STAT 420: Exam I Material

Summer 2018, Unger

    - Variation A
    - Variation B
    - Variation C
- Question 14
    - Variation A
    - Variation B
    - Variation C
- Question 15
    - Variation A
    - Variation B
    - Variation C

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. It would be best to copy and paste values that were returned using printing methods that do not round results. (Notably the direct output from calling `summary()` .) Also, do not modify the default digits option in the code blocks or your local `R` session.

All questions are worth **1 point**.

Question 1

Variation A

```
# Preamble
```

```
# Starter
set.seed(42)
# Your code here.
# Your code here.
```

After setting the given seed, generate 250 random observations from a normal distribution with a mean of 2 and a variance of 10. What proportion of these observations are larger than 5? (Your answer should be a number between 0 and 1.)

```
# Solution
set.seed(42)
sims = rnorm(n = 250, mean = 2, sd = sqrt(10))
mean(sims > 5)
```

```
## [1] 0.168
```

Variation B

```
# Preamble
```

```
# Starter
set.seed(42)
# Your code here.
# Your code here.
```

After setting the given seed, generate 350 random observations from a normal distribution with a mean of 3 and a variance of 11. What proportion of these observations are larger than 4? (Your answer should be a number between 0 and 1.)

```
# Solution
set.seed(42)
sims = rnorm(n = 350, mean = 3, sd = sqrt(11))
mean(sims > 4)
```

```
## [1] 0.3514286
```

Variation C

```
# Preamble
```

```
# Starter
set.seed(42)
# Your code here.
# Your code here.
```

After setting the given seed, generate 450 random observations from a normal distribution with a mean of 5 and a variance of 13. What proportion of these observations are larger than 6? (Your answer should be a number between 0 and 1.)

```
# Solution
set.seed(42)
sims = rnorm(n = 450, mean = 5, sd = sqrt(13))
mean(sims > 6)
```

```
## [1] 0.36
```

Question 2

Variation A

```
# Preamble
```

```
# Starter
(some_fun() + some_fun(arg1 = 3, arg2 = 1:10)) / some_fun(arg1 = 5, arg2 = 1:5)
```

Write a function named `some_fun`. The function should take two arguments as input:

- `arg1`, an integer larger than `1`, with a default value of `2`
- `arg2`, a vector of real numbers with a default value of `1`.

The function should output the average of the elements of `arg2` multiplied by `arg1`. Write and run your function, then report the value from running the starter code.

```
# Solution
some_fun = function(arg1 = 2, arg2 = 1) {
  arg1 * mean(arg2)
}

(some_fun() + some_fun(arg1 = 4, arg2 = 1:10)) / some_fun(arg1 = 5, arg2 = 1:5)
```

```
## [1] 1.6
```

Variation B

```
# Preamble
```

```
# Starter
(some_fun() + some_fun(arg1 = 3, arg2 = 1:10)) / some_fun(arg1 = 5, arg2 = 1:5)
```

Write a function named `some_fun`. The function should take two arguments as input:

- `arg1`, a vector of real numbers with a default value of `1`.
- `arg2`, an integer larger than `1`, with a default value of `2`

The function should output the average of the elements of `arg1` divided by `arg2`. Write and run your function, then report the value from running the starter code.

```
# Solution
some_fun = function(arg1 = 1, arg2 = 2) {
  mean(arg1) / arg2
}

(some_fun() + some_fun(arg1 = 1:4, arg2 = 4)) / some_fun(arg1 = 1:9, arg2 = 2)
```

```
## [1] 0.45
```

Question 3

Variation A

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

What proportion of the variance of `circumference` is explained by a linear relationship with `age`?

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
summary(orange_model)$r.squared
```

```
## [1] 0.8345167
```

Variation B

```
# Preamble
```

```
# Starter
```

For this question use the `iris` dataset, which is built into `R`.

What proportion of the variance of `Sepal.Length` is explained by a linear relationship with `Sepal.Width`?

