Question 1:

Select a publicly available dataset

Dataset of International rugby matches involving Ireland (2003-2018) was selected. It contains Ireland rugby match data from their first game of the 2003 world cup to their last game of 2018.

It contains 11 columns, two of them have text values, one has date values and the rest have numerical values.

a) Consider two variables of the dataset, and develop a decision-making strategy to check whether two averages of variables are equal at the significant level alpha=0.01.

Two columns with numerical values Rating and Opposition Rating are taken to test the hypothesis whether two averages of variables are equal at the significant level alpha = 0.01.

These columns show the highest correlation with the Result variable.

<u>Two-sided Hypothesis testing of the mean for two populations</u>

To test the hypothesis, 5 steps should be performed.

Step 1. Stating the hypothesis

H0: Mu1 = Mu2

H1: Mu1 != Mu2

Step 2. Setting significance level alpha = 0.01

Step 3. Computing the test value

$$test.value = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{(\sigma ¿¿ 1 ¿¿ 2/n_1) + (\sigma ¿¿ 2 ¿¿ 2/n_2) ¿¿¿¿}}$$

Step 4. Finding the critical value:

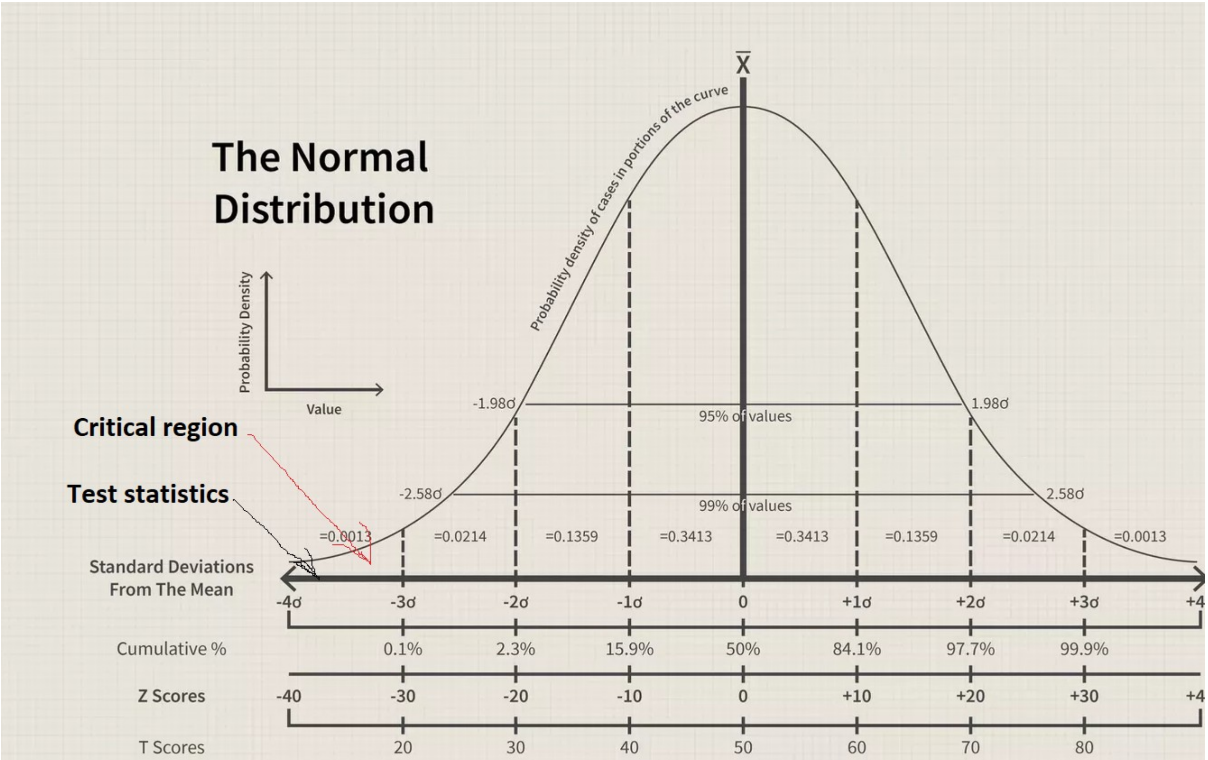
$$c.value = qnorm(1 - \alpha/2)$$

Critical value for alpha = 0.01 is equal to 2.58

Step 5. Specifying the decision rule:

if $¿ test.value \vee \geq c.value$ therefore $H_0 is rejected.$

Since test statistics = 4.63 which is greater than 2.58, then H0 is rejected.



Conclusion: The means of two populations are not equal at the significant level alpha = 0.01

b) Consider two variables of the dataset, and develop a decision-making strategy to check whether two averages of variables are different at the significant level alpha=0.10.

Two-sided Hypothesis testing of the mean for two populations

To test the hypothesis, 5 steps should be performed.

Step 1. Stating the hypothesis

H0: Mu1 = Mu2

H1: Mu1 != Mu2

Step 2. Setting significance level alpha = 0.10

Step 3. Computing the test value

$$test.value = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{(\sigma ¿¿ 1 ¿¿ 2/n_1) + (\sigma ¿¿ 2 ¿¿ 2/n_2) ¿¿¿¿}}$$

Step 4. Finding the critical value:

$$c.value = qnorm(1 - \alpha/2)$$

Critical value for alpha = 0.10 is equal to 1.64

Step 5. Specifying the decision rule:

if $¿ test.value \vee \geq c.value$ therefore $H_0 is rejected.$

Since test statistics = 4.63 which is greater than 1.64, then H0 is rejected.

c) Consider one variable in the dataset, and apply the test of the mean for a proposed candidate of μ at the significant level alpha=0.05.

A variable for Result was chosen. It will be tested for a mean = 9.5

Two-sided Hypothesis testing of the mean

To test the hypothesis, 5 steps should be performed.

Step 1. Stating the hypothesis

H0: Mu = 9.5

H1: Mu != 9.5

Step 2. Setting significance level alpha = 0.05

Step 3. Computing the test value

$$test.value = \frac{(\bar{X} - \mu_0)}{\sqrt{\sigma^2/n}}$$

Step 4. Finding the critical value:

$$c.value = qnorm(1 - \alpha/2)$$

Critical value for alpha = 0.05 is equal to 1.96

Step 5. Specifying the decision rule:

if $¿ test.value \vee \geq c.value$ therefore $H_0 is rejected.$

Since test statistics = -1.75 which is by absolute value less than critical value, the hypothesis is accepted.

Question 2

a) Build an ordinary least square model (OLS) for your dataset and show the summary information**.**

For the regression Result variable was chosen as independent variable (target output) and Rating and Opposition Rating as dependent variables.

A constant was added to the model at the initial stage.

The p-value (**P>|t|**) for the constant was greater than a significance level of 0.05, therefore the constant was insignificant in the model.

The value of R-squared for the model was 0.425

b) Find the intercept and the coefficients, and write the model equation that considers only significant features.

The insignificant constant was excluded and model was run again.

Now both variable show level of p-values less than 0.05 and value of R-squared increased to 0.484 which shows moderate level of correlation.