STAT 3202: Homework 08

Autumn 2018, OSU

Due: Friday, November 9

Please see the **detailed homework policy document** for information about homework formatting, submission, and grading.

Exercise 1

Consider the simple linear regression model

$$Y = -5 + 2.2x + \epsilon$$

where
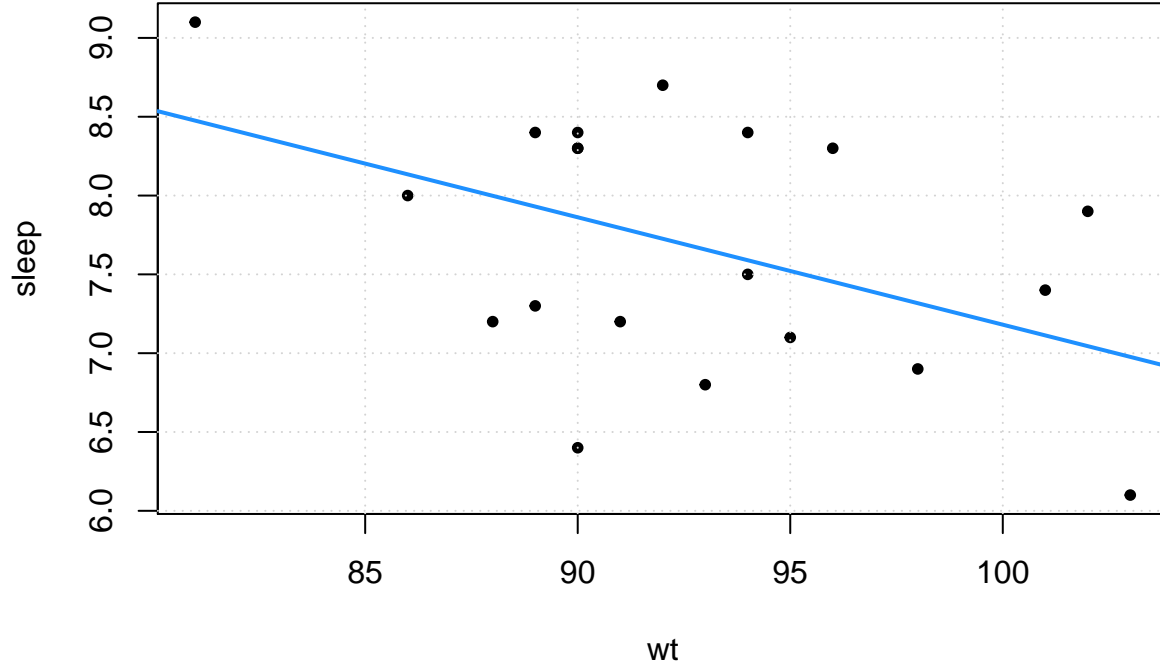
$$\epsilon \sim N(0, \sigma^2 = 16).$$

Calculate two probabilities using this model:

$$P[Y > 8 \mid X = 2]$$

$$P[Y > 8 \mid X = 4]$$

Exercise 2

The above (simulated) data shows the relationship between sleep (in hours) and weight (in kilograms) of a random sample of adult males on a particular night. A simple linear regression model was fit to this data. The fitted line is added to the above plot. Use the following information to estimate the mean sleep (in hours) of men that weigh:

- 85 kilograms
- 120 kilograms

Which of these estimates do you feel more confident about?

```
summary(sleep_wt_mod)
```

```
##
## Call:
## lm(formula = sleep ~ wt, data = sleep_wt_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4621 -0.6031  0.0988  0.5596  0.9741
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.99305    2.86363   4.886 0.000119 ***
## wt          -0.06812    0.03087  -2.206 0.040580 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7389 on 18 degrees of freedom
## Multiple R-squared:  0.2129, Adjusted R-squared:  0.1692
## F-statistic: 4.869 on 1 and 18 DF,  p-value: 0.04058
```

Exercise 3

The following output is repeated from the previous exercise.

```
summary(sleep_wt_mod)
```

```
##
## Call:
## lm(formula = sleep ~ wt, data = sleep_wt_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4621 -0.6031  0.0988  0.5596  0.9741
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.99305    2.86363   4.886 0.000119 ***
## wt          -0.06812    0.03087  -2.206 0.040580 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7389 on 18 degrees of freedom
## Multiple R-squared:  0.2129, Adjusted R-squared:  0.1692
## F-statistic: 4.869 on 1 and 18 DF,  p-value: 0.04058
```

Use this data to do two things:

- Report the proportion of the variance in sleep time that is explained by a linear relationship with weight.
- Estimate σ, the standard deviation of the errors in the simple linear regression model.

