

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, `mrte`. 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 `mrte` = 0,50, então uma contribuição de US\$ 1 do trabalhador corresponde a uma contribuição de US\$ 0,50 da empresa.

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

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

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

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

```
mean(k401k$mrte) # 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.

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

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

```
##
## Call:
## lm(formula = prate ~ mrte, 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 ***
## mrte          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 `mrte`.

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

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

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

```
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 `mrte` maiores ou iguais a 3,5.

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

Aproximadamente 7,4% da variação de `prate` é explicada por `mrte`. 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.

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

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

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

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

```
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?

$$\log(salary) = \beta_0 + \beta_1 ceoten + \mu$$

```
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
```

$$\widehat{\log(salary)} = 6,5055 + 0,0097 * ceoten$$

$$n = 177, \quad R^2 = 0,01316$$

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