PS3_Econometrics

Isha Shrivastava

10/14/2021

```
install.packages("sandwich")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
install.packages("lmtest")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.5     ✓ dplyr   1.0.7
## ✓ tidyr   1.1.4     ✓ stringr 1.4.0
## ✓ readr   2.0.2     ✓ forcats 0.5.1
```

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

```
library(dplyr)
library(haven)
library(sandwich)
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(Matrix)
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
#1a

caschool <- read_dta("caschool.dta")
#computing the sample size
n<- nrow(caschool)
x<- rep(1, 420)
#selecting the relevant explanatory variables
m1<- cbind(x, caschool$str, caschool$el_pct, caschool$meal_pct)
#creating dependant variable vector and assigning dimension
X<- as.matrix(m1)
y<- caschool$testscr
Y <- as.matrix(y)
colnames(X) <- NULL #we remove column names not to get confused between data frames a
nd matrices
colnames(Y) <- NULL
k <- ncol(X)

#running regression
reg1 <- lm(testscr~str+el_pct+meal_pct, caschool)
summary(reg1)
```

```
##
## Call:
## lm(formula = testscr ~ str + el_pct + meal_pct, data = caschool)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.849  -5.151  -0.308   5.243  31.501
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 700.14996    4.68569 149.423  < 2e-16 ***
## str          -0.99831    0.23875  -4.181 3.54e-05 ***
## el_pct       -0.12157    0.03232  -3.762 0.000193 ***
## meal_pct     -0.54735    0.02160 -25.341  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.08 on 416 degrees of freedom
## Multiple R-squared:  0.7745, Adjusted R-squared:  0.7729
## F-statistic: 476.3 on 3 and 416 DF,  p-value: < 2.2e-16
```

```
XprimeX <- t(X) %*% X
XprimeXinverse <- solve(XprimeX)
XprimeY <- t(X) %*% Y
beta_hat <- XprimeXinverse %*% XprimeY
prediction <- X %*% beta_hat
resid <- Y - prediction
#Calculating SSR
SSR <- t(resid) %*% resid
SSR
```

```
##         [,1]
## [1,] 34298.3
```

```
#Calculating TSS
TSS <- sum((Y-mean(Y))^2)
TSS
```

```
## [1] 152109.6
```

```
#Calculating ESS
ESS <- TSS - SSR
ESS
```

```
##          [,1]
## [1,] 117811.3
```

```
#R-Squared
Rsq <- 1 - (SSR/TSS)
Rsq
```

```
##           [,1]
## [1,] 0.7745159
```

```
#Adjusted R-Squared
adj_Rsq <- 1 - (SSR/ (n-k)) / (TSS/ (n-1))
adj_Rsq
```

```
##           [,1]
## [1,] 0.7728898
```

```
#1b

XprimeXinverse
```

```
##               [,1]          [,2]          [,3]          [,4]
## [1,]  0.2662975710 -1.333622e-02  2.877214e-04 -1.459435e-04
## [2,] -0.0133362191  6.913899e-04 -1.239063e-05 -1.064636e-06
## [3,]  0.0002877214 -1.239063e-05  1.266748e-05 -5.460364e-06
## [4,] -0.0001459435 -1.064636e-06 -5.460364e-06  5.658247e-06
```

```
#1c

proj <- X%*%XprimeXinverse%*% t(X)
maker <- diag(n) - proj

#1d
install.packages('matrixcalc')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(matrixcalc)  #package needed for checking idempotency
proj_symmetric <- isSymmetric(proj)
proj_symmetric
```

```
## [1] TRUE
```

```
proj_idempotent <- is.idempotent.matrix(proj)
proj_idempotent
```

```
## [1] TRUE
```

```
maker_symmetric <- isSymmetric(maker)
maker_symmetric
```

```
## [1] TRUE
```

```
maker_idempotent <- is.idempotent.matrix(maker)
maker_idempotent
```

```
## [1] TRUE
```

```
#they all evaluate to TRUE hence, the matrices are symmetric and idempotent

#Rank of Projection
proj_rank <- rankMatrix(proj)
proj_rank
```

```
## [1] 4
## attr(,"method")
## [1] "tolNorm2"
## attr(,"useGrad")
## [1] FALSE
## attr(,"tol")
## [1] 9.325873e-14
```

```
#Rank of Maker
maker_rank <- rankMatrix(maker)
maker_rank
```

```
## [1] 417
## attr(,"method")
## [1] "tolNorm2"
## attr(,"useGrad")
## [1] FALSE
## attr(,"tol")
## [1] 9.325873e-14
```

```
#1e

#Show Py = yˆ
Py <- proj %*% Y
Y_hat <- fitted(reg1)
#on calling variables Py and Y_hat, we get the same values
#We can also take the difference to see if they are equal
