MANB 1123 ASSIGNMENT #2
(HYPOTHESIS TESTING, CORRELATION ANALYSIS AND REGRESSION ANALYSIS)

RULES:

1. This assignment should be conducted in group. Each group should only consist of FOUR (4) members.
2. Assignment should be done in complete (answer all the questions given) and neat.
3. You may use any statistical packages/tools as you prefer to get the result.
4. Submission should follow the date given and should be in hardcopy.

Question 1:

At a recent meeting, the manager of a national call center for a major Internet bank made the statement that the average past-due amount for customers who have been called previously about their bills is now no larger than $20.00. Other bank managers at the meeting suggested that this statement may be in error and that it might be worthwhile to conduct a test to see if there is statistical support for the call center manager's statement. The file called **Bank Call Center** contains data for a random sample of 67 customers from the call center population. Assuming that the population standard deviation for past due amounts is known to be $60.00, what should be concluded based on the sample data? Test using α = 0.10.

Question 2:

The Center on Budget and Policy Priorities reported that average out-of-pocket medical expenses for prescription drugs for privately insured adults with incomes over 200% of the poverty level was $173 in 2002. Suppose an investigation was conducted in 2009 to determine whether the increased availability of generic drugs, Internet prescription drug purchases, and cost controls have reduced out-of-pocket drug expenses. The investigation randomly sampled 196 privately insured adults with incomes over 200% of the poverty level, and the respondents' 2009 out-of-pocket medical expenses for prescription drugs were recorded. These data are in the file **Drug Expenses**. Based on the sample data, can it be concluded that 2009 out-of-pocket prescription drug expenses are lower than the 2002 average reported by the Center on Budget and Policy Priorities? Use a level of significance of 0.01 to conduct the hypothesis test.

Question 3:

The makers of Mini-Oats Cereal have an automated packaging machine that can be set at any targeted fill level between 12 and 32 ounces. Every box of cereal is not expected to contain exactly the targeted weight, but the average of all boxes filled should. At the end of every shift (eight hours), 16 boxes are selected at random and the mean and standard deviation of the sample are computed. Based on these sample results, the production control manager determines whether the filling machine needs to be readjusted or whether it remains all right to operate. Use α = 0.05.

a. Establish the appropriate null and alternative hypotheses to be tested for boxes that are supposed to have an average of 24 ounces.
b. At the end of a particular shift during which the machine was filling 24-ounce boxes of Mini Oats, the sample mean of 16 boxes was 24.32 ounces, with a standard deviation of 0.70 ounce. Assist the production control manager in determining if the machine is achieving its targeted average.

c. Why do you suppose the production control manager would prefer to make this hypothesis test a two-tailed test? Discuss.
d. Conduct the test using a *p*-value.

Question 4:
The director of a state agency believes that the average starting salary for clerical employees in the state is less than $30,000 per year. To test her hypothesis, she has collected a simple random sample of 100 starting clerical salaries from across the state and found that the sample mean is $29,750.

a. State the appropriate null and alternative hypotheses.
b. Assuming the population standard deviation is known to be $2,500 and the significance level for the test is to be 0.05, what is the critical value (stated in dollars)?
c. Referring to your answer in part b, what conclusion should be reached with respect to the null hypothesis?
d. Referring to your answer in part c, which of the two statistical errors might have been made in this case? Explain.

Question 5:
The practice of "phishing," or using the Internet to pilfer personal information, has become an increasing concern, not only for individual computer users but also for online retailers and financial institutions. *The Wall Street Journal* reported 28% of people who bank online have cut back on their Internet use. The North Central Educators Credit Union instituted an extensive online security and educational program six months ago in an effort to combat phishing before the problem became extreme. The credit union's managers are certain that while Internet use may be down, the rate for their customers is much less than 28%. However, they believe that if more than 10% of their customers have cut back on their Internet banking transactions, they will be required to take more stringent action to lower this percentage. The credit union's Information Technology department analyzed 200 randomly selected accounts and determined that 24 indicated they had cut back on their Internet banking transactions.

a. State the appropriate null and alternative hypotheses for this situation.
b. Using α = 0.05 and the *p*-value approach, indicate whether the sample data support the managers' contention.

Question 6:
Cell phones are becoming an integral part of our daily lives. Commissioned by Motorola, a new behavioral study took researchers to nine cities worldwide from New York to London. Using a combination of personal interviews, field studies, and observation, the study identified a variety of behaviors that demonstrate the dramatic impact cell phones are having on the way people interact. The study found cell phones give people a newfound personal power, enabling unprecedented mobility and allowing them to conduct their business on the go. Interesting enough, gender differences can be found in phone use. Women see their cell phone as a means of expression and social communication, whereas males tend to use it as an interactive toy. A cell phone industry spokesman stated that half of all cell phones in use are registered to females.

a. State the appropriate null and alternative hypotheses for testing the industry claim.
b. Based on a random sample of cell phone owners shown in the data file called **Cell Phone Survey**, test the null hypothesis. (Use α = 0.05.)

