**Assignment 2**

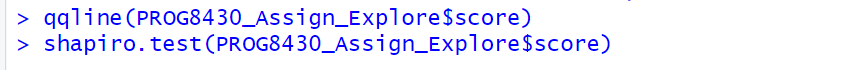
**Author:** Vidhya Venugopal, 8902970

**Course:** PROG8430

1. PROG8430\_Assign\_Explore\_23S

1.1 Inference

1.1.1 Normality - #Created a QQ Normal plot of the Political Awareness Test Score:



A picture containing text, screenshot, plot, line

Description automatically generated

1.1.2 # Conducted a statistical test for normality on the Political Awareness Test Score

A screenshot of a computer

Description automatically generated with low confidence

1.1.3 # Are the Political Awareness Test Scores normally distributed? What led you to this conclusion?

In the QQ Plot the points of the Political Awareness Test Scores in the plot fall along a straight diagonal line and the p-value of the Shapiro-wilk test is greater than 0.05. Therefore, the Political Awareness Test scores are normally distributed.

1.2 Statistically Significant Differences

1.2.1 # Compared the Political Awareness Test Scores between the treatment and control group using a suitable hypothesis test:

A picture containing text, screenshot, font

Description automatically generated

A screenshot of a computer program

Description automatically generated with medium confidence

1.2.2 # Explain why you chose the test you did.

Based on the previous QQ Plot, it has been concluded that the Political Awareness Test Scores is been normally distributed. Therefore, I used f.test and t.test to compare and check the variance of the scores against each group.

1.2.3 # Do you have strong evidence that the average test scores are different between the treatment and control groups?

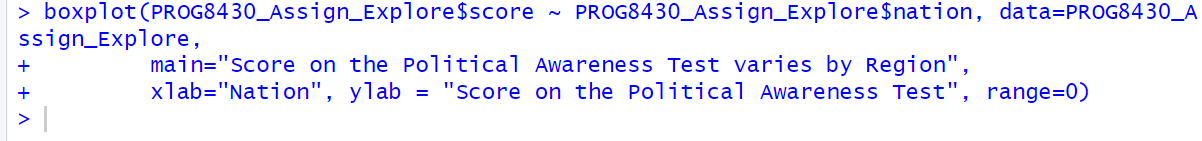
Yes, I have strong evidence that the average test scores are different between the treatment and control groups because based on the t.test, p-value < 0.05, therefore there is a significant difference in means and, there is a 95 percent confidence interval.

1.3 Multiple Statistical Differences

1.3.1 # The Score on the Political Awareness Test varies by Region using ANOVA (statistical) and a sequence of boxplots (graphical)

A screenshot of a computer

Description automatically generated with low confidence



A picture containing text, screenshot, diagram, rectangle

Description automatically generated

1.3.2 # The Measure of Political Involvement (Pol) varies by Political Affiliation using ANOVA and a sequence of boxplots.

A screenshot of a computer

Description automatically generated with low confidence

A picture containing text, font, screenshot, electric blue

Description automatically generated

A picture containing text, screenshot, diagram, rectangle

Description automatically generated

1. Ayr Temperature

1.4 Data Transformation

1.4.1 # Read the Ayr data and transformed it into an appropriate time series datatype.

A picture containing text, screenshot, font

Description automatically generated

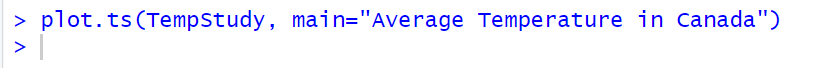
1.5 Descriptive Analysis

1.5.1 # Summarized the information (mean, std dev, etc.)

A screenshot of a computer

Description automatically generated

1.5.2 # Plot the time series data.



A picture containing text, screenshot, diagram, line

Description automatically generated

1.5.3 # Decomposing the times series data into the constituent components. Comment on each (any trends you observe, etc.)

