Chapter 4 - Inference

Thalles Quinaglia Liduares

11/03/2022

Exercise 4.2

Upload packages

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

Upload database

```
data<-wooldridge::lawsch85
```

Use the data in LAWSCH85.RAW for this exercise.

(i) Using the same model as in Problem 4 in Chapter 3, state and test the null hypothesis that the rank of law schools has no ceteris paribus effect on median starting salary.

The hypothesis test on variable `rank` is stated as follows:

$H_0 : \beta_5 = 0$ and $H_1 : \beta_5 \neq 0$

```
lm1<-lm(lsalary~LSAT+GPA+llibvol+lcost+rank, data)

summary(lm1)
```

```
## 
## Call:
## lm(formula = lsalary ~ LSAT + GPA + llibvol + lcost + rank, data = data)
## 
## Residuals:
##       Min       1Q    Median       3Q      Max
## -0.301356 -0.084982 -0.004359  0.077935  0.288614
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.3432260  0.5325192  15.667  < 2e-16 ***
## LSAT         0.0046965  0.0040105   1.171  0.24372
## GPA          0.2475239  0.0900370   2.749  0.00683 **
## llibvol      0.0949932  0.0332543   2.857  0.00499 **
## lcost        0.0375538  0.0321061   1.170  0.24427
## rank        -0.0033246  0.0003485  -9.541  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.1124 on 130 degrees of freedom
##   (20 observations deleted due to missingness)
## Multiple R-squared:  0.8417, Adjusted R-squared:  0.8356
## F-statistic: 138.2 on 5 and 130 DF,  p-value: < 2.2e-16
```

The estimated coefficient of variable `rank` has p-value<2e-16, so is statistically different from zero.

(ii) Are features of the incoming class of students—namely, `LSAT` and `GPA`—individually or jointly significant for explaining salary? (Be sure to account for missing data on `LSAT` and `GPA`.)

Indidually, `LSAT` do not show statistic significance in explaining the variable `salary`.

But, for an unit increase `GPA` point, the `salary` increases by aproximately 24.7%, to the significance level of 1%.

Now, testing the joint significance of `LSAT` and `GPA`.

```
linearHypothesis(lm1, c("LSAT=0", "GPA=0"))
```

```
## Linear hypothesis test
## 
## Hypothesis:
## LSAT = 0
## GPA = 0
## 
## Model 1: restricted model
## Model 2: lsalary ~ LSAT + GPA + llibvol + lcost + rank
## 
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    132 1.8942
## 2    130 1.6427  2   0.25151 9.9518 9.518e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of statistic is less than 0.05, so the null hypothesis of restricted model is rejected to the 1% level of confidence.

(iii) Test whether the size of the entering class (clsize) or the size of the faculty (faculty) needs to be added to this equation; carry out a single test. (Be careful to account for missing data on clsize and faculty.)

```
lm2<-lm(lsalary~LSAT+GPA+llibvol+lcost+rank+faculty+clsize, data)

summary(lm2)
```

```
##
## Call:
## lm(formula = lsalary ~ LSAT + GPA + llibvol + lcost + rank +
##     faculty + clsize, data = data)
##
## Residuals:
##       Min       1Q    Median       3Q      Max
## -0.299686 -0.082282 -0.009353  0.078702  0.269288
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.416e+00  5.523e-01  15.239  < 2e-16 ***
## LSAT         5.582e-03  4.180e-03   1.336  0.18416
## GPA          2.661e-01  9.325e-02   2.853  0.00508 **
## llibvol      5.516e-02  4.040e-02   1.365  0.17466
## lcost        2.967e-02  3.468e-02   0.856  0.39393
## rank        -3.428e-03  3.573e-04  -9.594  < 2e-16 ***
## faculty      6.748e-05  3.999e-04   0.169  0.86629
## clsize       1.342e-04  1.535e-04   0.874  0.38379
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1131 on 123 degrees of freedom
##   (25 observations deleted due to missingness)
## Multiple R-squared:  0.844,  Adjusted R-squared:  0.8351
## F-statistic: 95.05 on 7 and 123 DF,  p-value: < 2.2e-16
```

The inclusion of variables `faculty` and `clsize` shows that theses two variables are not statistically significant. I will proceed a test of joint significance, to analyse if in fact, these two variables are not good predictors of `salary` .

```
linearHypothesis(lm2, c("faculty=0","clsize=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## faculty = 0
## clsize = 0
##
## Model 1: restricted model
## Model 2: lsalary ~ LSAT + GPA + llibvol + lcost + rank + faculty + clsize
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    125 1.5974
## 2    123 1.5732  2  0.024259 0.9484 0.3902
```

In fact, the p-value of 0.39, reveals that these two variables probabily are equal to 0 in estimated model.

(iv) What factors might influence the rank of the law school that are not included in the salary regression?

The variable `Age` of Law School, might influence the results, because older institutions have more reputation and tends to atract the better students.