Homework #10

Joanna Yu (W203 Tuesday 4pm Fall 2018)

11/19/2018

1. Recall that the slope coefficient in a simple regression of Yi on Xi can be expressed as,

$$\beta_1 = \frac{\hat{cov}(X_i, Y_i)}{\hat{var}(X_i)}$$

Suppose that you were to add a random variable, M_i, representing measurement error, to each X_i. You may assume that Mi is uncorrelated with both X_i and Y_i. You then run a regression of Y_i on $X_i + M_i$ instead ofon X_i. Does the measurement error increase or decrease your slope coefficient?

$$\begin{aligned}
\beta_1 &= \frac{\hat{cov}(X_i + M_i, Y_i)}{\hat{var}(X_i + M_i)} \\
&= \frac{\hat{cov}(X_i, Y_i) + \hat{cov}(M_i, Y_i)}{\hat{var}(X_i) + \hat{var}(M_i) + 2\hat{cov}(X_i, M_i)} \\
&= \frac{\hat{cov}(X_i, Y_i)}{\hat{var}(X_i) + \hat{var}(M_i)}
\end{aligned}$$

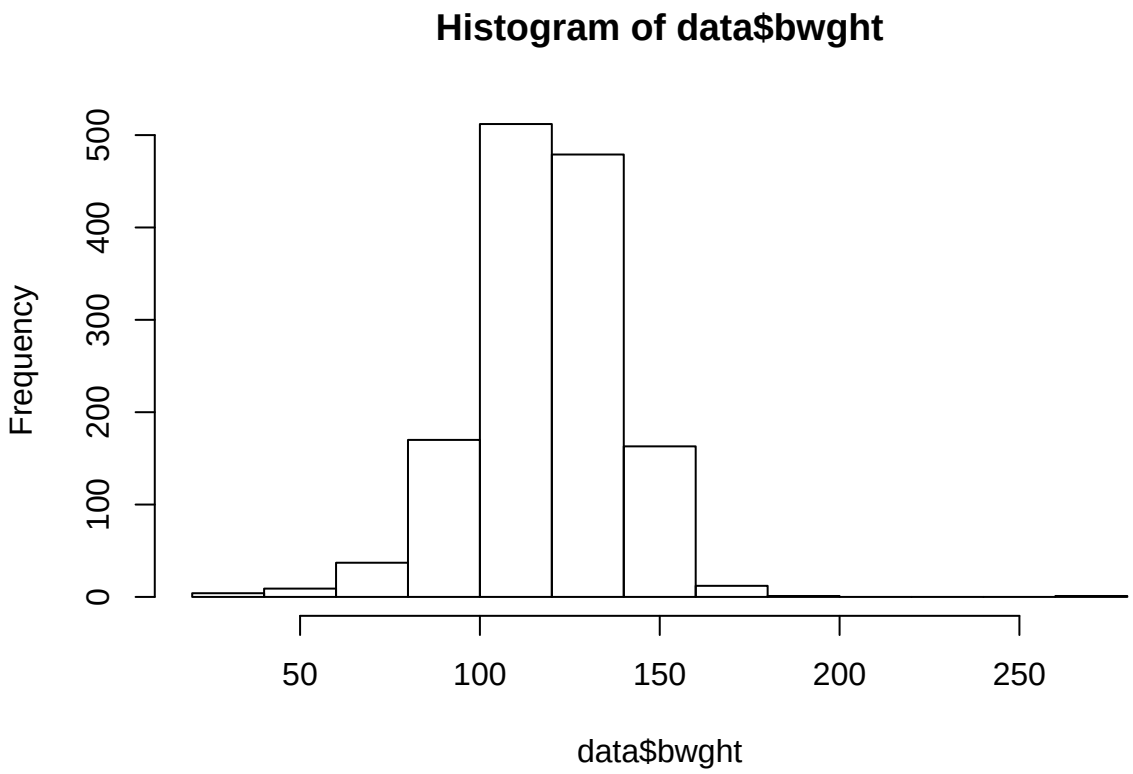
After simplifying the formula, we can see that β_1 is expected to become smaller since the denominator is now the sum of the variances of X_i and M_i. Therefore the measurement error decreases the slope coefficient.

