C1M2_peer_reviewed_2021_06_16_00_41_42

June 26, 2021

1 Module 2: Peer Reviewed Assignment

1.0.1 Outline:

The objectives for this assignment:

1. Mathematically derive the values of $\hat{\beta}_0$ and $\hat{\beta}_1$
2. Enhance our skills with linear regression modeling.
3. Learn the uses and limitations of RSS, ESS, TSS and R^2.
4. Analyze and interpret nonidentifiability.

General tips:

1. Read the questions carefully to understand what is being asked.
2. This work will be reviewed by another human, so make sure that you are clear and concise in what your explanations and answers.

[2]:
```
# Load Required Packages
library(RCurl) #a package that includes the function getURL(), which allows for
↪reading data from github.
library(ggplot2)
library(tidyverse)
```

```
  Attaching packages                          tidyverse
1.3.0

  tibble  3.0.1      dplyr   0.8.5
  tidyr   1.0.2      stringr 1.4.0
  readr   1.3.1      forcats 0.5.0
  purrr   0.3.4

  Conflicts
tidyverse_conflicts()
  tidyr::complete() masks
RCurl::complete()
  dplyr::filter()   masks
stats::filter()
  dplyr::lag()      masks stats::lag()
```

1.1 Problem 1: Maximum Likelihood Estimates (MLEs)

Consider the simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ for $i = 1, ..., n, \quad \varepsilon_i \sim N(0, \sigma^2)$. In the videos, we showed that the least squares estimator in matrix-vector form is $\widehat{\beta} = (\beta_0, \beta_1)^T = (X^T X)^{-1} X^T \mathbf{Y}$. In this problem, you will derive the least squares estimators for simple linear regression without (explicitly) using linear algebra.

Least squares requires that we minimize

$$f(\mathbf{x}; \beta_0, \beta_1) = \sum_{i=1}^{n} \Big(Y_i - [\beta_0 + \beta_1 x_i] \Big)^2$$

over β_0 and β_1.

1. (a) Taking Derivatives Find the partial derivative of $f(\mathbf{x}; \beta_0, \beta_1)$ with respect to β_0, and the partial derivative of $f(\mathbf{x}; \beta_0, \beta_1)$ with respect to β_1. Recall that the partial derivative with respect to x of a multivariate function $h(x, y)$ is calculated by taking the derivative of h with respect to x while treating y constant.

Part I

Original:

$$f(\mathbf{x}; \beta_0, \beta_1) = \sum_{i=1}^{n} \Big(Y_i - [\beta_0 + \beta_1 x_i] \Big)^2$$

Chain Rule:

$$\frac{\partial f}{\partial \beta_0} = \sum_{i=1}^{n} 2 * \Big(Y_i - [\beta_0 + \beta_1 x_i] \Big) * -1$$

Simplify:

$$\frac{\partial f}{\partial \beta_0} = -2 \sum_{i=1}^{n} \Big(Y_i - [\beta_0 + \beta_1 x_i] \Big)$$

$$\frac{\partial f}{\partial \beta_0} = -2 \sum_{i=1}^{n} Y_i + 2n\beta_0 + 2\beta_1 \sum_{i=1}^{n} x_i$$

We don't know the individual values of x_i, but we know that the *sum* of the individual x terms divided by the *total* number of x terms is $\bar{x}$. Likewise, we can use the same technique with Y_i and $\bar{Y}$:

$$\frac{\partial f}{\partial \beta_0} = -2n\bar{Y} + 2n\beta_0 + 2n\beta_1 \bar{x}$$

Part II

Original:

$$f(\mathbf{x}; \beta_0, \beta_1) = \sum_{i=1}^{n} \Big(Y_i - [\beta_0 + \beta_1 x_i] \Big)^2$$

Chain Rule:

$$\frac{\partial f}{\partial \beta_1} = \sum_{i=1}^{n} 2 * \Big(Y_i - [\beta_0 + \beta_1 x_i]\Big) * -x_i$$

Simplify:

$$\frac{\partial f}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i Y_i + 2\beta_0 \sum_{i=1}^{n} x_i + 2\beta_1 \sum_{i=1}^{n} (x_i)^2$$

1. (b) Solving for $\hat{\beta}_0$ and $\hat{\beta}_1$ Use **1. (a)** to find the minimizers, $\widehat{\beta}_0$ and $\widehat{\beta}_1$, of f. That is, set each partial derivative to zero and solve for β_0 and β_1. In particular, show

