Provas de Contabilometria

Equipe de monitores de Contabilometria - UFPE 2024.2

21 de fevereiro de 2025

1 Prova da Tarde

Letra a

N	Ativo (Y)	NProcessos (X)	$X \times Y$	X^2
1	20	8	160	64
2	40	5	200	25
3	5	15	75	225
4	25	8	200	64
5	2	20	40	400
$\sum$	92	56	675	778

β_1 pode ser definido da seguinte forma:

$$\beta_1 = \frac{n \times \sum(X_iY_i) - \sum Y_i \sum X_i}{n \times \sum X^2 - (\sum X_i)^2}$$

Agora podemos utilizar os valores da tabela criada para calcular β_1.

$$\beta_1 = \frac{5 \times 675 - 92 \times 56}{5 \times 778 - (56)^2} = -2,35676$$

β_0 pode ser definido como $\beta_0 = \bar{Y} - \beta_1 \times \bar{X}$, logo:

$$\beta_0 = \frac{92}{5} + 2,35676 \times \frac{56}{5} = 44,79576$$

Agora que temos os valores de β_0 e β_1 podemos interpretar o seguinte: na ausência de processos (quando o número de processos = 0) o valor do ativo da empresa é, em média, de R$ 44,79576 milhões de reais. O β_1 nos informa que, quando o número de processos aumenta em uma unidade, o valor do ativo decresce, em média, em R$ 2,35676 milhões de reais.

letra b

A variável setor é uma variável qualitativa nominal, porém pode ser expressa como uma variável binária. Para tanto podemos escrever que *serviço* = 0 e *comércio* = 1.

Com isso temos a seguinte tabela:

Ativo	N Processos	Setor
20	8	0
40	5	1
5	15	0
25	8	1
2	20	0

A partir disso podemos construir a matriz X.

$$X = \begin{bmatrix} 1 & 8 & 0 \\ 1 & 5 & 1 \\ 1 & 15 & 0 \\ 1 & 8 & 1 \\ 1 & 20 & 0 \end{bmatrix}$$

Também podemos construir a matriz Y.

$$Y = \begin{bmatrix} 20 \\ 40 \\ 5 \\ 25 \\ 2 \end{bmatrix}$$

Para proceder ao cálculo de $(X'X)^{-1}$ primeiro precisamos escrever X'.

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 8 & 5 & 15 & 8 & 20 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Logo, multiplicando X' por X, obtemos:

$$X'X = \begin{bmatrix} 5 & 56 & 2 \\ 56 & 778 & 13 \\ 2 & 13 & 2 \end{bmatrix}$$

Invertendo a matriz $X'X$ (aqui também podemos utilizar as frações obtidas) obtemos:

$$(X'X)^{-1} = \begin{bmatrix} 2,99568 & 0,18574 & -1,78834 \\ 0,18574 & 0,01296 & 0,10151 \\ -1,78834 & 0,10151 & 1,62851 \end{bmatrix}$$

Multiplicamos agora a matriz X' pela matriz Y:

$$X'Y = \begin{bmatrix} 92 \\ 675 \\ 65 \end{bmatrix}$$

Por fim obtemos o produto de $(X'X)^{-1}$ por $X'Y$:

$$(X'X)^{-1}X'Y = \beta = \begin{bmatrix} 33,98272 \\ -1,74298 \\ 9,84665 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Após calculados os valores de β podemos interpretar que: para empresas do setor de *serviços* e que não possuem nenhum processo, o valor do ativo é, em média, de R$ 33,98272 (β_0) milhões de reais. Quando o número de processos aumenta em 1 unidade, o valor do ativo diminui, em média, em R$ 1,74298 (β_1) milhões de reais. As empresas do setor de *comércio* têm, em média, um ativo R$ 9,85665 (β_2) milhões de reais maior que os das empresas do setor de *serviços*.

