

Econometria I Lista 8

Profa. Lorena Hakak

Entrega: 27/11/2022

Capítulo 6

Exercício 6

Para saber se vamos incluir no modelo ou não, vamos encontrar o R-quadrado ajustado:

$$\begin{aligned}\bar{R}^2 &= 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1} \\ \bar{R}^2 &= 1 - \frac{(1 - 0,232)(680 - 1)}{680 - 8 - 1} \\ \bar{R}^2 &= 0,0223\end{aligned}$$

Para testar ao nível de significância de 10%, faremos o teste F:

$$F = \frac{\frac{(R_{ir}^2 - R_r^2)}{q}}{\frac{1 - R_{ir}^2}{(n - k - 1)}} = \frac{\frac{(0,232 - 0,229)}{673 - 671}}{\frac{1 - 0,232}{(671)}} = 1,31$$

Como $F > 1,18$ então rejeita-se a hipótese nula. Eu incluiria essas variáveis pelo teste estatístico aplicado, apesar de alterar o R-quadrado e o R-quadrado ajustado em valores muito pequenos. Há como defender a não inclusão das variáveis pelo critério da parcimônia, em escolher modelos mais simples.

Exercício 8

(i)

Sim, devemos inserir a variável *attend* na regressão juntamente com *alcohol*, mas não na mesma variável. Isso porque cada uma dessas variáveis possuem impactos em direção contrárias: enquanto o aumento da frequência tende a aumentar as notas, o aumento do consumo de álcool tende a diminuir as notas.

(ii)

Essas variáveis são importantes de serem incluídas pois ajuda no controle do nível prévio dos alunos. Caso essas variáveis não sejam inseridas, o efeito do álcool nas notas de graduação podem estar superestimadas.

Capítulo 7

Exercício 2

(i)

Para cada cigarro a mais que a mãe fuma por dia durante a gravidez esperasse que se reduza 0,5% do peso da criança. Caso a mãe fume pelo menos 10 cigarros por dia, esperasse uma redução do peso da criança em 5%.

(ii)

Espera-se que uma criança branca pese aproximadamente 4,5% a mais que uma criança não branca. A diferença é estatisticamente significativa, já que sua estatística t será de 3 (0,045/0,015), sendo significativa a 1%.

(iii)

O que o efeito nos diz é que quanto maior os anos de escolaridade da mãe, menor é o peso da criança. Essa variável não se mostrou estatisticamente significativa.

(iv)

Não será possível calcular a estatística F por conta de que os dois modelos usam observações diferentes, já que possui um número menor do que a regressão restrita (a primeira). Para poder dar certo, teríamos que utilizar a mesma base de dados para as duas equações.

Exercício 3

(i)

Sim, a variável $hsize^2$ deve ser incluída, já que sua estatística t é de 4,13, sendo estatisticamente significativa a 1%. Para calcular o tamanho ótimo, temos que observar o efeito marginal do tamanho da sala:

$$\frac{\Delta sat}{\Delta hsize} = 19,30 - 4,38hsize = 0$$

$$hsize \approx 4,4$$

(ii)

A diferença passa a ser de aproximadamente 45,09 pontos no SAT . É estatisticamente significativa a 1%, já que sua estatística t foi de 10,5.

(iii)

Sendo a regressão populacional escrita por $sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 female + \beta_4 black + \beta_5 female.black$, então o teste de hipótese seria:

$$H_0 : \beta_3 = 0 \text{ e } \beta_4 = 0$$

$$H_1 : \beta_3 = 0 \text{ e } \beta_4 \neq 0$$

Além disso, a diferença entre homens negros e não negros é de 169,81 pontos no SAT .

(iv)

A diferença de notas entre mulheres negras e não negras é de 107,05 pontos no SAT (62,31- 169,81). Para ver se é estatisticamente significativa, eu precisaria fazer um teste t restringindo β_4 e β_5 do modelo populacional.

Exercício 6

Dado que a correlação entre a variável omitida (aptidão) e o treinamento é negativa, já que o treinamento ocorreu para aqueles com menor aptidão, e o coeficiente da variável omitida será positiva (quanto maior a aptidão, maior o salário esperado) então o viés é negativo, tendo um modelo

Exercício 7

(i)

Sendo a regressão populacional feita em (7.29):

$$inlf = \beta_0 + \beta_1nwifeinc + \beta_2educ + \beta_3exper + \beta_4exper^2 + \beta_5age + \beta_6kidslt6 + \beta_7kidsge6$$

E agora utilizando como variável dependente $outlf$, temos que:

$$outlf = \beta_0 + \beta_1nwifeinc + \beta_2educ + \beta_3exper + \beta_4exper^2 + \beta_5age + \beta_6kidslt6 + \beta_7kidsge6 + u$$