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{and} \qquad \widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{x}$$

Solving for $\hat{\beta}_0$:

$$\frac{\partial f}{\partial \beta_0} = -2n\bar{Y} + 2n\beta_0 + 2n\beta_1\bar{x} = 0$$

$$-2n\hat{\beta}_0 = -2n\bar{Y} + 2n\hat{\beta}_1\bar{x}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x}$$

Solving for $\hat{\beta}_1$:

$$\frac{\partial f}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i Y_i + 2\beta_0 \sum_{i=1}^{n} x_i + 2\beta_1 \sum_{i=1}^{n} x_i^2 = 0$$

$$-2\sum_{i=1}^{n} x_i Y_i + 2\hat{\beta}_0 \sum_{i=1}^{n} x_i + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

Substitute for $\hat{\beta}_0$:

$$-2\sum_{i=1}^{n} x_i Y_i + 2(\bar{Y} - \hat{\beta}_1\bar{x}) \sum_{i=1}^{n} x_i + 2\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

Distribute and divide by 2:

$$-\sum_{i=1}^{n} x_i Y_i + \bar{Y}\sum_{i=1}^{n} x_i - \hat{\beta}_1\bar{x}\sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0$$

Arrange like terms:

$$-\sum_{i=1}^{n} x_i Y_i + \bar{Y}\sum_{i=1}^{n} x_i - \hat{\beta}_1\bar{x}\sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} (x_i)^2 = 0$$

$$\bar{Y}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i Y_i = \hat{\beta}_1(\bar{x}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2)$$

Isolate:

$$\hat{\beta}_1 = \frac{\bar{Y}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i Y_i}{\bar{x}\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^2}$$

$$\hat{\beta}_1 = \frac{n\bar{Y}\bar{x} - \sum_{i=1}^n x_i Y_i}{n\bar{x}^2 - \sum_{i=1}^n x_i^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i Y_i - n\bar{Y}\bar{x}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i Y_i - 2n\bar{Y}\bar{x} + n\bar{Y}\bar{x}}{\sum_{i=1}^n x_i^2 - 2n\bar{x}^2 + n\bar{x}^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i Y_i - x_i\bar{Y} - \bar{x}Y_i + \bar{x}\bar{Y})}{\sum_{i=1}^n x_i^2 - \sum_{i=1}^n 2\bar{x}x_i + \sum_{i=1}^n \bar{x}^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

1.2 Problem 2: Oh My Goodness of Fit!

In the US, public schools have been slowly increasing class sizes over the last 15 years [https://stats.oecd.org/Index.aspx?DataSetCode=EDU_CLASS]. The general cause for this is because it saves money to have more kids per teacher. But how much money does it save? Let's use some of our new regression skills to try and figure this out. Below is an explanation of the variables in the dataset.

Variables/Columns:
School
Per-Pupil Cost (Dollars)
Average daily Attendance
Average Monthly Teacher Salary (Dollars)
Percent Attendance
Pupil/Teacher ratio

Data Source: E.R. Enlow (1938). "Do Small Schools Mean Large Costs?," Peabody Journal of Educaltion, Vol. 16, #1, pp. 1-11

[3]:
```
school.data = read_table("school.dat", col_names = FALSE)
names(school.data) = c("school", "cost", "avg.attendance", "avg.salary", "pct.
↪attendance", "pup.tch.ratio")
head(school.data)
dim(school.data)
```

```
Parsed with column specification:
cols(
  X1 = col_character(),
  X2 = col_double(),
  X3 = col_double(),
  X4 = col_double(),
  X5 = col_double(),
  X6 = col_double()
)
```

