Stat/Math 415 Homework 6

Due Friday Nov 1; Joseph Sepich (jps6444)

1 Problem 7.6-3

```
midterms <- c(70, 74, 80, 84, 80, 67, 70, 64, 74, 82)
finals <- c(87, 79, 88, 98, 96, 73, 83, 79, 91, 94)

meanx <- sum(midterms) / 10
meany <- sum(finals) / 10

beta <- sum((finals - meany) * (midterms - meanx)) / sum((midterms - meanx)^2)

epsilon <- finals - meany - beta * (midterms - meanx)
variance_est <- sum(epsilon^2) / 10

print(paste0('Sample mean x: ',meanx))
```

```
## [1] "Sample mean x: 74.5"
```

```
print(paste0('Sample mean y(alpha): ',meany))
```

```
## [1] "Sample mean y(alpha): 86.8"
```

```
print(paste0('Beta: ',beta))
```

```
## [1] "Beta: 1.01568154402895"
```

```
print(paste0('SigmaSquared: ',variance_est))
```

```
## [1] "SigmaSquared: 17.9998069963812"
```

Suppose x denotes the midterm score, which indicates the final score y. This gives us the least squares regression line:

$$y = 86.8 + 1.0157(x - 74.5) = 11.132 + 1.0157x$$

1.1 Part a

To find a 95% confidence interval we use the following equation:

$$\hat{\alpha} + \hat{\beta}(x - \overline{x}) + -t_{\alpha/2}\sqrt{\frac{n}{n-2}\hat{\sigma^2}(\frac{1}{n} + \frac{(x-\overline{x})^2}{\Sigma_{i=1}^{n}(x_i - \overline{x})^2})}$$

Looking up the value for $t_{0.025}$ in the t distribution table for n - 2 = 10 - 2 = 8 degrees of freedom, we get t = 2.306. We plug the values into our equation for each value of x we want to get a CI for and we get:

```r
xVals <- c(68, 75, 82)
n <- 10

for (x_val in xVals) {
  y_hat <- meany + beta*(x_val - meanx)
  t <- 2.306
  sqrt_coefficient <- (n * variance_est) / (n-2) #n * sigmasquared / n-2
  sqrt_val <- (1 / n) + ((x_val - meanx)^2 / (sum((midterms - meanx)^2)))

  lower_bound <- y_hat - (t * sqrt(sqrt_coefficient * sqrt_val))
  upper_bound <- y_hat + (t * sqrt(sqrt_coefficient * sqrt_val))

  print(paste0('Lower bound for ',x_val,' is ',lower_bound))
  print(paste0('Upper bound for ',x_val,' is ',upper_bound))
}
```

```
## [1] "Lower bound for 68 is 75.2827850670534"
## [1] "Upper bound for 68 is 85.1133548605703"
## [1] "Lower bound for 75 is 83.8384438112727"
## [1] "Upper bound for 75 is 90.7772377327562"
## [1] "Lower bound for 82 is 89.1071380547519"
## [1] "Upper bound for 82 is 99.7280851056824"
```

This makes the following Confidence Intervals:

- x=68 y=[75.283, 85.113]
- x=75 y=[83.838, 90.777]
- x=82 y=[89.107, 99.728]

1.2 Part b

For the prediction interval we do the same thing, but add in variance due to random error resulting in the following equation:

$$\hat{\alpha} + \hat{\beta}(x - \overline{x}) + -t_{\alpha/2}\sqrt{\frac{n}{n-2}\hat{\sigma^2}(1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{\Sigma_{i=1}^{n}(x_i - \overline{x})^2})}$$

```r
for (x_val in xVals) {
  y_hat <- meany + beta*(x_val - meanx)
  t <- 2.306
  sqrt_coefficient <- (n * variance_est) / (n-2) #n * sigmasquared / n-2

  # DIFFERENCE HERE FROM PART A
  sqrt_val <- 1 + (1 / n) + ((x_val - meanx)^2 / (sum((midterms - meanx)^2)))

  lower_bound <- y_hat - (t * sqrt(sqrt_coefficient * sqrt_val))
  upper_bound <- y_hat + (t * sqrt(sqrt_coefficient * sqrt_val))

  print(paste0('Lower bound for ',x_val,' is ',lower_bound))
  print(paste0('Upper bound for ',x_val,' is ',upper_bound))
}
```

```
## [1] "Lower bound for 68 is 68.20617467105"
## [1] "Upper bound for 68 is 92.1899652565736"
## [1] "Lower bound for 75 is 75.8325524371854"
## [1] "Upper bound for 75 is 98.7831291068436"
## [1] "Lower bound for 82 is 82.2583905459186"
## [1] "Upper bound for 82 is 106.576832614516"
```

This makes the following Prediction Intervals:

- x=68 y=[68.206, 92.190]
- x=75 y=[75.833, 98.783]
- x=82 y=[82.258, 106.577]

2 Problem 7.6-7

3 Problem 8.3-1

4 Problem 8.3-3

5 Problem 8.3-7

6 Problem 8.3-11