ETC 2420/5242 Lab 7 2017

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.

Warmup

A dictionary of variables that we will use further (in addition to the `science` variable we just created) is as follows:

Variable name	Description
ST004D01T	Gender
OUTHOURS	Out-of-School Study Time
ANXTEST	Personality: Test Anxiety (WLE)
EMOSUPP	Parental emotional support (WLE)
PARED	Index highest parental education in years of schooling
JOYSCIE	Enjoyment of science (WLE)
WEALTH	Family wealth (WLE)
ST013Q01TA	How many books are there in your home? 1 0-10 books, 2 11-25 books, 3 26-100 books, 4 101-200 books,
ST012Q01TA	How many in your home: Televisions; 1 None, 2 One, 3 Two, 4 Three or more
SENWT	Weight

Model building will be done using:

- Response: `science` (standardised)
- Explanatory variables: `ST004D01T, JOYSCIE, ST013Q01TA, ST012Q01TA`.

Subset the data to contain just these variables, remove cases with missing values, change gender to be a factor and standardise the science scores.

Question 1

a. Compute and report the 95% confidence interval for the parameter for the number of books in the household (nbooks), using classical t-interval methods.
b. Use this to test the hypothesis that `nbooks` is not important for the model. Write down the null and alternative hypotheses you are using.

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 `nbooks` 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$SENWT)
stat
sort(stat$t)[25]
sort(stat$t)[975]
```

b. How does the bootstrap interval compare with the t-interval? Compare and constrast them.

Question 3

Now make a 95% bootstrap confidence interval for predicted value for a girl, enjoys science (JOYSCIE=1.0) 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 science score range.

Question 4

Do the same 95% bootstrap confidence interval for predicted value for a boy, all other values remain the same. Compare and contrast the intervals for the girl and boy, on the raw science score scale.

Question 5

Compute a bootstrap 95% prediction interval for the girl as in question 3. Be sure to convert the values back into the actual science score range.

TURN IN

- Your `.Rmd` file
- Your `html` file that results from knitting the Rmd.

Resources

- Bootstrapping with the boot package
- OECD PISA