**Assignment 1, part 2, SCMA 632 (30 points)**

**Due 9 AM, Monday, October 31, 2016**

**E-mail Excel file to scma.stat@gmail.com with subject line SCMA632 A1P2**

*R.L. Andrews*

Assignment 1, Part 2 has individual and group work. Each group will electronically submit one Excel spreadsheet containing the individual and group work. Data for this are found on my homepage http://www.people.vcu.edu/~randrews/ under the title [**U.S. Census Bureau Data (summary for states)**](http://www.people.vcu.edu/~randrews/fast_track/US_Census_Bureau_State_Data_2006.xls).

Individually you are to examine the list of variables on the Column Variable Description tab. You are to use the variable in **column 49, People of all ages in poverty - percent 2004**, as the response or dependent variable. Then using these data, each person individually (15 points) will

1. without doing any analysis of the data, select five variables from the list that you think would be a good predictor of the response variable, People of all ages in poverty.
2. use the appropriate data for the five variables selected above and build what you think is the best model to predict the response variable selecting predictors from your five selected variables. Submit your variable list and regression output for your best model that includes VIF values for each variable to the other team members by 9 AM Saturday, October 22. This submission to them is to be labeled with your name and put on a tab for the final group submission. Do not make changes to your individual file after you confer with the other group members.

Each group (15 points) will

1. put each team member’s answers to the parts above on individual tabs at the beginning of an Excel file with the person’s name on the tab and put all of their work on one tab.
2. confer among the team members and select an appropriate set of variables and data values to be used for building the best model to predict the response variable. You may create new variables from the information in this data set. Clearly define each created variable.
3. build the model that you think is the best model and describe how you arrived at the model, as well as giving the regression output for this final best model that includes VIF values for each variable.

The final submission for each team is to be sent to me in an Excel spreadsheet via e-mail to **scma.stat@gmail.com** by **9 AM, Monday, October 31**. Each person on the team is to have a tab with their name on it and their responses to the individual parts 1 & 2. The team will have a tab describing any created variables and how they arrived at the final model and a tab with the final model regression output for this final best model that includes VIF values for each variable.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group 1** | **Group 2** | **Group 3** | **Group 4** | **Group 5** | **Group 6** | **Group 7** | **Group 8** | **Group 9** |
| **Lnu** | **Jain** | **Boyce** | **Bates** | **Erskine** | **Ahmad** | **Gordon** | **Shah** | **Howse** |
| **Wiley** | **Mann** | **Ciccolo** | **Rapp** | **Janjua** | **Halun** | **Harris** | **Stells** | **Pandya** |
| **Wood** | **Mattauch** | **Sharma** | **Zhu** | **Mahoney** | **Madsen** | **Rancic** | **Tunney** |  |