**Statistical Analysis**

A panel dataset is a dataset that contains observations on multiple individuals (also known as "panel members") over multiple time periods. Some common descriptive statistics that can be used on a panel dataset include:

**Means and standard deviations**: These statistics can be used to describe the overall level and variability of the variables in the dataset. These statistics provide a general sense of the average level and variability of the variables in the dataset, which can be useful for comparing different groups or time periods.

**Frequency distributions:** These can be used to describe the distribution of the variables across the different individuals and time periods, which can be useful for identifying patterns or outliers in the data.

**Cross-tabulations**: These can be used to examine the relationship between two or more variables in the dataset, which can be useful for identifying patterns or associations in the data.

**Autocorrelation and heteroscedasticity tests**: These tests can be used to check for temporal dependencies and non-constant variances in the data, which can affect the validity of certain statistical models

**Panel-specific statistics**: Like fixed effects, random effects, and first differences can be used to account for unobserved individual-specific characteristics and temporal dependencies in the data. These statistics are useful for accounting for unobserved individual-specific characteristics and temporal dependencies in the data, which can improve the accuracy and interpretability of the estimates.

**Panel Data Hausman Test**

This method is used to test whether a panel data fixed effects model or a panel data random effects model is more appropriate for a given dataset. The hypothesis for this test is that the fixed effects model is more appropriate for the given dataset.

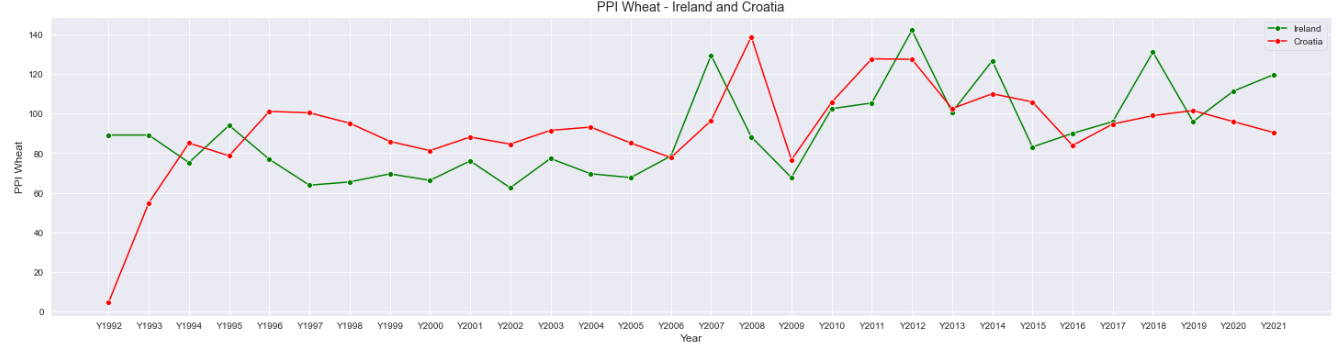
The results of this test show that the p-value for the Hausman test is less than 0.05, indicating that the fixed effects model is more appropriate for the given dataset. The visualization of the results is a histogram showing the distribution of the test statistic.

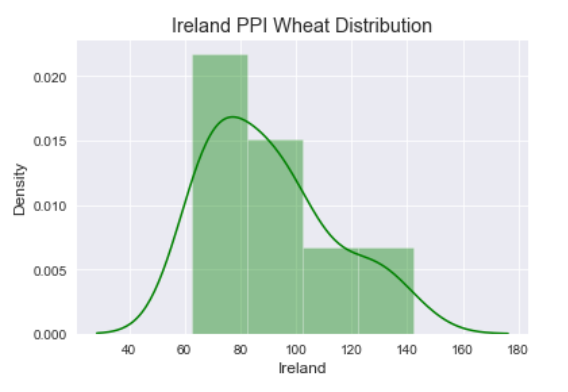
Conclusion: The panel data Hausman test results indicate that the fixed effects model is more appropriate for the given dataset.

