Exercícios em computador

C1 Os dados do arquivo `401K` são um subconjunto de dados analisados por Papke (1995) para estudar a relação entre a participação em um plano de pensão `401k` e a generosidade do plano. A variável `prate` é a porcentagem de trabalhadores aptos e com uma conta ativa; esta é a variável que gostaríamos de explicar. A medida da generosidade é a taxa de contribuição do plano, `mrate`. Esta variável mostra a quantia média com que a empresa contribui para o fundo trabalhista a cada US$ 1 de contribuição do trabalhador. Por exemplo, se a `mrate` = 0,50, então uma contribuição de US$ 1 do trabalhador corresponde a uma contribuição de US$ 0,50 da empresa.

```r
library(wooldridge)
data("k401k")
```

(i) Encontre a taxa de participação e a taxa de contribuição médias na amostra de planos.

```r
mean(k401k$prate) # media taxa de participacao
```

```
## [1] 87.36291
```

```r
mean(k401k$mrate) # media taxa de contribuicao
```

```
## [1] 0.7315124
```

(ii) Agora, estime a equação de regressão simples, e relate os resultados ao lado do tamanho da amostra e do R-quadrado.

$$\widehat{prate} = \hat{\beta}_0 + \hat{\beta}_1 mrate$$

```r
model <- lm(prate ~ mrate, data = k401k)
summary(model) # resultados
```

```
##
## Call:
## lm(formula = prate ~ mrate, data = k401k)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -82.303  -8.184   5.178  12.712  16.807
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  83.0755     0.5633  147.48   <2e-16 ***
## mrate         5.8611     0.5270   11.12   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 16.09 on 1532 degrees of freedom
## Multiple R-squared:  0.0747, Adjusted R-squared:  0.0741
## F-statistic: 123.7 on 1 and 1532 DF,  p-value: < 2.2e-16
```

(iii) Interprete o intercepto de sua equação. Interprete o coeficiente de `mrate`.

O intercepto $\beta_0 = 83,07$ indica o valor esperado ou estimado de `prate` quando `mrate` é igual a zero. Já o coeficiente β_1 indica que existe uma relação positiva entre `mrate` e `prate`, e que também a cada unidade acrescidade de `mrate` estima-se um aumento em `prate` de 5,86 unidades.

(iv) Encontre a prate prevista quando `mrate` = 3,5. Esta é uma previsão razoável? Explique o que está ocorrendo aqui.

$$\widehat{prate} = 83,07 + 5,86 * 3,5$$

```r
83.07 + (5.86 * 3.5)
```

```
## [1] 103.58
```

Este valor é impossível, dado que a taxa máxima de participação é de 100%. Isto ilustra que, especialmente quando as variáveis dependentes são limitadas, um modelo de regressão simples pode fornecer previsões estranhas para valores extremos da variável independente. Na amostra, existem apenas 34 valores de `mrate` maiores ou iguais a 3,5.

(v) Quanto da variação da prate é explicada pela `mrate`? Na sua opinião, isso é bastante?

Aproximadamento 7,4% da variação de `prate` é explicada por `mrate`. Este não é um valor alto, indicando que provavelmente existem outros fatores que influenciam a taxa de participação.

C2 O conjunto de dados do arquivo `CEOSAL2` contém informações sobre CEOs de corporações norte-americanas. A variável `salary` é a compensação anual, em milhares de dólares, e `ceoten` é o número prévio de anos como CEO da empresa.

```r
library(wooldridge)
data("ceosal2")
```

(i) Encontre o salário médio e a permanência média na amostra.

```r
mean(ceosal2$salary) # salario medio
```

```
## [1] 865.8644
```

```r
mean(ceosal2$ceoten) # permanencia media
```

```
## [1] 7.954802
```

(ii) Quantos CEOs estão em seu primeiro ano no cargo (isto é, ceoten = 0)? Qual é a permanência mais longa como CEO?

```
max(subset(ceosal2, ceoten == 0)$comten) # permanencia mais longa quando ceoten igual a 0
```

```
## [1] 33
```

(iii) Estime o modelo de regressão simples e registre seus resultados da forma usual. Qual é o aumento percentual previsto (aproximado) no salário quando se tem um ano a mais como CEO?

$$log(salary) = \beta_0 + \beta_0 ceoten + \mu$$

```
model <- lm(lsalary ~ ceoten, data = ceosal2) # estimando modelo
summary(model) # resultados
```

```
##
## Call:
## lm(formula = lsalary ~ ceoten, data = ceosal2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.15314 -0.38319 -0.02251  0.44439  1.94337
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.505498   0.067991  95.682   <2e-16 ***
## ceoten      0.009724   0.006364   1.528    0.128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6038 on 175 degrees of freedom
## Multiple R-squared:  0.01316,    Adjusted R-squared:  0.007523
## F-statistic: 2.334 on 1 and 175 DF,  p-value: 0.1284
```

$$\widehat{log(salary)} = 6,5055 + 0,0097 * ceoten$$
$$n = 177, \qquad R^2 = 0,01316$$

Um aumento de um ano a mais como CEO gera um aumento no salário em 0,97%.

