FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE DE RIBEIRÃO PRETO
FEA-RP/USP

Ribeirão Preto - Set/2018

GABARITO - Lista 3

Professor: Luiz Guilherme Dácar da Silva Scorzafave

REC2301 - Econometria I

Monitora: Laura Ogando

1 Questões teóricas

Questão 1

a) A estimação por MQO se dá a partir da minimização da soma dos quadrados dos resíduos da regressão. Assim,

$$(\hat{\delta}_{MQO}, \hat{\gamma_1}_{MQO}) = \underset{(\delta,\gamma_1)}{argmin} \sum_{i=1}^{N} (u_i)^2 = \underset{(\delta,\gamma_1)}{argmin} \sum_{i=1}^{N} (Y_i - \delta - \gamma_1 X_i)^2.$$

Derivando em relação a $\hat{\delta}$ e $\hat{\gamma_1}$ e igualando a zero, chegamos às Condições de Primeira Ordem:

$$\frac{\partial(\cdot)}{\partial\hat{\delta}} = \sum_{i=1}^{N} (Y_i - \hat{\delta} - \hat{\gamma_1} X_i) = \sum_{i=1}^{N} \hat{u_i} = 0 \text{ e} \tag{1}$$

$$\frac{\partial(\cdot)}{\partial\hat{\gamma_1}} = \sum_{i=1}^{N} \Big[(Y_i - \hat{\delta} - \hat{\gamma_1} X_i) X_i \Big] = \sum_{i=1}^{N} \hat{u_i} X_i = 0. \tag{2}$$

Resolvendo primeiro (1) utilizando propriedades do operador de somatório, chegamos em:

$$\sum_{i=1}^{N} Y_i - \sum_{i=1}^{N} \hat{\delta} - \sum_{i=1}^{N} \hat{\gamma_1} X_i = 0$$

$$\sum_{i=1}^{N} Y_i - N\hat{\delta} - \hat{\gamma_1} \sum_{i=1}^{N} X_i = 0$$

$$\sum_{i=1}^{N} Y_i - \hat{\gamma_1} \sum_{i=1}^{N} X_i = N\hat{\delta}$$

Finalmente chegamos em:

$$\hat{\delta} = \bar{Y} - \hat{\gamma_1} \bar{X} \tag{3}$$

Substituindo esse valor em (2):

$$\sum_{i=1}^{N} \big[(Y_i - (\bar{Y} - \hat{\gamma_1} \bar{X}) - \hat{\gamma_1} X_i) X_i \big] = 0$$

$$\sum_{i=1}^{N} \big[(Y_i - \bar{Y} + \hat{\gamma_1} \bar{X} - \hat{\gamma_1} X_i) X_i \big] = 0$$

$$\sum_{i=1}^{N} \big[(Y_i - \bar{Y} - \hat{\gamma_1} (X_i - \bar{X})) X_i \big] = 0$$

$$\sum_{i=1}^{N} (Y_i - \bar{Y}) X_i - \hat{\gamma_1} \sum_{i=1}^{N} (X_i - \bar{X}) X_i = 0$$

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y}) X_i}{\sum_{i=1}^{N} (X_i - \bar{X}) X_i}$$

Finalmente, chegamos em:

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})X_i}{\sum_{i=1}^{N}(X_i - \bar{X})X_i} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})Y_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2} \tag{4}$$

b) Para obter o estimador de MQO de γ_1 são necessárias as três primeiras hipóteses: linearidade nos parâmetros (RLS 1), amostragem aleatória (RLS 2) e variação amostral em X_i (RLS 3). Sabemos que $\hat{\gamma}_1$ pode ser escrito, a partir da última expressão de (4) (basta substituir Y_i por $\delta + \gamma_1 X_i + u_i$), como:

$$\hat{\gamma}_1 = \gamma_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})u_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2}.$$