```
# Solution
iris_model = lm(Sepal.Length ~ Sepal.Width, data = iris)
summary(iris_model)$r.squared
```

```
## [1] 0.01382265
```

Variation C

```
# Preamble
```

```
# Starter
```

For this question use the `trees` dataset, which is built into `R`.

What proportion of the variance of `Volume` is explained by a linear relationship with `Height`?

```
# Solution
tree_model = lm(Volume ~ Height, data = trees)
summary(tree_model)$r.squared
```

```
## [1] 0.3579026
```

Question 4

Variation A

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

Use a simple linear regression model to create predictions for the circumference of orange trees in millimeters when their age is 400 days and 2500 days. Report the value of the prediction that you feel is more valid.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
# predict(orange_model, newdata = data.frame(age = c(400, 2500)))
# range(Orange$circumference)
predict(orange_model, newdata = data.frame(age = 400))
```

```
##        1
## 60.10778
```

Variation B

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

Use a simple linear regression model to create predictions for the circumference of orange trees in millimeters when their age is 500 days and 2500 days. Report the value of the prediction that you feel is more valid.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
# predict(orange_model, newdata = data.frame(age = c(500, 2500)))
# range(Orange$circumference)
predict(orange_model, newdata = data.frame(age = 500))
```

```
##        1
## 70.78481
```

Variation C

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

Use a simple linear regression model to create predictions for the circumference of orange trees in millimeters when their age is 600 days and 2500 days. Report the value of the prediction that you feel is more valid.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
# predict(orange_model, newdata = data.frame(age = c(600, 2500)))
# range(Orange$circumference)
predict(orange_model, newdata = data.frame(age = 600))
```

```
##        1
## 81.46185
```

Question 5

Variation A

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

Use a simple linear regression model to create a 90% confidence interval for the change in mean circumference of orange trees in millimeters when age is increased by 1 day. Report the lower bound of this interval.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
confint(orange_model, parm = "age", level = 0.90)[, 1]
```

```
## [1] 0.0927633
```

Variation B

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

Use a simple linear regression model to create a 99% confidence interval for the change in mean circumference of orange trees in millimeters when age is increased by 1 day. Report the upper bound of this interval.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
confint(orange_model, parm = "age", level = 0.99)[, 2]
```

```
## [1] 0.1293926
```

Question 6

Variation A

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

Use a simple linear regression model to create a 90% confidence interval for the mean circumference of orange trees in millimeters when the age is 250 days. Report the lower bound of this interval.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
predict(orange_model, interval = "confidence", newdata = data.frame(age = 250), level = 0.90)[,
"lwr"]
```

```
## [1] 32.48418
```

Variation B

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

Use a simple linear regression model to create a 95% confidence interval for the mean circumference of orange trees in millimeters when the age is 330 days. Report the upper bound of this interval.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
predict(orange_model, interval = "confidence", newdata = data.frame(age = 330), level = 0.95)[,
"upr"]
```

```
## [1] 65.52032
```

Question 7

Variation A

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

Use a simple linear regression model to create a 99% prediction interval for an observation of the circumference of an orange tree in millimeters when its age is 400 days. Report the upper bound of this interval.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
predict(orange_model, interval = "prediction", newdata = data.frame(age = 400), level = 0.99)[,
"upr"]
```

```
## [1] 126.9615
```

Variation B

```
# Preamble
```

```
# Starter
```

For this question use the `Orange` dataset, which is built into `R`.

Use a simple linear regression model to create a 95% prediction interval for an observation of the circumference of an orange tree in millimeters when its age is 500 days. Report the lower bound of this interval.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
predict(orange_model, interval = "prediction", newdata = data.frame(age = 500), level = 0.95)[,
"lwr"]
```

```
## [1] 21.29195
```

Question 8

Variation A

```
# Preamble
```

```
# Starter
```

Consider the SLR model

$$Y = \beta_0 + \beta_1 x + \epsilon$$

where Y is circumference and x is age.

