Dr. Sudhakar Raju

BIA 6309

SPRING 2017

**ANSWERS TO ASSIGNMENT 3**

1.

**STEP 1: Set up the form of the hypothesis test**

NULL HYPOTHESIS (H0 ): 

(That is, true population mean continues to be equal to or lower than $1800. This is the status quo).

ALTERNATIVE HYPOTHESIS (HA ): 

The true population mean has actually increased and is now greater than $1800.

**STEP 2: Identify the t value corresponding to the critical significance level of 5%.**

The degrees of freedom are equal to (n- 1) or (40 -1) = 39. Since there is no 39 in the t table, simply round of the degrees of freedom to 40. With 40 df, the critical t value at 5% is 1.68.

**STEP 3: CALCULATE THE TEST STATISTIC**

The mean value of the sample (sample mean or ) is $1950 and sample standard deviation (s) equals $500. Thus:

t = 

t =  = 

Thus, the t statistic has a value of 1.90 with degree of freedom = n – 1 or 39.

The basic logic is this. If the true population mean is actually $1800, what is the probability of ending up with a value like $1950? If the probability of a value such as $1950 is very low (5% or lower) then we can assert that observing such an unlikely value constitutes evidence that the null should be rejected. If t = 1.68 corresponds to 5% of the t distribution, then t = 1.90 (with df=39) lies even further to the right and encompasses an even smaller area. (You can calculate the exact value in Excel with the command: =T.DIST.RT (1.90, 39) which returns a value of .0324 or 3.24%. Since this probability value is lower than 5%, we reject the null hypothesis and effectively conclude that claim values had increased beyond $1800.

**2. STEP 1: Set up the form of the hypothesis test**

NULL HYPOTHESIS (H0 ): 

(That is, true population mean equals 100 inches).

ALTERNATIVE HYPOTHESIS (HA ): 

The true population mean has actually increased and is now greater than 100 inches.

(Note that the alternative hypothesis here could also be . But since the sample mean is greater than 100 inches, the concern is that the true mean has increased beyond 100 inches).

**STEP 2: Identify the t value corresponding to the critical significance level of 5%.**

The degrees of freedom are equal to (n- 1) or (36 -1) = 35. Since there is no 35 in the t table, simply round of the degrees of freedom. With 40 df, the critical t value at 5% is approximately 1.68.

**STEP 3: CALCULATE THE TEST STATISTIC**

The mean value of the sample (sample mean or ) is 100.50 and sample standard deviation (s) equals 1 inch. Thus:

t = 

t =  = 

The basic logic is this. If the true population mean value is actually 100 inches, what is the probability of ending up with a value like 100.50 inches? If the probability of a value such as 100.50 inches is very low (5% or lower) then we can assert that observing such an unlikely value actually means that the true population mean is not 100 inches. Thus, if the probability value is 5% or lower, we reject the null hypothesis.

In this situation, the probability of observing a t value like 3.0 is extremely small. (You can calculate the exact value in Excel with the command: =T.DIST.RT (3.0, 35) which returns a value of .0025 or .25%. Since this probability value is substantially lower than 5%, we reject the null hypothesis and effectively conclude that mean measurement has increased beyond 100 inches and the sample is not meeting quality standards.

3.

**STEP 1: Set up the form of the hypothesis test**

NULL HYPOTHESIS (H0 ): 

(That is, true population mean continues to be equal to or lower than $20. This is the status quo).

ALTERNATIVE HYPOTHESIS (HA ): 

The true population mean has increased to more than $20.

**STEP 2: Identify the t value corresponding to the critical significance level of 5%.**

The degrees of freedom are equal to (n- 1) or (49 -1) = 48. Since there is no 48 in the t table, simply round of the degrees of freedom to 50. With 50 df, the critical t value at 5% is 1.68.

**STEP 3: CALCULATE THE TEST STATISTIC**

t = 

t =  = 

This is a 7 sigma event! In other words, if the true value had really been $20 or lower, observing a value like $22.60 is extremely unlikely. The unlikeliness of this event means that the true value could not have really been $20 or lower. We must reject the null hypothesis and conclude that there is extremely significant evidence that the typical amount spent per customer is more than $20.

4.

**STEP 1: Set up the form of the hypothesis test**

NULL HYPOTHESIS (H0 ): 

(That is, true population mean continues to be equal to or greater than 15 minutes. This is the status quo).

ALTERNATIVE HYPOTHESIS (HA ): 

The true population mean has decreased to 15 minutes.

**STEP 2: Identify the t value corresponding to the critical significance level of 5%.**

The degrees of freedom are equal to (n- 1) or (25 -1) = 24. With 24 df, the critical t value at 5% is 1.71.

