

# HW 04: Multiple linear regression

Jwalin Patel

03/24/2021

```
library(tidyverse)
library(broom)
library(knitr)
```

```
sitting <- read.csv("data/sitting.csv")
```

## Part 1

### Question 1

We will use SLR to model MET and sitting as predictor and response variables, respectively.

*Simple Linear Regression Model :  $sitting = \beta_0 + \beta_1 MET$*

```
m1 <- lm(sitting ~ MET, data = sitting)
m1 %>%
  tidy(conf.int = TRUE) %>%
  kable(digits = 3, caption = "Prediction of the Reported hours per day spent sitting for subjects given reported metabolic equivalent unit minutes per week")
```

Table 1: Prediction of the Reported hours per day spent sitting for subjects given reported metabolic equivalent unit minutes per week

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	7.502	0.915	8.203	0.000	5.641	9.362
MET	0.000	0.000	-0.421	0.676	-0.001	0.001

Now we calculate  $R^2$  to consider how well the model fits the relationship between reported hours per day spent sitting and the reported metabolic equivalent unit minutes per week.

```
anova(m1) %>%
  kable(digits = 3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
MET	1	2.008	2.008	0.177	0.676
Residuals	33	373.592	11.321	NA	NA

We calculate  $R^2$  using the formula:

$$R^2 = \frac{SS_{Model}}{SS_{Total}}$$

```
2.008/(2.008+373.592)
```

```
## [1] 0.005346113
```

Thus,  $R^2 = 0.005346113$ . This means only around 0.53 percent of the variation in the reported hours per day spent sitting is explained by the reported metabolic equivalent unit minutes per week. Since the model only considers one predictor variable to explain variation in sitting hours, this is not a good representation of sitting hours.

## Part 2

### Question 2

```
m2 <- lm(MTL ~ sitting, data = sitting)
m2 %>%
  tidy(conf.int = TRUE) %>%
  kable(digits = 3, caption = "Prediction of the Medial temporal lobe thickness in mm for subjects given Reported hours per day spent sitting")
```

Table 3: Prediction of the Medial temporal lobe thickness in mm  
for subjects given Reported hours per day spent sitting

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	2.700	0.073	36.933	0.000	2.551	2.848
sitting	-0.023	0.009	-2.476	0.019	-0.042	-0.004

The coefficient is estimated as -0.023, with expected value ranging from lower bound of -0.042 and upper bound of -0.004. This means with every 1 hour increase in Reported hours per day spent sitting for subjects, we can expect the Medial temporal lobe thickness to reduce by 0.004 to 0.042 mm, and reduce by 0.023 on average.

...