Calculate the p-value of the test $H_0 : \beta_1 = 0.123 \text{ vs } H_1 : \beta_1 \neq 0.123$.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
n = length(resid(orange_model))
est = summary(orange_model)$coefficients[2, "Estimate"]
se = summary(orange_model)$coefficients[2, "Std. Error"]
t = (est - 0.123) / se
2 * pt(abs(t), df = n - 2, lower.tail = FALSE)
```

```
## [1] 0.05837492
```

Variation B

```
# Preamble
```

```
# Starter
```

Consider the SLR model

$$Y = \beta_0 + \beta_1 x + \epsilon$$

where Y is circumference and x is age.

Calculate the p-value of the test $H_0 : \beta_1 = 0.125 \text{ vs } H_1 : \beta_1 \neq 0.125$.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
n = length(resid(orange_model))
est = summary(orange_model)$coefficients[2, "Estimate"]
se = summary(orange_model)$coefficients[2, "Std. Error"]
t = (est - 0.125) / se
2 * pt(abs(t), df = n - 2, lower.tail = FALSE)
```

```
## [1] 0.03472037
```

Variation C

```
# Preamble
```

```
# Starter
```

Consider the SLR model

$$Y = \beta_0 + \beta_1 x + \epsilon$$

where Y is circumference and x is age.

Calculate the p-value of the test $H_0 : \beta_1 = 0.127 \text{ vs } H_1 : \beta_1 \neq 0.127$.

```
# Solution
orange_model = lm(circumference ~ age, data = Orange)
n = length(resid(orange_model))
est = summary(orange_model)$coefficients[2, "Estimate"]
se = summary(orange_model)$coefficients[2, "Std. Error"]
t = (est - 0.127) / se
2 * pt(abs(t), df = n - 2, lower.tail = FALSE)
```

```
## [1] 0.02002805
```

Question 9

Variation A

```
# Preamble
```

```
# Starter
```

Consider the true model

$$Y = 2 + 1.5x_1 - 2.1x_2 + 3.2x_3 + \epsilon$$

where

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

What is the probability that Y is less than 3 given that:

- $x_1 = 0$
- $x_2 = 0$
- $x_3 = 0$

```
# Solution
mu = 2 + 1.5 * 0 - 2.1 * 0 + 3.2 * 0
pnorm(3, mean = mu, sd = 2)
```

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

Variation B

```
# Preamble
```

```
# Starter
```

Consider the true model

$$Y = 3 + 1.4x_1 - 2.3x_2 + 3.1x_3 + \epsilon$$

where

$$\epsilon \sim N(0, \sigma^2 = 9)$$

What is the probability that Y is less than 5 given that:

- $x_1 = 0$
- $x_2 = 0$
- $x_3 = 0$

```
# Solution
mu = 3 + 1.4 * 0 - 2.3 * 0 + 3.1 * 0
pnorm(5, mean = mu, sd = 3)
```

```
## [1] 0.7475075
```

Variation C

```
# Preamble
```

```
# Starter
```

Consider the true model

$$Y = 4 + 1.3x_1 - 2.4x_2 + 3.7x_3 + \epsilon$$

where

$$\epsilon \sim N(0, \sigma^2 = 25)$$

What is the probability that Y is less than 3 given that:

- $x_1 = 0$
- $x_2 = 0$
- $x_3 = 0$

```
# Solution
mu = 4 + 1.3 * 0 - 2.4 * 0 + 3.7 * 0
pnorm(3, mean = mu, sd = 5)
```

```
## [1] 0.4207403
```

Question 10

Variation A

```
# Preamble
```

```
# Starter
```

Consider the true model

$$Y = 2 + 1.5x_1 - 2.1x_2 + 3.2x_3 + \epsilon$$

where

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

What is the probability that Y is greater than 11 given that:

- $x_1 = 1$
- $x_2 = 2$
- $x_3 = 3$

```
# Solution
mu = 2 + 1.5 * 1 - 2.1 * 2 + 3.2 * 3
pnorm(11, mean = mu, sd = 2, lower.tail = FALSE)
```

```
## [1] 0.1468591
```