Problem 2 - Examine the dependent variable, infant birth weight in ounces (bwght) and the independent variable, the number of cigarettes smoked by the mother each day during pregnacy (cigs).

```
load("bwght.RData")
summary(data$bwght)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    23.0   107.0   120.0   118.7   132.0   271.0
```

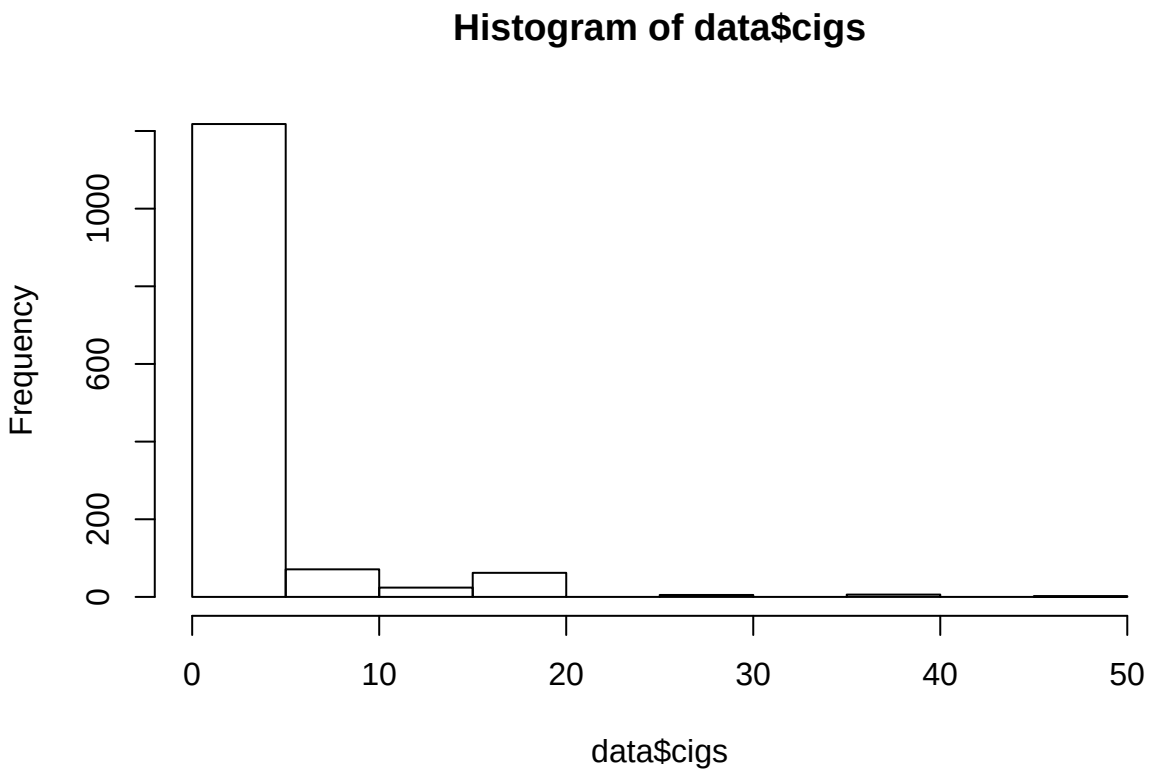
```
hist(data$bwght)
```




```
summary(data$cigs)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   2.087   0.000  50.000
```

```
hist(data$cigs)
```