Para obtermos a matriz de variância-covariância precisamos primeiro calcular S^2. Sabemos que:

$$\beta_0 = 33,99272$$
$$\beta_1 = -1,74298$$
$$\beta_2 = 9,84665$$

Primeiro calcularemos a soma do quadrado dos resíduos (SQR).

$$[20-(\beta_0+\beta_1\times 8+\beta_2\times 0)]^2 = 0,00151$$
$$[40-(\beta_0+\beta_1\times 5+\beta_2\times 1)]^2 = 23,86840$$
$$[5-(\beta_0+\beta_1\times 15+\beta_2\times 0)]^2 = 8,05436$$
$$[25-(\beta_0+\beta_1\times 8+\beta_2\times 1)]^2 = 23,86840$$
$$[2-(\beta_0+\beta_1\times 20+\beta_2\times 0)]^2 = 8,25644$$
$$\sum u^2 = 64,06911$$

Sabemos que:

$$S^2 = \frac{\sum u^2}{n-k}, \text{ onde } k = 3$$
$$S^2 = \frac{64,06911}{5-3} = 32,03456$$

Sabemos ainda que:

$$(X'X)^{-1} = \begin{bmatrix} 2,99568 & 0,18574 & -1,78834 \\ 0,18574 & 0,01296 & 0,10151 \\ -1,78834 & 0,10151 & 1,62851 \end{bmatrix}$$

A partir disso podemos calcular a variância de β_1

$$Var(\beta_1) = S^2 \times 0,01296 = 32,03456 \times 0,01296 = 0,4151$$

Podemos calcular o nosso intervalo a partir disso:

$$[\beta_1^* - Var(\beta_1^*); \beta_1^* + Var(\beta_1^*)]$$

$$[-1,74298 - 0,4151; -1,74298 + 0,4151]$$

$$[-2,1581; -1,3278]$$

O β_1 da regressão simples é igual a $-2,35676$ e está, portanto, fora do intervalo definido. Logo, concluímos que **há** evidências de que H_1 foi violada.

letra c

Sabemos que $R^2 = 1 - \frac{SQR}{SQT}$.

E da letra b) sabemos que $SQR = 64,06911$.

Calculemos agora SQT.

$$\bar{Y} = \frac{92}{5} = 18,4$$

$$SQT = \sum(Y - \bar{Y})^2$$

$$(Y_1 - \bar{Y})^2 = (20 - 18,4)^2 = 2,56$$

$$(Y_2 - \bar{Y})^2 = (40 - 18,4)^2 = 466,56$$

$$(Y_3 - \bar{Y})^2 = (5 - 18,4)^2 = 179,56$$

$$(Y_4 - \bar{Y})^2 = (25 - 18,4)^2 = 43,56$$

$$(Y_5 - \bar{Y})^2 = (2 - 18,4)^2 = 268,96$$

$$SQT = \sum(Y - \bar{Y})^2 = 961,2$$

Portanto podemos calcular SQT da seguinte forma:

$$R^2 = 1 - \frac{SQR}{SQT} = 1 - \frac{64,06911}{961,2} = 0,93334$$

Isso nos diz que 93,33% da variabilidade de Y é explicada pelos seus regressores, ou seja, quão bom é o ajuste da nossa reta de regressão aos dados.

letra d

Para atender ao Teorema de Gauss-Markov precisamos **necessariamente** atender $H_1 - H_4$. Vimos na letra b) que possivelmente o modelo da letra a) violou H_1. Entretanto nada sugere que o modelo da letra b) também violou H_1. Como foi necessário assumirmos $H_1 - H_2$ para construção do modelo, e caso não haja correlação entre os erros de diferentes observações, bem como a variância dos erros deverá ser constante, podemos dizer que o modelo da letra; b) atende ao teorema de Gauss-Markov, diferentemente do modelo da letra a).

letra e

Como o modelo possui um regressor a mais, não podemos comparar diretamente os dois modelos, pois R^2 é uma função não decrescente em X. Para tanto precisamos utilizar o R^2 ajustado.

O R^2 ajustado é dado por:

$$\bar{R}^2 = 1 - \frac{SQR}{SQT} \times \frac{n-1}{n-k}$$

Logo,

$$\bar{R}^2_{multipla} = 1 - \frac{64,06911}{961,2} \times \frac{5-1}{5-3} = 0,86669$$

Como informado na questão, $\bar{R}^2_{simples} = 0,8285$, portanto o modelo que melhor se ajusta aos dados é o da regressão múltipla.