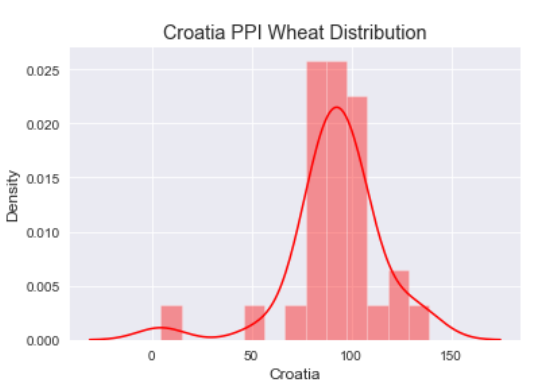
similarities between some country(s) against Ireland, apply parametric and non-parametric inferential statistical techniques to compare them (e.g., t-test, analysis of variance, Wilcoxon test, chi-squared test, among others). You must justify your choices and verify the applicability of the tests. Hypotheses and conclusions must be clearly stated. You are expected to use at least 5 different inferential statistics tests.

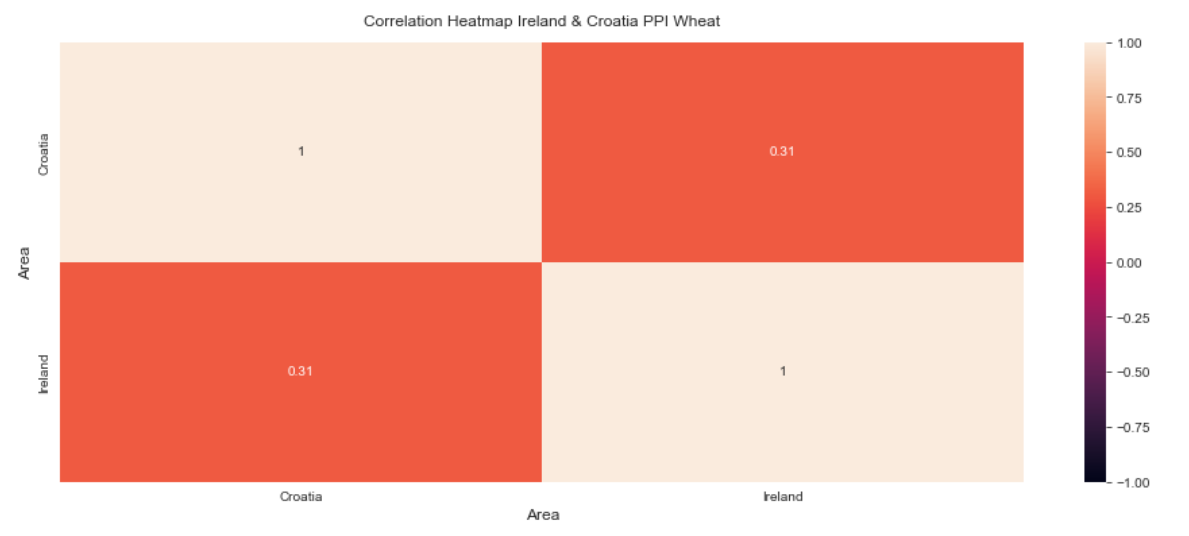
Q: Price of Wheat in Ireland vs Croatia (similar in size to Ireland by population).

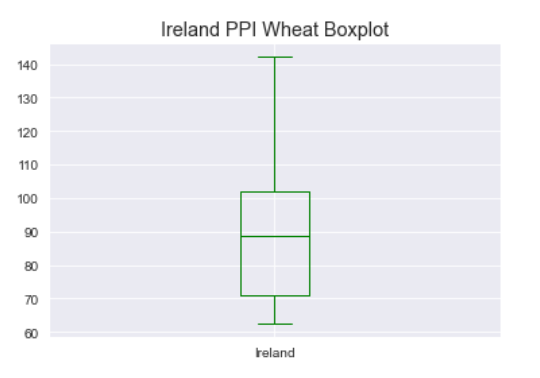
Note: The Croatian War of Independence was fought from 1991 to 1995