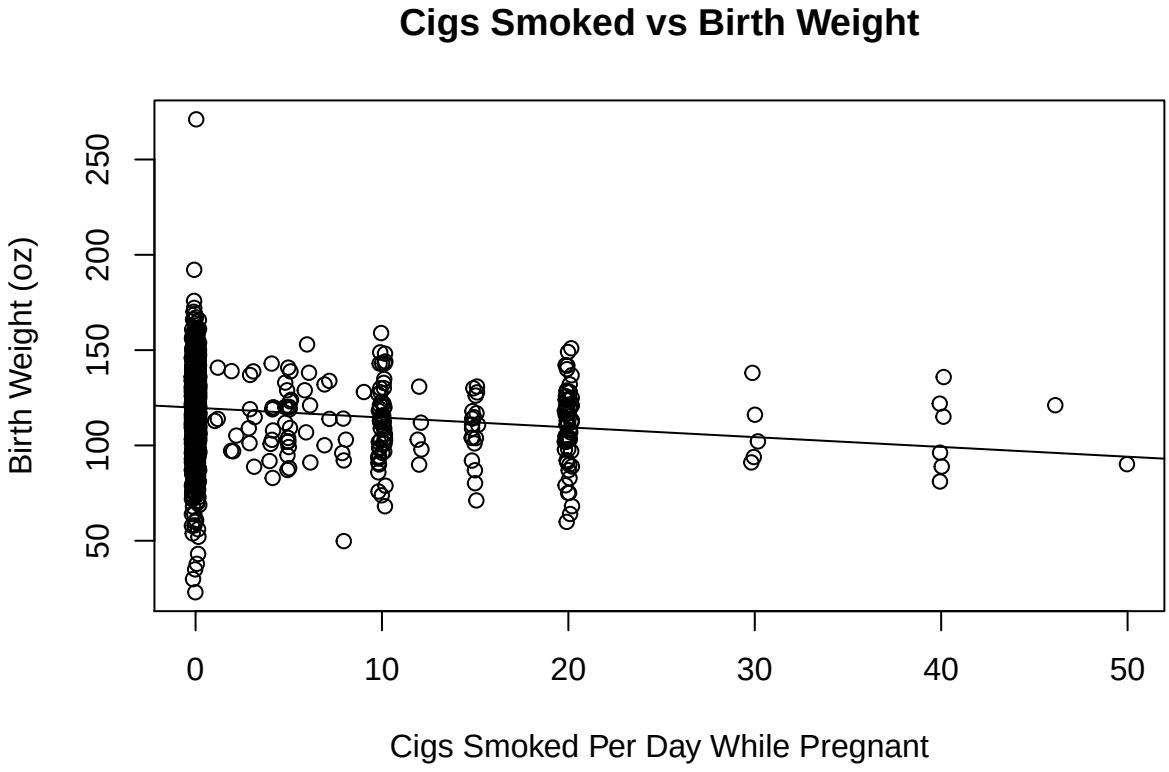


2. Fit a linear model that predicts bwght as a function of cigs. Superimpose your regression line on a scatterplot of your variables.

```
cig_bwght_model = lm(bwght ~ cigs, data = data)
cig_bwght_model
```

```
##
## Call:
## lm(formula = bwght ~ cigs, data = data)
##
## Coefficients:
## (Intercept)         cigs
##    119.7719      -0.5138
```

```
plot(jitter(data$cigs), jitter(data$bwght), xlab = "Cigs Smoked Per Day While Pregnant",
     ylab = "Birth Weight (oz)", main = "Cigs Smoked vs Birth Weight")
abline(cig_bwght_model)
```




```
summary(cig_bwght_model)$r.square
```

```
## [1] 0.02272912
```

3. Examine the coefficients of your fitted model. Explain, in particular, how to interpret the slope coefficient on cigs. Is it practically significant?

The slope is -0.51. It means that for each extra cigarette smoked during pregnancy, the model predicts a half ounce smaller baby at birth. It has low practical significance because the r square is very low.

4. Write down the two moment conditions for this regression. Use R to verify that they hold for your fitted model.

Condition #1

$$E(U_i) = 0$$

Condition #2

$$cov(U_i, X_i) = E(U_i * X_i) = 0$$

```
bwght_residual = data$bwght - (119.7719-0.5138*data$cigs)
mean(bwght_residual)
```

```
## [1] 5.864553e-05
```

```
mean(bwght_residual*data$cigs)
```

```
## [1] 0.001117219
```

5. Does this simple regression capture a causal relationship between smoking and birthweight? Explain why or why not.

Definitely not. The r is very low and the graph itself suggests no strong correlation, so definitely not causal.

6. Does your scatterplot show evidence of measurement error in cigs? If so, what does this say about the true relationship between cigarettes and birthweight?

I believe there is measurement error in cigs. There is a lot of clustering around values 10, 20. Many smokers probably round their numbers to the nearest 10. This would make modeling more difficult even if there is stronger association between the two. And values over 40 seem very extreme and unlikely.

7. Using your coefficients, what is the predicted birthweight when cigs is 0? When cigs is 20?

cigs = 0 -> birthweight = 120 oz
cigs = 20 -> birthweight = 119.7719-0.5138*20 = 109.50 oz

8. Use R's predict function to verify your previous answers. You may insert your linear model object into the command below.

```
predict(cig_bwght_model, data.frame(cigs = c(0, 20) ) )
```

```
##        1        2
## 119.7719 109.4965
```

9. To predict a birthweight of 100 ounces, what would cigs have to be?

100 = 119.7719-0.5138*cigs
cigs = 38