2 Prova da Noite

letra a

N	Ativo (Y)	NProcessos (X)	$X \times Y$	X^2
1	20	8	160	64
2	40	5	200	25
3	5	15	75	225
4	25	8	200	64
5	2	20	40	400
$\sum$	92	56	675	778

β_1 pode ser definido da seguinte forma:

$$\beta_1 = \frac{n \times \sum(X_i Y_i) - \sum Y_i \sum X_i}{n \times \sum X^2 - (\sum X_i)^2}$$

Agora podemos utilizar os valores da tabela criada para calcular β_1.

$$\beta_1 = \frac{5 \times 675 - 92 \times 56}{5 \times 778 - (56)^2} = -2,35676$$

β_0 pode ser definido como $\beta_0 = \bar{Y} - \beta_1 \times \bar{X}$, logo:

$$\beta_0 = \frac{92}{5} + 2,35676 \times \frac{56}{5} = 44,79576$$

Agora que temos os valores de β_0 e β_1 podemos interpretar o seguinte: na ausência de processos (quando o número de processos = 0) o valor do ativo da empresa é, em média, de R$ 44,79576 milhões de reais. O β_1 nos informa que, quando o número de processos aumenta em uma unidade, o valor do ativo decresce, em média, em R$ 2,35676 milhões de reais.

letra b

A variável setor é uma variável qualitativa nominal, porém pode ser expressa como uma variável binária. Para tanto podemos escrever que *serviço* $= 0$ e *comércio* $= 1$. Com isso temos a seguinte tabela:

Ativo	N Processos	Setor
20	8	0
40	5	0
5	15	1
25	8	1
2	20	0

A partir disso podemos construir a matriz X.

$$X = \begin{bmatrix} 1 & 8 & 0 \\ 1 & 5 & 0 \\ 1 & 15 & 1 \\ 1 & 8 & 1 \\ 1 & 20 & 0 \end{bmatrix}$$

Também podemos construir a matriz Y.

$$Y = \begin{bmatrix} 20 \\ 40 \\ 5 \\ 25 \\ 2 \end{bmatrix}$$

Para proceder ao cálculo de $(X'X)^{-1}$ primeiro precisamos escrever X'.

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 8 & 5 & 15 & 8 & 20 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Logo, multiplicando X' por X, obtemos:

$$X'X = \begin{bmatrix} 5 & 56 & 2 \\ 56 & 778 & 23 \\ 2 & 23 & 2 \end{bmatrix}$$

Invertendo a matriz $X'X$ (aqui também podemos utilizar as frações obtidas) obtemos:

$$(X'X)^{-1} = \begin{bmatrix} 1,13732 & -0,07390 & -0,29679 \\ -0,07390 & 0,00664 & -0,00332 \\ -0,29679 & -0,00332 & 0,83499 \end{bmatrix}$$

Multiplicamos agora a matriz X' pela matriz Y:

$$X'Y = \begin{bmatrix} 92 \\ 675 \\ 30 \end{bmatrix}$$

Por fim obtemos o produto de $(X'X)^{-1}$ por $X'Y$:

$$(X'X)^{-1}X'Y = \beta = \begin{bmatrix} 46,39424 \\ -2,33887 \\ -4,49723 \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

Após calculados os valores de β podemos interpretar que: para empresas do setor de *serviços* e que não possuem nenhum processo, o valor do ativo é, em média, de R$ 46,39424 (β_0) milhões de reais. Quando o número de processos aumenta em 1 unidade, o valor do ativo diminui, em média, em R$ 2,33887 (β_1) milhões de reais. As empresas do setor de *comércio* têm, em média, um ativo R$ 4,49723 (β_2) milhões de reais menor que os das empresas do setor de *serviços*.