Como $outlf = 1 - inlf$, então:

$$1 - inlf = \beta_0 + \beta_1nwifeinc + \beta_2educ + \beta_3exper + \beta_4exper^2 + \beta_5age + \beta_6kidslt6 + \beta_7kidsge6 + u$$

$$-inlf = -1 + \beta_0 + \beta_1nwifeinc + \beta_2educ + \beta_3exper + \beta_4exper^2 + \beta_5age + \beta_6kidslt6 + \beta_7kidsge6 + u$$

$$inlf = (1 - \beta_0) - \beta_1nwifeinc - \beta_2educ - \beta_3exper - \beta_4exper^2 - \beta_5age - \beta_6kidslt6 - \beta_7kidsge6 - u$$

(ii)

Não acontece nada como erro padrão das variáveis, sendo que a mudança dos sinais dos betas irão mudar o sinal apenas da estatística t . O mesmo ocorre para o intercepto, já que:

$$Var(1 - \beta_0) = Var(-\beta_0) = Var(\beta_0)$$

(iii)

O R-quadrado não irá se alterar. Isso porque as mesmas variáveis estão sendo modeladas, alterando apenas a estrutura da regressão. Ao invés de ver como as variáveis influenciam a probabilidade de $inlf$ ser 1, no segundo modelo eles mostram a tendência da probabilidade de ser 0.

Exercícios Práticos

```
require(wooldridge)
```

```
## Loading required package: wooldridge
```

C2

(i)

```
lwage <- log(wage2$wage)
```

```
reg1 <- lm(lwage ~ educ + exper + tenure + married + black + south + urban, data = wage2)
```

```
summary(reg1)
```

```
##
```

```
## Call:
```

```
## lm(formula = lwage ~ educ + exper + tenure + married + black +  
##      south + urban, data = wage2)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -1.98069 -0.21996  0.00707  0.24288  1.22822
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  5.395497   0.113225  47.653 < 2e-16 ***  
## educ         0.065431   0.006250  10.468 < 2e-16 ***  
## exper        0.014043   0.003185   4.409 1.16e-05 ***  
## tenure       0.011747   0.002453   4.789 1.95e-06 ***  
## married      0.199417   0.039050   5.107 3.98e-07 ***  
## black        -0.188350   0.037667  -5.000 6.84e-07 ***  
## south        -0.090904   0.026249  -3.463 0.000558 ***  
## urban        0.183912   0.026958   6.822 1.62e-11 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 0.3655 on 927 degrees of freedom
```

```
## Multiple R-squared:  0.2526, Adjusted R-squared:  0.2469
```

```
## F-statistic: 44.75 on 7 and 927 DF, p-value: < 2.2e-16
```

A diferença percentual entre os salários de negros e não negros é de aproximadamente 18,8%. Além disso, é estatisticamente significante.

(ii)

```

exper2 <- (wage2$exper)^2

tenure2 <- (wage2$tenure)^2

reg2 <- lm(lwage ~ educ + exper + tenure + married + black + south + urban + exper2 + tenure2, data = wage2)

r2.ir <- summary(reg2)$r.squared

r2.r <- summary(reg1)$r.squared

F <- (r2.ir - r2.r) / (1 - r2.ir) * (925/2)

F

## [1] 1.489806
qf(1-0.2,2,925)

## [1] 1.612241

```

(iii)

```

reg3 <- lm(lwage ~ educ + exper + tenure + married + black + south + urban + black*educ, data = wage2)

summary(reg3)

##
## Call:
## lm(formula = lwage ~ educ + exper + tenure + married + black +
##      south + urban + black * educ, data = wage2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.97782 -0.21832  0.00475  0.24136  1.23226
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.374817   0.114703  46.859 < 2e-16 ***
## educ         0.067115   0.006428  10.442 < 2e-16 ***
## exper        0.013826   0.003191   4.333 1.63e-05 ***
## tenure       0.011787   0.002453   4.805 1.80e-06 ***
## married      0.198908   0.039047   5.094 4.25e-07 ***
## black        0.094809   0.255399   0.371 0.710561
## south       -0.089450   0.026277  -3.404 0.000692 ***
## urban        0.183852   0.026955   6.821 1.63e-11 ***
## educ:black  -0.022624   0.020183  -1.121 0.262603
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3654 on 926 degrees of freedom
## Multiple R-squared:  0.2536, Adjusted R-squared:  0.2471
## F-statistic: 39.32 on 8 and 926 DF, p-value: < 2.2e-16

```

(iv)

```
reg4 <- lm(lwage ~ educ + exper + tenure + married + black + south + urban, data = wage2)
summary(reg4)

##
## Call:
## lm(formula = lwage ~ educ + exper + tenure + married + black +
##      south + urban, data = wage2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.98069 -0.21996  0.00707  0.24288  1.22822
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.395497   0.113225  47.653  < 2e-16 ***
## educ         0.065431   0.006250  10.468  < 2e-16 ***
## exper        0.014043   0.003185   4.409 1.16e-05 ***
## tenure       0.011747   0.002453   4.789 1.95e-06 ***
## married      0.199417   0.039050   5.107 3.98e-07 ***
## black       -0.188350   0.037667  -5.000 6.84e-07 ***
## south       -0.090904   0.026249  -3.463 0.000558 ***
## urban        0.183912   0.026958   6.822 1.62e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3655 on 927 degrees of freedom
## Multiple R-squared:  0.2526, Adjusted R-squared:  0.2469
## F-statistic: 44.75 on 7 and 927 DF,  p-value: < 2.2e-16
```

C8

(i)

O sinal de β_1 será positivo.

