

Análise de Regressão Linear Múltipla II

Aula 05

Gujarati e Porter, 2011 – Capítulos 7 e 8

Heij et al., 2004 – Capítulo 3

Exemplo

A senhorita Jolie, gerente do departamento de RH da empresa TEMCO, objetiva estudar o comportamento médio dos salários dos funcionários dos mais diversos setores da empresa. Para tanto, baseando-se numa amostra aleatória de 46 funcionários da empresa, ela propôs os seguintes modelos de regressão:

$$\text{salario} = \beta_0 + \beta_1 \text{educ} + \varepsilon \quad (1)$$

$$\text{salario} = \alpha_0 + \alpha_1 \text{anosemp} + v \quad (2)$$

$$\text{salario} = \delta_0 + \delta_1 \text{educ} + \delta_2 \text{anosemp} + \xi \quad (3)$$

Como a gerente pode avaliar a qualidade de ajuste dos modelos?

COEFICIENTE DE EXPLICAÇÃO

ou

COEFICIENTE DE DETERMINAÇÃO

Coeficiente de Explicação

Resultado: $SST = SSE + SSR$

Parcela da variabilidade de y que é explicada pelas variáveis do modelo

Parcela da variabilidade de y que **não** é explicada pelas variáveis do modelo

$$R^2 = \frac{SSE}{SST}$$

Proporção da variabilidade de y que é explicada pelo conjunto de variáveis explicativas.

Voltando ao Exemplo

$$\text{salario} = \beta_0 + \beta_1 \text{educ} + \varepsilon$$

Dependent Variable: SALARIO
 Method: Least Squares
 Date: 08/26/12 Time: 14:31
 Sample: 1 46
 Included observations: 46
 SALARIO=C(1)+C(2)*EDUC

| | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C(1) | 24581.70 | 2129.189 | 11.54510 | 0.0000 |
| C(2) | 3009.878 | 367.6294 | 8.187262 | 0.0000 |
| R-squared | 0.603715 | Mean dependent var | 39827.39 | |
| Adjusted R-squared | 0.594709 | S.D. dependent var | 10999.24 | |
| S.E. of regression | 7002.393 | Akaike info criterion | 20.58840 | |
| Sum squared resid | 2.16E+09 | Schwarz criterion | 20.66790 | |
| Log likelihood | -471.5331 | Hannan-Quinn criter. | 20.61818 | |
| F-statistic | 67.03125 | Durbin-Watson stat | 1.334781 | |
| Prob(F-statistic) | 0.000000 | | | |

$$\text{salario} = \alpha_0 + \alpha_1 \text{anosemp} + v$$

Dependent Variable: SALARIO
 Method: Least Squares
 Date: 08/26/12 Time: 14:23
 Sample: 1 46
 Included observations: 46
 SALARIO=C(1)+C(2)*ANOSEMP

| | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C(1) | 28394.16 | 1793.951 | 15.82772 | 0.0000 |
| C(2) | 1107.218 | 140.4476 | 7.883500 | 0.0000 |
| R-squared | 0.585491 | Mean dependent var | 39827.39 | |
| Adjusted R-squared | 0.576070 | S.D. dependent var | 10999.24 | |
| S.E. of regression | 7161.598 | Akaike info criterion | 20.63336 | |
| Sum squared resid | 2.26E+09 | Schwarz criterion | 20.71286 | |
| Log likelihood | -472.5673 | Hannan-Quinn criter. | 20.66314 | |
| F-statistic | 62.14957 | Durbin-Watson stat | 1.081824 | |
| Prob(F-statistic) | 0.000000 | | | |

Dependent Variable: SALARIO
 Method: Least Squares
 Date: 08/26/12 Time: 15:45
 Sample: 1 46
 Included observations: 46
 SALARIO=C(1)+C(2)*EDUC+C(3)*ANOSEMP

| | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|--------|
| C(1) | 23177.47 | 1769.732 | 13.09660 | 0.0000 |
| C(2) | 1916.489 | 379.2670 | 5.053139 | 0.0000 |
| C(3) | 672.3250 | 141.6725 | 4.745629 | 0.0000 |
| R-squared | 0.739927 | Mean dependent var | 39827.39 | |
| Adjusted R-squared | 0.727830 | S.D. dependent var | 10999.24 | |
| S.E. of regression | 5738.291 | Akaike info criterion | 20.21070 | |
| Sum squared resid | 1.42E+09 | Schwarz criterion | 20.32996 | |
| Log likelihood | -461.8462 | Hannan-Quinn criter. | 20.25538 | |
| F-statistic | 61.16907 | Durbin-Watson stat | 1.229794 | |
| Prob(F-statistic) | 0.000000 | | | |

$$\text{salario} = \delta_0 + \delta_1 \text{educ} + \delta_2 \text{anosemp} + \xi$$