**STEP 3: CALCULATE THE TEST STATISTIC**

t = 

t =  = 

You can simply look at the absolute value of 1 rather than the negative value of -1. (Since the t distribution is symmetric positive/negative values are essentially the same). The calculated test statistic of -1 (or just simply 1) does not lie in the null rejection zone. Thus, there is insufficient evidence to conclude that the average evening long distance call has decreased.

You can calculate the exact probability value corresponding to 1 or -1. On Excel type in =T.DIST.RT(1,24) which will give a value of 16.36%.

**V. The Georgia Public Service Commission Case**

Note that customers and dollars in the dataset are both measured in 1000s.

a.) b.) See attached R code

c.) Line\_Maintenance\_Expenses = 33.3205 + 15.0159 Customers

Call:

lm(formula = line\_maintenance\_expense ~ customers)

Residuals:

Min 1Q Median 3Q Max

-354.49 -133.00 36.30 99.62 238.79

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 33.321 138.704 0.240 0.815

customers 15.016 1.807 8.311 8.42e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

*Residual standard error: 187.7* on 10 degrees of freedom

Multiple R-squared: 0.8735, Adjusted R-squared: 0.8609

F-statistic: 69.07 on 1 and 10 DF, p-value: 8.42e-06

d.) Adjusted R2 = 0.86. Approximately 86% of the total variation in line maintenance expense is accounted for the number of customers served by the telephone company.

e.) 

f.) According to ( e ) above, a company with 75 (i.e. 75,000) customers should, on average, incur a line maintenance expense of $1160 or $1,160,000 approximately. To determine if $1,500,000 is unreasonable, we can ask what would 5% of the highest line maintenance expenses correspond to in terms of Z values. 5% of the highest (or lowest) values correspond to a 90% confidence interval:

90% CONFIDENCE INTERVAL: $1160 ± (1.65) (SE)

$1160 ± (1.65) (188)

$1160 ± 310

$850 & $1470

The standard error above is the SE of the entire regression. See “Residual standard error: 187.7” generated by R ).

This means that companies that are at the lower end of the expense spectrum (5% of the cheapest line maintenance expense companies) will have line expenses that are lower than $850,000. On the other hand, 5% of the companies that have the highest line maintenance expenses will correspond to companies with expenses that exceed $1470. Since this company has expenses of $1500,000 it is among the 5% of companies with the highest line maintenance expenses. Thus, these charges doe seem very high.

g. Using the abline function it is obvious that some values are under estimated while some are over-estimated. There seems to be systematic variation that this ***not*** being accounted for by the linear model. The standard linear model may not be appropriate.

h. Line Maintenance Expenses= 707.47 – 7.39 Customers + .1543 [Customers]2

Call:

lm(formula = line\_maintenance\_expense ~ customers + SQUARED\_TERM)

Residuals:

Min 1Q Median 3Q Max

-220.32 -59.47 -16.24 67.12 191.37

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 707.47474 230.49267 3.069 0.0134 \*

customers -7.39221 7.03366 -1.051 0.3207

SQUARED\_TERM 0.15430 0.04761 3.241 0.0101 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

***Residual standard error: 134.4*** on 9 degrees of freedom

Multiple R-squared: 0.9416, Adjusted R-squared: 0.9287

F-statistic: 72.61 on 2 and 9 DF, p-value: 2.801e-06

i. and j.) This new model accounts for 93% of the total variation in line maintenance expense as compared to the original regression model which gave an adjusted R2 of .86 implying that the addition of the quadratic term in the model served a useful purpose.

k.) 

l.) See predicted (fitted) values generate by R. The quadratic model seems to fit the data much better.

m.) Suppose we use the same argument as we did before to check if our new predicted value of $1021 from the Quadratic Regression is unreasonable:

90% CONFIDENCE INTERVAL: $1021 ± (1.65) (134.4)

$1021 ± (222)

$799 & $1243

Only 5% of the companies have line expenses greater than $1.243 million. Since this company has line expenses of $1.50 million it is clearly in the top 5% of most expensive companies.

In fact, suppose we construct a 95% confidence interval. In this case, we have:

95% CONFIDENCE INTERVAL: $1021 ± (1.96) (134.4)

$1021 ± (263)

$758 & $1284

By this standard, only 2.50% of companies have line expenses greater than $1.284 million. Since this company has line expenses of $1.50 million, it is clearly in the top 2.50% of most expensive companies. Thus, it ***would*** appear unusual for a company with 75,000 customers to show a line maintenance charge of $1,500,000.

n.) The quadratic model appears to be best.