A tibble: 6 × 6

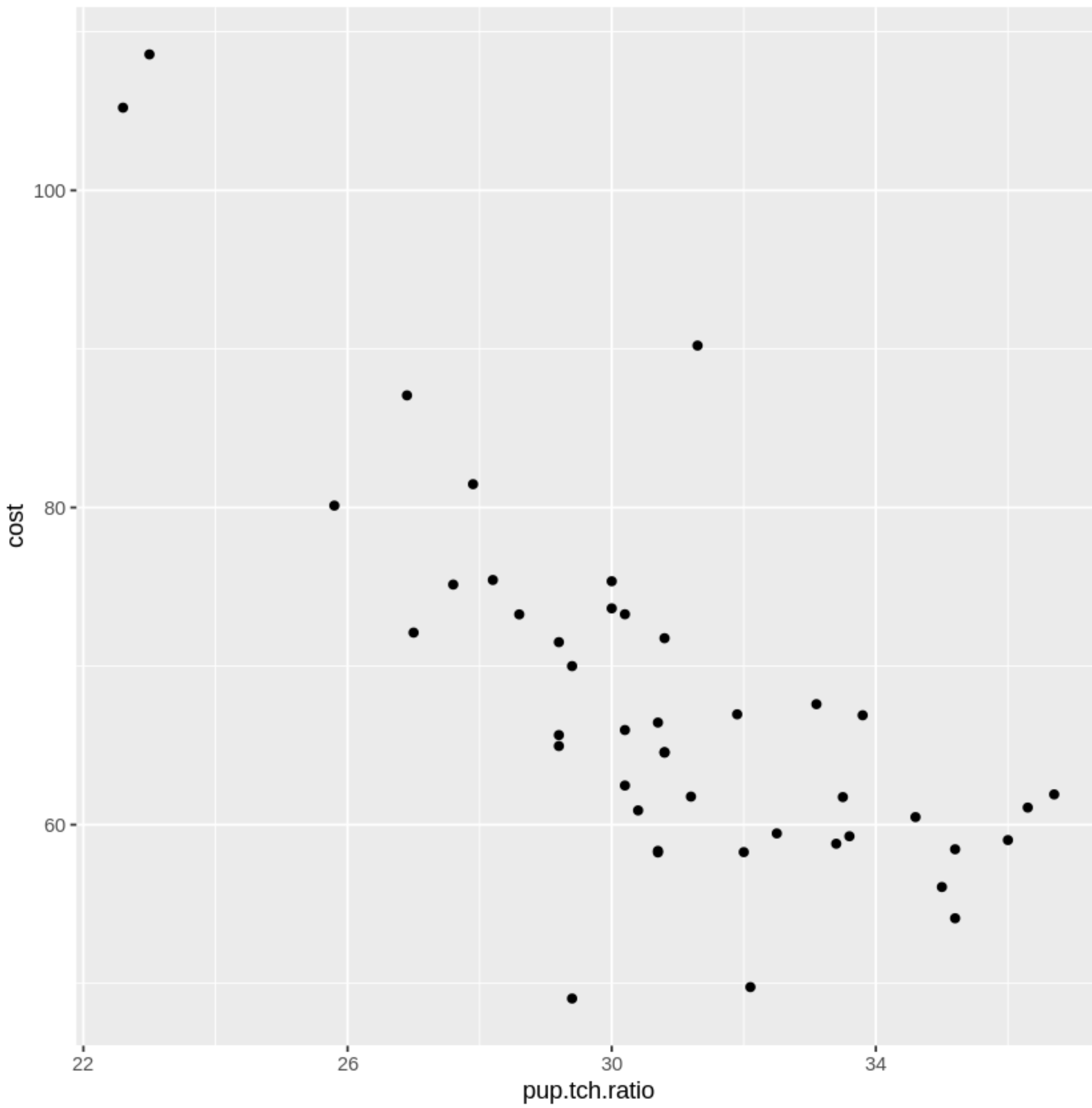
school	cost	avg.attendance	avg.salary	pct.attendance	pup.tch.ratio
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Adair	66.90	451.4	160.22	90.77	33.8
Calhoun	108.57	219.1	161.79	89.86	23.0
Capitol View	70.00	268.9	136.37	92.44	29.4
Connally	49.04	161.7	106.86	92.01	29.4
Couch	71.51	422.1	147.17	91.60	29.2
Crew	61.08	440.6	146.24	89.32	36.3