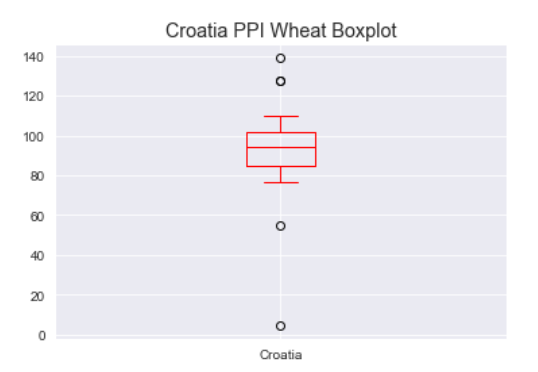


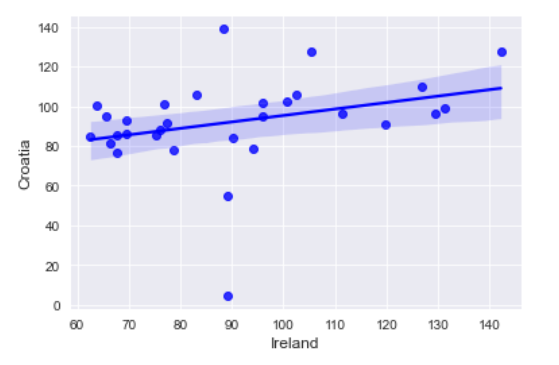












**Statistical Tests:**

**T-Test, two populations, at 95% siginifgance level.**

**Referece:** [**https://support.minitab.com/en-us/minitab/20/help-and-how-to/statistics/basic-statistics/how-to/2-sample-t/interpret-the-results/key-results/**](https://support.minitab.com/en-us/minitab/20/help-and-how-to/statistics/basic-statistics/how-to/2-sample-t/interpret-the-results/key-results/)



###### “ Because the p-value is less than 0.000, which is less than the significance level of 0.05, the decision is to reject the null hypothesis and conclude that the ratings of the countries are different. “

Because the p-value is more than 0.05, which is the significance level, the decision is to accept the null hypothesis, therefore we are unable to conclude that the PPI of Wheat in Ireland and Croatia are different.

To determine whether the difference between the population means is statistically significant, compare the p-value to the significance level. Usually, a significance level (denoted as α or alpha) of 0.05 works well. A significance level of 0.05 indicates a 5% risk of concluding that a difference exists when there is no actual difference.

P-value ≤ α: The difference between the means is statistically significantly (Reject H0)

If the p-value is less than or equal to the significance level, the decision is to reject the null hypothesis. You can conclude that the difference between the population means does not equal the hypothesized difference. If you did not specify a hypothesized difference, Minitab tests whether there is no difference between the means (Hypothesized difference = 0). Use your specialized knowledge to determine whether the difference is practically significant. For more information, go to Statistical and practical significance.

P-value > α: The difference between the means is not statistically significant (Fail to reject H0)

If the p-value is greater than the significance level, the decision is to fail to reject the null hypothesis. You do not have enough evidence to conclude that the difference between the population means is statistically significant. You should make sure that your test has enough power to detect a difference that is practically significant. For more information, go to Power and Sample Size for 2-Sample t.

T-Value DF P-Value

6.31 32 0.000

Test

Null hypothesis H₀: μ₁ - µ₂ = 0

Alternative hypothesis H₁: μ₁ - µ₂ ≠ 0

Key Result: P-Value

In these results, the null hypothesis states that the difference in the mean rating between two countries is 0. Because the p-value is less than 0.000, which is less than the significance level of 0.05, the decision is to reject the null hypothesis and conclude that the ratings of the countries are different.

Step 3: Check your data for problems

Problems with your data, such as skewness and outliers can adversely affect your results. Use the graphs to look for skewness (by examining the spread of each sample) and to identify potential outliers.

