Chapter 7 - Multiple Regression Analysis with Qualitative Information

Thalles Quinaglia Liduares

2022-04-20

Upload packages

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

```
data<-wooldridge::mlb1

attach(data)
```

Exercise 7.3

A model that allows major league baseball player salary to differ by position is

$$log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + \beta_5 rbisyr + \beta_6 runsyr + \beta_7 fldperc + \beta_8 allstar + \beta_9 frstbase + \beta_{10} scndbase + \beta_{11} thrdbase + \beta_{12} shrtstop + \beta_{13} catcher + u$$

where outfield is the base group

(i) State the null hypothesis that, controlling for other factors, catchers and outfielders earn, on average, the same amount. Test this hypothesis using the data in MLB1.RAW and comment on the size of the estimated salary differential.

The null hypothesis is that $H_0 : \beta_{13} = 0$ against the alternative $H_1 : \beta_{13} \neq 0$

```
summary(lm1<-lm(lsalary~years+gamesyr+bavg+hrunsyr+rbisyr+runsyr+fldperc
                +allstar+frstbase+scndbase+thrdbase+shrtstop+catcher))
```

```
##
## Call:
## lm(formula = lsalary ~ years + gamesyr + bavg + hrunsyr + rbisyr +
##     runsyr + fldperc + allstar + frstbase + scndbase + thrdbase +
##     shrtstop + catcher)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.42088 -0.42665 -0.03092  0.47925  2.74975
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.1295538  2.3044544   4.830 2.07e-06 ***
## years        0.0584178  0.0122732   4.760 2.87e-06 ***
## gamesyr      0.0097670  0.0033776   2.892  0.00408 **
## bavg         0.0004814  0.0011411   0.422  0.67340
## hrunsyr      0.0191459  0.0159638   1.199  0.23124
## rbisyr       0.0017875  0.0074755   0.239  0.81116
## runsyr       0.0118707  0.0045264   2.623  0.00912 **
## fldperc      0.0002833  0.0023078   0.123  0.90239
## allstar      0.0063351  0.0028828   2.198  0.02866 *
## frstbase    -0.1328009  0.1309243  -1.014  0.31115
## scndbase    -0.1611011  0.1414296  -1.139  0.25547
## thrdbase     0.0145271  0.1430352   0.102  0.91916
## shrtstop    -0.0605672  0.1302031  -0.465  0.64210
## catcher      0.2535592  0.1313128   1.931  0.05432 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7092 on 339 degrees of freedom
## Multiple R-squared:  0.6535, Adjusted R-squared:  0.6403
## F-statistic: 49.19 on 13 and 339 DF,  p-value: < 2.2e-16
```

```
linearHypothesis(lm1, c("catcher=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## catcher = 0
##
## Model 1: restricted model
## Model 2: lsalary ~ years + gamesyr + bavg + hrunsyr + rbisyr + runsyr +
##     fldperc + allstar + frstbase + scndbase + thrdbase + shrtstop +
##     catcher
##
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    340 172.40
## 2    339 170.52  1    1.8755 3.7286 0.05432 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It's possible to reject the null hypothesis at the 10% significance level. Hence, there's a gap between the salaries of two analysed categories.

(ii) State and test the null hypothesis that there is no difference in average salary across positions, once other factors have been controlled for.

In this case, the null hypothesis is given by

$$H_0 : \beta_i = 0 \quad i = 9, 10, 11, 12, 13$$

```
linearHypothesis(lm1, c("frstbase=0","scndbase=0","thrdbase=0","shrtstop=0","catcher=0"))
```

```
## Linear hypothesis test
##
## Hypothesis:
## frstbase = 0
## scndbase = 0
## thrdbase = 0
## shrtstop = 0
## catcher = 0
##
## Model 1: restricted model
## Model 2: lsalary ~ years + gamesyr + bavg + hrunsyr + rbisyr + runsyr +
##     fldperc + allstar + frstbase + scndbase + thrdbase + shrtstop +
##     catcher
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    344 174.99
## 2    339 170.52  5    4.4703 1.7774 0.1168
```

It's not possible to reject the null hypothesis even at the 10% significance level.

(iii) Are the results from parts (i) and (ii) consistent? If not, explain what is happening.

In the first test, there's some evidence of differentiation in salary between categories. Otherwise, in second case, the results suggests absence of discrimination, so the results must be investigated more depthly.