**Project Documentation**

**First of all : We read the dataset and analyze the skim through the data variables.**

## 1.Descriptive statistics:

## A)Data Summary, and calculating mean , median , minimum, maximum, first and third quartile for each variable of our data(PWD)

A screenshot of a computer

Description automatically generated

## First: We used the function summary to summarize all the data provided in the data set.

## Then: we calculated mean, median, minimum, maximum, first and third quartile for all variables of our data (PWD).

## A screenshot of a computer Description automatically generatedFor example, Feeder variable output after running the above code is:

0.25= first quartile, 0.75= third quartile.

**B) For the categorical variable existing, calculate a frequency table**

## The table for categorical variables existing :The output of the frequency table shows up that our data contains two dependent variables 20 males and 20 females.

**A picture containing text, screenshot, font, line

Description automatically generated**

* **Next :** **calculating a frequency table for the categorical variables**
* **Based on the output , there are 5 categories in the variable Treatment: A, B, C, D, and E. The output shows that there are 8 observations in each category, for a total of 40 observations in the data frame.**

## A picture containing font, text, screenshot, white Description automatically generated

## C) Calculate the correlation coefficient (ADWG0021 and ADWG2150*)* and (ADWG0021and ADWG0050)

## \* The cor.test() is a function that returns a list containing the correlation coefficient and the p-value of the test.

## \*  The $estimate element of the Correlation object is used to extract the correlation coefficient,

## *-First : calculating* the correlation coefficient (ADWG0021 and ADWG2150)

## *A picture containing text, screenshot, font Description automatically generated*

## *Based on output the correlation coefficient is 0.227035 which is larger than zero that’s why* ADWG0021 and ADWG2150 have positive correlation

## *-Second:* *calculating* the correlation coefficient of ADWG0021and ADWG0050

## *A picture containing text, screenshot, font Description automatically generated*

## *Based on output the correlation coefficient is 0.4427165 which is larger than zero that’s why* ADWG0021and ADWG0050

## have positive correlation

## Graphics:

## 1st: we made the frequency table for males and females:­­

## The code:

## A picture containing text, screenshot, diagram Description automatically generated

## The output:

## The total sample=40

## Males= 20

## Females=20

**2nd: We made a bar chart graph with mean ADWG0021in males and females:**

A bar chart is a graphical representation of categorical data in which the categories are displayed as bars of equal width and height proportional to the counts or percentages of each category. In a bar chart, the height of each bar represents the frequency or relative frequency of the category, and the bars are usually separated by small spaces to emphasize that the categories are distinct.

Bar charts are useful for visualizing the distribution of categorical data and for comparing the frequencies or relative frequencies of different categories.

## The code:

## The output:

## A screenshot of a computer Description automatically generatedA screenshot of a computer Description automatically generatedThe result shows the mean value of ADWG0021 in males and females.

## The mean of males is slightly greater than females.

## Males’ mean = 144.07

## Females’ mean = 142.15

**3rd: Make a histogram of a continuous variable: “ADWG2150” as well as “ADWG0021”.**

Histograms are useful for visualizing the shape of the distribution of a continuous variable, as well as its center, spread, and skewness. A symmetric distribution, such as a normal distribution, will have a roughly bell-shaped histogram, with the highest frequency in the center of the distribution and lower frequencies towards the tails. It can be used to check normality, yet it is not so accurate. As a result, we can use the Shapiro test (as mentioned later).

The Code:

A screenshot of a computer

Description automatically generated

## The Output:

## a) hist(data$ADWG2150):

## The histogram shows normal data distribution in ADWG2150.

## -This will be proven in details later-

## A screenshot of a computer Description automatically generatedb) hist(data$ADW0021):

## The histogram shows normal data distribution in ADWG0021.

**4th: We made a scatterplot of 2 continuous variables ADWG0050 and ADWG0021, and added the regression lines for each gender:**

**A screenshot of a computer

Description automatically generated**A scatterplot is a graph that shows how two continuous variables relate to one another. A pair of data points are plotted together, with each point denoting the value of one variable on the x-axis and the other variable's value on the y-axis. The resulting figure displays the general shape of the relationship between the two variables, along with any potential trends, clusters, or outliers.