Assim, passando a esperança condicional em X, temos que:

$$\mathbb{E}[\hat{\gamma}_1|X_i] = \mathbb{E}\left[\gamma_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})u_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\middle| X_i\right] = \mathbb{E}[\gamma_1|X_i] + \mathbb{E}\left[\frac{\sum_{i=1}^{N}(X_i - \bar{X})u_i}{\sum_{i=1}^{N}(X_i - \bar{X})^2}\middle| X_i\right].$$

Usando o fato de que γ_1 é uma constante e de que condicional a X_i, todos os termos que envolvem X_i são constantes, temos que:

$$\mathbb{E}[\hat{\gamma}_1|X_i] = \gamma_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})\mathbb{E}[u_i|X_i]}{\sum_{i=1}^{N}(X_i - \bar{X})^2}.$$

Ora, se for válido que $\mathbb{E}[u_i|X_i] = 0$ (RLS 4), então teremos que:

$$\mathbb{E}[\hat{\gamma}_1|X_i] = \gamma_1 + \frac{\sum_{i=1}^{N}(X_i - \bar{X})\cdot 0}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \gamma_1.$$

Passando a esperança nos dois lados e usando a Lei das Expectativas Iteradas[1](L.E.I.),temos que:

$$\mathbb{E}\left[\mathbb{E}[\hat{\gamma}_1|X_i]\right] = \mathbb{E}[\hat{\gamma}_1] = \gamma_1 = \mathbb{E}[\gamma_1].$$

E, portanto, temos que o estimador de γ_1 é não viesado.

c) A hipótese de que $\mathbb{E}(u_i|X_i) = 0$ é a hipótese que de o valor médio de u_i não depende de X_i, também conhecida como hipótese de média condicional zero. Essa hipótese tem papel fundamental na discussão da diferença entre os conceitos de *correlação* e *causalidade*. O primeiro trata apenas de uma dependência linear, ou seja, uma ligação entre dois eventos ou variáveis. Já o segundo é mais forte: trata do efeito de um evento/variável em cima de outro, ou seja, porque X ocorreu, então Y necessariamente ocorrerá.

Em Econometria buscamos sempre relações de causalidade entre variáveis e precisamos ter o cuidado de não tirar conclusões equivocadas. As vezes as variáveis são

[1]L.E.I.: Sejam Z e W duas V.A. Então vale que: $\mathbb{E}\left[\mathbb{E}(Z|W)\right] = \mathbb{E}(Z)$

apenas correlacionadas, não possuindo nenhum tipo de causalidade, ou então uma terceira variável que causa as duas simultaneamente. Até mesmo é possível que as duas possuam causalidade uma na outra (simultaneidade). É muito importante ter a noção de que muitos fatores podem afetar a variável de interesse e, nesse sentido, torna-se complicado isolar o efeito de uma única variável sobre a variável de interesse. Por isso, a hipótese $\mathbb{E}(u_i|X_i) = 0$ é tão importante: mantendo as demais variáveis explicativas fixas, o valor esperado do termo de erro é nulo, isto é, não pode haver nada no termo de erro que esteja correlacionado com as variáveis explicativas para que a interpretação de causalidade aconteça, pois estaremos isolando o efeito "limpo" daquela variável sobre a variável de interesse.

Questão 2

a) Sabemos as fórmulas dos coeficiente de intercepto e inclinação, calculados na questão 1, item a. Sendo assim, basta substituirmos os valores da tabela nas fórmulas. Precisamos primeiro calcular $\bar{X}$ e $\bar{Y}$.