diff1 <- Py - prediction

#Since the difference between Py and Y_hat is very small and close to zero, we can sa
y that they are equal

#Show My = εˆ
e_hat <- resid
My <- maker %*% Y
#on calling variables My and e_hat, we get the same values
#We can also take the difference to see if they are equal
diff2 <- My - e_hat

#1f
#Show MX = 0

MX <- maker %*% X
#on calling variable MX we get a 0 matrix

#Show PM = 0

PM <- proj %*% maker
#on calling variable PM we get a 0 matrix


#1g
reg1 <- lm(testscr ~ str + el_pct + meal_pct, data = caschool)
summary(reg1)
```

```
##
## Call:
## lm(formula = testscr ~ str + el_pct + meal_pct, data = caschool)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.849  -5.151  -0.308   5.243  31.501
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 700.14996    4.68569 149.423  < 2e-16 ***
## str          -0.99831    0.23875  -4.181 3.54e-05 ***
## el_pct       -0.12157    0.03232  -3.762 0.000193 ***
## meal_pct     -0.54735    0.02160 -25.341  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.08 on 416 degrees of freedom
## Multiple R-squared:  0.7745, Adjusted R-squared:  0.7729
## F-statistic: 476.3 on 3 and 416 DF,  p-value: < 2.2e-16
```

```
#incorporating robust standard errors
reg1 %>%
    vcovHC() %>%
    diag() %>%
    sqrt()
```

```
## (Intercept)         str      el_pct    meal_pct
##  5.64104056  0.27375999  0.03324308  0.02432103
```

```
coeftest(reg1, vcov = vcovHC(reg1))
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 700.149965   5.641041 124.1172 < 2.2e-16 ***
## str          -0.998309   0.273760  -3.6467 0.0002994 ***
## el_pct       -0.121573   0.033243  -3.6571 0.0002878 ***
## meal_pct     -0.547346   0.024321 -22.5050 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
install.packages('estimatr')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
install.packages('car')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(estimatr)
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
## The following object is masked from 'package:purrr':
##
##     some
```

```
linearHypothesis(reg1, "str=0")
```

```
## Linear hypothesis test
##
## Hypothesis:
## str = 0
##
## Model 1: restricted model
## Model 2: testscr ~ str + el_pct + meal_pct
##
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1    417 35740
## 2    416 34298  1    1441.5 17.483 3.536e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Since the p value for str is less than 0.05, the beta we obtain is statistically sig
nificant. Additionally the t-value is -3.680. We can reject Ho because the absolute v
alue is greater than 1.96 and Ho = 0, so it does not lie withing our confidence inter
val. This coefficient of str is statistically significant


#1h

reg2 <- lm(testscr ~ str + expn_stu + el_pct + meal_pct, data = caschool)
summary(reg2)
```

```
##
## Call:
## lm(formula = testscr ~ str + expn_stu + el_pct + meal_pct, data = caschool)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.366  -5.683   0.281   5.288  30.266
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.660e+02  9.460e+00  70.398  < 2e-16 ***
## str         -2.354e-01  2.983e-01  -0.789     0.43
## expn_stu     3.622e-03  8.766e-04   4.132 4.36e-05 ***
## el_pct      -1.283e-01  3.175e-02  -4.042 6.32e-05 ***
## meal_pct    -5.464e-01  2.119e-02 -25.780  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.91 on 415 degrees of freedom
## Multiple R-squared:  0.7834, Adjusted R-squared:  0.7813
## F-statistic: 375.3 on 4 and 415 DF,  p-value: < 2.2e-16
```

```
#it is probable that the student teacher ratio is correlated with some other predicto
r variable - for this case it could be the new term, which is expenditure per student
. This would mean the model has collinearity.
#next I can perform an f-test to check for collinearity in the model. If it does, the
n we would need to perform analysis for highly correlated variables.
```