**applied statistics assignment -07**

**Name: Shiva Kumar Peddapuram**

**KSU ID:- 811235874** [**speddapu@kent.edu**](mailto:speddapu@kent.edu) **Submitted to Jun Li**

***Question: - 6.1 (Data file: UN11) With the UN data, perform a test of NH : (6.6) against the alternative AH :(6.7), and summarize results. This is an overall F-test for a model with one factor and no additional regressors***

***Solution:***

***Code:***

Graphical user interface, text, application

Description automatically generated

Execution part:

Graphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated

***Observation:***

Text

Description automatically generated

***Explanation:***

**In the above-mentioned problem we have initially installed library by using install. Packages () now we have installed alr4 package .and we have checked the installed library whether the library is installed or not by using the function library(“alr4”). Now we have defined our data file that is UN11.**

***Observation: -***

**So, from the above 2 models we can observe the following content:**

**For the first model Res.Df is 198 and for RSS is 20293.2 and Sum of squares is 0**

**For the second model Res.Df = 196, RSS=7730.2**

***Conclusion: -***

**So from the above code executions we can conclude that by using annova function in R we can find and calculate desired F test when used with two fitted models.so from the above we can see that p\_value is almost nearly 0.**

**The residuals for the NH model is 10.12 and it decreases to 6.28 for the 2nd model. The intercept for the AH model is 82.446 which is larger than the intercept for the NH model**

**So, we can conclude that log(ppgdp) cannot be eliminated.**

**6.4 (Data file: UN11) With the UN data, consider testing**

**NH : lifeExpF ~log(ppgdp) group:log(ppgdp)**

**AH: lifeExpF ~group l + + og(ppgdp)+ group:log(ppgdp)**

**The AH model is the most general model given at (6.10), but the NH was not given previously.**

***6.4.1 Explain in a sentence or two the meaning of the NH model.***

***Solution:***

1. The NH model is a log linear regression model.
2. It uses log of ppgdp variable with group variable which is mentioned above.
3. By using this we will be knowing whether any group and log of ppgdp as per requirements like individually or both at a time
4. ) There is a common intercept and a range of slopes in each group.
5. Each group has a common catchment and several slopes.
6. Here, when group is absent, every distinct intercept is eliminated

***I have also included the code for the above models:***

**Text, letter

Description automatically generated**

***6.4.2 Perform the test and summarize results.***

***Answer:***

***Coding part: -***

Graphical user interface, text, application, email

Description automatically generated

**Code execution model wise:**

**Graphical user interface

Description automatically generated with medium confidence**

**First model:**

**A screenshot of a computer

Description automatically generated with medium confidence**

**2nd model:**

**Table

Description automatically generated with medium confidence**

**Annova function:**

**Text, letter

Description automatically generated**

**Summary:**

**here I have taken summary of annova function so that we can get some overview of observations**

**Table

Description automatically generated**

***Explanation:***

**The residual factor for the first model is 195 and for the 2nd model is 193. RSS FOR 1STmodel is 5232.0 RSS for the 2nd model is 5077.7.**

**Sum of squares = 154.3**

**Important observation and conclusion:**

**Here, we find that none of the variables are significant when run individually on the NH model's linear regression results (p-values for all variables are all > 0.05 at the 5% level of significance).**

**But when variables are combined into a model, they significantly improve the model' ability to predict female life expectancy.   
So, we can conclude that the Alternate Hypothesis that the model is significant is accepted and that the Null Hypothesis that the model is unimportant is rejected by the p-value = 2.2e- 16 <<< 0.05 at the 5% significance level.**