Regression lines: used to visualize the relationship between two continuous variables in a scatterplot. It is a straight line that represents the best-fit relationship between two variables, based on a statistical model that describes the relationship between them.

A screenshot of a computer

Description automatically generatedThe Code:

A screenshot of a computer

Description automatically generated

The Output:

The scatter plot shows a very weak positive correlation (or no correlation) between the two variables.

**5th: Make a boxplot of ADWG0021 and separate boxplots per Treatment (as. factors):**

## A boxplot is a graphical representation of the distribution of a continuous variable or variables. Boxplots display a box that represents the middle 50% of the data, with a line inside the box representing the median value and "whiskers" that extend from the box to the minimum and maximum values, excluding outliers.

## Boxplots are particularly useful for comparing the distributions of multiple groups or variables, and for identifying potential outliers or extreme values in the data.

## The Code:

## A picture containing text, font, screenshot, line Description automatically generated

## A picture containing diagram, screenshot, rectangle, plan Description automatically generatedThe Output:

## In this output we are using separate boxplots to compare the effect of each treatment(A, B,C,D,E) on the variable ADWG0021 ,which represents the average daily weight gain between day 0 to day 21.

## A picture containing diagram, screenshot, rectangle, plan Description automatically generated

## 6. Hypothesis testing:

* **A)1)We hypothesis that ADWG0021 is different between male vs female.**

## First, we create two new variables Male and Female which contain the values of ADWG0021 for males and females(Male= 2 and female= 1).

## The code is used to do a two-sample t-test to compare the means of two groups in the data frame PWD based on a continuous variable ADWG0021 and a categorical variable Sex. And the output, including the t-statistic, degrees of freedom, p-value, and confidence interval.

## \*t.test performs a Welch two-sample t-test to compare the means of ADWG0021 between males and females.

## \*The sample estimates: report the means of ADWG0021 for males and females, which are 144.0724 and 142.1528, This explains that males have slightly higher values of ADWG0021 than females, but the difference is not statistically significant.

## A picture containing text, screenshot, font, algebra Description automatically generated

## Our Null hypothesis is that male and female are not different

## Alternative hypothesis : Male and female are different

## P-VALUE= 0.7 > 0.05 { Since p-value is greater than alpha value(0.05) which means that we don’t have enough evidence to reject null hypothesis and supporting alternative hypothesis }.

## 2) Assuming normality and homoscedasticity, can you test this hypothesis using statistical hypothesis framework, assess whether the previous test assumptions have been met for the test.

## \*Assess whether the previous test assumptions have been met for the test: By testing normality using ( Histogram, qq-plot, Shapiro-test)

## Histogram: The code is used to create two histograms of a continuous variable, separated by a categorical variable, to visually assess any differences in the distribution of the variable between the two groups.

## 

## A picture containing text, screenshot, diagram, line Description automatically generated

## Histogram doesn’t describe whether the data is normally distributed or not, so we used Shapiro-test to check the normality of the data.

## 1) qq-plot and Shapiro-test for males:

## A picture containing text, font, screenshot Description automatically generated

## Output:

## 

## In qq-plot : Since points on the plot fall close to the line, it indicates that the data is normally distributed.

## A picture containing text, font, white, screenshot Description automatically generated

## Our The null hypothesis : data is normally distributed

## The alternative hypothesis : data is not normally distributed

## Shapiro-test output : P-value= 0.9513 Since p-value is greater than alpha value(0.05) which means that we don’t have enough evidence to reject null hypothesis and supporting alternative hypothesis }.

## 2) Histogram, qq-plot and Shapiro-test for Females:

## 

## Output:

## A picture containing text, line, diagram, plot Description automatically generated

## The straight line fit most of the points, so this shows that the data is normally distributed.

## 

## Histogram doesn’t describe whether the data is normally distributed or not so we used Shapiro-test to check the normality of the data.

## 

## Our The null hypothesis : data is normally distributed

## The alternative hypothesis : data is not normally distributed

## Shapiro-test output : P-value= 0.98432 Since p-value is greater than alpha value(0.05) which means that we don’t have enough evidence to reject null hypothesis and supporting alternative hypothesis }.