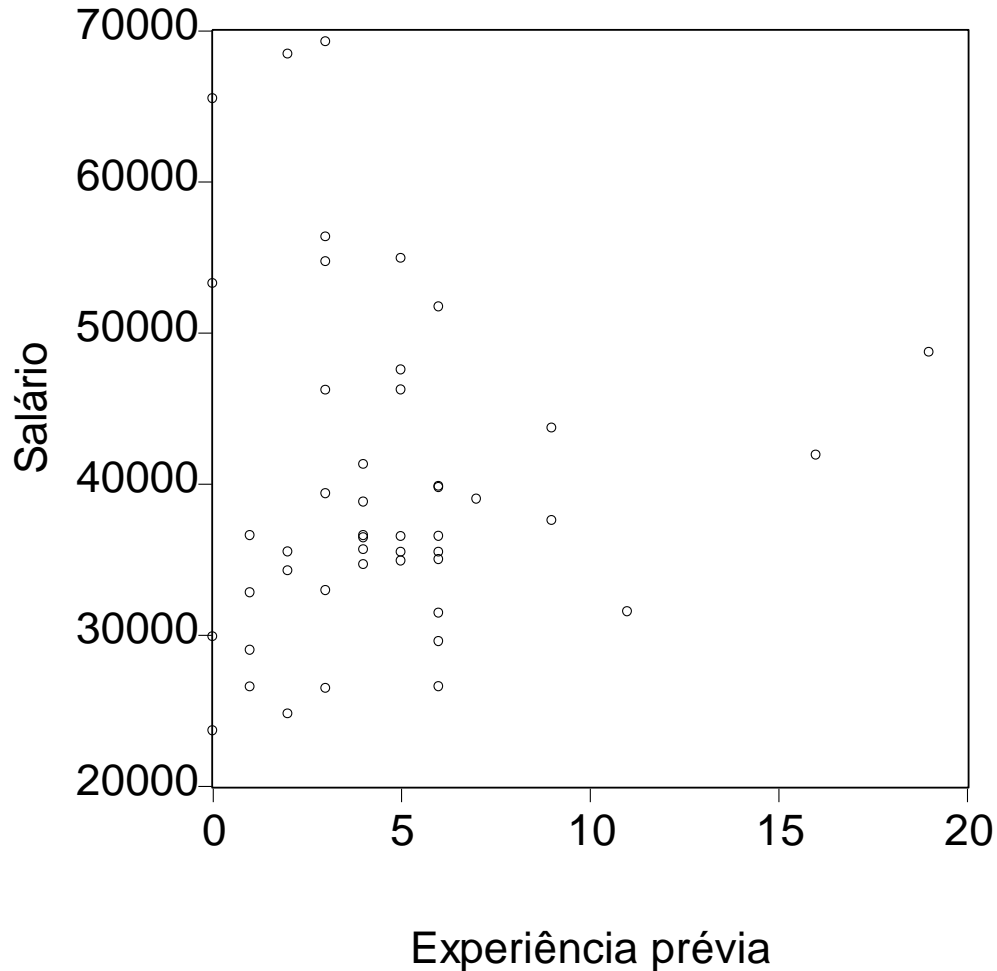
Voltando ao Exemplo

| Variáveis explicativas no modelo | R ² |
|----------------------------------|----------------|
| Educ | 60,4% |
| Anosemp | 58,6% |
| Educ e Anosemp | 74,0% |

Voltando ao Exemplo

O departamento de RH desconfia que a variável *EXPPREV* (experiência anterior, em anos) não é importante para explicar o salário dos funcionários, uma vez que os recém-contratados passam por um treinamento antes de iniciar as atividades na empresa. Pede-se, então: acrescente a variável ao modelo de regressão linear múltipla e verifique o que acontece com o R^2 ?

Salário vs EXPPREV



Correlação: **0,03**

Salário vs EXPPREV

Dependent Variable: SALARIO

Method: Least Squares

Date: 08/22/03 Time: 17:42

Sample: 1 46

Included observations: 46

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| C | 23480.46 | 2027.696 | 11.57987 | 0.0000 |
| EDUC | 1925.882 | 384.4395 | 5.009586 | 0.0000 |
| ANOSEMP | 671.3254 | 143.2125 | 4.687618 | 0.0000 |
| EXPPREV | -73.82734 | 232.7840 | -0.317150 | 0.7527 |
| R-squared | 0.740548 | Mean dependent var | | 39827.39 |
| Adjusted R-squared | 0.722016 | S.D. dependent var | | 10999.24 |
| S.E. of regression | 5799.262 | Akaike info criterion | | 20.25179 |
| Sum squared resid | 1.41E+09 | Schwarz criterion | | 20.41080 |
| Log likelihood | -461.7912 | F-statistic | | 39.95994 |
| Durbin-Watson stat | 1.250596 | Prob(F-statistic) | | 0.000000 |

Coeficiente de Determinação

Fato: Quanto maior o número de variáveis independentes, maior será o valor de R^2 .

Isso pode vir a ser um problema ao se comparar modelos, já que modelos com um número maior de variáveis tenderão a ter um R^2 maior do que um modelo, eventualmente equivalente, em termos de qualidade, com um número menor de variáveis.

R^2 – ajustado

Valor ajustado pelo número de variáveis

$$R_a^2 = 1 - (1 - R^2) \frac{n-1}{n-(k+1)}$$

O acréscimo de variáveis não acarreta necessariamente um aumento em R_a^2 .

Voltando ao Exemplo

| Variáveis explicativas no modelo | R^2 | R_a^2 |
|----------------------------------|-------|---------|
| Educ | 60,3% | 59,5% |
| Anosemp | 58,6% | 57,6% |
| Educ e Anosemp | 73,9% | 72,8% |
| Educ, Anosemp e Expprev | 74,1% | 72,2% |