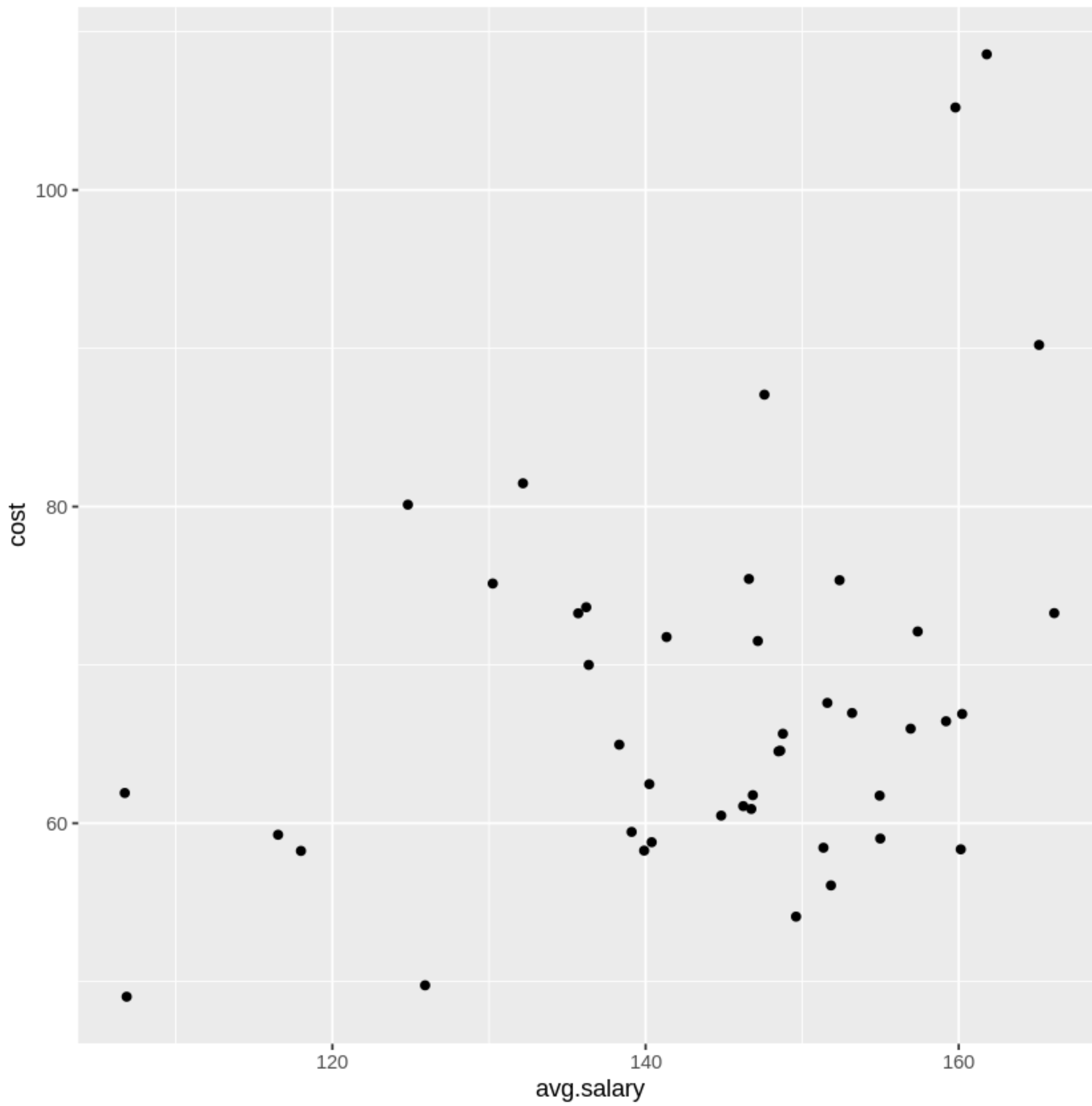
1. 44 2. 6

2. (a) Create a model Begin by creating two figures for your model. The first with `pup.tch.ratio` on the x-axis and `cost` on the y-axis. The second with `avg.salary` on the x-axis and `cost` on the y-axis. Does there appear to be a relation between these two predictors and the response.

Then fit a multiple linear regression model with `cost` as the response and `pup.tch.ratio` and `avg.salary` as predictors.

[4]:
```
fig_1 <- ggplot(school.data, aes(x=pup.tch.ratio, y = cost)) +
geom_point()
fig_1

fig_2 <- ggplot(school.data, aes(x = avg.salary, y = cost)) +
geom_point()
fig_2

mlr_school <- lm(cost ~ pup.tch.ratio + avg.salary, data = school.data)
```


100
80
60
cost
22
26
30
34
pup.tch.ratio



There appears to be a weakly positive linear relationship between the average salary and the cost. There appears to be a moderately negative linear relationship between the pupil/teacher ratio and the cost.