Variation B

```
# Preamble
```

```
# Starter
```

Consider the true model

$$Y = 3 + 1.4x_1 - 2.3x_2 + 3.1x_3 + \epsilon$$

where

$$\epsilon \sim N(0, \sigma^2 = 9)$$

What is the probability that Y is greater than 11 given that:

- $x_1 = 1$
- $x_2 = 2$
- $x_3 = 3$

```
# Solution
mu = 3 + 1.4 * 1 - 2.3 * 2 + 3.1 * 3
pnorm(12, mean = mu, sd = 3, lower.tail = FALSE)
```

```
## [1] 0.1668553
```

Variation C

```
# Preamble
```

```
# Starter
```

Consider the true model

$$Y = 4 + 1.3x_1 - 2.4x_2 + 3.7x_3 + \epsilon$$

where

$$\epsilon \sim N(0, \sigma^2 = 25)$$

What is the probability that Y is greater than 13 given that:

- $x_1 = 1$
- $x_2 = 2$
- $x_3 = 3$

```
# Solution
mu = 4 + 1.3 * 1 - 2.4 * 2 + 3.7 * 3
pnorm(13, mean = mu, sd = 5, lower.tail = FALSE)
```

```
## [1] 0.3897388
```

Question 11

Variation A

```
# Preamble
```

```
# Starter
set.seed(420)
x_values = data.frame(
  x1 = rnorm(10),
  x2 = runif(10),
  x3 = runif(10),
  x4 = runif(10)
)
```

Consider the true model

$$Y = 4 + 2.5x_1 + 3x_2 - 5x_3 + \epsilon$$

where

$$\epsilon \sim N(0, \sigma^2 = 9)$$

What is $\text{SD}[\hat{\beta}_2]$ given the values of predictors above?

```
# Solution
X = as.matrix(cbind(rep(1, 10), x_values))
C = solve(t(X) %*% X)
sqrt(9 * C[2 + 1, 2 + 1])
```

```
## [1] 4.45513
```

Variation B

```
# Preamble
```

```
# Starter
set.seed(420)
x_values = data.frame(
  x1 = rnorm(15),
  x2 = runif(15),
  x3 = runif(15),
  x4 = runif(15)
)
```

Consider the true model

$$Y = 5 + 1.5x_1 + 2x_2 - 3x_3 + \epsilon$$

where

$$\epsilon \sim N(0, \sigma^2 = 25)$$

What is $\text{SD}[\hat{\beta}_1]$ given the values of predictors above?

```
# Solution
X = as.matrix(cbind(rep(1, 15), x_values))
C = solve(t(X) %*% X)
sqrt(25 * C[1 + 1, 1 + 1])
```

```
## [1] 1.679908
```

Question 12

Variation A

```
# Preamble
```

```
# Starter
```

Calculate the critical value used for a 90% confidence interval about a single β parameter of a multiple linear regression model with 4 predictors that is fit to 15 observations. (Your answer should be a positive value.)

```
# Solution
conf_level = 0.90
sig_level = 1 - conf_level
n = 15
p = 4 + 1
abs(qt(sig_level / 2, df = n - p))
```

```
## [1] 1.812461
```

Variation B

```
# Preamble
```

```
# Starter
```

Calculate the critical value used for a 95% confidence interval about a single β parameter of a multiple linear regression model with 5 predictors that is fit to 16 observations. (Your answer should be a positive value.)

```
# Solution
conf_level = 0.95
sig_level = 1 - conf_level
n = 16
p = 5 + 1
abs(qt(sig_level / 2, df = n - p))
```

```
## [1] 2.228139
```

Variation C

```
# Preamble
```

```
# Starter
```

Calculate the critical value used for a 99% confidence interval about a single β parameter of a multiple linear regression model with 6 predictors that is fit to 17 observations. (Your answer should be a positive value.)

```
# Solution
conf_level = 0.99
sig_level = 1 - conf_level
n = 17
p = 6 + 1
abs(qt(sig_level / 2, df = n - p))
```

```
## [1] 3.169273
```

Question 13

Variation A

```
# Preamble
```

```r
# Starter
library(MASS)
```

For this question use the `Boston` dataset from the `MASS` package.