**Examine the spread of your data to determine whether your data appear to be skewed.**

**Poisson Distribution** = Elements, in a place, in a period of time.

For Ireland, what is the probability that the Wheat PPP will be greater than the Irish average in a given year?

0.4892

For Ireland, what is the probability that the Wheat PPP will be less than or equal to the Irish average in a given year?

0.5107

We can confirm this is correct by adding these two probabilities together, to get 1. The Wheat PPP in Ireland must either be greater than, less or equal to the average, which covers all possibilities.

**Normal Distribution:**

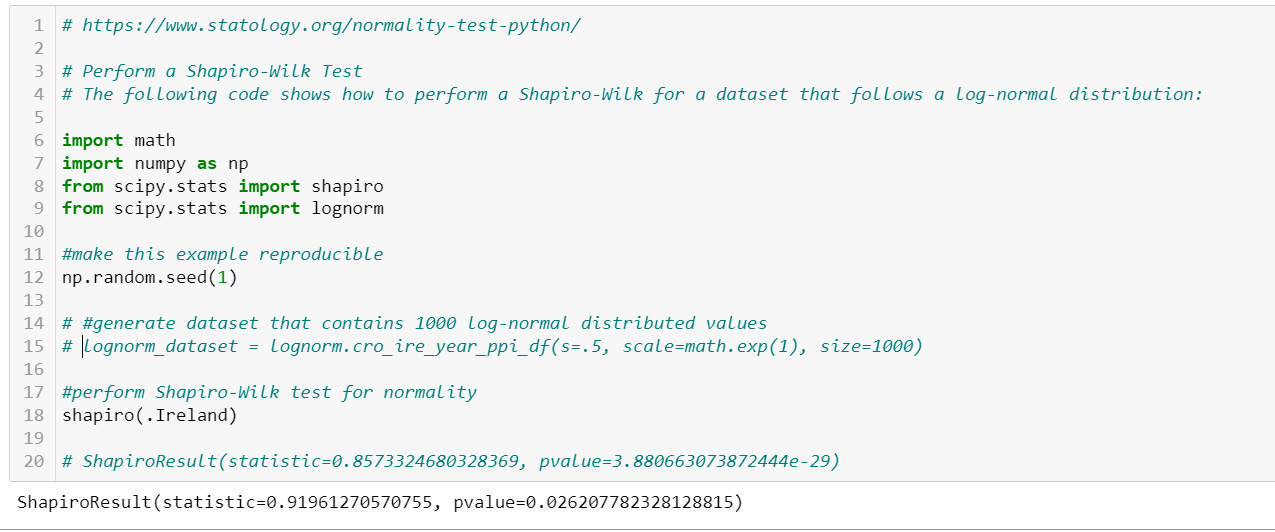
**We need to check the skewness of Ireland**

* The variables with skewness > 1 are highly positively skewed.
* The variables with skewness < -1 are highly negatively skewed.
* The variables with 0.5 < skewness < 1 are moderately positively skewed.
* The variables with -0.5 < skewness < -1 are moderately negatively skewed.
* Variables with -0.5 < skewness < 0.5 are symmetric i.e normally, which is the case for Ireland.

**3. (Formal Statistical Test) Perform a Shapiro-Wilk Test.**

* If the p-value of the test is greater than α = .05, then the data is assumed to be normally distributed.

Ireland – Test says Ireland data is not normally distributed (not sure if correct, looks normally distributed)



**Wilcoxon signed-rank test,** also known as Wilcoxon matched pair test is a non-parametric hypothesis test that compares the median of two paired groups and tells if they are identically distributed or not.