The function decompose decomposes the data to the three constituent elements which is ‘seasonal’, ‘trend’, ‘random’. Because the fluctuations seem constant over time, an additive model probably  
describes it best. I have observed in the plot that the seasonal is an averaged value based on the 4 cycles and the trend is a moving average.

A screenshot of a computer

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated with low confidence

A picture containing text, screenshot, font, number

Description automatically generated

A picture containing text, font, screenshot, number

Description automatically generated

A picture containing text, screenshot, font, number

Description automatically generated

A close-up of a logo

Description automatically generated with low confidence

A picture containing text, font, handwriting, line

Description automatically generated

1.5.4 # Deseasonalize the information and plotting the result.

Seasonal:

A picture containing text, font, screenshot, line

Description automatically generated

A picture containing text, screenshot, diagram, line

Description automatically generated

Trend:

A picture containing text, screenshot, font, line

Description automatically generated

A picture containing text, screenshot, font

Description automatically generated

Random:

A picture containing text, font, screenshot, line

Description automatically generated

A picture containing text, diagram, font, line

Description automatically generated

1.5.5 In order to deseasonalize, estimate the seasonal component by subtracting it from the other two components which is trend and random. I have observed in the seasonal plot that the change or low fluctuation between the years 1960 to 1975 and after 1985 to 1995. There is highest fluctuation between the year 1975 to 1985 and from 1995 to 2003. On the other hand, the trend component makes changes in the overall time series. Finally, in random component there is a random changes in the time series data except the changes observed in seasonal and trend components.

1.5.6 # Smooth the temperature chart using a moving average. Try 3 different values for the moving average and choose the one you think best shows the trend (if any).

n=5

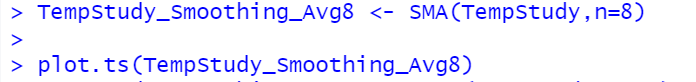
A picture containing text, font, screenshot, line

Description automatically generated

A picture containing text, line, diagram, plot

Description automatically generated

n=8



A picture containing text, diagram, line, plot

Description automatically generated

n=15

A picture containing text, font, handwriting, line

Description automatically generated

A graph of a mountain

Description automatically generated with low confidence

1.5.7 # Determine if the time series is stationary.

A picture containing text, screenshot, font, line

Description automatically generated

After smoothing with a moving average, the p-value which is generated from the adf.test is not less than 0.05. Therefore, the time series seems to be not stationary.

1.5.8 # Creating an autocorrelation chart (using acf) and comment on which lags are significant. Do previous values seem to influence current values?

A blue text on a white background

Description automatically generated with low confidence

A picture containing text, screenshot, line, diagram

Description automatically generated

Autocorrelations calculate the correlation between observation 𝑡𝑛and 𝑡𝑛−𝑖 𝑤ℎ𝑒𝑟𝑒 𝑖 = 0 𝑡𝑜 𝑛 𝑜𝑏𝑠  
Values above the blue line are significant. This suggests that recent values are more highly correlated with outcomes that older values.

1.6.1 # Create a simple moving average forecast of temperature in Ayr for five years beyond the data provided. Graph your results along with a 75% prediction interval.

A screenshot of a computer

Description automatically generated

A picture containing text, screenshot, font

Description automatically generated

A picture containing text, line, font, plot

Description automatically generated

1.6.2 # Create an exponentially smoothed forecast of temperature in Ayr for five years beyond the data provided. Graph your results along with a 75% prediction interval.

A screenshot of a computer program

Description automatically generated with low confidence

1.6.3 # Compare the two forecasts you created in steps 1 and 2 above. Which forecast seems superior? Why?

Basically, the loss function value represents the measure of the discrepancy between the forecasted values and the actual values in a forecasting model. Therefore, SMA(57) model seems to be superior as the discrepancy between the actual value is less comparatively with ETS(MNN) model.

A picture containing text, screenshot, font, algebra

Description automatically generated

A picture containing text, line, plot, screenshot

Description automatically generated