Math 300 Lesson 15 Notes

Multiple Regression - Two Numerical

YOUR NAME HERE

June, 2022

Contents

Objectives

1. For two numerical explanatory variables in a linear regression model, conduct exploratory analysis and explain the relationship between the variables.
2. Fit a linear regression model to two numerical explanatory variables using the `lm()` function and interpret the output.

Reading

Chapter 6.2

Lesson

Work through the learning checks LC6.2 - 6.3. Complete the code as necessary.

- We use `select()` to change the names of the variables as well as to select them.
- Collinearity (or multicollinearity) is a phenomenon where one explanatory variable in a multiple regression model is highly correlated with another. This course doesn't discuss multicollinearity but it impacts the inference portion of the analysis cycle. Math 378 is a course that presents methods to handle multicollinearity.
- We preface our interpretation with the statement, "taking into account all the other explanatory variables in our model" in this section. This means we have to treat the other variables as at a constant value even though collinearity in practice may not allow this. It is only from an interpretation point of view that we use that statement.
- A phenomenon known as Simpson's Paradox, whereby overall trends that exist in aggregate either disappear or reverse when the data are broken down into groups. The next lesson discusses this in more depth.

Setup

```
library(tidyverse)
library(moderndive)
library(skimr)
library(ISLR)
```

Recreate the analysis done in the book.

```
my_skim <- skim_with(numeric = sfl(hist = NULL))
```

```
# Complete code and remove comment symbol
#credit_ch6 <- Credit %>% as_tibble() %>%
#  select(ID, debt = ________, credit_limit = Limit,
#         income = _______, credit_rating = ____________, age = Age)
```

Let's look at 5 random rows of data.

```
# Complete code and remove comment symbol
#set.seed(507)
#credit_ch6 %>%
#  sample_n(size = __________) %>%
#  print()
```

```
# Complete code and remove comment symbol
#glimpse(___________)
```

```
# Complete code and remove comment symbol
#credit_ch6 %>%
#  select(debt, ___________, income) %>%
#  my_skim() %>%
#  print()
```

```
# Complete code and remove comment symbol
#credit_ch6 %>%
#  select(debt, credit_limit, income) %>%
#  _______() %>%
#  print()
```

LC 6.2 (Objective 1)

(LC 6.2) Conduct a new exploratory data analysis with the same outcome variable y being `debt` but with `credit_rating` and `age` as the new explanatory variables x_1 and x_2. Remember, this involves three things:

- Most crucially: Looking at the raw data values.
- Computing summary statistics, such as means, medians, and interquartile ranges.
- Creating data visualizations.

What can you say about the relationship between a credit card holder's debt and their credit rating and age?

Solution:

- Most crucially: Looking at the raw data values.

Computing summary statistics, such as means, medians, and interquartile ranges.

- Creating data visualizations.

```
# Complete code and remove comment symbol
# ggplot(credit_ch6, aes(x = _____________, y = debt)) +
#   geom_point() +
#   labs(
#     x = "Credit rating", y = "Credit card debt (in $)",
#     title = "Debt and credit rating"
#   ) +
#   geom_smooth(method = "lm", se = ________)
```

```
# Repeat for age and debt
```

It seems that there is a positive relationship between one's credit rating and their debt, and very little relationship between one's age and their debt. The is a slight linear relationship between `age` and `credit_rating`.

LC 6.3 (Objective 2)

(LC 6.3) Fit a new simple linear regression using `lm(debt ~ credit_rating + age, data = credit_ch6)` where `credit_rating` and `age` are the new numerical explanatory variables x_1 and x_2. Get information about the "best-fitting" regression plane from the regression table by applying the `get_regression_table()` function. How do the regression results match up with the results from your previous exploratory data analysis?

```
# Complete code and remove comment symbol
# Fit regression model:
# debt_model_2 <- lm(_________ ~ ___________ + _________, data = credit_ch6)
```

```
# Get regression table:
# Complete code and remove comment symbol
# print(get_regression_table(_________________))
```

Documenting software

- File creation date: 2022-06-04
- R version 4.1.3 (2022-03-10)
- `tidyverse` package version: 1.3.1
- `skimr` package version: 2.1.4
- `ISLR` package version: 1.4

- `moderndive` package version: 0.5.4