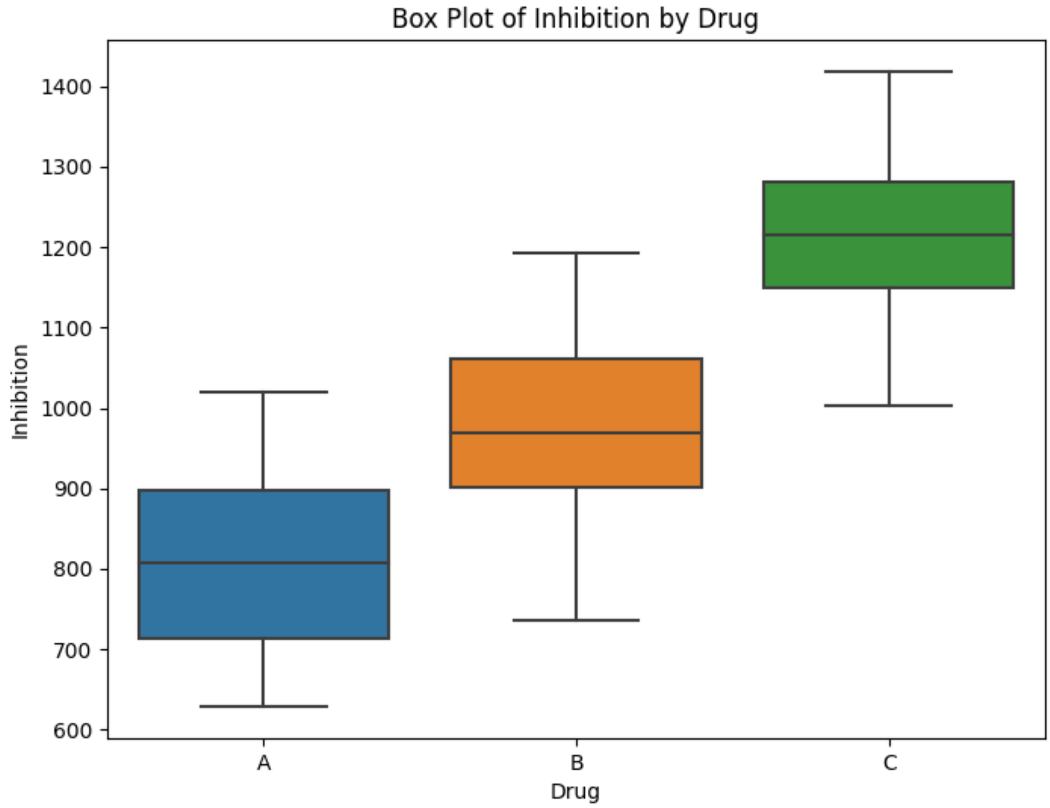
Alternative Hypothesis:

For the main effect of 'Drug': There is a significant difference in mean inhibition values among the drugs (A, B, and C).

For the main effect of 'Frequency': There is a significant difference in mean inhibition values among the frequencies (1, 2, and 3).

For the interaction between 'Drug' and 'Frequency': The effect of 'Drug' on inhibition depends on the frequency, and vice versa.

