**SCMA524 Process Data Collection Assignment (**R.L. Andrews)

**40 points for Quiz 1, Due 9 AM Tuesday, October 24**

**E-mail an Excel file to scma.stat@gmail.com with subject “Project 1.” Use the letters A through I as tab labels.**

**Put all information for each lettered section together and ordered A through I. Put sections of text in a textbox.**

This project provides practical experience in planning for and gathering process data then answering specific questions through visualization and analysis of the data and providing a self-contained written summary of findings.

This project will involve an activity in your daily life. Consider the activities that you do at least once daily during a 7-day calendar week or 5-day work week. You must **obtain data for 30 or more days**. You are to design a data collection process that will provide data that can be used to obtain knowledge of the outcome variable and its relationship to other variables. It is not necessary for you to choose an outcome variable that you want to change.

**A. Identify the activity chosen** **and describe how the time it takes to complete the activity will be measured (Label this variable T). Provide operational definitions for starting time and finishing time of the activity and give the units of time that will be used for the measurement**. **(*Minimum of 20 or more different realistically possible values.*)** The amount of time required to complete the activity is a difference (finishing time - starting time). Select a measurement unit for time that will provide a minimum of twenty realistic possible values for this variable. Do not measure the time in hours, when minutes or seconds would be more appropriate. An **Operational Definition** is a definition that clearly and precisely defines exactly what is to be observed and how it is to be measured.

**B. Let C be a categorical (qualitative) variable with two possible nominal categories. Name the variable & provide operational definitions for each category. C must measure something different from the variable T defined in part A.** This variable’s measurement will be recorded using one of two possible categories. If the variable naturally has more than two categories, combine categories to get two groups. Choose something for which one would reasonably expect **at least 5 observations out of 30 total for each of the two categorical groups**.

**C. Let Q be a quantitative (numerical) variable. Name the variable & provide operational definitions for the measurement including the scale used to measure it.** Choose something for which one would reasonably expect a **Minimum of 8 or more different anticipated possible values.** Do not use a time that is a portion of the time to complete the activity in part A. Use a variable that measures something different from what is being measured for T and C.

**D. Record data in chronological order on each day for all five variables obtaining at least 30 days of data**

 **(n ≥ 30) for: 1. the date, 2. the day of the week, 3. C = the selected categorical or qualitative variable,**

 **4. Q = the selected quantitative variable,** and **5. T = the time to complete the activity.**

If you perform the activity more than once daily it is acceptable to either record data each time the activity is performed or only the first time. **RECORD DATA FOR ALL FIVE VARIABLES FOR 30 DAYs!**

Below is an example for check sheet column headings for doing e-mail each day.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. **Date** | 2. **Day**  | 3. **C:** Did at home  | 4. **Q:** # communications | 5. **T:** Minutes to complete |
| 8/29/2017 | Tuesday | YES  | 15  | 76  |

**E. Create two run charts** (one for T & one for Q).  **Using each chart determine process stability relative to location, variability, randomness and overall writing your conclusions for each category for both T & Q .**

**F. Create a scatter diagram with Q on the horizontal or X axis & T on the vertical or Y axis. Fit a least square regression line reporting the Intercept, Slope, R2 and se =the standard deviation of the residuals. Calculate the correlation between Q and T. Do Q and T appear to be linearly related?** (Give a reason.)

**G. Graphically compare the distribution of values for T for the two categories of C. Using the values for C, sort and separate the T data into two groups (one for each category). For each group, find the mean, standard deviation and five-number summary for T. Based on your analysis, does time T appear to be related to the categorical variable C?** Use your graph and statistics to support your written answer.

**H. Does T, the time to complete the activity, appear to be related to the day of the week?** (Give a written statement that is based an appropriate graph or analysis of the data collected and not on your personal belief.)

**I. Give an executive summary (200 words or less) of your findings relative to the relationship between T and any other observed variable.** (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 letters A through H to guide the reader to details or visualizations.)