Question 7:
A large number of complaints have been received in the past six months regarding airlines losing fliers' baggage. The airlines claim the problem is nowhere near as great as the newspaper articles have indicated. In fact, one airline spokesman claimed that less than 1% of all bags fail to arrive at the destination with the passenger. To test this claim, 800 bags were randomly selected at various airports in the United States when they were checked with this airline. Of these, 6 failed to reach the destination when the passenger (owner) arrived.

a. Is this sufficient evidence to support the airline spokesman's claim? Test using a significance level of 0.05. Discuss.
b. Estimate the proportion of bags that fail to arrive at the proper destination using a technique for which 95% confidence applies.

Question 8:
A survey by the Pew Internet & American Life Project found that 21% of workers with an e-mail account at work say they are getting more spam than a year ago. Suppose a large multinational company, after implementing a policy to combat spam, asked 198 randomly selected employees with e-mail accounts at work whether they are receiving more spam today than they did a year ago. The results of the survey are in the file **Spam**. At the 0.025 level of significance, can the company conclude that the percentage of its employees receiving more spam than a year ago is smaller than that found by the Pew study?

Question 9:
Airlines were severely affected by the oil price increases of 2008. Many airlines began charging for services that had previously been free, such as baggage and meals. One national airline had as an objective getting an additional $5 to $10 per trip from its customers. Surveys could be used to determine the success of the company's actions. The file entitled **AirRevenue** contains results of samples gathered before and after the company implemented its changes.

a. Produce a 95% confidence interval for the difference in the average fares paid by passengers before and after the change in policy. Based on the confidence interval, is it possible that revenue per passenger increased by at least $10? Explain your response.
b. Conduct a test of hypothesis to answer the question posed in part a. Use a significance level of 0.025.
c. Did you reach the same conclusion in both parts (a) and (b)? Is this a coincidence or will it always be so? Explain your response.

Question 10:
The United Way raises money for community charity activities. Recently, in one community, the fundraising committee was concerned about whether there is a difference in the proportion of employees who give to United Way depending on whether the employer is a private business or a government agency. A random sample of people who had been contacted about contributing last year was selected. Of those contacted, 70 worked for a private business and 50 worked for a government agency. For the 70 private-sector employees, the mean contribution was $230.25 with a standard deviation equal to $55.52. For the 50 government employees in the sample the mean and standard deviation were $309.45 and $61.75, respectively.

a. Based on these sample data and α = 0.05, what should be concluded? Be sure to show the decision rule.

b. Construct a 95% confidence interval for the difference between the mean contributions of private business and government agency employees who contribute to United Way. Do the hypothesis test and the confidence interval produce compatible results? Explain and give reasons for your answer.

Question 11:

A treadmill manufacturer has developed a new machine with softer tread and better fans than its current model. The manufacturer believes these new features will enable runners to run for longer times than they can on its current machines. To determine whether the desired result is achieved, the manufacturer randomly sampled 35 runners. Each runner was measured for one week on the current machine and for one week on the new machine. The weekly total number of minutes for each runner on the two types of machines was collected. The results are contained in the file **Treadmill**. At the 0.02 level of significance, can the treadmill manufacturer conclude that the new machine has the desired result?

Question 12:

Although not all students have debt after graduating from college, more than half do. The College Board's *2008 Trends in Student Aid* addresses, among other topics, the difference in the average college debt accumulated by undergraduate Bachelor of Arts degree recipients by type of college for the 2006–2007 academic years. Samples might have been used to determine this difference in which the private, for profit colleges' average was $38,300 and the public college average was $11,800. Suppose the respective standard deviations were $2,050 and $2,084. The sample sizes were 75 and 205, respectively.

a. Examine the sample standard deviations. What do these suggest is the relationship between the two population standard deviations? Support your assertion.
b. Conduct a hypothesis test to determine if the average college debt for Bachelor of Arts degree recipients is at least $25,000 more for graduates from private colleges than from public colleges. Use α = 0.05significance level and a *p*-value approach for this hypothesis test.

Question 13:

As the number of air travelers with time on their hands increases, logic would indicate spending on retail purchases in airports would increase as well. A study by *Airport Revenue News* addressed per person spending at select airports for merchandise, excluding food, gifts, and news items. A file entitled **Revenues** contains sample data selected from airport retailers in 2001 and again in 2004.

a. Produce a scatter plot for per person spending at selected airports for merchandise, excluding food, gifts, and news items, for the years 2001 and 2004. Does there appear to be a linear relationship between spending in 2001 and spending in 2004? Explain your response.
b. Calculate the correlation coefficient between per person spending in 2001 and per person spending in 2004. Does it appear that an increase in per person spending in 2001 would be associated with an increase in spending in 2004? Support your assertion.
c. Conduct a hypothesis test to determine if a positive correlation exists between the per person spending in 2001 and that in 2004. Use a significance level of 0.05 and assume that these figures form a random sample.