What proportion of the median home value (`medv`) is explained by a linear relationship with the average number of rooms per dwelling and the per capita crime rate?

```r
# Solution
library(MASS)
boston_mod = lm(medv ~ rm + crim, data = Boston)
summary(boston_mod)$r.squared
```

```
## [1] 0.5419592
```

Variation B

```r
# Preamble
```

```r
# Starter
library(MASS)
```

For this question use the `Boston` dataset from the `MASS` package.

What proportion of the median home value (`medv`) is explained by a linear relationship with the average number of rooms per dwelling and the nitrogen oxide concentration in parts per 10 million?

```r
# Solution
library(MASS)
boston_mod = lm(medv ~ rm + nox, data = Boston)
summary(boston_mod)$r.squared
```

```
## [1] 0.535438
```

Variation C

```r
# Preamble
```

```r
# Starter
library(MASS)
```

For this question use the `Boston` dataset from the `MASS` package.

What proportion of the median home value (`medv`) is explained by a linear relationship with the average number of rooms per dwelling and the full-value property-tax rate per $10,000?

```
# Solution
library(MASS)
boston_mod = lm(medv ~ rm + tax, data = Boston)
summary(boston_mod)$r.squared
```

```
## [1] 0.5605639
```

Question 14

Variation A

```
# Preamble
```

```
# Starter
library(MASS)
```

For this question use the `Boston` dataset from the `MASS` package.

Consider the multiple regression model

$$Y = \beta_0 + \beta_{\texttt{lstat}} x_{\texttt{lstat}} + \beta_{\texttt{rm}} x_{\texttt{rm}} + \beta_{\texttt{crim}} x_{\texttt{crim}} + \beta_{\texttt{tax}} x_{\texttt{tax}} + \beta_{\texttt{nox}} x_{\texttt{nox}} + \epsilon$$

with the usual assumptions on the error term and Y is `medv`.

Report the p-value for testing $H_0 : \beta_{\texttt{crim}} = 0 \text{ vs } H_1 : \beta_{\texttt{crim}} \neq 0$.

```
# Solution
library(MASS)
boston_mod = lm(medv ~ lstat + rm + crim + tax + nox, data = Boston)
summary(boston_mod)$coefficients["crim", "Pr(>|t|)"]
```

```
## [1] 0.09102573
```

Variation B

```
# Preamble
```

```
# Starter
library(MASS)
```

For this question use the `Boston` dataset from the `MASS` package.

Consider the multiple regression model

$$Y = \beta_0 + \beta_{\texttt{lstat}} x_{\texttt{lstat}} + \beta_{\texttt{rm}} x_{\texttt{rm}} + \beta_{\texttt{crim}} x_{\texttt{crim}} + \beta_{\texttt{tax}} x_{\texttt{tax}} + \beta_{\texttt{nox}} x_{\texttt{nox}} + \epsilon$$

with the usual assumptions on the error term and Y is `medv` .

Report the p-value for testing $H_0 : \beta_{\texttt{tax}} = 0$ vs $H_1 : \beta_{\texttt{tax}} \neq 0$.

```
# Solution
library(MASS)
boston_mod = lm(medv ~ lstat + rm + crim + tax + nox, data = Boston)
summary(boston_mod)$coefficients["tax", "Pr(>|t|)"]
```

```
## [1] 0.003344395
```

Variation C

```
# Preamble
```

```
# Starter
library(MASS)
```

For this question use the `Boston` dataset from the `MASS` package.

Consider the multiple regression model

$$Y = \beta_0 + \beta_{\texttt{lstat}} x_{\texttt{lstat}} + \beta_{\texttt{rm}} x_{\texttt{rm}} + \beta_{\texttt{crim}} x_{\texttt{crim}} + \beta_{\texttt{tax}} x_{\texttt{tax}} + \beta_{\texttt{nox}} x_{\texttt{nox}} + \epsilon$$

with the usual assumptions on the error term and Y is `medv` .

