Week 3 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 t distribution with 7 degrees of freedom. Calculate $P[X > 1.3]$.

```
1 - pt(1.3, df = 7)
```

```
## [1] 0.1173839
```

```
pt(1.3, df = 7, lower.tail = FALSE)
```

```
## [1] 0.1173839
```

- Hint: By default `pt()` considers area under the curve to the left of the given value.

Exercise 2

```
#
```

Consider a random variable Y that has a t distribution with 9 degrees of freedom. Find c such that $P[X > c] = 0.025$.

```
qt(1 - 0.025, df = 9)
```

```
## [1] 2.262157
```

```
qt(0.025, df = 9, lower.tail = FALSE)
```

```
## [1] 2.262157
```

- Hint: Like `pt()`, by default `qt()` deals in areas “to the left.”

Exercise 3

```
#
```

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 p-value for testing $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$?

```
tree_model = lm(Girth ~ Height, data = trees)
summary(tree_model)$coefficients["Height", "Pr(>|t|)"]
```

```
## [1] 0.002757815
```

- Hint: Utilize the `summary()` function.

Exercise 4

```
#
```

Continue using the SLR model you fit in Exercise 3. What is the length of a 90% confidence interval for β_1?

```
tree_model = lm(Girth ~ Height, data = trees)
ci_beta_1 = confint(tree_model, parm = "Height", level = 0.90)
ci_beta_1[2] - ci_beta_1[1]
```

```
## [1] 0.2656018
```

- Hint: Remember to specify the confidence level.

Exercise 5

```
#
```

Continue using the SLR model you fit in Exercise 3. Calculate a 95% confidence interval for the mean tree girth of a tree that is 79 feet tall. Report the upper bound of this interval.

```
tree_model = lm(Girth ~ Height, data = trees)
predict(tree_model, newdata = data.frame(Height = 79), interval = "confidence")[, "upr"]
```

```
## [1] 15.12646
```

Graded

Exercise 1

```
#
```

Consider a random variable X that has a t distribution with 5 degrees of freedom. Calculate $P[|X| > 2.1]$.

```
pt(-2.1, df = 5) + pt(2.1, df = 5, lower.tail = FALSE)
```

```
## [1] 0.08975325
```

```
2 * pt(2.1, df = 5, lower.tail = FALSE)
```

```
## [1] 0.08975325
```

Exercise 2

```
#
```

Calculate the critical value used for a 90% confidence interval about the slope parameter of a simple linear regression model that is fit to 10 observations. (Your answer should be a positive value.)

```
conf_level = 0.90
sig_level = 1 - conf_level
n = 10
abs(qt(sig_level / 2, df = n - 2))
```

```
## [1] 1.859548
```

Exercise 3

```
#
```

Consider the true simple linear regression model

$$Y_i = 5 + 4x_i + \epsilon_i \qquad \epsilon_i \sim N(0, \sigma^2 = 4) \qquad i = 1, 2, \ldots 20$$

Given $S_{xx} = 1.5$, calculate the probability of observing data according to this model, fitting the SLR model, and obtaining an estimate of the slope parameter greater than 4.2. In other words, calculate

$$P[\hat{\beta}_1 > 4.2]$$

```
Sxx = 1.5
beta_1 = 4
sigma = 2
e_beta_1_hat = 4
sd_beta_1_hat = sqrt(sigma ^ 2 / Sxx)
pnorm(4.2, mean = e_beta_1_hat, sd = sd_beta_1_hat, lower.tail = FALSE)
```

```
## [1] 0.4512616
```

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$$

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 value of $\text{SE}[\hat{\beta}_1]$?

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
summary(faithful_model)$coefficients["waiting", "Std. Error"]
```

```
## [1] 0.002218541
```

Exercise 5

```
#
```

What is the value of the test statistic for testing $H_0 : \beta_0 = 0$ vs $H_1 : \beta_0 \neq 0$?

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
summary(faithful_model)$coefficients["(Intercept)", "t value"]
```

```
## [1] -11.70212
```

Exercise 6

```
#
```

What is the value of the test statistic for testing $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$?

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
summary(faithful_model)$coefficients["waiting", "t value"]
```

```
## [1] 34.08904
```

Exercise 7

```
#
```

Test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ with $\alpha = 0.01$. What decision do you make?

- Fail to reject H_0
- **Reject** H_0
- Reject H_1
- Not enough information

Exercise 8

```
#
```

Calculate a 90% confidence interval for β_0. Report the upper bound of this interval.

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
confint(faithful_model, parm = "(Intercept)", level = 0.90)[, 2]
```

```
## [1] -1.609697
```

Exercise 9

```
#
```

Calculate a 95% confidence interval for β_1. Report the length of the margin of this interval.

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
ci_beta_1 = confint(faithful_model, parm = "waiting")
(ci_beta_1[2] - ci_beta_1[1]) / 2
```

```
## [1] 0.00436784
```

Exercise 10

```
#
```

Create a 90% confience interval for the mean eruption duration for a waiting time of 81 minutes. Report the lower bound of this interval.

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
predict(faithful_model, interval = "confidence",
        level = 0.90, newdata = data.frame(waiting = 81))[, "lwr"]
```

```
## [1] 4.189899
```

Exercise 11

```
#
```

Create a 99% prediction interval for a new obersation's eruption duration for a waiting time of 72 minutes. Report the upper bound of this interval.

```
faithful_model = lm(eruptions ~ waiting, data = faithful)
predict(faithful_model, interval = "prediction",
        level = 0.99, newdata = data.frame(waiting = 72))[, "upr"]
```

```
## [1] 4.861612
```

Exercise 12

Consider a 90% confidence interval for the mean response and a 90% prediction interval, both at the same x value. Which interval is narrower?

- **Confidence interval**
- Prediction interval
- No enough information, it depends on the value of x

Exercise 13

Suppose you obtain a 99% confidence interval for β_1 that is $(-0.4, 5.2)$. Now test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ with $\alpha = 0.01$. What decision do you make?

- **Fail to reject H_0**
- Reject H_0
- Reject H_1
- Not enough information

Exercise 14

Suppose you test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ with $\alpha = 0.01$ and fail to reject H_0. Indicate all of the following that must always be true:

- There is no relationship between the response and the predictor.
- **The probability of observing the estimated value of β_1 (or something more extreme) is greater than 0.01 if we assume that $\beta_1 = 0$.**
- The value of $\hat{\beta}_1$ is very small. For example, it could not be 1.2.
- The probability that $\beta_1 = 0$ is very high.
- We would also fail to reject at $\alpha = 0.05$.

Exercise 15

Consider a 95% confidence interval for the mean response calculated at $x = 6$. If instead we calculate the interval at $x = 7$, mark each value that would change:

- **Estimate**
- Critical Value
- **Standard Error**