MGMT302 Term Project, definition and description section

R.L. Andrews

This project will be done as a team and the team members will be assigned after the withdraw date.

**1. List all contributing team members on the first tab.** For each team member create a tab with his/her name on it containing his/her individual information for the work laid out in the **Preliminary work for MGMT302 Course project** document.

**2.** Each team will use a data set for the 50 U.S. states with measurements for five variables meeting the criteria below. You are not restricted to the variables that team members submit individually, but these data at least provide a starting point.

**C1 Categorical Variable 1** is to have two categories (≥ 10 per group). (Location: Coastal border or Inland )

**C2 Categorical Variable 2** is to have three categories (≥ 10 per group). (Time zone: Eastern, Central or Other)

**Q1 Quantitative Variable 1** is to have a quantitative measurement. It is a number that truly measures a quantity.

**Q2 Quantitative Variable 2** is to have **a quantitative measurement that is different from Q1**.

**R Response Variable**, like the two above, it is to have a quantitative measurement. Select **a variable different from those above** and **one for which your team would like to see if its value may be related in some way to the values of the other variables**.

The quantitative variables should have many possible values (**at least 20 possible values for Q1, Q2, & R.** The variables **C1, C2, Q1, Q2 & R must be measurements of separate variables (Q1 & Q2 can’t be subdivisions of R)**. Choose the variables Q1, Q2 and R so that their **value is not zero for any state**.

 **Provide a separate operational definition for C1, C2, Q1, Q2 & R that includes the source of the values for each variable.** An **Operational Definition** is a definition that clearly defines a measurement and it must be clear so that all persons involved will have exactly the same understanding of what is being defined. It clearly describes the units of measurement, such as square miles. For the qualitative or categorical variables provide clear definitions for each of the categories. If the source for the values of a variable was a website then give the URL for that website.

**3. Report the values for all five variables for each of the 50 states listed alphabetically (see sample template below).**

**MAKE SURE THAT THE DATA ON FIVE VARIABLES MEET THE ABOVE CRITERIA!**

***Have at least 10 observations for each category of a categorical variable.***

Below is a sample for sheet column headings.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **State** | **C1**.Var. name  | **C2.**Var. name | **Q1.**Var. name | **Q2.**Var. name | **R**. Var. name |
| State’s Name | Value of C1 for the state | Value of C2 for the state | Value of Q1 for the state | Value of Q2 for the state | Value of R for the state |

**Save the data in a secure place! Do not save all your data & analysis on only one storage device.**

**4. State what you believe will be the relationships between each of these pairs of variables** (Be Specific!)**.**

 **a. C1 vs. R, b. C2 vs. R, c. C1 vs. C2, d. Q1 vs. R, e. Q2 vs. R, and f. Q1 vs. Q2**

**These statements are to be written based on your believe prior to doing your analysis of the data.**

Examples:

a. C1 vs. R: We think there would be no relationship between (Coastal or Inland ) and R (% unemployment).

b. C2 vs. R: We think there is no relationship between time zone (Eastern, Central or Other) and R (% unemployment).

c. C1 vs. C2: We think that that the Central time zone states will have a smaller proportion of states that are Coastal.

d. Q1 vs. R: We think that as Q1 (% attending college) increases, R (% unemployment) will decrease.

**5. Screen the to determine if there are there any invalid data values that you are discarding for further analysis because they are not indicative of the phenomenon being studied?** If you have any such values, list the state and the specific values and for each value tell why you think it is not indicative and should not be retained in the data set for analysis.

