# Chapter 6 - Multiple Regression Analysis: Further Issues

Thalles Quinaglia Liduares

2022-04-13

# Exercise 6.9

Upload packages

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

Upload database

```
data<-wooldridge::nbasal
attach(data)
```

**The data set NBASAL.RAW contains salary information and career statistics for 269 players in the National Basketball Association (NBA).**

**(i) Estimate a model relating points-per-game (points) to years in the league (exper), age, and years played in college (coll). Include a quadratic in exper; the other variables should appear in level form. Report the results in the usual way.**

```
summary(lm1<-lm(points~exper+age+coll+expersq))
```

```
##
## Call:
## lm(formula = points ~ exper + age + coll + expersq)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.4736  -3.9849  -0.8246   3.6536  20.9790
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 35.21831    6.98673   5.041 8.62e-07 ***
## exper        2.36363    0.40550   5.829 1.62e-08 ***
## age         -1.07396    0.29507  -3.640 0.000328 ***
## coll        -1.28625    0.45059  -2.855 0.004651 **
## expersq     -0.07703    0.02348  -3.280 0.001177 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.509 on 264 degrees of freedom
## Multiple R-squared:  0.1412, Adjusted R-squared:  0.1282
## F-statistic: 10.85 on 4 and 264 DF,  p-value: 3.687e-08
```

The estimated equation is expressed as follows

$$\widehat{points} = 35.21 + 2.36exper - 1.07age - 1.28coll - 0.07exper^2$$

**(ii) Holding college years and age fixed, at what value of experience does the next year of experience actually reduce points-per-game? Does this make sense?**

Deriving the estimated equation in relation to the variable `exper`, to find the point of inflection.

$$\frac{\partial \widehat{points}}{\partial exper} = 2.36 - 2(0.07)exper = 0$$

$$exper^* = \frac{2.36}{2(0.07)} \approx 17$$

Hence, the number of points starts to decline with 17 years of experience. However, this value do not make practical sense.

**(iv) Add a quadratic in age to the equation. Is it needed? What does this appear to imply about the effects of age, once experience and education are controlled for?**

```
summary(lm2<-lm(points~exper+age+coll+expersq+agesq))
```

```
##
## Call:
## lm(formula = points ~ exper + age + coll + expersq + agesq)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2520  -4.0283  -0.8208   3.4792  20.8332
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 73.59034   35.93341   2.048  0.04156 *
## exper        2.86383    0.61272   4.674 4.72e-06 ***
## age         -3.98369    2.68908  -1.481  0.13969
## coll        -1.31260    0.45108  -2.910  0.00392 **
## expersq     -0.12807    0.05244  -2.442  0.01525 *
## agesq        0.05355    0.04919   1.089  0.27732
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.508 on 263 degrees of freedom
## Multiple R-squared:  0.1451, Adjusted R-squared:  0.1288
## F-statistic: 8.925 on 5 and 263 DF,  p-value: 7.615e-08
```

In this case, the inclusion of variable `agesq` demonstrates that the points in the match increases until certain age, and then becomes to decreasing.

**(v) Now regress `log(wage)` on `points`, `exper`, `exper^2`, `age`, and `coll`. Report the results in the usual format.**

```
summary(lm3<-lm(lwage~points+exper+expersq+age+coll))
```

```
##
## Call:
## lm(formula = lwage ~ points + exper + expersq + age + coll)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9564 -0.2907  0.1141  0.4102  1.7078
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.779039   0.845421   8.019 3.53e-14 ***
## points       0.077730   0.007113  10.928  < 2e-16 ***
## exper        0.217845   0.049788   4.375 1.75e-05 ***
## expersq     -0.007082   0.002769  -2.558   0.0111 *
## age         -0.048137   0.034947  -1.377   0.1695
## coll        -0.040271   0.052872  -0.762   0.4469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6367 on 263 degrees of freedom
## Multiple R-squared:  0.4878, Adjusted R-squared:  0.4781
## F-statistic:  50.1 on 5 and 263 DF,  p-value: < 2.2e-16
```

The estimated equation is expressed as follows

$$\widehat{log(wage)} = 6.77 + 0.07points + 0.21exper - 0.007exper^2 - 0.04age - 0.04coll$$

**(vi) Test whether `age` and `coll` are jointly significant in the regression from part (v). What does this imply about whether age and education have separate effects on wage, once productivity and seniority are accounted for?**

```
linearHypothesis(lm3, c("age=0","coll=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## age = 0
## coll = 0
##
## Model 1: restricted model
## Model 2: lwage ~ points + exper + expersq + age + coll
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    265 107.59
## 2    263 106.63  2   0.96416 1.1891 0.3061
```

It's not possible to reject the null hypothesis that both coefficients are equal to zero.