Report the p-value for testing $H_0 : \beta_{\texttt{nox}} = 0$ vs $H_1 : \beta_{\texttt{nox}} \neq 0$.

```
# Solution
library(MASS)
boston_mod = lm(medv ~ lstat + rm + crim + tax + nox, data = Boston)
summary(boston_mod)$coefficients["nox", "Pr(>|t|)"]
```

```
## [1] 0.1724987
```

Question 15

Variation A

```
# Preamble
```

```
# Starter
library(MASS)
```

For this question use the `Boston` dataset from the `MASS` package. Use `medv` as the response variable.

Compare the model

$$Y = \beta_0 + \beta_{\texttt{lstat}} x_{\texttt{lstat}} + \beta_{\texttt{rm}} x_{\texttt{rm}} + \beta_{\texttt{crim}} x_{\texttt{crim}} + \beta_{\texttt{tax}} x_{\texttt{tax}} + \beta_{\texttt{nox}} x_{\texttt{nox}} + \epsilon$$

to the model

$$Y = \beta_0 + \beta_{\texttt{lstat}} x_{\texttt{lstat}} + \beta_{\texttt{rm}} x_{\texttt{rm}} + \epsilon$$

using an F test with an α of 0.05. Report the RSS of the model you prefer.

```
# Solution
library(MASS)
full_mod = lm(medv ~ lstat + rm + crim + tax + nox, data = Boston)
null_mod = lm(medv ~ lstat + rm, data = Boston)
# anova(null_mod, full_mod)
sum(resid(full_mod) ^ 2)
```

```
## [1] 14869.32
```

Variation B

```
# Preamble
```

```
# Starter
library(MASS)
```

For this question use the `Boston` dataset from the `MASS` package. Use `medv` as the response variable.

Compare the model

$$Y = \beta_0 + \beta_{\texttt{rm}} x_{\texttt{rm}} + \beta_{\texttt{crim}} x_{\texttt{crim}} + \beta_{\texttt{tax}} x_{\texttt{tax}} + \beta_{\texttt{nox}} x_{\texttt{nox}} + \epsilon$$

to the model

$$Y = \beta_0 + \beta_{\texttt{rm}} x_{\texttt{rm}} + \beta_{\texttt{nox}} x_{\texttt{nox}} + \epsilon$$

using an F test with an α of 0.05. Report the RSS of the model you prefer.

```
# Solution
library(MASS)
full_mod = lm(medv ~ rm + crim + tax + nox, data = Boston)
null_mod = lm(medv ~ rm + nox, data = Boston)
# anova(null_mod, full_mod)
sum(resid(full_mod) ^ 2)
```

```
## [1] 18134.23
```

Variation C

```
# Preamble
```

```
# Starter
library(MASS)
```

For this question use the `Boston` dataset from the `MASS` package. Use `medv` as the response variable.

Compare the model

$$Y = \beta_0 + \beta_{\texttt{lstat}} x_{\texttt{lstat}} + \beta_{\texttt{crim}} x_{\texttt{crim}} + \beta_{\texttt{tax}} x_{\texttt{tax}} + \beta_{\texttt{nox}} x_{\texttt{nox}} + \epsilon$$

to the model

$$Y = \beta_0 + \beta_{\texttt{lstat}} x_{\texttt{lstat}} + \beta_{\texttt{tax}} x_{\texttt{tax}} + \epsilon$$

using an F test with an α of 0.05. Report the RSS of the model you prefer.

```
# Solution
library(MASS)
full_mod = lm(medv ~ lstat + crim + tax + nox, data = Boston)
null_mod = lm(medv ~ lstat + tax, data = Boston)
# anova(null_mod, full_mod)
sum(resid(null_mod) ^ 2)
```

```
## [1] 19197.98
```