2. (b) RSS, ESS and TSS In the code block below, manually calculate the RSS, ESS and TSS for your MLR model. Print the results.

[5]:
```
RSS <- sum(resid(mlr_school)^2)
RSS

ESS <- sum((fitted(mlr_school) - mean(school.data$cost))^2)
ESS
```

```
TSS <- with(school.data, sum((cost - mean(cost))^2))
TSS
```

2396.7384822516

4177.41819729385

6574.15667954545

2. (c) Are you Squared? Using the values from **2.b**, calculate the R^2 value for your model. Check your results with those produced from the `summary()` statement of your model.

In words, describe what this value means for your model.

[6]:
```
r_squared <- 1 - (RSS / TSS)
r_squared

summary(mlr_school)
```

0.635430276599779

```
Call:
lm(formula = cost ~ pup.tch.ratio + avg.salary, data = school.data)

Residuals:
     Min       1Q   Median       3Q      Max 
-13.8538  -5.3484  -0.6884   3.5671  19.7207 

Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)   118.44679   17.11396   6.921 2.13e-08 ***
pup.tch.ratio  -2.79829    0.36853  -7.593 2.43e-09 ***
avg.salary      0.24770    0.08168   3.033  0.00419 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.646 on 41 degrees of freedom
Multiple R-squared:  0.6354,Adjusted R-squared:  0.6176 
F-statistic: 35.73 on 2 and 41 DF,  p-value: 1.039e-09
```

The R^2 value is the proportion of the variance in the dependent variable that is explained by the independent variables. In the case of our multiple linear regression model, ~63.54% of the variance in the response (school's cost per pupil) is explained by the model predictors (pupil to teacher ratio and average teacher salary).

2. (d) Conclusions Describe at least two advantages and two disadvantages of the R^2 value.

Two advantages of the coefficient of determination are that the interpretation of the statistic is straightforward and that the value gives an indication of the goodness of fit of the model.

Two of the disadvantages of the coefficient of determination are that adding more predictors to the model will always cause the R^2 value to stay the same or increase, regardless of the impact of the additional predictors and that the R^2 statistic says nothing about a causal relationship between the predictors and the response.

2 Problem 3: Identifiability

This problem might require some outside-of-class research if you haven't taken a linear algebra/matrix methods course.

Matrices and vectors play an important role in linear regression. Let's review some matrix theory as it might relate to linear regression.

Consider the system of linear equations

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{i,j} + \varepsilon_i, \tag{1}$$

for $i = 1, ..., n$, where n is the number of data points (measurements in the sample), and $j = 1, ..., p$, where

1. $p + 1$ is the number of parameters in the model.
2. Y_i is the i^{th} measurement of the *response variable*.
3. $x_{i,j}$ is the i^{th} measurement of the j^{th} *predictor variable*.
4. ε_i is the i^{th} *error term* and is a random variable, often assumed to be $N(0, \sigma^2)$.
5. β_j, $j = 0, ..., p$ are *unknown parameters* of the model. We hope to estimate these, which would help us characterize the relationship between the predictors and response.

3. (a) MLR Matrix Form Write the equation above in matrix vector form. Call the matrix including the predictors X, the vector of Y_is $\mathbf{Y}$, the vector of parameters β, and the vector of error terms ε. (This is more LaTeX practice than anything else...)**

Short form:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

Long form:

$$(Y_1, Y_2, \cdots, Y_n)^T = \begin{pmatrix} 1 & x_{11} & ... & x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & ... & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + (\epsilon_1, \epsilon_2, \cdots, \epsilon_n)^T$$

3. (b) Properties of this matrix In lecture, we will find that the OLS estimator for β in MLR is $\widehat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \mathbf{Y}$. Use this knowledge to answer the following questions:

1. What condition must be true about the columns of X for the "Gram" matrix $X^T X$ to be invertible?
2. What does this condition mean in practical terms, i.e., does X contain a deficiency or redundancy?
3. Suppose that the number of measurements (n) is less than the number of model parameters ($p+1$). What does this say about the invertibility of $X^T X$? What does this mean on a practical level?
4. What is true about about $\widehat{\boldsymbol{\beta}}$ if $X^T X$ is not invertible?