In this example, the Wilcoxon Signed-Rank Test uses the following null and alternative hypotheses:

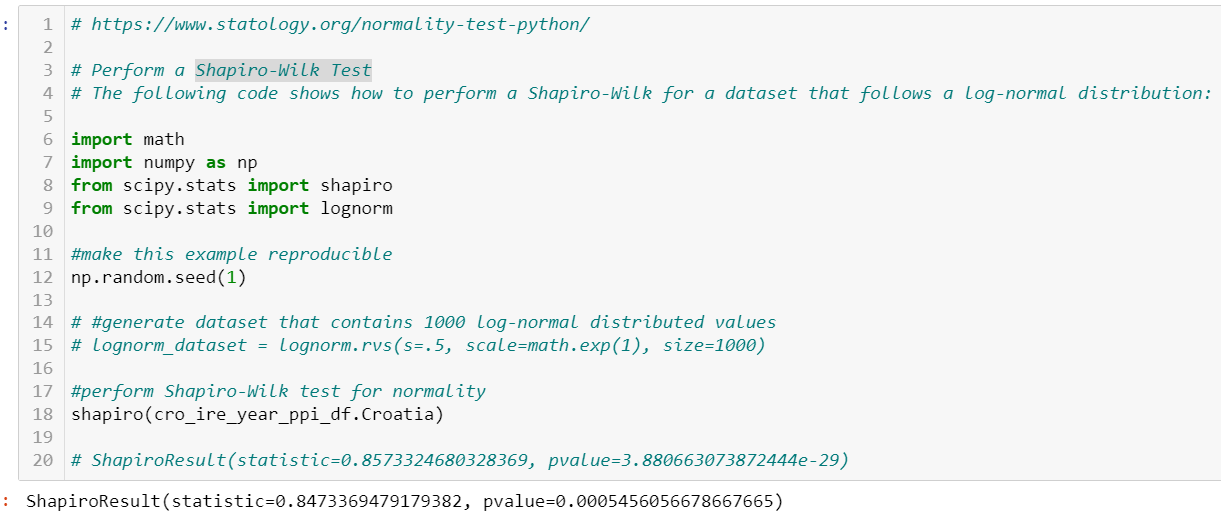
H0: The true mean is equal between the two groups

HA: The true mean is not equal between the two groups

Since the p-value is less than 0.05, we reject the null hypothesis.

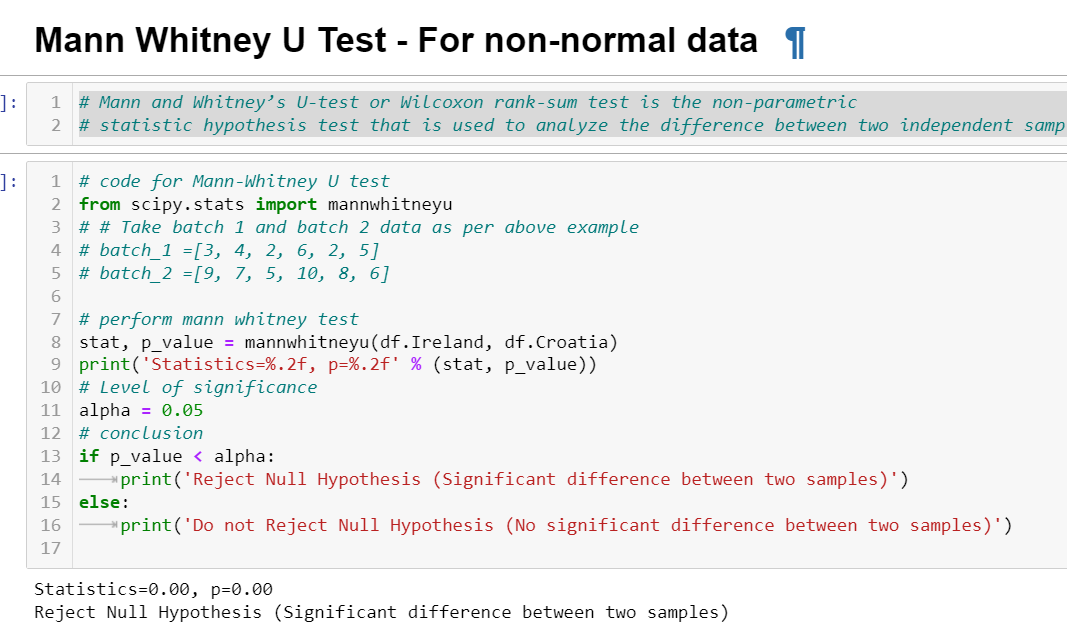
We have sufficient evidence to say that the true mean is not equal between the two countries.

