Lesson 10 Companion

Confidence interval for a slope parameter:

For this example I will use the same data set that we used for the Lesson 9 Companion: the non-academic time vs. GPA data set.

```
library(tidyverse)

GPA = read_table2("http://www.isi-stats.com/isi/data/chap10/GPA.txt")

head(GPA)
```

```
## # A tibble: 6 x 2
##   hours   gpa
##   <dbl> <dbl>
## 1   7.7  3.17
## 2  18.9  3.28
## 3  19.4  3.72
## 4  34.7  3.38
## 5  42.8  3.17
## 6  42.8  2.87
```

You will recall from the discussion in that Lesson 9 document that the slope calculated by the least squares regression model (and available to you in the summary of your **lm()** model) is a statistic that is based on your sample of data. Because it is merely an estimate of the true association between your explanatory and response variable, the best we can hope for in terms of learning that true slope value is likely a confidence interval.

The process for finding the confidence interval should be very familiar as it looks a lot like finding the confidence interval for the other parameters we have discussed. The slope estimate is used as a center point of the interval and a quantity is added and subtracted from that center to find the upper and lower values of the interval.

$\beta_1 \pm t^* \times SE_{\beta_1}$ (from Section 4.2 of *Advanced Statisical Introductions with Probability*)

In the Lesson 9 document I demonstrated the "long way" of finding the standard error of the slope estimate, here I will pull this value from of the model's summary output instead. Let's find a 90% confidence interval for the slope parameter.

```
model = lm(GPA$gpa ~ GPA$hours)

beta_1 = model$coefficients[[2]]

t_crit = abs(qt(0.025, df = nrow(GPA) - 2))

standard_error = summary(model)[[4]][[4]]

lower = beta_1 - t_crit * standard_error

upper = beta_1 + t_crit * standard_error
```

```r
paste("(", lower, ",", upper, ")")
```

```
## [1] "( -0.0120883844185627 , 0.000320637671644718 )"
```

From this interval we can say that we are 90% confident that the true slope parameter falls between -0.01209 and 0.00032. Does this make sense? Well if you will recall for our data exploration, there didn't appear to be a strong relationship between our explanatory variable (non-academic time) and response variable (GPA) on the scatter plot of our sample. What sort of true slope would we expect to see if our sample was representative? No association would correspond to a slope of zero and therefore this range of values seems plausible.