1. The column vectors of the design matrix $\mathbf{X}$ must be linearly independent for the Gram matrix to have an inverse.
2. If all of the column vectors are linearly independent, then there are no redundancies between the column vectors and the matrix will have a full rank. Linearly dependence would lead to a rank deficiency and the model would be non-identifiable.
3. The amount of available information is deficient if n < p + 1, which means that the model parameters cannot be uniquely estimated. This is an example of an underdetermined linear system, which generally has either no solution or infinitely many solutions. In this case, there will be infinitely many solutions. Since there is no unique solution, the Gram matrix will not be invertible.
4. If the Gram Matrix $X^T X$ is not invertible, then there is no unique solution for $\widehat{\boldsymbol{\beta}}$. There will be infinitely many solutions instead.

2.1 Problem 4: Downloading...

The following data were collected to see if time of day madea difference on file download speed. A researcher placed a file on a remote server and then proceeded to download it at three different time periods of the day. They downloaded the file 48 times in all, 16 times at each Time of Day (`time`), and recorded the Time in seconds (`speed`) that the download took.

4. (a) Initial Observations The `downloading` data is loaded in and cleaned for you. Using `ggplot`, create a boxplot of `speed` vs. `time`. Make some basic observations about the three categories.

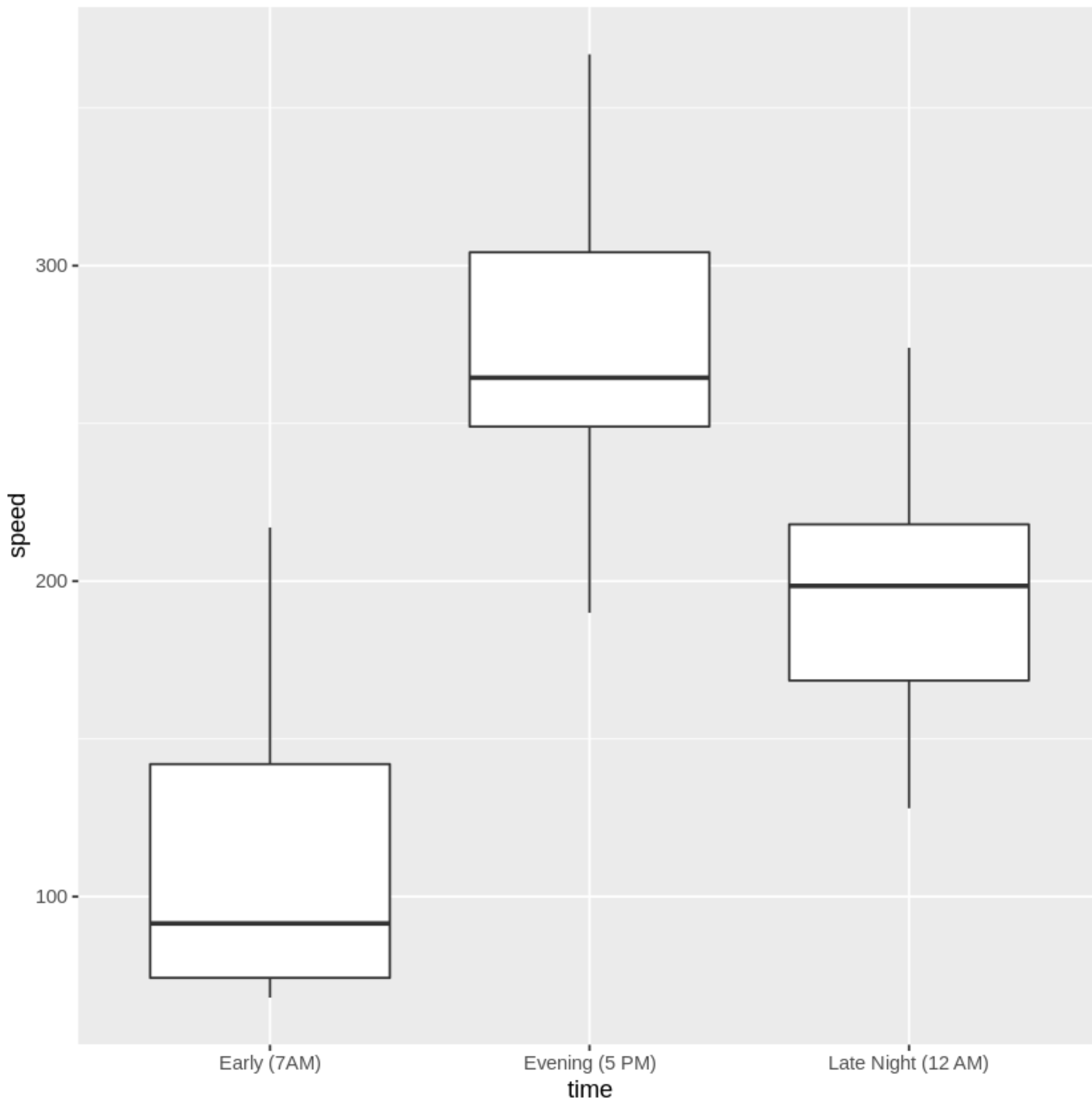
[7]:
```
# Load in the data and format it
downloading = read.csv("downloading.txt", sep="\t")
names(downloading) = c("time", "speed")
# Change the types of brand and form to categories, instead of real numbers
downloading$time = as.factor(downloading$time)
summary(downloading)
tapply(downloading$speed, downloading$time, summary)
```

```
                 time        speed
Early (7AM)       :16   Min.   : 68.0
Evening (5 PM)    :16   1st Qu.:129.8
```

```
 Late Night (12 AM):16   Median :198.0
                         Mean   :193.2
                         3rd Qu.:253.0
                         Max.   :367.0


$`Early (7AM)`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  68.00   74.25   91.50  113.38  142.00  217.00

$`Evening (5 PM)`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  190.0   249.0   264.5   273.3   304.2   367.0

$`Late Night (12 AM)`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  128.0   168.5   198.5   193.1   218.0   274.0
```