C3 Use os dados do arquivo SLEEP75, de Biddle e Hamermesh (1990), para estudar se há uma compensação entre o tempo gasto dormindo por semana e o tempo gasto em um trabalho remunerado. Podemos usar qualquer variável como a variável dependente. Para materializar, estime o modelo

$$sleep = \beta_0 + \beta_1 totwrk + u,$$

em que `sleep` são os minutos dormidos à noite por semana e `totwrk` é o total de minutos trabalhados durante a semana.

```r
library(wooldridge)
data("sleep75")
```

```r
model <- lm(sleep ~ totwrk, data = sleep75) # estimando modelo
summary(model) # resultados
```

```
##
## Call:
## lm(formula = sleep ~ totwrk, data = sleep75)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2429.94  -240.25     4.91   250.53  1339.72
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3586.37695   38.91243  92.165   <2e-16 ***
## totwrk        -0.15075    0.01674  -9.005   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 421.1 on 704 degrees of freedom
## Multiple R-squared:  0.1033, Adjusted R-squared:  0.102
## F-statistic: 81.09 on 1 and 704 DF,  p-value: < 2.2e-16
```

(i) Registre seus resultados em uma equação junto com o número de observações e o R^2. O que o intercepto desta equação significa?

$$\widehat{sleep} = 3586,38 - 0,15 * totwrk$$
$$n = 706, \qquad R^2 = 0,1033$$

O intercepto da equação indica os minutos dormidos por semana quando a quantidade de minutos trabalhados na semana, `totwrk` for igual a zero.

(ii) Se `totwrk` aumentar 2 horas, quanto você estima que `sleep` cairá? Você acha que este é um efeito grande?

```r
model$coefficients[2] * 2 * 60 # totwrk é medido em horas, portanto precisa fazer a conversão
```

```
##   totwrk
## -18.0895
```

Um aumento de duas horas de trabalho na semana, reduzirá em 18 minutos dormidos durante a semana. Isso não parece ser um valor muito alto.

C4 Use os dados do arquivo `WAGE2` para estimar uma regressão simples que explique o salário mensal (`wage`) em termos da pontuação do QI (`IQ`).

```r
library(wooldridge)
data("wage2")
```

(i) Encontre o salário médio e o `IQ` médio da amostra. Qual é o desvio padrão amostral do `IQ`? (Pontuações de `IQ` são padronizadas, por isso, a média na população é 100 com um desvio padrão igual a 15.)

```r
mean(wage2$wage) # salário médio
```

```
## [1] 957.9455
```

```r
mean(wage2$IQ) # QI médio
```

```
## [1] 101.2824
```

```r
sd(wage2$IQ) # desvio-padrão de QI
```

```
## [1] 15.05264
```

(ii) Estime um modelo de regressão simples em que um aumento de um ponto em `IQ` altere `wage` em uma quantia constante de dólares. Use este modelo para encontrar o aumento previsto do salário para o caso de um acréscimo de 15 pontos de `IQ`. O `IQ` explica a maior parte da variação em `wage`?

```r
model <- lm(wage ~ IQ, data = wage2) # estimando o modelo
summary(model) # resultados
```

```
##
## Call:
## lm(formula = wage ~ IQ, data = wage2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -898.7 -256.5  -47.3  201.1 2072.6
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 116.9916    85.6415   1.366    0.172
## IQ            8.3031     0.8364   9.927   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 384.8 on 933 degrees of freedom
## Multiple R-squared:  0.09554,    Adjusted R-squared:  0.09457
## F-statistic: 98.55 on 1 and 933 DF,  p-value: < 2.2e-16
```

$$\widehat{salary} = 116,9 + 8,30(IQ)$$
$$n = 935, \qquad R^2 = 0,09554$$

Um aumento de 15 no `IQ` aumenta o salário mensal previsto em $8,30 * (15) = \$124,50$ (em dólares de 1980). A pontuação de `IQ` não explica nem 10% da variação salarial.

(iii) Agora, estime um modelo em que cada acréscimo de um ponto em `IQ` tenha o mesmo efeito percentual em `wage`. Se `IQ` aumentar 15 pontos, qual será o aumento percentual previsto aproximado em `wage`?

```r
model <- lm(lwage ~ IQ, data = wage2) # estimando modelo
summary(model)
```

```
##
## Call:
## lm(formula = lwage ~ IQ, data = wage2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.09324 -0.25547  0.02261  0.27544  1.21487
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.8869944  0.0890206   66.13   <2e-16 ***
## IQ          0.0088072  0.0008694   10.13   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3999 on 933 degrees of freedom
## Multiple R-squared:  0.09909,    Adjusted R-squared:  0.09813
## F-statistic: 102.6 on 1 and 933 DF,  p-value: < 2.2e-16
```

$$log(\widehat{salary}) = 5,89 + 0,008(IQ)$$
$$n = 935, \qquad R^2 = 0,09$$

Se variar em 15 então $\Delta log(salário) = 0,0088(15) = 0,132$, que é a mudança proporcional (aproximada) no salário previsto. O aumento percentual é, portanto, de aproximadamente 13,2.