Question 14:

The file **Online** contains a random sample of 48 customers who made purchases last quarter from an online retailer. The file contains information related to the time each customer spent viewing the online catalog and the dollar amount of purchases made. The retailer would like to analyze the sample data to determine whether a relationship exists between the time spent viewing the online catalog and the dollar amount of purchases.

a. Compute the regression equation based on these sample data and interpret the regression coefficients.
b. Compute the coefficient of determination and interpret its meaning.
c. Test the significance of the overall regression model using a significance level of 0.01.
d. Test to determine whether the true regression slope coefficient is equal to 0. Use a significance level of 0.01 to conduct the hypothesis test.

Question 15:

A manufacturer produces a wash-down motor for the food service industry. The company manufactures the motors to order by modifying a base model to meet the specifications requested by the customer. The motors are produced in a batch environment with the batch size equal to the number ordered. The manufacturer has recently sampled 27 customer orders. The motor manufacturer would like to determine if there is a relationship between the cost of producing the order and the order size so that it could estimate the cost of producing a particular size order. The sampled data are contained in the file **Washdown Motors**.

a. Use the sample data to estimate the least squares regression model.
b. Provide an interpretation of the regression coefficients.
c. Test the significance of the overall regression model using a significance level of 0.01.
d. The company has just received an order for 30 motors. Use the regression model developed in part (a) to estimate the cost of producing this particular order.
e. Referring to part d, what is the 90% confidence interval for an average cost of an order of 30 motors?

Question 16:

An article in BusinessWeek presents a list of the 100 companies perceived as having "hot growth" characteristics. A company's rank on the list is the sum of 0.5 times its rank in return on total capital and 0.25 times its sales and profit-growth ranks. The file entitled **Growth** contains sales ($million), sales increase (%), return on capital, market value ($million), and recent stock price of the top 20 ranked companies.

a. Produce a correlation matrix for the variables contained in the file entitled **Growth**.
b. Select the two variables that are most highly correlated with the recent stock price and produce the regression equation to predict the recent stock price as a function of the two variables you chose.
c. Determine if the overall model is significant. Use a significance level of 0.10.
d. Examine the coefficient of determination and the adjusted coefficient of determination. Does it seem that either of the independent variables' addition to R^2 does not justify the reduction in degrees of freedom that results from its addition to the regression model? Support your assertions.
e. Select the variable that is most correlated with the stock price and test to see if it is a significant predictor of the stock price. Use a significance level of 0.10 and the p-value approach.

Question 17:

The Gilmore Accounting firm, in an effort to explain variation in client profitability, collected the data found in the file called **Gilmore**, where:

y = Net profit earned from the client

x_1 = Number of hours spent working with the client

x_2 = Type of client;

1, if manufacturing
2, if service
3, if governmental

a. Develop a scatter plot of each independent variable against the client income variable. Comment on what, if any, relationship appears to exist in each case.
b. Run a simple linear regression analysis using only variable x_1 as the independent variable. Describe the resulting estimate fully.
c. Test to determine if the number of hours spent working with the client is useful in predicting client profitability.
d. Incorporate the client type into the regression analysis using dummy variables. Describe the resulting multiple regression estimate.
e. Test to determine if this model is useful in predicting the net profit earned from the client.
f. Test to determine if the number of hours spent working with the client is useful in this model in predicting the net profit earned from a client.
g. Considering the tests you have performed, construct a model and its estimate for predicting the net profit earned from the client.
h. Predict the average difference in profit if the client is governmental versus one in manufacturing. Also state this in terms of a 95% confidence interval estimate.

Question 18:

An investment analyst collected data of 20 randomly chosen companies. The data consisted of the 52-week-high stock prices, PE ratio, and the market value of the company. These data are in the file entitled **INVESTMENT**. The analyst wishes to produce a regression equation to predict the market value using the 52-week-high stock price and the PE ratio of the company. He creates a complete second-degree polynomial.

a. Construct an estimate of the regression equation using the indicated variables.
b. Determine if any of the quadratic terms are useful in predicting the average market value. Use a p-value approach with a significance level of 0.10.
c. Determine if any of the PE ratio terms are useful in predicting the average market value. Use a test statistic approach with a significance level of 0.05.

SUBMISSION DATE: 11 MAY 2018 before 5PM (HARD COPY)