## In conclusion: (For proving normality of our data)

## The null hypothesis : data is normally distributed

## The alternative hypothesis : data is not normally distributed

## \*Since p-values of both males and females > alpha(0.05) , then we don't have enough evidence to reject the null hypothesis and support the alternative hypothesis.

## \*We assume normality(no enough evidence to support non normality) ,so previous test assumptions have been meet for the normality test, Therefore we proved normality using qq-plot and Shapiro-test.

## \*By comparing the two histograms, we can visually assess any differences in the distribution of ADWG0021 between males and females. For example, if one histogram has a higher peak or is more spread out than the other, it may indicate a difference in the mean or variance of ADWG0021 between the two groups, but the two histograms show that both distributions of ADWG0021 between males and females are normally distributed.

## Testing homoscedasticity

## \*Assess whether the previous test assumptions have been met for the test: By testing homoscedasticity using Bartlett's-test of homoscedasticity.

## A picture containing text, font, screenshot, line Description automatically generated

## \*The code is used to fit a linear regression model to the relationship between ADWG0021 and Sex and create a set of four diagnostic plots of the model to assess the assumptions of the model and identify any potential issues with the model fit or the data.

## Output: The scatterplot of the fitted values against the dependent variable can help identify patterns in the relationship between the variables, such as linearity or curvature.

## The Q-Q plot of residuals can help assess the normality of the residuals.

## The plot of residuals against the fitted values can help identify heteroscedasticity (unequal variances) or nonlinear relationships.

## A picture containing text, diagram, line, parallel Description automatically generated

## In Normal Q-Q plot: we can use it to check normality of the data and as shown in the output most of the points are near to form a straight line that’s why the data is normally distributed.

## In Scale-Location :checks the Variance and from the output it indicates that the data is homoscedastic.

## In Bartlett's test

## The null hypothesis (HO) means there is no difference in variance, the alternative (H1) means there is differences in variance (Heteroscedastic).

## From the test results : we can see that p-value equals to 0.4129 which is larger than the significant level alpha(0.05) so we do not have enough evidence to reject the null hypothesis and in support of alternative hypothesis. This concludes that we can assume our data is homoscedastic (they do have the same variance).

## We hypothesis that ADWG0021is “different” in the group receiving *Treatment A (*normal feed + ZnO)compared to the Treatment B (normal feed + nutraceuticals). Can you test this hypothesis assuming heteroscedasticity , Assess the previous test assumption.

## First: We hypothesis that ADWG0021is “different” in the group receiving Treatment A compared to the Treatment B

## A screenshot of a computer program Description automatically generated with low confidence

## The Output:

## \*diff = (A-B), creates a new vector diff that contains the differences between the values of ADWG0021 for the A and B treatment groups. This vector represents the paired differences between the two groups, which will be used in the paired t-test.

## \*Then we applied a two-sided paired t-test on the differences between the A and B treatment groups.

## \*The t.test() function is used to perform a t-test on two sets of observations.

## A picture containing text, line, font, diagram Description automatically generated

## Most of the points fitted the straight line as shown in the Q-Q plot this shows that there is no difference in variance(Homoscedastic) and the data is normally distributed (it indicates that the differences are normally distributed).

## A picture containing text, screenshot, font, number Description automatically generated

## The output of t-test shows that the value of P-value(0.06534) is greater that alpha value (0.05) so we don’t have enough evidence to reject null hypothesis and in support of alternative hypothesis.

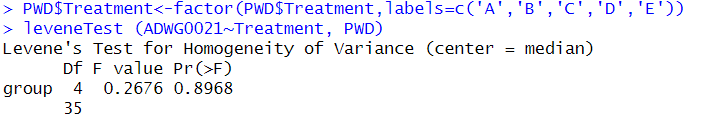
## mean difference is 24.59077 this means that mean difference is not equal to 0 which indicates that there is no significant difference between treatment A and treatment B.