(ii)

```
reg5 <- lm(approve ~ white, data = loanapp)
summary(reg5)

##
## Call:
## lm(formula = approve ~ white, data = loanapp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.90839  0.09161  0.09161  0.09161  0.29221
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
##
```

```
## (Intercept) 0.70779 0.01824 38.81 <2e-16 ***
## white 0.20060 0.01984 10.11 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3201 on 1987 degrees of freedom
## Multiple R-squared: 0.04893, Adjusted R-squared: 0.04845
## F-statistic: 102.2 on 1 and 1987 DF, p-value: < 2.2e-16
```

Ele é estatisticamente significativo e grande do ponto de vista prático, sendo 20% mais provável que brancos consigam empréstimo do que não brancos.

(iii)

```
reg6 <- lm(approve ~ white + hrat + obrat + loanprc + unem + male + married + dep + sch + cosign + chis
summary(reg6)
```

```
##
## Call:
## lm(formula = approve ~ white + hrat + obrat + loanprc + unem +
##     male + married + dep + sch + cosign + chist + pubrec + mortlat1 +
##     mortlat2 + vr, data = loanapp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06482  0.00781  0.06387  0.13673  0.71105
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.936731   0.052735  17.763 < 2e-16 ***
## white        0.128820   0.019732   6.529 8.44e-11 ***
## hrat         0.001833   0.001263   1.451  0.1469
## obrat       -0.005432   0.001102  -4.930 8.92e-07 ***
## loanprc     -0.147300   0.037516  -3.926 8.92e-05 ***
## unem        -0.007299   0.003198  -2.282  0.0226 *
## male        -0.004144   0.018864  -0.220  0.8261
## married      0.045824   0.016308   2.810  0.0050 **
## dep         -0.006827   0.006701  -1.019  0.3084
## sch          0.001753   0.016650   0.105  0.9162
## cosign       0.009772   0.041139   0.238  0.8123
## chist        0.133027   0.019263   6.906 6.72e-12 ***
## pubrec      -0.241927   0.028227  -8.571 < 2e-16 ***
## mortlat1    -0.057251   0.050012  -1.145  0.2525
## mortlat2    -0.113723   0.066984  -1.698  0.0897 .
## vr          -0.031441   0.014031  -2.241  0.0252 *
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3021 on 1955 degrees of freedom
## (18 observations deleted due to missingness)
## Multiple R-squared: 0.1656, Adjusted R-squared: 0.1592
## F-statistic: 25.86 on 15 and 1955 DF, p-value: < 2.2e-16
```

A *dummy white* continua estatisticamente significativa, reduzindo seu peso para 13%.

(iv)

```
reg7 <- lm(approve ~ white + hrat + obrat + loanprc + unem + male + married + dep + sch + cosign + chis  
summary(reg7)
```

```
##  
## Call:  
## lm(formula = approve ~ white + hrat + obrat + loanprc + unem +  
##     male + married + dep + sch + cosign + chist + pubrec + mortlat1 +  
##     mortlat2 + vr + white * obrat, data = loanapp)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -1.05523  0.01253  0.06320  0.12692  0.83284  
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)  1.180648   0.086808  13.601  < 2e-16 ***  
## white        -0.145975   0.080263  -1.819  0.069109 .  
## hrat          0.001790   0.001260   1.421  0.155521  
## obrat        -0.012226   0.002216  -5.518  3.88e-08 ***  
## loanprc      -0.152536   0.037436  -4.075  4.79e-05 ***  
## unem         -0.007528   0.003189  -2.360  0.018352 *  
## male         -0.006015   0.018817  -0.320  0.749241  
## married       0.045536   0.016260   2.800  0.005154 **  
## dep          -0.007630   0.006686  -1.141  0.253905  
## sch           0.001777   0.016601   0.107  0.914787  
## cosign        0.017709   0.041081   0.431  0.666458  
## chist         0.129855   0.019227   6.754  1.90e-11 ***  
## pubrec       -0.240325   0.028149  -8.538  < 2e-16 ***  
## mortlat1     -0.062782   0.049891  -1.258  0.208400  
## mortlat2     -0.126845   0.066891  -1.896  0.058071 .  
## vr           -0.030540   0.013993  -2.183  0.029188 *  
## white:obrat  0.008088    0.002290   3.531  0.000423 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.3012 on 1954 degrees of freedom  
## (18 observations deleted due to missingness)  
## Multiple R-squared:  0.1709, Adjusted R-squared:  0.1641  
## F-statistic: 25.17 on 16 and 1954 DF,  p-value: < 2.2e-16
```

É significativo.

(v)

Será igual a $-0,146 + 0,008 \cdot 35 = 0,134$.