**(Carefully follow the report and analysis requirements on both sides of this sheet!)**

**MGMT302 Term Project, report guide & analysis section**

**E-mail an Excel file to randrews@vcu.edu by noon, Friday, June 20, 2014**

**Copy all team members with the submission e-mail. *I suggest that you do each item as we cover the material.***

**Prepare a report as an Excel file**. The report will have 14 items. Use the item numbers 1-15 **(**1-5are from on the front of this page) to **organize and order the report with each tab labeled by the item number(s) on the tab.** **Assemble charts, calculations and conclusions together in a coherent presentation on a worksheet (Use adjacent sheets if more than one tab is needed for a number). Do NOT place these at the end as appendices!** **Insert sentences and longer statements into a textbox.** Give support for each answer! No credit will be given for answers without a supporting reason. Your team is to do the work for this project. You may obtain technical assistance and advice on any portion of the analyses, but **you must do the work.**  (**You may use consultants but you may not use subcontractors**!) W**ork submitted with your name(s) on it is a pledge that all non-referenced work was done by you**. Use graphic illustrations, if possible, along with the analysis procedures we have learned in this class to answer questions. If the requested analysis does not seem to be totally appropriate for your data, perform the analysis anyway and explain why it would not be appropriate for your data. If you would be willing to allow me to use your data in the future, please indicate that I have permission to use your data. You will be credited for collecting the data if I do use your data in the future.

**6. Is the mean of the response variable R the same for both values of C1?** [**14.2-5**]Include your believed relationship statement 3a, appropriate null and alternate hypotheses to test 4a, appropriate graph(s), analysis results, and your conclusions about the relationship between C1 and R.

**7. Is the mean of the response variable R the same for all three values of C2? [21.6&7**] Include your believed relationship statement 4b, appropriate null and alternate hypotheses to test 4b, appropriate graph(s), analysis results, and your conclusions about the relationship between C2 and R. **If you detect a statistically significant difference in the means, tell which of the three you think are different and why.**

**8. Is there a relationship between the two categorical variables?** [15.6] Create a cross-classification table, state appropriate null and alternate hypotheses to statistically test for a relationship between C1 and C2, test this hypothesis with =.10 showing analysis results and describe the nature of any significant relationship. Check for any violation of assumptions for hypothesis testing with this procedure. Do the data support your believed relationship statement in 4c?

**9. Is there a linear relationship between Q1 and R? [16.4]** Include your believed relationship statement 4d, the appropriate null and alternate hypotheses to test 4d, graph with fitted line, analysis results with a 90% confidence interval for the fitted slope with Y=R & X=Q1, and your conclusions about the belief expressed in 4d using =.10.

**10. Is there a linear relationship between Q2 and R? [16.4]** Include your believed relationship statement 4e, the appropriate null and alternate hypotheses to test 4e, graph with fitted line, analysis results with a 90% confidence interval for the fitted slope with Y=R & X=Q1, and your conclusions about the belief expressed in 4e using =.10.

**11. Select the quantitative variable, Q1 or Q2, with the stronger linear relation with the response variable.** **Let xν (x0) = average of the first two values for this variable. Find a 90% confidence interval for the mean of R (Y) and the 90% prediction interval for an individual response R (Y) when the predictor variable value is xν (x0). [16.6&7].**

**12. Find the correlations between all three of the quantitative variables.** Based on these, is there a linear relationship between Q1 and Q2? Give your believed relationship statement 4f. Does the correlation support 4f? **[6.3 & 16.5]**

**13. Give multiple regression output for the response variable R based on C1, C2, Q1 and Q2 as predictor variables.** Create a dummy variable to include the categorical variable C1 and two dummy variables to include C2. **[19.1]** **Give the regression results using all five independent variables** (three dummy variables, Q1 and Q2). Use plots of the residuals versus each variable to **assess any perceived violation of assumptions**. Next g**ive the results for the model you believe would be the best regression model for predicting R**. (For the best model you may add interaction terms [19.2] or quadratic terms [19.6] if you think they adequately improve the model to justify inclusion.)

**14. Use JMP to build a logistic regression model [18.6] for predicting C1 using Q1, Q2 and R as predictor variables. Tell which category of C1 is being predicted (the category of C1 for which p^ is being calculated), give the coefficients for each the variables & conclusion about the significance of each. Use the values of Q1, Q2 & R for the state of Virginia and the model to calculate p^, then tell why you think this model is useful or not.**

**15. Give an executive summary (200 words or less) of your findings relative to the relationships between your variables.** (The executive summary should be a standalone document that summarizes your findings and does not require reading the rest of the document for understanding. However, you may reference tabs using the appropriate numbers 1 through 14 to guide the reader to details or visualizations.)