## So, in conclusion after doing t-test and Q-Q plot and we hypothesis that ADWG0021is “different” in the group receiving Treatment A compared to the Treatment B we found that our data is homoscedastic (they have the same variance)

C) **We hypothesis that ADWG0021is different between the different Treatments . Can you perform comparison between the different groups, after assessing the assumptions and performing post-hoc testing (assuming normality and homoscedasticity).**

**Firstly: ADWG0021is different between the different Treatments perform comparison between the different groups.**

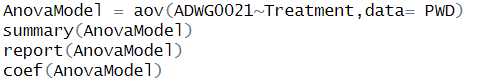
**\*Levene's test for homogeneity of variance .The test is used to assess whether the variances of a continuous variable are equal across different groups or levels of a categorical variable.**

****

**The null hypothesis of Levene's test is that the variances of the groups are equal**

**The alternative hypothesis is that at least one of the variances is different from the others.**

**Since the result of levene-test is p-value(0.8968) > alpha value(0.05), so we don't have enough evidence to reject the null hypothesis and they have same variance , therefore we can state that they are homoscedastic. as there is not enough evidence to prove that the variances of the ADWG0021 values are different across the five treatment groups.**

****

**\*Anova test is a statistical formula used to compare variances across the means (or average) of different grou**ps.

**\*The aov() function is used to fit an ANOVA model.**

**\*Anova Model produces a summary of the ANOVA model.**

Output of anova test :

**A picture containing text, font, screenshot, line

Description automatically generated**

**\*The summary() function is used to display a summary of the ANOVA model which includes the degrees of freedom, the sums of squares, the mean squares, the F-statistic, and the p-value and this information help us to assess the significance of the relationship between ADWG0021 and Treatment.**

A black text on a white background

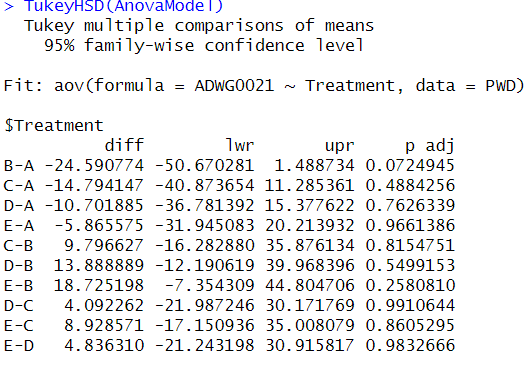
Description automatically generated with low confidence

**\*The coef() function is used to extract the coefficients of the model, which include the intercept and slope coefficients for each level of the categorical variable Treatment.**

**Since the result of anova test is p-value(0.101) > alpha value(0.05) so, we don't have enough evidence to reject the null hypothesis this indicates that they have same variance , therefore we can state that they are homoscedastic.**

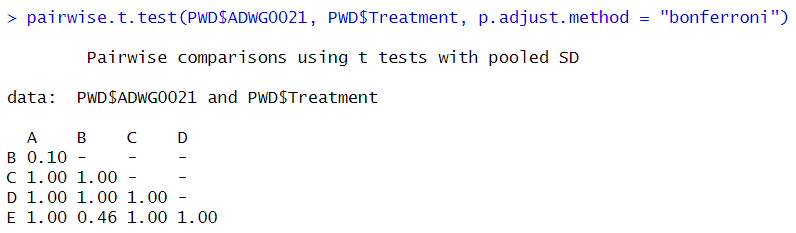
**Performing post-hoc testing (assuming normality and homoscedasticity).**

**Post-hoc testing is a statistical procedure that is used to make pairwise comparisons between groups after an overall ANOVA or regression analysis has been performed, in order to determine which specific groups, differ significantly from each other.**

****

**The output of the TukeyHSD() function display a table of comparisons between each pair of levels of Treatment, along with the difference between the means, the standard error, the p-value, and the 95% confidence interval for the difference.**

**\* But post-hoc tests can increase the chance of making a Type I error (rejecting the null hypothesis when it is true).**

****

**\*We used the Bonferroni correction method to account for multiple comparisons**

**The output of the pairwise.t.test() function would display a table of comparisons between each pair of levels of Treatment, along with the difference between the means, the t-statistic, the degrees of freedom, the p-value, and the adjusted p-value (after applying the Bonferroni correction).**

**\*** **It is important to note that the pairwise t-test with Bonferroni adjustment is a conservative approach, which increases the chance of making a Type II error especially when the sample size is small, or the variances are unequal.**

**The table indicates which pairs of levels are significantly different from each other based on p-values.**