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{55+60+70+75}{4} = 65$$

$$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{12+15+17+24}{4} = 17$$

Calculando o denominador de $\hat{\gamma}_1$:

$$\begin{aligned}\sum_{i=1}^{4}(X_i - \bar{X})^2 &= (55-65)^2 + (60-65)^2 + (70-65)^2 + (75-65)^2 \\ &= (-10)^2 + (-5)^2 + 5^2 + 10^2 \\ &= 100 + 25 + 25 + 100 \\ &= 250\end{aligned}$$

Calculando o numerador de $\hat{\gamma}_1$:

$$\begin{aligned}\sum_{i=1}^{4}(X_i - \bar{X})Y_i &= (55-65)12 + (60-65)15 + (70-65)17 + (75-65)24 \\ &= (-120) + (-75) + 85 + 240 \\ &= 130\end{aligned}$$

Portanto, o valor de $\hat{\gamma}_1$ é dado por:

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^{4}(X_i - \bar{X})Y_i}{\sum_{i=1}^{4}(X_i - \bar{X})^2} = \frac{130}{250} = 0,52$$

Por (3), sabemos que $\hat{\delta}$ é igual a:

$$\hat{\delta} = \bar{Y} - \hat{\gamma}_1\bar{X} = 17 - (0,52)(65) = -16,8$$

Portanto, $\hat{\delta} = -16,8$ e $\hat{\gamma}_1 = 0,52$.

b) Sabemos que $\hat{u_i} = Y_i - \hat{\delta} - \hat{\gamma}_1 X_i$. Logo:

$$\begin{aligned}
\hat{u}_1 &= 12 - (-16,8) - (0,52)(55) = 0,2;\\
\hat{u}_2 &= 15 - (-16,8) - (0,52)(60) = 0,6;\\
\hat{u}_3 &= 17 - (-16,8) - (0,52)(70) = -2,6;\\
\hat{u}_4 &= 24 - (-16,8) - (0,52)(75) = 1,8.
\end{aligned}$$

c) $\sum_{i=1}^{4} \hat{u}_i = 0,2 + 0,6 - 2,6 + 1,8 = 0$

d) $\sum_{i=1}^{4} X_i \hat{u}_i = 55(0,2) + 60(0,6) + 70(-2,6) + 75(1,8) = 11 + 36 - 182 + 135 = 0$

e) Sabemos que o SQT é igual a:

$$\begin{aligned}
SQT &= \sum_{i=1}^{4} (Y_i - \bar{Y})^2 = (12-17)^2 + (15-17)^2 + (17-17)^2 + (24-17)^2\\
&= 25 + 4 + 0 + 49 = 78
\end{aligned}$$

f) Sabemos que o SQR é igual a:

$$\begin{aligned}
SQR &= \sum_{i=1}^{4} \hat{u_i}^2 = (0,2)^2 + (0,6)^2 + (-2,6)^2 + (1,8)^2\\
&= 0,04 + 0,36 + 6,76 + 3,24 = 10,4
\end{aligned}$$

g) $R^2 = \dfrac{SQE}{SQT} = 1 - \dfrac{SQR}{SQT} = 1 - \dfrac{10,4}{78} = 0,87$

Questão 3

a) A estimativa para o intercepto indica que para alunos que não estudaram ao longo do curso a nota é, em média, de 2 pontos. A estimativa para o coeficiente de inclinação indica que um minuto a mais de estudo eleva a nota do aluno, em média, em 0,008 pontos.

b) Vamos aplicar a fórmula de MQO para os dois estimadores. Antes veja que:

$$\bar{y^*} = \frac{\sum_{i=1}^{N} y_i^*}{N} = \frac{\sum_{i=1}^{N} 10y_i}{N} = 10 \cdot \frac{\sum_{i=1}^{N} y_i}{N} = 10\bar{y},$$

onde $y = Nota$. Agora, aplicando as fórmulas usuais (ver questão $1(a)$ dessa lista), temos:

$$\hat{\beta_1}^* = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) y_i^*}{\sum_{i=1}^{N} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{N} (x_i - \bar{x}) 10y_i}{\sum_{i=1}^{N} (x_i - \bar{x})^2} = 10 \cdot \frac{\sum_{i=1}^{N} (x_i - \bar{x}) y_i}{\sum_{i=1}^{N} (x_i - \bar{x})^2} = 10\hat{\beta}_1,$$

onde $x = Estudo$.

