Chapter 7 - Multiple Regression Analysis with Qualitative Information

Thalles Quinaglia Liduares

2022-04-20

Exercise 7.4

Upload packages

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

```
data<-wooldridge::gpa2

attach(data)
```

Use the data in GPA2.RAW for this exercise.

(i) Consider the equation

$$colgpa = \beta_0 + \beta_1 size + \beta_2 size^2 + \beta_3 hsperc + \beta_4 sat + \beta_5 female + \beta_6 athlete + u$$

where `colgpa` is cumulative college grade point average, `hsize` is size of high school graduating class, in hundreds, `hsperc` is academic percentile in graduating class, `sat` is combined SAT score, `female` is a binary gender variable, and `athlete` is a binary variable, which is one for student-athletes. What are your expectations for the coefficients in this equation? Which ones are you unsure about?

Negative expected signs: $hsize^2$

Positive expected signs: `hsize` , `hsperc` , `sat` , `female` , `athlete` .

(ii) Estimate the equation in part (i) and report the results in the usual form. What is the estimated GPA differential between athletes and nonathletes? Is it statistically significant?

```
options(scipen=999)

summary(lm1<-lm(colgpa~hsize+hsizesq+hsperc+sat+female+athlete))
```

```
##
## Call:
## lm(formula = colgpa ~ hsize + hsizesq + hsperc + sat + female +
##     athlete)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.69216 -0.34954  0.03416  0.38806  1.90159
##
## Coefficients:
##                Estimate  Std. Error t value             Pr(>|t|)
## (Intercept)  1.24136451  0.07949233  15.616 < 0.0000000000000002 ***
## hsize       -0.05685426  0.01635134  -3.477             0.000512 ***
## hsizesq      0.00467540  0.00224938   2.079             0.037723 *
## hsperc      -0.01321258  0.00057278 -23.068 < 0.0000000000000002 ***
## sat          0.00164641  0.00006682  24.640 < 0.0000000000000002 ***
## female       0.15488141  0.01800465   8.602 < 0.0000000000000002 ***
## athlete      0.16930636  0.04234921   3.998             0.000065 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5544 on 4130 degrees of freedom
## Multiple R-squared:  0.2925, Adjusted R-squared:  0.2915
## F-statistic: 284.6 on 6 and 4130 DF,  p-value: < 0.00000000000000022
```

The athlets students have a GPA of 0.169 higher than non-athlete students, as others factors remain fixed. The estimate is statistically significant at the 1% level. The SAT score, is one of the prerequisites for students claim an athlete fellowship.

(iii) Drop sat from the model and reestimate the equation. Now, what is the estimated effect of being an athlete? Discuss why the estimate is different than that obtained in part (ii).

```
summary(lm2<-lm(colgpa ~ hsize + hsizesq + hsperc + female +
    athlete))
```

```
##
## Call:
## lm(formula = colgpa ~ hsize + hsizesq + hsperc + female + athlete)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5164 -0.3819  0.0205  0.4204  1.8809
##
## Coefficients:
##               Estimate Std. Error t value             Pr(>|t|)
## (Intercept)  3.0476980  0.0329148  92.594 < 0.0000000000000002 ***
## hsize       -0.0534038  0.0175092  -3.050              0.00230 **
## hsizesq      0.0053228  0.0024086   2.210              0.02716 *
## hsperc      -0.0171365  0.0005892 -29.086 < 0.0000000000000002 ***
## female       0.0581231  0.0188162   3.089              0.00202 **
## athlete      0.0054487  0.0447871   0.122              0.90318
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5937 on 4131 degrees of freedom
## Multiple R-squared:  0.1885, Adjusted R-squared:  0.1875
## F-statistic: 191.9 on 5 and 4131 DF,  p-value: < 0.00000000000000022
```

Now, the coefficient associated to `athlete` becomes much small and non significant.

(iv) In the model from part (i), allow the effect of being an athlete to differ by gender and test the null hypothesis that there is no ceteris paribus difference between women athletes and women nonathletes.

In progress..