

Module3Test

Markus Jury

2018 M02 16

Contents

Sample Solutions to Module 3 Exercises	1
Preliminaries	1
Data Preparation	1
Analysis	3

Sample Solutions to Module 3 Exercises

Author: Andrew Proctor

Preliminaries

Initialize packages

```
library(rio)
library(tidyverse)

## -- Attaching packages -----
## v ggplot2 2.2.1      v purrr  0.2.4
## v tibble  1.4.2      v dplyr  0.7.4
## v tidyr   0.8.0      v stringr 1.2.0
## v readr   1.1.1      v forcats 0.2.0

## -- Conflicts -----
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Data Preparation

Import files

```
characteristics <- import("./fml1161x.sas7bdat") %>% as.tibble()
expenditures <- import("./mtbi161x.sas7bdat") %>% as.tibble()
```

Make column names lower case

```
colnames(characteristics) <- colnames(characteristics) %>% map(tolower)
colnames(expenditures) <- colnames(expenditures) %>% map(tolower)
```

Select only desired variables in each data set

```
characteristics <- characteristics %>% select(newid,hh_cu_q,educ_ref,
                                             creditx, region, fincbtxm)
expenditures <- expenditures %>% select(newid,cost,ref_mo,ref_yr)
```

Rename selected variables

```
characteristics <- characteristics %>% rename(hh_size=hh_cu_q,
                                             hh_income=fincbtxm)
```

Make columns except for “newid” numeric

```
characteristics[, -1] <- characteristics[, -1] %>% map(as.numeric)
expenditures[, -1] <- expenditures[, -1] %>% map(as.numeric)
```

Create 80% sample of data frames

```
characteristics <- sample_frac(characteristics, 0.8)
expenditures <- sample_frac(expenditures, 0.8)
```

Practice different joins

```
cex_data <- left_join(expenditures, characteristics, by="newid")
cex_data_inner <- semi_join(expenditures, characteristics, by="newid")
cex_data_semi <- semi_join(expenditures, characteristics, by="newid")
cex_data_anti <- anti_join(expenditures, characteristics, by="newid")
```

Create regions indicators

```
cex_data <- cex_data %>% mutate(region1 = ifelse(region == 1,1,0),
                                             region2 = ifelse(region == 2,1,0),
                                             region3 = ifelse(region == 3,1,0),
                                             region4 = ifelse(region == 4,1,0),
                                             region4 = ifelse(is.na(region),1,0)
                                             )
```

Define ols function

```
my_ols <- function(indvars,depvar){

  ### Keep only observations with no missing values for indvars and depvar
  X <- indvars[(!is.na(indvars)) & (!is.na(depvar))]
  y <- depvar[(!is.na(indvars)) & (!is.na(depvar))]

  ##### Create constant vector
  ones_vec <- matrix(rep(1), length(X))

  ### Create matrix X equal to constant vec and indvars
```

```

X <- cbind(ones_vec, X)

### Name constant column "constant"
colnames(X)[1] <- "constant"

### Solve for coefficients
beta <- solve(t(X)%*%X) %*% (t(X)%*%y)
colnames(beta) <- "Estimate"

### Convert to data frame
beta <- as.data.frame(beta)
}

```

Analysis

Estimate Coefficients

```

coeffs_schooling <- my_ols(cex_data$educ_ref, cex_data$cost)

```

Display results

```

coeffs_schooling
## # A tibble: 2 x 1
##   Estimate
## *      <dbl>
## 1    -570
## 2     104

```