Croatia: Test says Croatia data is not normally distributed



# Mann Whitney U Test – For non-normal data

Mann and Whitney’s U-test or Wilcoxon rank-sum testis the non-parametric statistic hypothesis test that is used to analyze the difference between two independent samples of ordinal data.

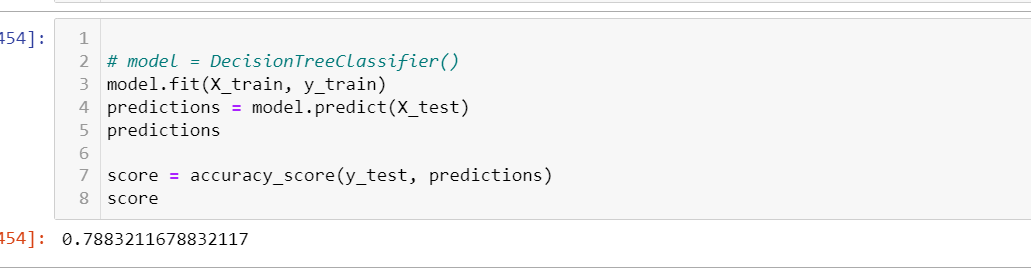


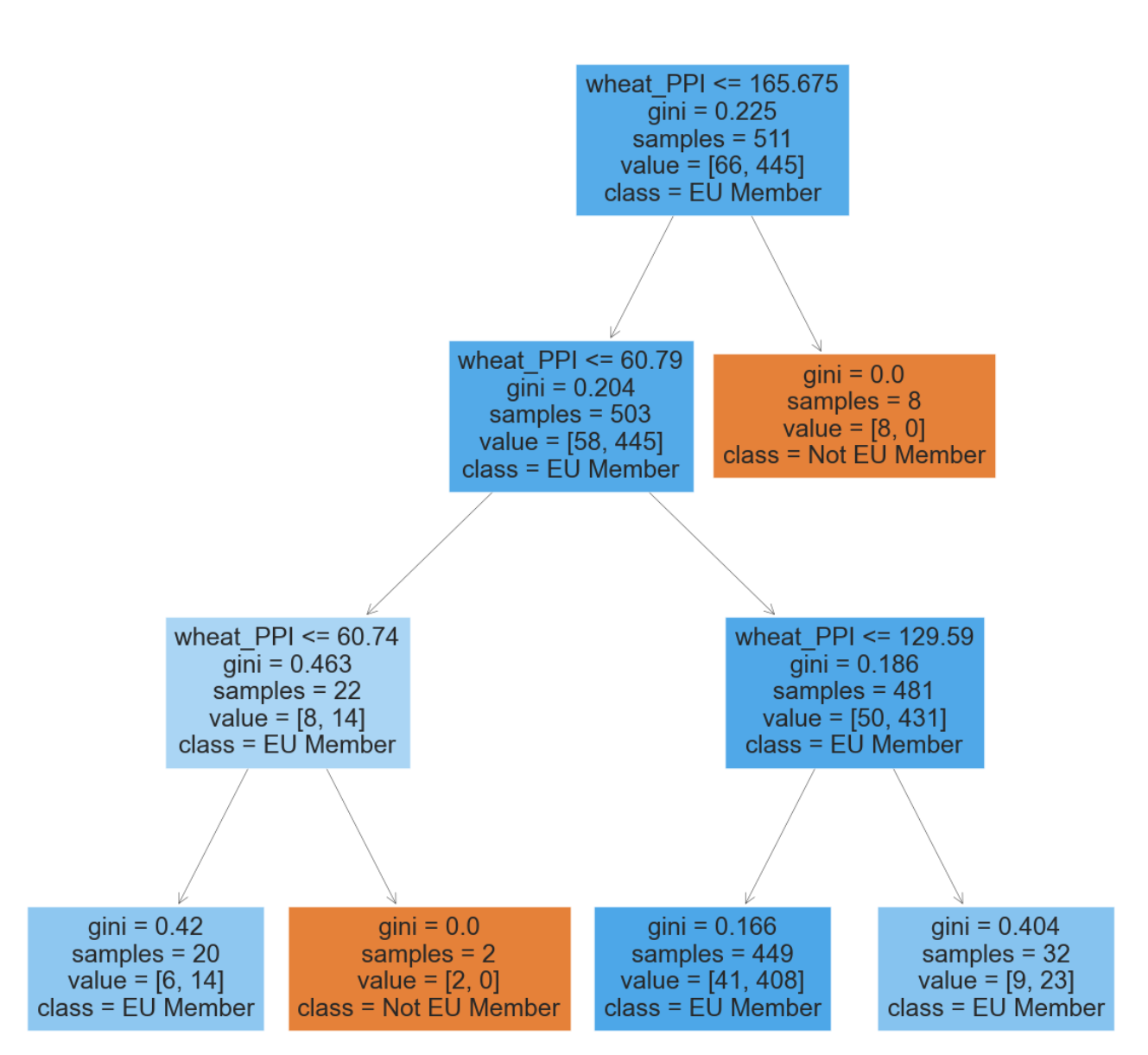
ANOVA Test:

