Chapter 6 -

Thalles Quinaglia Liduares

2022-04-19

Exercise 6.10

Upload packages

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

Upload database

```
data<-wooldridge::bwght2

attach(data)
```

Use the data in BWGHT2.RAW for this exercise.

(i) Estimate the equation

$$log(bwght) = \beta_0 + \beta_1 npvis + \beta_2 npvis^2 + u$$

by OLS, and report the results in the usual way. Is the quadratic term significant?

```
summary(lm1<-lm(lbwght~npvis+npvissq))
```

```
##
## Call:
## lm(formula = lbwght ~ npvis + npvissq)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.15564 -0.08375  0.02241  0.11417  0.45529
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.9578827  0.0273125 291.364  < 2e-16 ***
## npvis        0.0189167  0.0036806   5.140 3.06e-07 ***
## npvissq     -0.0004288  0.0001200  -3.573 0.000362 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2031 on 1761 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.02125,    Adjusted R-squared:  0.02014
## F-statistic: 19.12 on 2 and 1761 DF,  p-value: 6.097e-09
```

The estimated equation is expressed as follows

$$\widehat{log(bwght)} = 7.95 + 0.018npvis - 0.0004npvis^2$$

The coefficient associated to variable $npvis^2$ is significant at the 1% level.

(ii) Show that, based on the equation from part (i), the number of prenatal visits that maximizes `log(bwght)` is estimated to be about 22. How many women had at least 22 prenatal visits in the sample?

Maximizing the above equation in relation to `npvis`, we have that

$$\frac{\partial \widehat{log(bwght)}}{\partial npvis} = 0.018 - 2 \cdot (0.0004)npvis = 0$$

$$npvis^* = \frac{0.018}{2 \cdot (0.0004)} = 22.5$$

```
npvis_22<- data %>%
  filter(npvis >= 22) %>%
  summarise(n=n())

npvis_22
```

```
##    n
## 1 21
```

In the sample, 21 woman's had at least 22 prenatal visits.

(iv) Add mother's age to the equation, using a quadratic functional form. Holding npvis fixed, at what mother's age is the birth weight of the child maximized? What fraction of women in the sample are older than the "optimal" age?

```
summary(lm2<-lm(lbwght~npvis+npvissq+mage+magesq))
```

```
##
## Call:
## lm(formula = lbwght ~ npvis + npvissq + mage + magesq)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.16701 -0.08369  0.02141  0.11635  0.44489
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.5837127  0.1370568  55.333  < 2e-16 ***
## npvis        0.0180374  0.0037086   4.864 1.26e-06 ***
## npvissq     -0.0004079  0.0001204  -3.387 0.000721 ***
## mage         0.0253920  0.0092542   2.744 0.006134 **
## magesq      -0.0004119  0.0001548  -2.661 0.007863 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2027 on 1759 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.02562,    Adjusted R-squared:  0.02341
## F-statistic: 11.56 on 4 and 1759 DF,  p-value: 2.865e-09
```

Maximizing the above equation in relation to mother's age variable, we have that

$$\frac{\partial \widehat{log(bwght)}}{\partial mage} = 0.025 - 2 \cdot (0.0004) mage = 0$$

$$mage^* = \frac{0.025}{2 \cdot (0.0004)} = 31.25$$

Hence, the age's woman that maximize the `log(bwght)` is 31 years.

```
mage_31 <- data %>%
  filter(mage > 31) %>%
  summarise(n=n())

mage_31
```

```
##     n
## 1 605
```

In the sample, there's a total of 605 womans with age larger than 31 years.

(v) Would you say that mother's age and number of prenatal visits explain a lot of the variation in log(bwght)?

No. In terms of R-Squared, these two variables presents a poor capacity of explain the variability in `log(bwght)`.

(vi) Using quadratics for both npvis and age, decide whether using the natural log or the level of bwght is better for predicting bwght.

Without log

```
summary(lm3<-lm(bwght~npvis+npvissq+mage+magesq))
```

```
##
## Call:
## lm(formula = bwght ~ npvis + npvissq + mage + magesq)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2915.78  -327.31    19.22   351.92  1773.70
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1860.3805   388.9770   4.783 1.87e-06 ***
## npvis         37.4739    10.5253   3.560  0.00038 ***
## npvissq       -0.7863     0.3418  -2.301  0.02154 *
## mage          81.6055    26.2640   3.107  0.00192 **
## magesq        -1.3272     0.4393  -3.021  0.00255 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 575.4 on 1759 degrees of freedom
##   (68 observations deleted due to missingness)
## Multiple R-squared:  0.01916,    Adjusted R-squared:  0.01693
## F-statistic:  8.59 on 4 and 1759 DF,  p-value: 7.286e-07
```

The model performs well with dependent variable in log form.