STAT 308 – Homework 3

For the problems in which calculations are needed, please include your `R` code with your answers, otherwise you will not be given full credit. Please upload your assignment by Thursday, September 22, 11:59 pm in a pdf file to Sakai.

- 1. Suppose we perform a simple linear regression where

$$n = 50, \bar{x} = -0.208, \bar{y} = 1.516, s_x = 2.354, s_y = 3.185$$

$$\hat{\beta}_0 = 1.745, \hat{\beta}_1 = 1.102, s_{Y|X} = 1.868, s_{\hat{\beta}_0} = 0.265, s_{\hat{\beta}_1} = 0.113$$

– a. Calculate a 95% confidence interval for β_1.

```
n <- 50
x_bar <- -0.208
y_bar <- 1.516
s_x <- 2.354
s_y <- 3.185
beta_0 <- 1.745
beta_1 <- 1.102
s_yx <- 1.868
s_beta_0 <- 0.265
s_beta_1 <- 0.113

alpha <- 0.05
crit <- qt(1-alpha/2,df=n-2)
beta_1 + c(-1,1)*crit*s_beta_1
```

```
## [1] 0.8747983 1.3292017
```

– b. Find a test statistic and p-value for testing $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 > 0$.

```
t <- beta_1/s_beta_1
t
```

```
## [1] 9.752212
```

```
p.val <- 1 - pt(t,df=n-2)
p.val
```

```
## [1] 2.861045e-13
```

– c. Calculate a 95% confidence interval for $\mu_{Y|X}$ when $x = 3$.

```r
x0 <- 3
yhat_x0 <- beta_0 + beta_1*x0
s_yx0 <- sqrt(s_yx^2*(1/n + (x0-x_bar)^2/((n-1)*s_x^2)))

yhat_x0 + c(-1,1)*crit*s_yx0
```

```
## [1] 4.147234 5.954766
```

– d. Calculate a 95% prediction interval for Y when $x = 3$.

```r
yhat_x0 + c(-1,1)*crit*sqrt(s_yx^2+s_yx0^2)
```

```
## [1] 1.187929 8.914071
```

- 2. Suppose I have a dataset, for which I perform a simple linear regression of Y on X.

– a. I calculate 95% confidence and prediction intervals for our predicted value of Y for $x = 2$, which gives us $\hat{Y}_{x=2} = -2.199$, and our two intervals (-2.59,-1.80) and (-4.25,-0.14). Which one is the prediction interval? Explain your reasoning.

The second is the prediction interval because prediction intervals are always wider than their corresponding confidence intervals.

– b. Now, suppose I wish to calculate a 95% confidence interval for the predicted value of Y for $x = \bar{x}$. Will this width of this confidence interval (the difference between upper and lower values of the interval) be larger or smaller? Explain your reasoning.

The width will go down because the width of any confidence or prediction interval on a least squares regression prediction is smallest when $x_0 = \bar{x}$.

- 3. Reconsider the dataset `AdRevenue.csv` as well as our simple linear regression model of ad revenue (in millions of dollars) based on circulation (in millions).

– a. Calculate a 95% confidence interval for β_1. Interpret this interval in the context of the problem.

```r
adrev <- read.csv("../data/AdRevenue.csv")
mod <- lm(AdRevenue ~ Circulation,adrev)
confint(mod,"Circulation")
```

```
##                2.5 %   97.5 %
## Circulation 20.95411 24.75267
```

We are 95% confident that when circulation increases by 1 million magazines, the expected ad revenue would increase by a value between 20.95 and 24.76 million dollars.

– b. Perform a hypothesis test for a linear relationship between ad revenue and circulation. Be sure to properly state your hypotheses, your test statistic, p-value, and a decision and conclusion at $\alpha = 0.05$.

$$H_0 : \beta_1 = 0,$$
$$H_1 : \beta_1 \neq 0$$

```
summary(mod)
```

```
##
## Call:
## lm(formula = AdRevenue ~ Circulation, data = adrev)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -147.694  -22.939   -7.845   13.810  131.130
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  99.8095     5.8547   17.05   <2e-16 ***
## Circulation  22.8534     0.9518   24.01   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 42.22 on 68 degrees of freedom
## Multiple R-squared:  0.8945, Adjusted R-squared:  0.8929
## F-statistic: 576.5 on 1 and 68 DF,  p-value: < 2.2e-16
```

$t = 24.01$, $p - value < 2.2 \times 10^{16}$.

Because $p{-}value < \alpha$, we would reject H_0 and conclude that there is a significant linear relationship between magazines in circulation and magazine's generated ad revenue.

– c. Calculate a 95% confidence interval for the predicted ad revenue when there are 2.5 million subscriptions in circulation. Interpret this interval in the context of the problem.

```
newdata <- data.frame(Circulation = 2.5)
predict(mod,newdata=newdata,interval="confidence")
```

```
##       fit      lwr      upr
## 1 156.943 146.8046 167.0814
```

We are 95% confident that the true expected ad revenue when there are 2.5 million subscriptions in circulation is between 146.80 and 167.08 million dollars.

– d. Calculate a 95% prediction interval for a randomly selected magazine with 2.5 million subscriptions in circulation. Interpret this interval in the context of the problem.

```
predict(mod,newdata=newdata,interval="prediction")
```

```
##       fit      lwr      upr
## 1 156.943 72.08257 241.8035
```

We are 95% confident that a randomly selected magazine with 2.5 million subscriptions in circulation will have between 72.08 and 241.80 million dollars in ad revenue.