[8]:
```
summary(lm(speed ~ time, data = downloading))

ggplot(downloading, aes(x = time, y = speed)) +
geom_boxplot()
```

```
Call:
lm(formula = speed ~ time, data = downloading)

Residuals:
    Min      1Q  Median      3Q     Max
-83.312 -34.328  -5.187  26.250 103.625

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)              113.37      11.79   9.619 1.73e-12 ***
timeEvening (5 PM)       159.94      16.67   9.595 1.87e-12 ***
timeLate Night (12 AM)    79.69      16.67   4.781 1.90e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 47.15 on 45 degrees of freedom
Multiple R-squared:  0.6717,Adjusted R-squared:  0.6571
F-statistic: 46.03 on 2 and 45 DF,  p-value: 1.306e-11
```



Download speeds in the early morning were the fastest by far, with the slowest times not even reaching the median of the evening download times. The evening download speeds were significantly slower than the other two categories' speeds, and the late night download speeds were at a pace inbetween the other two categories. None of the categories appear to have any outliers, as no points extend beyond the whiskers of the boxplot.

4. (b) How would we model this? Fit a regression to these data that uses `speed` as the response and `time` as the predictor. Print the summary. Notice that the result is actually *multiple* linear regression, not simple linear regression. The model being used here is:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \varepsilon_i$$

where

1. $X_{i,1} = 1$ if the i^{th} download is made in the evening (5 pm).
2. $X_{i,2} = 1$ if the i^{th} download is made at night (12 am).

Note: If $X_{i,1} = 0$ and $X_{i,2} = 0$, then the i^{th} download is made in the morning (7am).

To confirm this is the model being used, write out the explicit equation for your model - using the parameter estimates from part (a) - and print out it's design matrix.

[9]:
```
download_lm <- lm(speed ~ time, data = downloading)
download_lm
summary(download_lm)

#cat("Explicit equation: y_i = ", download_lm$coefficients[1], " + ",␣
↪download_lm$coefficients[2], "* x_i_1 + error_i")
model.matrix(download_lm)
```

```
Call:
lm(formula = speed ~ time, data = downloading)

Coefficients:
          (Intercept)      timeEvening (5 PM)  timeLate Night (12 AM)
               113.37                  159.94                   79.69



Call:
lm(formula = speed ~ time, data = downloading)

Residuals:
    Min      1Q  Median      3Q     Max
-83.312 -34.328  -5.187  26.250 103.625

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)              113.37      11.79   9.619 1.73e-12 ***
timeEvening (5 PM)       159.94      16.67   9.595 1.87e-12 ***
timeLate Night (12 AM)    79.69      16.67   4.781 1.90e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 47.15 on 45 degrees of freedom
Multiple R-squared:  0.6717,Adjusted R-squared:  0.6571
F-statistic: 46.03 on 2 and 45 DF,  p-value: 1.306e-11
```

