Chapter 6 - Multiple Regression Analysis: Further Issues

Thalles Quinaglia Liduares

2022-04-17

Exercise 6.11

Upload packages

```
library(wooldridge)
library(lmreg)
library(dplyr)
library(car)
```

Upload database

```
data<-wooldridge::apple

attach(data)
```

(i) Run the regression `ecolbs` **on** `ecoprc` , `regprc` **and report the results in the usual form, including the R-squared and adjusted R-squared. Interpret the coefficients on the price variables and comment on their signs and magnitudes.**

```
summary(lm1<-lm(ecolbs~ecoprc+regprc))
```

```
##
## Call:
## lm(formula = ecolbs ~ ecoprc + regprc)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.087 -1.087 -0.537  0.560 39.913
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.9653     0.3801   5.171 3.10e-07 ***
## ecoprc       -2.9265     0.5879  -4.978 8.23e-07 ***
## regprc        3.0289     0.7108   4.261 2.33e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.483 on 657 degrees of freedom
## Multiple R-squared:  0.03641,    Adjusted R-squared:  0.03348
## F-statistic: 12.41 on 2 and 657 DF,  p-value: 5.107e-06
```

The estimated equation is expressed as follows

$$\widehat{ecolbs} = 1.96 - 2.92 ecoprc + 3.02 regprc$$

The R-Squared and Adjusted R-Squared are equal to 3.6% and 3.3%, respectively.

(ii) Are the price variables statistically significant? Report the p-values for the individual t tests.

All coefficients presents statistical significance at the 1% level.

(iii) What is the range of fitted values for `ecolbs`? What fraction of the sample reports ecolbs = 0? Comment.

```
ecolbs0<-data %>%
  dplyr::filter(ecolbs == 0) %>%
  summarise(n=n())

share0<-round(100*ecolbs0/length(ecolbs),2)

share0
```

```
##           n
## 1 37.58
```

The percentual of values equal to zero are 37.58%.

```
summary(lm1$fitted.values)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.8552  0.9166  1.4712  1.4740  2.0257  2.0872
```

The interval for `ecolbs` are equal to [0.85, 2.08].

(iv) Do you think the price variables together do a good job of explaining variation in `ecolbs`? Explain.

In terms of R-Squared, these two variables, together, presents a low capacity of explaining the variability of dependent variable `ecolbs`.

(v) Add the variables `faminc`, `hhsize` (household size), `educ`, and `age` to the regression from part (i). Find the p-value for their joint significance. What do you conclude?

```
summary(lm2<-lm(ecolbs~ecoprc+regprc+faminc+hhsize+educ+age))
```

```
##
## Call:
## lm(formula = ecolbs ~ ecoprc + regprc + faminc + hhsize + educ +
##     age)
##
## Residuals:
##     Min     1Q Median     3Q     Max
## -2.570 -1.146 -0.595  0.515 39.940
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.056770   0.892650   1.184    0.237
## ecoprc      -2.861237   0.591991  -4.833 1.68e-06 ***
## regprc       3.006077   0.712308   4.220 2.79e-05 ***
## faminc       0.002195   0.002865   0.766    0.444
## hhsize       0.063094   0.067780   0.931    0.352
## educ         0.034313   0.045314   0.757    0.449
## age          0.001389   0.006763   0.205    0.837
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.486 on 653 degrees of freedom
## Multiple R-squared:  0.04022,    Adjusted R-squared:  0.0314
## F-statistic: 4.561 on 6 and 653 DF,  p-value: 0.0001519
```

```
linearHypothesis(lm2,
                c("ecoprc=0",
                  "regprc=0",
                  "faminc=0",
                  "hhsize=0",
                  "educ=0",
                  "age=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## ecoprc = 0
## regprc = 0
## faminc = 0
## hhsize = 0
## educ = 0
## age = 0
##
## Model 1: restricted model
## Model 2: ecolbs ~ ecoprc + regprc + faminc + hhsize + educ + age
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    659 4204.1
## 2    653 4035.0  6     169.1 4.561 0.0001519 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of F-Test is equal to 0.00015. Hence, we reject the null hypothesis that all coefficients are equal to 0.

(vi) Run separate simple regressions of `ecolbs` **on** `ecoprc` **and then** `ecolbs` **on** `regprc` **. How do the simple regression coefficients compare with the multiple regression from part (i)? Find the correlation coefficient between** `ecoprc` **and** `regprc` **to help explain your findings.**

```
summary(lm3<-lm(ecolbs~ecoprc))
```

```
##
## Call:
## lm(formula = ecolbs ~ ecoprc)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.889 -1.298 -0.467  0.533 40.618
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.3881     0.3717   6.426 2.52e-10 ***
## ecoprc        -0.8452     0.3315  -2.550    0.011 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.515 on 658 degrees of freedom
## Multiple R-squared:  0.009783,   Adjusted R-squared:  0.008279
## F-statistic: 6.501 on 1 and 658 DF,  p-value: 0.01101
```

```
summary(lm4<-lm(ecolbs~reglbs))
```

```
##
## Call:
## lm(formula = ecolbs ~ reglbs)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.655 -1.322 -0.542  0.345 40.345
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.65545    0.10610  15.603  < 2e-16 ***
## reglbs      -0.14151    0.03339  -4.239 2.57e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.494 on 658 degrees of freedom
## Multiple R-squared:  0.02658,    Adjusted R-squared:  0.0251
## F-statistic: 17.97 on 1 and 658 DF,  p-value: 2.572e-05
```

```
cor(ecolbs, ecoprc, use="complete.obs")
```

```
## [1] -0.09891153
```

```
cor(ecolbs, reglbs, use="complete.obs")
```

```
## [1] -0.1630243
```