Date Given: 24 March 2017 Date of Submission: 24 April 2017

BUSINESS STATISTICS CAPSTONE PROJECT

INSTRUCTION:

1. This project should be done in group of 4 people.
2. Answer all the questions given.
3. Project Deliverables:

   To successfully complete this capstone project, you are required to deliver a management report that addresses the management's requests. The final report should be presented in a professional format

Sub-Project 1 Description:

Assume that you are working as an intern for a financial management company. Your employer has a large number of clients who trust the company managers to invest their funds. In your position, you are responsible for producing reports for clients when they request information. Your company has two large data files with financial information for a large number of U.S. companies. The first is called **USCompanies 2003**, which contains financial information for the companies' 2001 or 2002 fiscal year-end. The second file is called **USCompanies 2005**, which has data for the fiscal 2003 or 2004 year-end. The 2003 file has data for 7,098 companies. The 2005 file has data for 6,992 companies. Thus, many companies are listed in both files but some are just in one or the other. The two files have many of the same variables, but the 2003 file has a larger range of financial variables than the 2005 file. For some companies, the data for certain variables are not available and a code of NA is used to so indicate. The 2003 file has a special worksheet that contains the description of each variable. These descriptions apply to the 2005 data file as well. You have been given access to these two data files for use in preparing your reports. Your role will be to perform certain statistical analyses that can be used to help convert these data into useful information in order to respond to the clients' questions.

This morning, one of the partners of your company received a call from a client who asked for a report that would compare companies in the financial services industry (SIC codes in the 6000s) to companies in production oriented businesses (SIC codes in the 2000s and 3000s). There are no firm guidelines on what the report should entail, but the partner has suggested the following:

- Start with the 2005 data file. Pull the data for all companies with the desired SIC codes into a new worksheet.
- Prepare a complete descriptive analysis of key financial variables using appropriate charts and graphs to help compare the two types of businesses.
- Determine whether there are differences between the two classes of companies in terms of key financial measures.
- Using data from the 2003 file for companies that have these SIC codes and that are also in the 2005 file, develop a comparison that shows the changes over the time span both within SIC code grouping and between SIC code groupings.

Sub Project 2 Description:

The athletic director of State University is interested in developing a multiple regression model that might be used to explain the variation in attendance at football games at his school. A sample of 16 games was selected from home games played during the past 10 seasons. Data for the following factors were determined:

y = Game attendance

x_1 = Team win/loss percentage to date

x_2 = Opponent win/loss percentage to date

x_3 = Games played this season

x_4 = Temperature at game time

The data collected are in the file called **Football**. Provide the regression analysis report for athletic director of State University. Your report should explain and summarize the following:

a) Scatter plots for each independent variable versus the dependent variable. Based on the scatter plots, produce a model that you believe represents the relationship between the dependent variable and the group of predictor variables represented in the scatter plots.

b) Based on the correlation matrix developed from these data, comment on whether you think a multiple regression model will be effectively developed from these data.

c) Use the sample data to estimate the multiple regression model that contains all four independent variables.

d) What percentage of the total variation in the dependent variable is explained by the four independent variables in the model?
e) Test to determine whether the overall model is statistically significant. Use $\alpha = 0.05$.
f) Which, if any, of the independent variables is statistically significant? Use a significance level of $\alpha = 0.08$ and the *p*-value approach to conduct these tests.
g) Estimate the standard deviation of the model error and discuss whether this regression model is acceptable as a means of predicting the football attendance at State University at any given game.
h) Develop a 95% confidence interval estimate for each of the regression coefficients and interpret each estimate. Comment on whether the interpretation of the intercept is relevant in this situation.

Sub Project 3 Description:

Glaser Machine Works has experienced a significant change in its business operations over the past 50 years. Glaser started business as a machine shop that produced specialty tools and products for the timber and lumber industry. This was a logical fit, given its location in the southern part of the United States. However, over the years Glaser looked to expand its offerings beyond the lumber and timber industry. Initially, its small size coupled with its rural location made it difficult to attract the attention of large companies that could use its products. All of that began to change as Glaser developed the ability not only to fabricate parts and tools but also to assemble products for customers who needed special components in large quantities. Glaser's business really took off when first foreign, and then domestic, automakers began to build automobile plants in the southern United States. Glaser was able to provide quality parts quickly for firms that expected high quality and responsive delivery. Many of Glaser's customers operated with little inventory and required that suppliers be able to provide shipments with short lead times.

As part of its relationship with the automobile industry, Glaser was expected to buy into the lean manufacturing and quality improvement initiatives of its customers. Glaser had always prided itself on its quality, but as the number and variety of its products increased, along with ever higher expectations by its customers, Glaser knew that it would have to respond by ensuring its quality and operations were continually improving. Of recent concern was the performance of its

manufacturing line 107B. This line produced a component part for a Japanese automobile company. The Japanese firm had initially been pleased with Glaser's performance, but lately the number of defects was approaching an unacceptable level. Managers of the 107B line knew the line and its workers had been asked to ramp up production to meet increased demand and that some workers were concerned with the amount of overtime being required. There was also concern about the second shift now being run at 107B. Glaser had initially run only one shift, but when demand for its product became so high that there was not sufficient capacity with one shift, additional workers were hired to operate a night shift.

Management was wondering if the new shift had been stretched beyond its capabilities. Glaser plant management asked you, the assistant production supervisor for line 107B, to conduct an analysis of product defects for the line. You have randomly selected several days of output and counted the number of defective parts produced on the 107B line. This information, along with other data, is contained in the file **Glaser Machine Works**. You need to provide a full report for the management team by 24th of April 2017. The full report should consist of the following:

1. Identify the primary issue of the case.
2. Identify a statistical model you might use to help analyze the case.
3. Develop a multiple regression model that can be used to analyze the product defects for line 107B. Be sure to carefully specify the dependent variable and the independent variables.
4. Discuss how the variables overtime hours, supervisor training, and shift will be modeled.
5. Run the regression model you developed and interpret the results.
6. Which variables are significant?
7. Provide a short report that describes your analysis and explains in managerial terms the findings of your model. Be sure to explain which variables, if any, are significant explanatory variables. Provide a recommendation to management.