# Chapter 4 - Inference

Thalles Quinaglia Liduares

14/03/2022

# Exercise 4.5

Upload packages

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

Upload database

```
data<-wooldridge::mlb1
```

**Use the data in MLB1.RAW for this exercise.**

**(i) Use the model estimated in equation (4.31) and drop the variable `rbisyr`. What happens to the statistical significance of `hrunsyr`? What about the size of the coefficient on `hrunsyr`?**

Original model (Example 4.31)

```
options(scipen = 999) # To avoid scientific notation

lm0<-lm(lsalary~years+gamesyr+bavg+hrunsyr+rbisyr, data)

summary(lm0)
```

```
## 
## Call:
## lm(formula = lsalary ~ years + gamesyr + bavg + hrunsyr + rbisyr,
##     data = data)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.02508 -0.45034 -0.04013  0.47014  2.68924
## 
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) 11.1924181  0.2888229  38.752 < 0.0000000000000002 ***
## years        0.0688626  0.0121145   5.684        0.0000000279 ***
## gamesyr      0.0125521  0.0026468   4.742        0.0000030886 ***
## bavg         0.0009786  0.0011035   0.887               0.376
## hrunsyr      0.0144295  0.0160570   0.899               0.369
## rbisyr       0.0107657  0.0071750   1.500               0.134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.7266 on 347 degrees of freedom
## Multiple R-squared:  0.6278, Adjusted R-squared:  0.6224
## F-statistic: 117.1 on 5 and 347 DF,  p-value: < 0.00000000000000022
```

```
lm1<-lm(lsalary~years+gamesyr+bavg+hrunsyr, data)

summary(lm1)
```

```
## 
## Call:
## lm(formula = lsalary ~ years + gamesyr + bavg + hrunsyr, data = data)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0642 -0.4614 -0.0271  0.4654  2.7216
## 
## Coefficients:
##              Estimate Std. Error t value            Pr(>|t|)
## (Intercept) 11.020912   0.265719  41.476 < 0.0000000000000002 ***
## years        0.067732   0.012113   5.592        0.0000000455 ***
## gamesyr      0.015759   0.001564  10.079 < 0.0000000000000002 ***
## bavg         0.001419   0.001066   1.331               0.184
## hrunsyr      0.035943   0.007241   4.964        0.0000010834 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.7279 on 348 degrees of freedom
## Multiple R-squared:  0.6254, Adjusted R-squared:  0.6211
## F-statistic: 145.2 on 4 and 348 DF,  p-value: < 0.00000000000000022
```

In the second estimated model, the variable `hrunsyr` becomes statistically significant at the 1% level. In the first model, this variable do not show statistical significance. The magnitude of coefficient increases from 0.01 to 0.03.

**(ii) Add the variables `runsyr` (runs per year), `fldperc` (fielding percentage), and `sbasesyr` (stolen bases per year) to the model from part (i). Which of these factors are individually significant?**

```
lm2<-lm(lsalary~years+gamesyr+bavg+hrunsyr+runsyr+
        fldperc+sbasesyr, data)

summary(lm2)
```

```
##
## Call:
## lm(formula = lsalary ~ years + gamesyr + bavg + hrunsyr + runsyr +
##     fldperc + sbasesyr, data = data)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -2.11554 -0.44557 -0.08808  0.48731  2.57872
##
## Coefficients:
##               Estimate Std. Error t value     Pr(>|t|)
## (Intercept) 10.4082680  2.0032546   5.196 0.0000003497 ***
## years        0.0699848  0.0119756   5.844 0.0000000118 ***
## gamesyr      0.0078995  0.0026775   2.950     0.003391 **
## bavg         0.0005296  0.0011038   0.480     0.631656
## hrunsyr      0.0232106  0.0086392   2.687     0.007566 **
## runsyr       0.0173922  0.0050641   3.434     0.000666 ***
## fldperc      0.0010351  0.0020046   0.516     0.605936
## sbasesyr    -0.0064191  0.0051842  -1.238     0.216479
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7176 on 345 degrees of freedom
## Multiple R-squared:  0.639,  Adjusted R-squared:  0.6317
## F-statistic: 87.25 on 7 and 345 DF,  p-value: < 0.00000000000000022
```

In this case, the only variable that shows statistical significance is the `runsyr` at the 1% level.

**(iii) In the model from part (ii), test the joint significance of `bavg`, `fldperc`, and `sbasesyr`.**

```
linearHypothesis(lm2, c("bavg=0","fldperc=0","sbasesyr=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## bavg = 0
## fldperc = 0
## sbasesyr = 0
##
## Model 1: restricted model
## Model 2: lsalary ~ years + gamesyr + bavg + hrunsyr + runsyr + fldperc +
##     sbasesyr
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    348 178.72
## 2    345 177.66  3    1.0583 0.685 0.5617
```

With base in the p-value of F-Test, we cannot reject the null hypothesis that this coefficients are jointly different from zero.