PSTAT 126 - Regression Analysis – Fall 2017

Lab 4 Handout

**Note**: These course materials are the sole property of Dr. Todd M. Gross. They are strictly for use by students enrolled in a course taught by Dr. Gross. They may not be altered, excerpted, distributed, or posted to any website or other document-sharing service.

Goals for this Lab

* Learn how to conduct a one-tailed test, and a test of a specific hypothesized value, for the slope, using R.
* Learn how to generate and interpret a confidence interval and a prediction interval for a predicted response in R.

Lab Exercise #1

A researcher has studied the relationship between movie ticket prices and movie theater attendance. She studied 15 theaters, and measured the average ticket price, and the average percentage of empty seats in the theater during one week. She found a sample slope of 15%, and a standard error of the slope of 2.6%.

1. The researcher wants to prove that for each dollar increase in ticket prices, the percentage of empty seats will increase by more than 10%. Test this hypothesis using a one-tailed test.
   1. State the Null and Alternative Hypotheses

H0: 1 ≤ 10%

H1: 1 > 10%

* 1. Calculate tobtained for this one-tailed test. Do the calculation using R.

tobtained = (b1 – β10)/se(b1)

> (15-10)/2.6

* 1. Determine the tcritical for this one-tailed test, using alpha=0.05 and df = n – 2. You can use the following R command to obtain tcritical:

> qt(.95,8)

* 1. Draw a statistical conclusion by comparing tobtained to tcritical. What can the researcher conclude?
  2. If this hypothesis is tested using a two-tailed test, would the statistical conclusion change? Note: you do NOT need to recalculate tobtained, but you do need to use R to get a new tcritical, using alpha = 0.025.  
       
     > qt(.975,8)

Lab Exercise #2

We will learn about confidence intervals and prediction intervals by continuing to look at the relationship between **bmi** and **diastolic** blood pressure in the **pima** dataset. The R commands are provided.

1. Open and the **pima**  dataset from the **faraway**  package, remove the zero values (as we did last week) and generate the summary of the linear model that predicts diastolic from bmi.

> data(pima,package="faraway")

> newpima<-subset(pima, bmi > 0 & diastolic > 0)

> attach(newpima)

> fit1<-lm(diastolic~bmi)

> summary(fit1)

1. Using the estimates of the slope and intercept from the R summary, calculate the predicted diastolic blood pressure for an individual with BMI = 29.

> 55.48694 + 0.51989 \* 29

1. Now repeat this calculation using the **prediction** command in R. You will need to use the data.frame option to specify the specific value of BMI.

> predict(fit1,data.frame(bmi=29))

1. Calculate a 95% confidence interval for the mean predicted blood pressure of all individuals with BMI = 29. Write a statement describing the results of the confidence interval. (Hint: “We are 95% confident that…”)

> predict(fit1,data.frame(bmi=29),interval="confidence",level=.95)

1. Now calculate a 95% prediction interval for the predicted blood pressure of an individual with BMI = 29. Write a statement describing the results of the confidence interval. (Hint: “We are 95% confident that…”). Is this interval bigger or smaller than the confidence interval in part 4? Why?

> predict(fit1,data.frame(bmi=29),interval="prediction",level=.95)

1. Increase the confidence level to 99% for the prediction interval in part 5. What does this do to the size of the interval? Why?

> predict(fit1,data.frame(bmi=29),interval="prediction",level=.99)