Para saber se existem evidências de que a hipótese de exogeneidade foi violada no modelo da letra a), podemos calcular o nosso intervalo da seguinte forma:

$$[\beta_1^* - \beta_1^* \times 25\%; \beta_1^* + \beta_1^* \times 25\%]$$

$$\beta_1^* \times 25\% = -2,33887 \times \frac{25}{100} = -0,58472$$

$$[-2,33887 - 0,58472; -2,33887 + 0,58472]$$

$$[-2,92359; -1,75415]$$

O β_1 da regressão simples é igual a $-2,35676$ e está, portanto, dentro do intervalo definido. Logo, concluímos que **não há** evidências de que H_1 foi violada.

letra c

Para obtermos R^2 usaremos $R^2 = 1 - \frac{SQR}{SQT}$

Sabendo que:

$$\bar{Y} = \frac{92}{5} = 18,4$$

$$\beta_0 = 46,39424$$

$$\beta_1 = -2,33887$$

$$\beta_2 = -4,49723$$

Primeiro calcularemos SQR

$$SQR = \sum (Y - \widehat{Y})^2$$

$$[20 - (\beta_0 + \beta_1 \times 8 + \beta_2 \times 0)]^2 = 59,03279$$

$$[40 - (\beta_0 + \beta_1 \times 5 + \beta_2 \times 0)]^2 = 28,09117$$

$$[5 - (\beta_0 + \beta_1 \times 15 + \beta_2 \times 1)]^2 = 3,29045$$

$$[25 - (\beta_0 + \beta_1 \times 8 + \beta_2 \times 1)]^2 = 3,29041$$

$$[2 - (\beta_0 + \beta_1 \times 20 + \beta_2 \times 0)]^2 = 5,67945$$

$$SQR = 99,38427$$

Agora calcularemos SQT

$$SQT = \sum(Y - \bar{Y})^2$$

$$(20 - 18,4)^2 = 2,56$$

$$(40 - 18,4)^2 = 466,56$$

$$(5 - 18,4)^2 = 179,56$$

$$(25 - 18,4)^2 = 43,56$$

$$(2 - 18,4)^2 = 268,96$$

$$SQT = 961,2$$

Logo,

$$R^2 = 1 - \frac{99,38427}{961,2} = 0,89660$$

Isso nos diz que 89,66% da variabilidade de Y é explicada pelos seus regressores. Ou seja, quão boa a nossa reta de regressão se ajusta aos dados.

Além disso, não é possível que R^2 da regressão simples seja maior que o R^2 da regressão múltipla, pois R^2 é uma função função que sempre cresce com a inclussão de novos regressores. Portanto, o R^2 de uma regressão múltipla **sempre** será maior que o R^2 de uma regressão simples. Caso queiramos comparar o poder avaliativo de dois modelos com um número distinto de regressores, precisaríamos utilizar o R^2 ajustado.

letra d

As variáveis setor e capital aberto são variáveis qualitativas nominais, porém podem ser expressas como variáveis binárias. Para tanto podemos escrever que $serviço = 0$, $comércio = 1$, $sim = 0$ e $não = 1$. Com isso temos a seguinte tabela:

Ativo	Setor	Cap. Abe.
20	0	0
40	0	1
5	1	0
25	1	1
2	0	0

Para que seja possível utilizar o MQO precisamos atender $H_1 - H_2$. Se considerarmos que os fatores não observáveis não estão correlacionados com as variáveis explicativas, H_1 estará atendida. Para tanto, basta que o modelo inclua uma constante (β_0) e esta hipótese estará atendida.

Para verificar H_2 precisamos mostrar que $det(X'X) \neq 0$

Faremos isso considerando o seguinte:

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Logo,

$$X'X = \begin{bmatrix} 5 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$$

E seu determinante é:

$$det(X'X) = 7$$

Portanto é possível estimar esse efeito utilizando as variáveis Setor e Capital Aberto.

letra e

Mostramos que o modelo anterior atende H_1 e H_2, logo, é possível construir um modelo de regressão **linear** a partir dos dados. Considerando não haver correlação entre os erros de diferentes observações e que a variância dos erros é constante, o modelo anterior atenderá ao teorema de Gauss-Markov ($H_1 - H_4$ atendidas).