Exercise 4
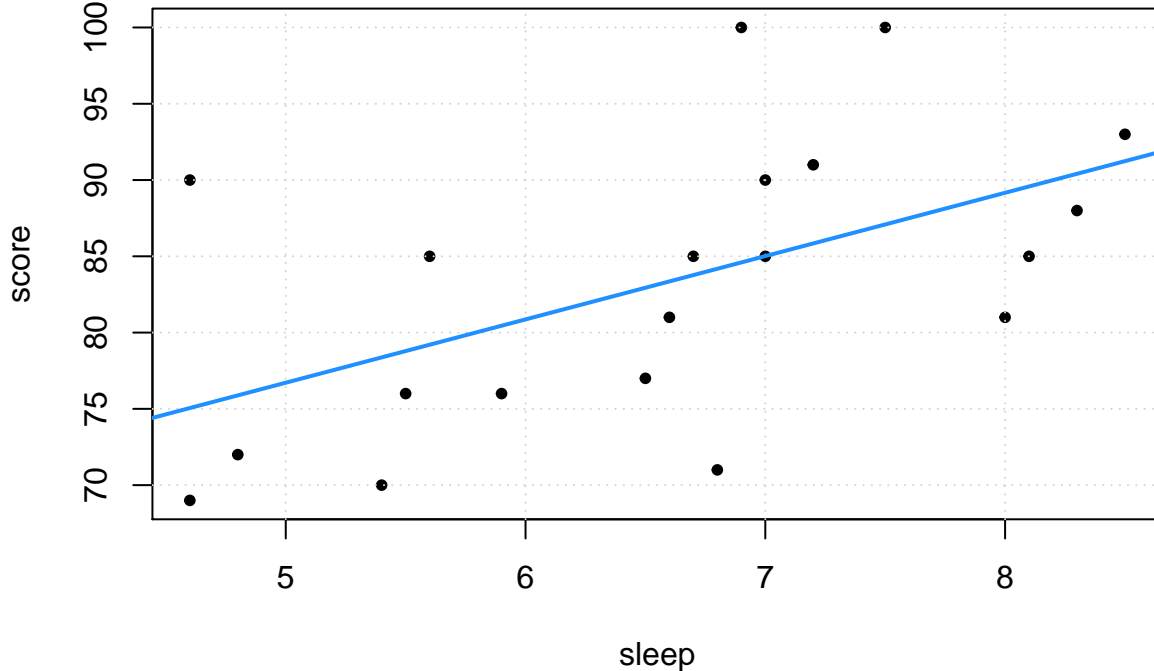
The above (simulated) data shows the relationship between exam scores and sleep (in hours) for a random sample of students in a large statistics course. A simple linear regression model was fit to this data. The fitted line is added to the above plot. Consider two students in this dataset:

- Student A, who slept 7.5 hours and scored 100 points
- Student B, who slept 8.0 hours and scored 81 points

For both, use the following information about the fitted simple linear regression model to predict their scores on the exam given how much they slept. (In other words, obtain the fitted values.) Calculate the residual for each of these.

```
summary(sleep_score_mod)
```

```
##
## Call:
## lm(formula = score ~ sleep, data = sleep_score_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.184  -4.919  -2.381   5.029  15.401
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   55.964     10.248   5.461 3.46e-05 ***
## sleep          4.150      1.534   2.705   0.0145 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.077 on 18 degrees of freedom
## Multiple R-squared:  0.289,  Adjusted R-squared:  0.2495
## F-statistic: 7.316 on 1 and 18 DF,  p-value: 0.0145
```

Exercise 5

Sometimes it can be reasonable to assume that β_0, the intercept of a regression model, should be 0. That is, the line should pass through the point $(0, 0)$. For example, if a car is traveling 0 miles per hour, its stopping distance should be 0!

We can simply define a model without an intercept,

$$Y_i = \beta x_i + \epsilon_i.$$

In the **Least Squares Approach** section of the text we saw the calculus behind the derivation of the regression estimates, and then we performed the calculation for the `cars` dataset using `R`.

Recreate this derivation for the model without the intercept. That is, use the method of least squares to derive an estimate for β using data points (x_i, y_i) for $i = 1, 2, \ldots n$. Simply put, find the value of β to minimize the function

$$f(\beta) = \sum_{i=1}^{n}(y_i - \beta x_i)^2.$$

Use your estimator to estimate β for the following simple dataset:

```
small_data = data.frame(x = c(2, 4, 6),
                        y = c(3, 6, 8))
small_data
```

```
##   x y
## 1 2 3
## 2 4 6
## 3 6 8
```