OPIM 5603 Statistics in Business Analytics

Homework Assignment #2
Due on January 31, 2018

This assignment is team work; students are allowed to work with one partner. Submit your answers in one Word file attachment via Husky CT or via email to Manuel.Nunez@uconn.edu, no later than 11:59 p.m. on the due date. Only one submission (email) per team is required. By submitting the assignment, you are confirming that this represents your team's work, and that no part of the work submitted has been copied from others. Consequences for cheating can be quite serious. It can lead to not only failing the course but also dismissal from the program. If determined that more than two students have worked on the same assignment submission, the participants will only get a fraction of the assignment's grade.

Problem Description: The manager of Best Deal, a local electronic appliance retailer, would like to assess how good each of their sales personnel is and, in particular, which individuals are high and low performers. Data have been gathered from a sample of 10 employees in the Sales Department and stored in an Excel file named "*Bestdeal.xlsx*". The file contains daily sales values (rounded to the nearest integer) from each of the sampled employees for the last six months. The average daily sales for the combined employee data is around $146. Some employees seem to be consistently selling above this average, but the manager wants more statistical evidence. The store is closed on Sundays and each of the employees had 10 paid time off days that they have used at their own discretion during the data collection period. To help the manager, complete the following tasks using R Studio:

Task 1. Import the Excel file into R Studio. Make sure that for each employee column you select type "numeric", otherwise the data will be imported as "character" and you will not be able to do the computations (the type can be set by using the drop-down menu underneath each column name).

Task 2. Merge all the employees' data into a single vector entitled "Sales.Vector". Compute the mean, median, range, sample standard deviation, coefficient of variation, skewness, and kurtosis of this vector. Comment on the centrality, dispersion, and distribution shape of the data.

Task 3. Research the "tapply" function in R. Use this function to compute the mean, sample standard deviation, and coefficient of variation of each employee. Put the results into a data frame named "Employee.Summary" with ten rows and three columns: the first column is the mean sales of each employee, the second column is the standard deviation of each employee and the third column is the coefficient of variation of each employee. Name the rows of the data

frame with the names of the corresponding employees. Identify which employee sales data has the highest dispersion and which employee sales data has the lowest dispersion.

Task 4. Because of the sampling error, the fact that an employee has sample mean sales greater than or lower than $146 (the combined average) does not necessarily imply that the employee is an overall "high performer" or "low performer". After consulting with a recent BAPM graduate, the manager decided to use the following rule to classify employees: if one employee has average sales $\bar{X}$ and sample standard deviation S, then we say that the employee is a "high performer" if $\bar{X} - 0.2 \times S > 146$. Similarly, the employee would be a "low performer" if $\bar{X} + 0.2 \times S < 146$. Use this rule to determine which employees are high or low performers (note that it will not be possible to clearly classify some of the employees). Add a column to the "Employee.Summary" frame where you enter the overall performance level of each employee as "high", "low", or NA (**Hint**: one way to do this is by using the ifelse command in R; another way is by partitioning the data frame into three groups, one group for each type of employee performance, and then combining the three data frames into a single one). Display your final data frame on the Word file with your answers.

Task 5. For each day of the week (Monday, Tuesday, …, Saturday), determine average sales across all employees in the sample. Also, for each day determine the total number of employee absences through the whole sampling period. Put the results into a frame entitled "Daily.Sales" with six rows (one for each day) and two columns (one for the average sales on each day and the other for the number of absences on each day). Name the rows of the data frame using the corresponding day names. Display your final data frame on the Word file with your answers.

Assignment Submission: Submit via email or Husky CT one Word document containing your written answers to all the required computations and questions. Also include in the Word file the corresponding R instructions you used to answer the questions (if any).