Além disso,

$$\hat{\beta_0}^* = \bar{y^*} - \hat{\beta_1}^* \bar{x} = 10\bar{y} - 10\hat{\beta}_1\bar{x} = 10(\bar{y} - \hat{\beta}_1\bar{x}) = 10\hat{\beta}_0.$$

c) Por definição, temos que $\hat{u^*} = y^* - \hat{y}^* = 10y - \hat{y}^*$. Vejemos então quanto vale $\hat{y}^*$:

$$\hat{y}^* = \hat{\beta_0}^* + \hat{\beta_1}^* x = 10\hat{\beta}_0 + 10\hat{\beta}_1 x = 10(\hat{\beta}_0 + \hat{\beta}_1 x) = 10\hat{y}.$$

Assim, $\hat{u^*} = 10y - 10\hat{y} = 10\hat{u}$.

d) Vamos aplicar a fórmula de MQO para os dois estimadores. Antes veja que:

$$\bar{\tilde{x}} = \frac{\sum_{i=1}^N \tilde{x}_i}{N} = \frac{\sum_{i=1}^N 2000 + x_i}{N} = \frac{\sum_{i=1}^N 2000}{N} + \frac{\sum_{i=1}^N x_i}{N} = 2000 + \bar{x},$$

onde $x = Estudo$. Agora, aplicando as fórmulas usuais, temos:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^N (\tilde{x}_i - \bar{\tilde{x}}) y_i}{\sum_{i=1}^N (\tilde{x}_i - \bar{\tilde{x}})^2} = \frac{\sum_{i=1}^N \left((\cancel{2000} + x_i) - (\cancel{2000} + \bar{x})\right) y_i}{\sum_{i=1}^N \left((\cancel{2000} + x_i) - (\cancel{2000} + \bar{x})\right)^2} = \frac{\sum_{i=1}^N (x_i - \bar{x}) y_i}{\sum_{i=1}^N (x_i - \bar{x})^2} = \hat{\beta}_1.$$

Além disso,

$$\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{\tilde{x}} = \bar{y} - \hat{\beta}_1(2000 + \bar{x}) = (\bar{y} - \hat{\beta}_1 \bar{x}) - 2000\hat{\beta}_1 = \hat{\beta}_0 - 2000\hat{\beta}_1.$$

e) Pela fórmula de R^2, $R^2 = \dfrac{SQE}{SQT}$ e, portanto, $\tilde{R^2} = \dfrac{\widetilde{SQE}}{\widetilde{SQT}}$. Primeiro, veja que $\widetilde{SQT} = \sum_{i=1}^N (y_i - \bar{y})^2 = SQT$. Por outro lado, $\widetilde{SQE} = \sum_{i=1}^N (\tilde{\hat{y}}_i - \bar{y})^2$, então devemos checar quanto vale $\tilde{\hat{y}}_i$:

$$\tilde{\hat{y}}_i = \tilde{\beta}_0 + \tilde{\beta}_1 \tilde{x}_i = (\hat{\beta}_0 - 2000\hat{\beta}_1) + \hat{\beta}_1(2000 + x_i) = \hat{\beta}_0 - \cancel{2000\hat{\beta}_1} + \cancel{2000\hat{\beta}_1} + \hat{\beta}_1 x_i = \hat{y}_i.$$

Logo, $\widetilde{SQE} = \sum_{i=1}^N (\tilde{\hat{y}}_i - \bar{y})^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 = SQE$. Portanto, $\tilde{R^2} = R^2$.

f) O coeficiente estimado nos diz que um aumento de um minuto nos estudos eleva em média $0,2\%$ a nota do aluno.

g) Não. Veja que RLS 1 diz respeito a linearidade nos parâmetros e não entre y e x. Portanto, desde que o modelo se mantenha linear nos parâmetros, teremos RLS 1 sendo válida.