

Math 300 Lesson 14 Notes

Multiple Linear Regression - Numerical & Discrete

YOUR NAME HERE

June, 2022

Contents

Objectives	1
Reading	1
Lesson	1
Documenting software	3

Objectives

1. Generate, plot, and explain the interaction model for two explanatory variables (one numerical and one categorical).
2. Generate, plot, and explain the parallel slopes model for two explanatory variables (one numerical and one categorical).
3. Generate a table of observations, fitted values, and residuals from a linear regression object.

Reading

Chapter 6 - 6.1

Lesson

Work through the learning check LC6.1. Complete the code blocks as necessary.

- Notice that the regression line for females stops at the extremes of the observed data in `ggplot2()`. You have to be careful about extrapolating. The assumption of linearity outside of the observed data is risky at best.
- The use of `+` and `*` in an R formula are not arithmetical operations but formula operations.

Setup

```
library(tidyverse)
library(moderndiver)
library(skimr)
library(ISLR)
```

Recreate the analysis done in the book.

```
# Complete code and remove comment symbol
# evals_ch6 <- evals %>%
#   select(ID, _____, age, _____)
```

Let's look at 5 random rows of data.

```
# Complete code and remove comment symbol
set.seed(941)
#evals_ch6 %>%
#   sample_n(size = _____)
```

- Interaction Model (Objective 1)

In this model we allow a different slope and intercept for each gender.

```
# Complete code and remove comment symbol
#ggplot(_____, aes(x = age, y = score, color = _____)) +
#   geom_point() +
#   labs(x = "_____", y = "Teaching Score", color = "_____") +
#   geom_smooth(method = "lm", se = FALSE) +
#   theme_bw()
```

To get the model in R, we use the * which is not multiplication but an interaction term in the model formula.

```
# Complete code and remove comment symbol
# Fit regression model:
#score_model_interaction <- lm(_____ ~ age * gender, data = _____)
```

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

As a check, answer the following questions. Note from the reading: *We say there is an interaction effect if the associated effect of one variable depends on the value of another variable.*

- What is the intercept for the males?
- What is the slope for the males?
- Interpret the slope for the female instructors?

-
- Parallel Slopes Model (Objective 2)

The parallel slopes model assumes that there is no interaction between the two explanatory variables. Their impact on the response is not related to the values of the other variable.

We will use the same data, but just build a different model.

```
# Complete code and remove comment symbol
#ggplot(evals_ch6, aes(x = _____, y = score, color = _____)) +
# geom_point() +
# labs(x = "Age", y = "Teaching Score", color = "_____") +
# geom_parallel_slopes(se = _____) +
# theme_bw()
```

- Notice that the line for females stops at the extremes of the observed data.

```
# Complete code and remove comment symbol
# Fit regression model:
# score_model_parallel_slopes <- lm(score ~ _____ + _____, data = evals_ch6)
```

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

- What is the intercept for the males?
- What is the slope for the males?
- Interpret the slope for the female instructors?

LC 5.4 (Objective 3)

(LC 5.4) Compute the observed values, fitted values, and residuals not for the interaction model as we just did, but rather for the parallel slopes model we saved in `score_model_parallel_slopes`.

Solution:

```
# Complete code and remove comment symbol
# regression_points_parallel <- _____
```

```
# Complete code and remove comment symbol
#head(_____)
```

Documenting software

- File creation date: 2022-06-24
- 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