

# Chapter 7 - Multiple Regression Analysis with Qualitative Information

Thalles Quinaglia Liduares

2022-04-19

## Exercise 7.1

Upload packages

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

Upload database

```
data<-wooldridge::gpa1

attach(data)
```

Use the data in GPA1.RAW for this exercise.

(i) Add the variables `mothcoll` and `fathcoll` to the equation estimated in (7.6) and report the results in the usual form. What happens to the estimated effect of PC ownership? Is PC still statistically significant?

```
summary(lm1<-lm(colGPA~PC+hsGPA+ACT+mothcoll+fathcoll))
```

```
##
## Call:
## lm(formula = colGPA ~ PC + hsGPA + ACT + mothcoll + fathcoll)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.78149 -0.25726 -0.02121  0.24691  0.74432
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.255554   0.335392   3.744 0.000268 ***
## PC           0.151854   0.058716   2.586 0.010762 *
## hsGPA        0.450220   0.094280   4.775 4.61e-06 ***
## ACT          0.007724   0.010678   0.723 0.470688
## mothcoll     -0.003758   0.060270  -0.062 0.950376
## fathcoll      0.041800   0.061270   0.682 0.496265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3344 on 135 degrees of freedom
## Multiple R-squared:  0.2222, Adjusted R-squared:  0.1934
## F-statistic: 7.713 on 5 and 135 DF,  p-value: 2.083e-06
```

Given the estimated coefficients, we observe that PC is statistically significant at the 5% level. The estimate for this coefficient remains almost equal in relation to the original case.

**(ii) Test for joint significance of mothcoll and fathcoll in the equation from part (i) and be sure to report the p-value.**

```
linearHypothesis(lm1,c("mothcoll=0","fathcoll=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## mothcoll = 0
## fathcoll = 0
##
## Model 1: restricted model
## Model 2: colGPA ~ PC + hsGPA + ACT + mothcoll + fathcoll
##
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1     137 15.149
## 2     135 15.094   2   0.054685 0.2446 0.7834
```

It's not possible to reject the null hypothesis that both `mothcoll` and `fathcoll` are equal to zero, because the p-value of F-Test is quite large.

**(iii) Add  $hsGPA^2$  to the model from part (i) and decide whether this generalization is needed.**

```
hsgpasq<-hsGPA*hsGPA
```

```
summary(lm2<-lm(colGPA ~ PC + hsGPA + ACT + mothcoll + fathcoll+hsgpasq))
```

```
##
## Call:
## lm(formula = colGPA ~ PC + hsGPA + ACT + mothcoll + fathcoll +
##     hsgpasq)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.78998 -0.24327 -0.00648  0.26179  0.72231
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.040328   2.443038   2.063   0.0410 *
## PC           0.140446   0.058858   2.386   0.0184 *
## hsGPA       -1.802520   1.443552  -1.249   0.2140
## ACT          0.004786   0.010786   0.444   0.6580
## mothcoll     0.003091   0.060110   0.051   0.9591
## fathcoll     0.062761   0.062401   1.006   0.3163
## hsgpasq      0.337341   0.215711   1.564   0.1202
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3326 on 134 degrees of freedom
## Multiple R-squared:  0.2361, Adjusted R-squared:  0.2019
## F-statistic: 6.904 on 6 and 134 DF, p-value: 2.088e-06
```

The inclusion of this variable does not increase the accuracy of the model.