Gabarito da Lista 5 – Tópico 4

a)

r_{yx1} = 0,7307

r_{yx2} = 0,8163

b)

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + e_j,$$

onde:

y_j é a quantidade de calor por grama de cimento;

x_{1j} é a quantidade de aluminato tricálcico;

x_{2j} é a quantidade de silicato tricálcico;

β_0 é quantidade de calor do cimento, quando as quantidades de aluminato tricálcico e silicato tricálcico são iguais a zero (X_1=0 e X_2=0);

β_1 é a taxa de variação a quantidade de calor, para cada unidade que se acrescenta na quantidade de aluminato tricálcico, numa quantidade fixa qualquer de silicato tricálcico;

β_2 é a taxa de variação a quantidade de calor, para cada unidade que se acrescenta na quantidade de silicato tricálcico, numa quantidade fixa qualquer de aluminato tricálcico;

e_j é o erro (variação aleatória) associado à observação j.

c) Se Y é uma variável aleatória, então, está sujeita a um erro de observação. Este erro (e_i) deverá ser adicionado ao modelo, desde que se admitam como verdadeiras as seguintes pressuposições:

1. As variáveis Xi são fixas, isto é, observados sem erro.
2. Os erros são aleatórios, têm média zero e variância constante, ou seja, $E(e_i)=0$ e $V(e_i)=\sigma^2$.
3. Os erros têm distribuição normal e são independentes entre si.
4. O modelo é adequado para todas as observações, não podendo haver nenhum valor de X que produza um valor de Y discrepante dos demais.

d)

$\hat{\beta}_0 = 52{,}58$

$\hat{\beta}_1 = 1{,}468$

$\hat{\beta}_2 = 0{,}6623$

$$\hat{\mu} = 52{,}58 + 1{,}468\, x_1 + 0{,}6623\, x_2$$

e)

1. Hipótese estatística

$$\begin{cases} H_0 : \beta_i = 0, \text{ sendo } i = 1, 2 \\ H_1 : \beta_i \neq 0, \text{ para pelo menos um } i \ (i = 1 \text{ e/ou } 2) \end{cases}$$

2. Taxa de erro: α = 0,05

3. Cálculo das somas de quadrados

SQ_{Total} = SQY = 2715,76

$SQ_{Reg} = \beta_1 \cdot SPYX_1 + \beta_2 \cdot SPYX_2$= 2657,86

$SQ_{Resíduo}$ = SQTotal – SQReg = 2715,76 – 2657,86 = 57,904

Tabela da análise da variação:

Fontes	GL	SQ	QM	F
Regressão	2	2657,86	1328,929	229,50
Resíduo	10	57,904	5,7904	-
Total	12	2715,763	-	-

Como $f_{\alpha(2;\,10)}$ = 4,10 < f = 229,50, rejeitamos H_0.

Concluímos ao nível de 5% de significância que pelo menos um dos coeficientes de regressão parciais difere de zero. Portanto, existe relação linear significativa entre a quantidade de calor e pelo menos uma das variáveis quantidade de aluminato tricálcico e quantidade de silicato tricálcico.

f)

r^2_c = 0,9744

g)

1. Hipótese estatística

$$\begin{cases} H_0^1 : \beta_1 = 0 \\ H_1^1 : \beta_1 \neq 0 \end{cases}$$

2. Taxa de erro: α = 0,05

3. Estatística do teste

t = 12,11

4. Decisão e conclusão

Como $t_{\alpha/2(10)}$ = 2,228 < t = 12,11, rejeitamos H_0.

Concluímos ao nível de 5% de significância o coeficiente de regressão β_1 difere de zero. Portanto, existe efeito significativo da quantidade de aluminato tricálcico sobre a quantidade de calor, adicional ao efeito da quantidade de silicato de tricálcico.

1. Hipótese estatística

$$\begin{cases} H_0^2 : \beta_2 = 0 \\ H_1^2 : \beta_2 \neq 0 \end{cases}$$

2. Taxa de erro: $\alpha = 0{,}05$

3. Estatística do teste

$t = 14{,}44$

4. Decisão e conclusão

Como $t_{\alpha/2(10)} = 2{,}228 < t = 14{,}44$, rejeitamos H_0.

Concluímos ao nível de 5% de significância o coeficiente de regressão β_2 difere de zero. Portanto, existe efeito significativo da quantidade de silicato tricálcico sobre a quantidade de calor, adicional ao efeito da quantidade de aluminato tricálcico.

h) A relação linear entre Y e (X_1, X_2) foi significativa, as contribuições adicionais das variáveis X_1 e X_2 para a explicação da variação de Y foram significativas. Assim, segundo os testes efetuados, o "melhor" modelo para exprimir a relação linear entre Y e (X_1, X_2) é o modelo de regressão linear múltipla com duas variáveis.

i) Valor esperado de y (x_1=15 e x_2=30) = 94,47.

j) Descreva resumidamente os principais métodos de seleção de variáveis.

$\Rightarrow$ Inclusão ascendente (*forward selection*): inicia-se com um modelo que possui somente o intercepto e, de acordo com o critério fixado, as variáveis preditoras são incluída no modelo, uma a uma. Uma vez incluída no modelo, a variável não sai mais.

$\Rightarrow$ Seleção descendente (*backward elimination*): começa com o modelo completo e, de acordo com o critério fixado, vai excluindo, uma a uma, as variáveis de menor contribuição não significativa, na presença das demais variáveis no modelo.

$\Rightarrow$ Seleção ascendente-descendente (*stepwise selection*) é uma aplicação conjunta dos critérios de inclusão e exclusão. O procedimento inicia do mesmo modo que a seleção ascendente, mas em cada passo verifica se, na presença das outras variáveis do modelo, alguma variável não agrega contribuição significativa à explicação da variação da resposta. Dentre as que não estão contribuindo significativamente, a de menor f parcial é eliminada. Por outro lado, uma variável que já foi excluída poderá retornar em um passo posterior.