Week 2 Quiz Material

When copy and pasting from a code block, or from your local `R` session, be sure to include all available digits for any numeric answer. Also, do not modify the default digits option in the code blocks or your local `R` session.

Practice

Exercise 1

```
#
```

Consider a random variable X that has a normal distribution with a mean of 5 and a variance of 9. Calculate $P[X > 4]$.

```
1 - pnorm(4, mean = 5, sd = 3)
```

```
## [1] 0.6305587
```

```
pnorm(4, mean = 5, sd = 3, lower.tail = FALSE)
```

```
## [1] 0.6305587
```

- Hint: `pnorm()` uses the standard deviation, not the variance.
- Hint: By default `pnorm()` considers area under the curve to the left of the given value.

Exercise 2

```
#
```

Consider the simple linear regression model

$$Y = -3 + 2.5x + \epsilon$$

where

$$\epsilon \sim N(0, \sigma^2 = 4).$$

What is the expected value of Y given that $x = 5$? That is, what is $\text{E}[Y \mid X = 5]$?

```
-3 + 2.5 * 5
```

```
## [1] 9.5
```

- Hint: Recall that the SLR model can be thought of as a conditional model.

Exercise 3

```
#
```

Given the SLR model in exercise 2, what is the standard deviation of Y when x is 10. That is, what is $\text{SD}[Y \mid X = 10]$?

```
sqrt(4)
```

```
## [1] 2
```

- Hint: Recall that the assumptions of the SLR model, specifically, “equal variance”. Don’t over-think it.

Exercise 4

```
#
```

For this Exercise, use the built-in `trees` dataset in `R`. Fit a simple linear regression model with `Girth` as the response and `Height` as the predictor. What is the slope of the fitted regression line?

```
coef(lm(Girth ~ Height, data = trees))[2]
```

```
##    Height
## 0.2557471
```

Exercise 5

```
#
```

Continue using the SLR model you fit in Exercise 4. What is the value of R^2 for this fitted SLR model?

```
summary(lm(Girth ~ Height, data = trees))$r.squared
```

```
## [1] 0.2696518
```

Graded

Exercise 1

```
#
```

Consider the simple linear regression model

$$Y = 10 + 5x + \epsilon$$

where

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

Calculate the probability that Y is less than 6 given that $x = 0$.

```
x = 0
mu = 10 + 5 * x
sigma = 4
pnorm(6, mean = mu, sd = sigma)
```

```
## [1] 0.1586553
```

$$Y \mid X = 0 \sim N(\mu = 10, \sigma^2 = 16)$$

Exercise 2

```
#
```

Using the SLR model in exercise 1, what is the probability that Y is greater than 3 given that $x = -1$?

```
x = -1
mu = 10 + 5 * x
sigma = 4
pnorm(3, mean = mu, sd = sigma, lower.tail = FALSE)
```

```
## [1] 0.6914625
```

$$Y \mid X = 0 \sim N(\mu = 5, \sigma^2 = 16)$$

Exercise 3

```
#
```

Using the SLR model in exercise 1, what is the probability that Y is greater than 3 given that $x = -2$?

```
x = -2
mu = 10 + 5 * x
sigma = 4
pnorm(3, mean = mu, sd = sigma, lower.tail = FALSE)
```

```
## [1] 0.2266274
```

$$Y \mid X = 0 \sim N(\mu = 0, \sigma^2 = 16)$$

Exercise 4

```
#
```

For exercises 4 - 11, use the `faithful` dataset, which is built into `R`.

Suppose we would like to predict the duration of an eruption of the Old Faithful geyser (http://www.yellowstonepark.com/about-old-faithful/) in Yellowstone National Park (https://en.wikipedia.org/wiki/Yellowstone_National_Park) based on the waiting time before an eruption. Fit a simple linear model in `R` that accomplishes this task.

What is the estimate of the intercept parameter?

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
coef(faithful_model)[1]
```

```
## (Intercept)
##   -1.874016
```

Exercise 5

```
#
```

What is the estimate of the slope parameter?

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
coef(faithful_model)[2]
```

```
##    waiting
## 0.07562795
```

Exercise 6

```
#
```

Use the fitted model to predict the duration of an eruption based on a waiting time of **80** minutes.

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
predict(faithful_model, data.frame(waiting = 80))
```

```
##       1
## 4.17622
```

Exercise 7

```
#
```

Use the fitted model to predict the duration of an eruption based on a waiting time of **120** minutes.

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
predict(faithful_model, data.frame(waiting = 120))
```

```
##        1
## 7.201338
```

Exercise 8

```
#
```

Of the predictions that you made, for 80 and 120 minutes, which is more reliable?

- **80**
- 120
- Both are equally reliable

```
range(faithful$waiting)
```

```
## [1] 43 96
```

Exercise 9

```
#
```

Calculate the RSS for the fitted model.

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
sum(resid(faithful_model) ^ 2)
```

```
## [1] 66.56178
```

Exercise 10

```
#
```

What proportion of the variation in eruption duration is explained by the linear relationship with waiting time?

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
summary(faithful_model)$r.squared
```

```
## [1] 0.8114608
```

Exercise 11

```
#
```

Calculate the standard deviation of the residuals of the fitted model.

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
sd(resid(faithful_model))
```

```
## [1] 0.495596
```

Exercise 12

Suppose both Least Squares and Maximum Likelihood are used to fit a simple linear regression model to the same data. The estimates for the slope and the intercept will be:

- **The same**
- Different
- Possibly the same or different depending on the data

Exercise 13

Consider the fitted regression model:

$$\hat{y} = -1.5 + 2.3x$$

Indicate all of the following that **must** be true:

- The difference between the y values of observations at $x = 10$ and $x = 9$ is 2.3.
- **A good estimate for the mean of Y when $x = 0$ is -1.5.**
- There are observations in the dataset used to fit this regression with negative y values.

Exercise 14

Indicate all of the following that are true:

- **The SLR model assumes that errors are independent.**
- The SLR model allows for a larger variances for larger values of the predictor variable.
- The SLR model assumes that the response variable follows a normal distribution.
- **The SLR model assumes taht the relationship between the response and the predictor is linear.**

Exercise 15

Suppose you fit a simple linear regression model and obtain $\hat{\beta}_1 = 0$. Does this mean that there is **no relationship** between the response and the predictor?

- Yes
- **No**
- Depends on the intercept