ETC 2420/5242 Lab 7 2016

Di Cook

Week 7

Purpose

For this lab we are going to compute bootstrap confidence intervals for the parameters and fitted values for the multiple linear model on education constructed last week.

Reading

- Read the code in the lecture notes on computing bootstrap confidence intervals for linear models from Week 6.

The variables that were used for modeling `math` was:

Variable name	Description	Coding
ST04Q01	Gender	1=Female, 2=Male
ST06Q01	Age when started school	Actual age, 9997-9999 indicate missing values
ST15Q01	Mother Current Job Status	1=Full-time, 2=Part-time, 3=Not working, but looking for a job, 4=Other (inc stay-at-home), 7-9 indicate missing values
ST19Q01	Father Current Job Status	1=Full-time, 2=Part-time, 3=Not working, but looking for a job, 4=Other (inc stay-at-home), 7-9 indicate missing values
ST26Q01	Possessions - desk	1=Yes, 2=No, 7-9 indicate missing values
ST26Q04	Possessions - computer	1=Yes, 2=No, 7-9 indicate missing values
ST26Q06	Possessions - Internet	1=Yes, 2=No, 7-9 indicate missing values
ST27Q02	How many - televisions	1=None, 2=One, 3=Two, 4=Three or more, 7-9 indicate missing values
ST28Q01	How many books at home	1=0-10, 2=11-25, 3=26-100, 4=101-200, 5=201-500, 6=More than 500, 7-9 indicate missing values
SENWGT_STU	Weight	Reflects how the student represents other students in Australia based on socioeconomic and demographic characteristics

Model building will be done using:

- Response: `math` (standardised)
- Explanatory variables: `ST04Q01, ST06Q01, ST15Q01, ST19Q01, ST26Q01, ST26Q04, ST26Q06, ST27Q02, ST28Q01`.

Question 1

a. Compute and report the 95% confidence interval for the parameter for the number of books in the household (ST28Q01), using classical t-interval methods.
b. Use this to test the hypothesis that ST28Q01 is not important for the model.

Question 2

a. The boot package can generate bootstrap samples for weighted data. To use the `boot` function for drawing samples, you need a function to compute the statistic of interest. Write the function to return the slope for ST28Q01 after fitting a `glm` to a bootstrap sample. The skeleton of the function `calc_stat` is below, where `d` is the data, and `i` is the vector of indices of the bootstrap sample.

```
library(boot)
calc_stat <- function(d, i) {
  x <- d[i,]
  mod <- FILL IN THE NECESSARY CODE
  stat <- FILL IN THE NECESSARY CODE
  return(stat)
}
stat <- boot(aus_nomiss, statistic=calc_stat, R=1000,
     weights=aus_nomiss$SENWGT_STU)
stat
sort(stat$t)[5]
sort(stat$t)[95]
```

b. How does the bootstrap interval compare with the t-interval?

Question 3

Now make a 95% bootstrap confidence interval for predicted value for a new student who is FEMALE, started school at 4, mother and father both work full-time, has a desk, computer and internet, two TVs and 26-100 books in the home. The weight for a student like this is 0.1041. Be sure to convert the values back into the actual math score range.

Question 4

Compute a bootstrap 95% prediction interval for the same student as in the previous question. Be sure to convert the values back into the actual math score range.

TURN IN

- Your `.Rmd` file
- Your Word (or pdf) file that results from knitting the Rmd.
- Make sure your group members are listed as authors, one person per group will turn in the report
- DUE: Wednesday after the lab, by 7am, loaded into moodle

Resources

- Bootstrapping with the boot package
- OECD PISA