A matrix: 48 × 3 of type dbl

	(Intercept)	timeEvening (5 PM)	timeLate Night (12 AM)
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
5	1	0	0
6	1	0	0
7	1	0	0
8	1	0	0
9	1	0	0
10	1	0	0
11	1	0	0
12	1	0	0
13	1	0	0
14	1	0	0
15	1	0	0
16	1	0	0
17	1	1	0
18	1	1	0
19	1	1	0
20	1	1	0
21	1	1	0
22	1	1	0
23	1	1	0
24	1	1	0
25	1	1	0
26	1	1	0
27	1	1	0
28	1	1	0
29	1	1	0
30	1	1	0
31	1	1	0
32	1	1	0
33	1	0	1
34	1	0	1
35	1	0	1
36	1	0	1
37	1	0	1
38	1	0	1
39	1	0	1
40	1	0	1
41	1	0	1
42	1	0	1
43	1	0	1
44	1	0	1
45	1	0	1
46	1	0	1
47	1	0	1
48	1	0	1

Actual equation: $Y_i = 113.37 + 159.94X_{i,1} + 79.69X_{i,2} + \varepsilon_i$

4. (c) Only two predictors? We have three categories, but only two predictors. Why is this the case? To address this question, let's consider the following model:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_2 X_{i,3} + \varepsilon_i$$

where

1. $X_{i,1} = 1$ if the i^{th} download is made in the evening (5 pm).
2. $X_{i,2} = 1$ if the i^{th} download is made at night (12 am).
3. $X_{i,3} = 1$ if the i^{th} download is made in the morning (7 am).

Construct a design matrix to fit this model to the response, `speed`. Determine if something is wrong with it. Hint: Analyze the design matrix.

[12]:
```
evening <- downloading$time == "Evening (5 PM)"
night <- downloading$time == "Late Night (12 AM)"
morning <- downloading$time == "Early (7AM)"

design_matrix <- model.matrix(~evening + night + morning)
colnames(design_matrix) <- c('Intercept', 'Evening', 'Night', 'Morning')
design_matrix
```

A matrix: 48 × 4 of type dbl

	Intercept	Evening	Night	Morning
1	1	0	0	1
2	1	0	0	1
3	1	0	0	1
4	1	0	0	1
5	1	0	0	1
6	1	0	0	1
7	1	0	0	1
8	1	0	0	1
9	1	0	0	1
10	1	0	0	1
11	1	0	0	1
12	1	0	0	1
13	1	0	0	1
14	1	0	0	1
15	1	0	0	1
16	1	0	0	1
17	1	1	0	0
18	1	1	0	0
19	1	1	0	0
20	1	1	0	0
21	1	1	0	0
22	1	1	0	0
23	1	1	0	0
24	1	1	0	0
25	1	1	0	0
26	1	1	0	0
27	1	1	0	0
28	1	1	0	0
29	1	1	0	0
30	1	1	0	0
31	1	1	0	0
32	1	1	0	0
33	1	0	1	0
34	1	0	1	0
35	1	0	1	0
36	1	0	1	0
37	1	0	1	0
38	1	0	1	0
39	1	0	1	0
40	1	0	1	0
41	1	0	1	0
42	1	0	1	0
43	1	0	1	0
44	1	0	1	0
45	1	0	1	0
46	1	0	1	0
47	1	0	1	0
48	1	0	1	0

An issue that is immediately apparent with the model is that one and only one of the predictors can have a value of 1 at any time, while the rest must be 0. This means that β_0 can never be isolated by setting all the X values to 0, so there is no interpretabilty of the intercept of the model. The downloads can only occur at three specific times for this dataset, and the intercept with all predictors set to 0 would basically mean that the download occurred at none of the time categories, which is nonsensical.

4. (d) Interpretation Interpret the coefficients in the model from **4.b**. In particular:

1. What is the difference between the mean download speed at 7am and the mean download speed at 5pm?
2. What is the mean download speed (in seconds) in the morning?
3. What is the mean download speed (in seconds) in the evening?
4. What is the mean download speed (in seconds) at night?

1. β_0 - (β_0 + β_1) = -β_1 = -159.94
2. β_0 = 113.37
3. β_0 + β_1 = 273.3
4. β_0 + β_2 = 193.1