Resultados adicionais 2

Lembrando, o coeficiente de correlação parcial é

$$r_{ZY|X} = r(Z^*, Y^*)$$

com Z^* - resíduo da regressão de Z em função de X

Y^* - resíduo da regressão de Y em função de X

i	Y	X	Z	$Z^* = Z - \hat{Z}$	$Y^* = Y - \hat{Y}$
1	y_1	x_1	z_1	z_1^*	y_1^*
2	y_2	x_2	z_2	z_2^*	z_2^*
$\vdots$	$\vdots$				
n	y_n	x_n	z_n	z_n^*	Y_n^*

$\hat{z}_i = a + b x_i$

$\hat{y}_i = c + d x_i$

resíduo $z_i^* = z_i - \hat{z}_i$

resíduo $Y_i^* = y_i - \hat{y}_i$

Z^* e Y^* são as variáveis originais Z e Y eliminado o efeito de X (que já foi adicionado ao modelo).

Modelo I $\quad \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \rightarrow SSR(I)$

Modelo III $\quad \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 z \rightarrow SSR(III)$

Os seguintes resultados estão nas duas últimas páginas do arquivo Aula 6:

Resultado 1

$$r^2_{ZY|X} = \frac{SSR(III) - SSR(I)}{SSE(I)}$$

Resultado 2

$$y = \beta_0 + \beta_1 x + \beta_2 z + \epsilon$$

$$F_p(Z|X) = \frac{r^2_{ZY|X}}{\frac{1 - r^2_{ZY|X}}{n-3}}$$

esta é a estatística do teste $\begin{cases} H_0: \beta_2 = 0 \text{ contra} \\ H_a: \beta_2 \neq 0 \end{cases}$.

No exemplo, tínhamos calculado

F_p para testar a introdução de Z dado que X

já está no modelo

$$F_p = \frac{SSR(III) - SSR(I)}{MSE(III)} = \frac{1138,73 - 810}{13,78} = 23,85$$

e

$r_{ZY|X} = -0,764 \qquad r^2_{ZY|X} = 0,5837$

Verificação do Resultado 1

$$\frac{SSR(III) - SSR(I)}{SSE(I)} = \frac{1138,73 - 810}{563} = 0,5838$$

$$= r^2_{ZY|X}$$

Verificação do Resultado 2

$$\frac{r^2_{ZY|X}}{\frac{1 - r^2_{ZY|X}}{n-3}} = \frac{0,5837}{\frac{1 - 0,5837}{17}} = 23,84 = F_p\,(Z|X)$$

Como obter $r_{ZY|X}$ somente a partir das somas de quadrados dos modelos I e III?

$$r_{ZY|X} = \pm\sqrt{\frac{SSR(III) - SSR(I)}{SSE(I)}}$$

Para decidir sobre o sinal, utilizar o resultado 2) dos comentários adicionais

$$\hat{\beta}_2 = r_{ZY|X}\sqrt{\frac{SSE(Y|X)}{SSE(Z|X)}}$$

Portanto, $r_{ZY|X}$ tem o mesmo sinal de $\hat{\beta}_2$ no modelo $\hat{y} = 126{,}59 + 0{,}650x - 0{,}454z$

$$\therefore\ r_{ZY|X} = -\sqrt{\frac{SSR(III) - SSR(I)}{SSE(I)}}$$

Gráfico da Variável Adicionada

Para avaliar a importância da introdução de uma nova variável explicativa no modelo, um gráfico de resíduos útil é o gráfico da variável adicionada.

Consiste no gráfico dos resíduos da regressão da variável resposta em função da variável já incluída (Y^*) em função dos resíduos da regressão da variável a ser adicionada em função da variável já incluída (Z^*).

$Y^* \times Z^*$ Y^* - variável Y eliminado o efeito de X

Z^* - variável Z eliminado o efeito de X.

Uma tendência nesse gráfico é indicativo que a variável resposta e a variável <u>a ser</u> incluída são correlacionadas eliminado o efeito da variável <u>já</u> incluída.

Verifica-se que o coeficiente angular da reta de mínimos quadrados de Y^* em função de Z^* é o valor de $\hat{\beta}_2$ na equação $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 z$.

Este fato é consequência do Resultado Adicional 2):

$$\hat{\beta}_2 = \frac{\sum (y - \hat{y}_2)(x_2 - \hat{x}_2)}{\sum (x_2 - \hat{x}_2)}.$$

Coeficiente de Explicação Ajustado para graus de liberdade

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-3}.$$

No exemplo

$$\bar{R}^2 = 1 - (1 - 0{,}83)\frac{19}{17} = 0{,}81.$$

Interações em Modelos de Regressão

Quando utilizamos o modelo $E(y|x,z) = \beta_0 + \beta_1 x + \beta_2 z$, estamos admitindo que o acréscimo médio em Y quando se aumenta uma unidade em X mantido o valor de Z constante é igual a β_1 qualquer que seja o valor de Z.

$$E(y|x+1,z) - E(y|x,z) = \beta_0 + \beta_1 x + \beta_1 + \beta_2 z - \beta_0 - \beta_1 x - \beta_2 z = \beta_1.$$

Vale uma interpretação semelhante para β_2.

Exemplo

Y - vendas de um produto em uma certa região

X_1 - gasto com propaganda na região

X_2 - gasto com propaganda na TV.

Consideremos que temos apenas $X_2 = 1$ e $X_2 = 3$.

Supondo o modelo

$E(y) = 10 + 2x_1 + 5x_2$,

segue que

$$E(y|x_2=1) = 10 + 2x_1 + 5 = 15 + 2x_1$$ e

$$E(y|x_2=3) = 10 + 2x_1 + 5 \cdot 3 = 25 + 2x_1 ,$$

que corresponde a duas retas paralelas.

Neste caso, o acréscimo médio em Y quando se aumenta uma unidade em X_1 seria igual a 2 para qualquer valor de X_2, ou seja, para $x_2=1$ e $x_2 = 3$.

No entanto, se a verdadeira relação entre Y e X_1, X_2 for da forma

$$E(y|x_1, x_2) = 10 + 2x_1 + 5x_2 + 0{,}5\,x_1 x_2$$ teríamos

$$E(y|x_2=1) = 15 + 2{,}5\,x_1$$ e

$$E(y|x_2=3) = 25 + 3{,}5\,x_1 .$$

Nesse caso, fixado X_2 (gasto com propaganda na TV), o acréscimo médio em Y quando se aumenta uma unidade na propaganda local seria 2,5 para $x_2=1$ e 3,5 para $x_2 = 3$. O efeito seria maior para maiores níveis de gastos com propaganda de TV.

É essa a interpretação da interação no modelo de regressão.

Além disso, aqui,

$$E(y \mid x_1+1, x_2) - E(y \mid x_1, x_2) =$$

$$= \beta_0 + \beta_1(x_1+1) + \beta_2 x_2 + \beta_3(x_1+1)x_2$$

$$- \beta_0 - \beta_1 x_1 - \beta_2 x_2 - \beta_3 x_1 x_2 =$$

$$= \beta_1 + \beta_3 x_2$$

e não mais β_1, como no modelo sem interação.