Chapter 6 & 7 - Linear Regression, Hypothesis Test and Confidence Interval in Multiple Regression

Ercio Munoz

September 19, 2018

Linear Regression with Multiple Regressors

The syntaxis for linear regression with more than one regressor is very similar, we just need to list the regressors separated by a plus sign "+". For example, we can estimate the model $testscr_i = \beta_0 + \beta_1 str_i + \beta_2 el\_pct_i + \epsilon_i$:

```
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(foreign)
a=
"http://fmwww.bc.edu/ec-p/data/stockwatson/caschool.dta"
data_set = read.dta(a)
regression = lm(testscr~str+el_pct,data=data_set)
summary(regression)
```

```
##
## Call:
## lm(formula = testscr ~ str + el_pct, data = data_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -48.845 -10.240  -0.308   9.815  43.461
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 686.03225    7.41131  92.566  < 2e-16 ***
## str          -1.10130    0.38028  -2.896  0.00398 **
## el_pct       -0.64978    0.03934 -16.516  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.46 on 417 degrees of freedom
## Multiple R-squared:  0.4264, Adjusted R-squared:  0.4237
## F-statistic:   155 on 2 and 417 DF,  p-value: < 2.2e-16
```

```
# With robust standard errors
coeftest(regression, vcov = vcovHC(regression, "HC1"))
```

```
##
```

```
## t test of coefficients:
##
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept) 686.032249   8.728224  78.5993  < 2e-16 ***
## str          -1.101296   0.432847  -2.5443  0.01131 *
## el_pct       -0.649777   0.031032 -20.9391  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hypothesis Tests and Confidence Intervals in Multiple Regression (Ch7)

Now adding a third regressor to the previous model $testscr_i = \beta_0 + \beta_1 str_i + \beta_2 expn\_stu_i + \beta_3 el\_pct_i + \epsilon_i$::

```
library(sandwich)
library(lmtest)
library(foreign)
a=
"http://fmwww.bc.edu/ec-p/data/stockwatson/caschool.dta"
data_set = read.dta(a)
regression = lm(testscr~str+expn_stu+el_pct,data=data_set)
coeftest(regression, vcov = vcovHC(regression, "HC1"))
```

```
##
## t test of coefficients:
##
##                Estimate  Std. Error  t value Pr(>|t|)
## (Intercept) 649.5779473  15.4583434  42.0212  < 2e-16 ***
## str          -0.2863992   0.4820728  -0.5941  0.55277
## expn_stu      0.0038679   0.0015807   2.4469  0.01482 *
## el_pct       -0.6560227   0.0317844 -20.6397  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can test the null hypothesis $H_0 : \beta_1 = \beta_2 = 0$ using a F test:

```
library(car)
# F test with robust variance
myH0 <- c("str","expn_stu")
linearHypothesis(regression, myH0, vcov = vcovHC(regression, "HC1"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## str = 0
## expn_stu = 0
##
## Model 1: restricted model
## Model 2: testscr ~ str + expn_stu + el_pct
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F   Pr(>F)
## 1    418
## 2    416  2 5.4337 0.004682 **
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can report multiple regressions in a table using the command "stargazer":

```
# Estimate four different models
model1 = lm(testscr~str,data=data_set)
model2 = lm(testscr~str+el_pct,data=data_set)
model3 = lm(testscr~str+el_pct+meal_pct,data=data_set)
model4 = lm(testscr~str+el_pct+meal_pct+calw_pct,data=data_set)

# Load package and display table of results
library(stargazer)
```

```
##
## Please cite as:
##  Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.1. https://CRAN.R-project.org/package=stargazer
```

```
stargazer(list(model1,model2,model3,model4),type="text",keep.stat=c("rsq","n"))
```

```
##
## ===============================================================
##                         Dependent variable:
##              ---------------------------------------------
##                                testscr
##                 (1)        (2)        (3)        (4)
## ---------------------------------------------------------------
## str          -2.280***  -1.101***  -0.998***  -1.014***
##               (0.480)    (0.380)    (0.239)    (0.240)
##
## el_pct                  -0.650***  -0.122***  -0.130***
##                          (0.039)    (0.032)    (0.034)
##
## meal_pct                           -0.547***  -0.529***
##                                     (0.022)    (0.032)
##
## calw_pct                                       -0.048
##                                                (0.061)
##
## Constant     698.933*** 686.032*** 700.150*** 700.392***
##               (9.467)    (7.411)    (4.686)    (4.698)
##
## ---------------------------------------------------------------
## Observations    420        420        420        420
## R2             0.051      0.426      0.775      0.775
## ===============================================================
## Note:                         *p<0.1; **p<0.05; ***p<0.01
```

The previous table reported homoskedasticity-only standard errors. In the following table we use the command "vcovHC" to compute standard errors, save them and then use them in the table:

```
library(sandwich)
library(lmtest)
rse_1 = sqrt(diag(vcovHC(model1, type = "HC1")))
rse_2 = sqrt(diag(vcovHC(model2, type = "HC1")))
rse_3 = sqrt(diag(vcovHC(model3, type = "HC1")))
```

```r
rse_4 = sqrt(diag(vcovHC(model4, type = "HC1")))

library(stargazer)
stargazer(list(model1,model2,model3,model4),
          type="text",keep.stat=c("rsq","n"),
          se=list(rse_1,rse_2,rse_3,rse_4) )
```

```
##
## ===============================================================
##                             Dependent variable:
##                 ----------------------------------------------
##                                    testscr
##                    (1)         (2)         (3)         (4)
## ---------------------------------------------------------------
## str             -2.280***   -1.101**    -0.998***   -1.014***
##                  (0.519)     (0.433)     (0.270)     (0.269)
##
## el_pct                      -0.650***   -0.122***   -0.130***
##                              (0.031)     (0.033)     (0.036)
##
## meal_pct                                -0.547***   -0.529***
##                                          (0.024)     (0.038)
##
## calw_pct                                             -0.048
##                                                      (0.059)
##
## Constant        698.933*** 686.032*** 700.150*** 700.392***
##                  (10.364)   (8.728)     (5.568)     (5.537)
##
## ---------------------------------------------------------------
## Observations      420         420         420         420
## R2                0.051       0.426       0.775       0.775
## ===============================================================
## Note:                           *p<0.1; **p<0.05; ***p<0.01
```

```r
# Check the first equation:
coeftest(model1, vcov = vcovHC(model1, "HC1"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 698.93295   10.36436 67.4362 < 2.2e-16 ***
## str          -2.27981    0.51949 -4.3886 1.447e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```