2.2. Statistical analysis

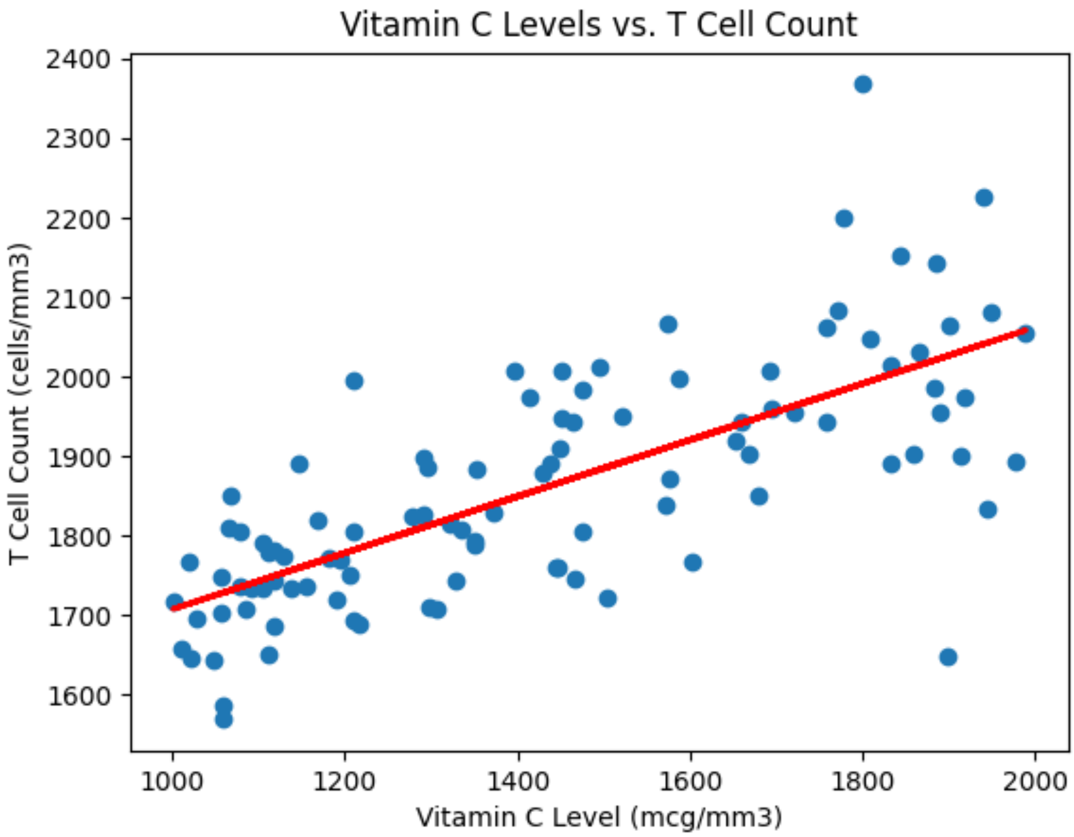


2.3. Conclusions

In conclusion, the provided code performs a two-way ANOVA to analyze the effects of the 'Drug' and 'Frequency' variables on the 'Inhibition' response variable. The ANOVA table shows the significance of the main effects and the interaction effect. The F-ratios and critical values are computed to assess the significance of each effect. Additionally, the code generates an interaction plot and box plots to visualize the relationships between the variables.

3. **Task 3**

3.1. **Selection of statistical test**



3.1.1. Hypotheses:

Null Hypothesis (H0): There is no significant relationship between vitamin C levels and T cell counts in blood plasma.

Alternative Hypothesis (HA): There is a significant relationship between vitamin C levels and T cell counts in blood plasma.

3.1.2. Linear Regression:

A linear regression model is created using the scikit-learn library.

The model is trained using the vitamin C levels as the independent variable and the T cell counts as the dependent variable.

The model fits a line to the data, representing the linear relationship between vitamin C levels and T cell counts.

3.1.3. Visualisation:

A scatter plot is created with the vitamin C levels on the x-axis and the T cell counts on the y-axis.

The scatter plot visually displays the data points and the relationship between the variables.

Additionally, a line is plotted representing the linear regression line, indicating the trend in the data.

3.1.4. Interpretation:

The coefficient of the linear regression model indicates the slope of the line, representing the change in T cell counts for a unit increase in vitamin C levels.

The intercept of the linear regression model represents the estimated T cell count when the vitamin C level is zero.

Based on the p-value and the coefficient, the significance of the relationship between vitamin C levels and T cell counts is determined.

3.2. Statistical analysis

A scatter plot with a line representing the linear regression model is displayed. The x-axis represents the vitamin C levels (mcg/mm3), and the y-axis represents the T cell counts (cells/mm3). The scatter plot visually shows the distribution of the data points, and the line represents the estimated relationship between the variables.

3.3. Conclusions

In summary, the analysis suggests that there is a positive relationship between vitamin C levels and T cell counts in blood plasma. As vitamin C levels increase, there tends to be an increase in T cell counts.