Math 300 Lesson 37 Notes

Regression Summary

YOUR NAME HERE

July, 2022

Contents

Objectives

1. Use the output of a regression table to conduct a hypothesis test and confidence interval of the slope parameter from a linear regression model.

Reading

Chapter 10.5

Lesson

There are no learning checks for this lesson.

- Notice that the mathematical normal-based inference calculations can be challenging. It is important to have some familiarity with them since they are still used by many researchers.
- Review material for GR4.

Libraries

```
library(tidyverse)
library(infer)
library(moderndive)
```

Normal-based inference

Let's get the regression table from the course evaluation data we have been working with.

Data and model

Let's get the data and model again.

```
evals_ch5 <- evals %>%
  select(ID, score, bty_avg, age)
glimpse(evals_ch5)
```

```
## Rows: 463
## Columns: 4
## $ ID      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,~
## $ score   <dbl> 4.7, 4.1, 3.9, 4.8, 4.6, 4.3, 2.8, 4.1, 3.4, 4.5, 3.8, 4.5, 4.~
## $ bty_avg <dbl> 5.000, 5.000, 5.000, 5.000, 3.000, 3.000, 3.000, 3.333, 3.333,~
## $ age     <int> 36, 36, 36, 36, 59, 59, 59, 51, 51, 40, 40, 40, 40, 40, 40, 40~
```

```
# Complete the code
#score_model <- lm(________ ~ _____________, data = evals_ch5)
```

And the regression summary.

```
# Complete the code
#get_regression_table(_____________)
```

Inference

The hypothesis test is

$$H_O : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

Under the null, there is no linear relationship between teaching scores and beauty scores.

The test statistic is

$$t = \frac{b_1 - \beta_1}{\sigma_{n-2}}$$

where b_1 is estimated slope, β_1 is the slope under the null hypothesis, σ is the standard error of the slope estimate, and $n - 2$ is the degrees of freedom, the number of data points, n, minus 2. In words, the test statistic measure the distance between observed and hypothesized values as measured in the number of standard deviations.

For our example, we have

```
# Complete the code
# get_regression_table(score_model) %>%
#   filter(term=="_____________") %>%
#   summarize(n=nrow(___________),
#           t_stat=(___________-0)/___________,
#           p_value=2*pt(___________,df=_____________,lower.tail = FALSE))
```

The regression table gives us this output without us having to perform the computations but it is important to have an idea of how it is calculated.

Notice that if we want to test the hypothesis where the null had a value other than zero. We could use the formula we just reviewed. As an example, suppose we wanted to test:

$$H_O : \beta_1 = 0.1$$

$$H_A : \beta_1 \neq 0.1$$

We just made up this number, there would have to be some science or subject matter expertise behind this choice. But for the purpose of learning, we will continue.

The code for the p-value is

```
# Complete the code
# get_regression_table(score_model) %>%
#   filter(term=="_____________") %>%
#   summarize(n=nrow(___________),
#           t_stat=(___________-0)/___________,
#           p_value=2*pt(___________,df=_____________,lower.tail = _____________))
```

In this example we would fail to reject the null hypothesis. Notice that since the test statistic is negative, we had to use the lower tail of the t distribution.

Confidence Interval

The use of the confidence interval makes the above analysis much easier. The regression table has the confidence interval, but here is the code to calculate it.

```
# Complete the code
# get_regression_table(score_model) %>%
#   filter(term=="________") %>%
#   summarize(n=nrow(evals_ch5),
#             lower_ci=_______-________*qt(_____,df=n-2),
#             upper_ci=_______+________*qt(_____,df=n-2))
```

This interval gives us a range of plausible values of the slope. We see that zero is not in the interval but 0.1 is.

What are the assumptions needed to use a normal-based inference method?

Documenting software

- File creation date: 2022-07-05
- R version 4.1.3 (2022-03-10)
- `tidyverse` package version: 1.3.1
- `moderndive` package version: 0.5.4
- `infer` package version: 1.0.0