Ian Bach

MGMT 670 – Business Analytics

Purdue University

Professor Alexander

# Exercise Problem #4

Homework exercises must be prepared individually and submitted online through Brightspace prior to the posted deadline. Up to three submissions are allowed but only the last submission will be graded. See the Syllabus for more information on Homework Exercise requirements and expectations. Any necessary modifications to this assignment will be posted to Brightspace as an announcement.

For this exercise, submit only one Microsoft Word document with all appropriate output results and graphs from Minitab and Excel into the single Word document.

## Question 1

**Descent, Inc., produces a variety of climbing and mountaineering equipment. One of its products is a traditional three-strand climbing rope. An important characteristic of any climbing rope is its tensile strength. Descent produces the three-strand rope on two separate production lines: one in Bozeman and the other in Challis. The Bozeman line has recently installed new production equipment. Descent regularly tests the tensile strength of its ropes by randomly selecting them to various tests. The most recent random sample of ropes, taken after the new equipment was installed at the Bozeman plant, revealed the following:**

**Bozeman Challis**

**x̅1 = 7,200 lb x̅2 = 7,087 lb**

**s1 = 425 s2 = 415**

**n1 = 25 n2 = 20**

**Descent’s production managers are willing to assume that the population of tensile strengths for each plant is approximately normally distributed with equal variances.**

**Based on the sample results, can Descent’s managers conclude that there is a difference between the mean tensile strengths of ropes produced in Bozeman and Challis? Conduct the appropriate hypothesis test at the 0.05 level of significance.**

* Null Hypothesis: The mean tensile strength of ropes from Bozeman is equal to ropes of Challis. (*μ*1​=*μ*2​)
* Alternative Hypothesis: The mean tensile strength of ropes from Bozeman is not equal compared to Challis. (*μ*1≠*μ*2​)
* Data
  + Bozeman Mean = 7200 Ib
  + Bozeman Standard Deviation = 425
  + Bozeman Sample Size = 25
  + Challis Mean = 7,087
  + Challis Standard Deviation = 415
  + Challis Sample Size = 20
  + Level of significance: 0.05
* 
* According to the information, The absolute value of the t statistic is less than the critical t value shows we do not reject the null hypothesis (t statistic = 0.8955 and critical t value = 2.017). This shows that there is not sufficient evidence to reject the null hypothesis showing no difference in mean between the two productions.

## Question 2

**A marketing research firm is interested in determining whether there is a difference between the proportion of households in Chicago and the proportion of households in Milwaukee who purchase groceries online. The research firm decided to randomly sample households earning over $50,000 a year in the two cities and ask them if they purchased any groceries online last year. The random sample involved 150 Chicago households and 135 Milwaukee households. The results of the sample can be found in the file *Online Groceries***.

1. Construct a 95% confidence interval estimate for the difference between the two population proportions.



* The CI ranges range from a negative (-.016) to a positive (0.1218).
* This shows that there isn't strong evidence to show that one city has a higher proportion of households purchasing online groceries compared to the other.

1. At the 0.10 level of significance, can the marketing research firm conclude that a greater proportion of households in Chicago earning over $50,000 annually buys more groceries online than do similar households in Milwaukee? Support your answer with the appropriate hypothesis test.

* Data
  + 0.10 = level of significance
  + Shown below:



* Show in the information the P value = 0.071 which is less than the significance level. This indicates to rejection of the Null Hypothesis which shows I'm in favor of the alternative hypothesis. The alternative hypothesis explained that a greater proportion of households in Chicago buy groceries online compared to Milwaukee.

## Question 3

**A shipping company believes that the variation in the cost of a customer’s shipment can be explained by differences in the weight of the package being shipped. To investigate whether this relationship is useful, a random sample of 20 customer shipments was selected, and the weight (in lb) and the cost (in dollars, rounded to the nearest dollar) for each shipment were recorded. The following results were obtained:**

Weight (x) Cost (y)

8 11

6 8

5 11

7 11

12 17

9 11

17 27

13 16

8 9

18 25

17 21

17 24

10 16

20 24

9 21

5 10

13 21

6 16

6 11

12 20







* 1. Construct a scatter plot for these data. What, if any, relationship appears to exist between the two variables?
* The plot indicates there is a linear relationship between weight and cost.
  1. Compute the linear regression model based on the sample data. Interpret the slope and intercept coefficients.
* The slope of 1.10 shows that for each additional IB, the shipping cost increases at $1.10.
* The intercept of 4.499 or 4.50 shows that the base cost when weight is 0.
  1. Test the significance of the overall regression model using a significance level equal to 0.05. (P value)
     + Null Hypothesis: Indicating there is no linear relationship between the weight of package and cost of shipping.
     + Alternative Hypothesis: Indicating there is a linear relationship between the weight of package and the cost of shipping.
     + The P value shown above is less than the significance level of 0.05. This shows a strong rejection of the null hypothesis indicating that there is a strong linear relationship between weight of packages and cost of shipping.
  2. What percentage of the total variation in shipping cost can be explained by the regression model you developed in part b?
     + Looking at the plot, shown is R^2 = 0.768. This indicates that the regression can be explain 76.80% of the total variation in the shipping.

## Question 4

The National Football League (NFL) is arguably the most successful professional sports league in the United States. Following the recent season, the commissioner’s office staff performed an analysis in which a simple linear regression model was developed with average home attendance used as the dependent variable and the total number of games won during the season as the independent variable. The staff was interested in determining whether games won could be used as a predictor for average attendance. Develop the simple linear regression model. The data are in the file called *NFL*.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 1 | 93224985.89 | 93224985.89 | 1.233557653 | 0.275539149 |
| Residual | 30 | 2267222428 | 75574080.92 |  |  |
| Total | 31 | 2360447414 |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| 63260.5323 | 3993.499925 | 15.84087479 | 4.07798E-16 | 55104.7174 | 71416.34721 | 55104.7174 | 71416.34721 |
| 511.7303371 | 460.7458564 | 1.110656406 | 0.275539149 | -429.2382347 | 1452.698909 | -429.2382347 | 1452.698909 |

1. What percentage of total variation in average home attendance is explained by knowing the number of games the team won?
   * The percentage of total variation is explained R Square or the Coefficient of Determination which is 0.3949 or 3.949%. This indicates that 3.949% of the total variation in average home attendance is explained by the number of games won.
2. What is the standard error of the estimate for this regression model?
   * The standard error of the estimate is 8693.33.
3. Using alpha = 0.05, test to determine whether the regression slope coefficient is significantly different from 0.
   * Slope Coefficient = 511.73
   * Standard Error of the Slope Coefficient = 460.75
   * T stat = 1.11
   * P value = 0.276
   * The P value is greater than the significance level, so we fail to reject the null hypothesis that the slope coefficient is equal to zero. This explains why the number of games won during the season does not resemble statistically a prediction of the average home attendance.
4. After examining the regression analysis results, what should the NFL staff conclude about how the average attendance is related to the number of games the team won?
   * Null Hypothesis = The slope coefficient is equal to zero
   * Alternative Hypothesis = The slope coefficient is not equal to zero
   * This indicates there is not enough confirmation to show that winning more games has a large effect on increasing home game attendance.