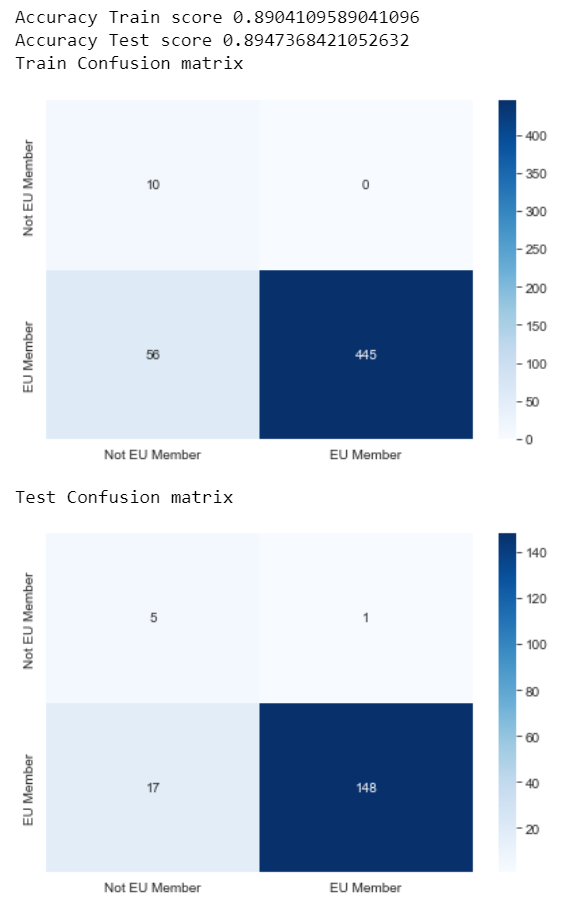
The null hypothesis (H0) is that there is no difference between the groups and equality between means (walruses weigh the same in different months). The alternative hypothesis (H1) is that there is a difference between the means and groups (walruses have different weights in different months.



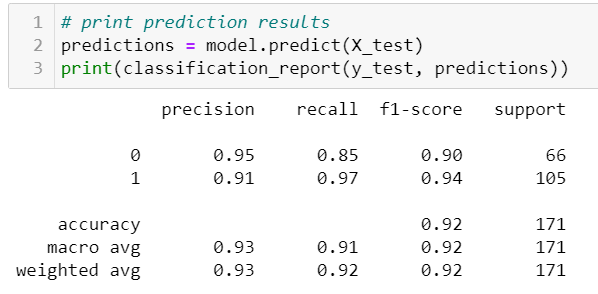
Decision Tree







GridSearchCV:



Then aftern running GridSerchCV

