

**FACULDADE DE ECONOMIA,
ADMINISTRAÇÃO E CONTABILIDADE
FEA/USP**

São Paulo - Set/2017

Lista 4 - Econometria I

Professora: Paula Pereda

EAE0324

Monitores: Bernardo Ostrovski e Igor Barreto

ALUNO: _____

ENTREGA: *Dia 22.09.17, até o início da monitoria*

Esta lista de exercícios tratará mais a fundo sobre Regressão Múltipla, com uma discussão de tópicos relacionados. Trabalharemos melhor com as hipóteses clássicas no arcabouço matricial.

1 Regressão Múltipla

Questão 1 Um dos ramos mais estudados em microeconometria é o da Economia da Educação, com suas origens em Mincer (1974).¹ Uma questão bastante discutida quando se tenta estimar o efeito de educação em salários é um possível viés causado pela omissão de variável omitida, tal como discutido em Blackburn e Neumark (1993)² e Gibbons e Katz (1992)³. Suponha os seguintes modelos:

$$Sal_i = \beta_0 + \beta_1 Educ_i + \beta_2 Hab_i + u_i; \quad (1)$$

$$Educ_i = \delta_0 + \delta_1 Hab_i + \eta_i; \quad (2)$$

$$Sal_i = \alpha_0 + \alpha_1 Educ_i + \epsilon_i, \quad (3)$$

onde Sal_i é o salário (em R\$/hora) da pessoa i ; $Educ_i$ são os anos de escolaridade da pessoa i ; Hab_i é uma medida da habilidade da pessoa i ; e u_i , η_i e ϵ_i são termos de erros dos modelos.

1. Mostre que podemos estimar $\hat{\beta}_1$ da seguinte maneira:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N \hat{\eta}_i Sal_i}{\sum_{i=1}^N \hat{\eta}_i^2} \quad (4)$$

Dica: veja sua resposta da Questão 4, itens 1 e 2, para ajudar.

¹Mincer, J. (1974). Schooling, Experience, and Earnings. Human Behavior Social Institutions No. 2.

²<http://www.jstor.org/stable/pdf/2535084.pdf?refreqid=excelsior%3Ac8d9ab4dafa062c490aab8da2b9b6a55>

³<http://www.jstor.org/stable/pdf/2297862.pdf?refreqid=excelsior%3A0e5033fa5bd707aef05a9f29af066967>

2. Suponha que você não observe Hab_i , então você estima seus coeficientes a partir de (3). Mostre formalmente o viés em cima de $\hat{\alpha}_1$ por omitir a variável Hab_i . Argumente se ele é positivo, neutro ou negativo.
3. No mesmo contexto do item anterior, mostre quanto vale $Var[\hat{\alpha}_1|Educ_i, Hab_i]$.
4. A partir de sua resposta no item anterior, mostre que sempre vale $Var[\hat{\alpha}_1|Educ_i, Hab_i] < Var[\hat{\beta}_1|Educ_i, Hab_i]$.
5. A partir da sua resposta no item 2 e no item 4, discuta sobre o *tradeoff* entre viés e variância dos estimadores. Argumente sob quais condições seria aceitável um viés na estimação de um parâmetro. Faça um gráfico para ilustrar o que está sendo discutido por você.

Questão 2 Seja dado o seguinte modelo:

$$y_i = x_i'\beta + u_i. \quad (5)$$

1. Usando notação matricial, enuncie as hipóteses do Teorema de Gauss-Markov.
2. Demonstre que $\hat{\beta}_{MQO}$ é *BLUE*, indicando onde entra cada hipótese listada por você no item anterior.

Questão 3 Suponha que você esteja interessado em pesquisar os maiores fatores que influenciam o aprendizado do aluno em uma universidade. Primeiramente, você imagina um modelo do tipo:

$$Nota_i = \beta_0 + \beta_1 Estudo_i + \beta_2 Dist_i + \beta_3 Monit_i + u_i, \quad (6)$$

onde $Nota_i$ é a nota (de 0 a 10) do aluno i em Econometria (medida em pontos); $Estudo_i$ é o quanto o aluno i estudou ao longo do semestre (em minutos); $Dist_i$ é a distância (em quilômetros) da residência do aluno i para a universidade em que estuda; $Monit_i$ é o quanto (em minutos) de monitorias o aluno i assistiu; e, por fim, u_i são outros fatores não observados que influenciam a nota do aluno, independentes as variáveis incluídas no modelo. Além disso, suponha que, condicional as variáveis independentes, o erro possui variância constante. Com sua base de 5 dados, você obteve estimativas para os coeficientes rodando um MQO no *Stata*.

1. Primeiro você deseja organizar melhor sua expressão dada pelo modelo acima, usando um pouco de notação matricial que aprendeu nas aulas e monitoria. Escreva duas expressões, sendo a primeira o modelo para um único indivíduo (ou seja, uma expressão do tipo $Nota_i = \dots$) e a segunda o modelo geral, incorporando todos indivíduos (ou seja, uma expressão do tipo $Nota = \dots$). Deixe claro o que é cada matriz expressa por você, os elementos que a constituem e suas dimensões (lembre que você possui o tamanho da amostra).
2. Demonstre que o vetor de estimadores de β feito por MQO é não viesado para estimar β .
3. Obtenha uma expressão para $Var[\hat{\beta}_{MQO}|X]$, colocando $Var[U|X]$ como $\sigma^2 I_5$.
4. Demonstre que $\hat{\beta}_{MQO}$ é consistente.
5. Enuncie as hipóteses clássicas necessárias dentro do arcabouço matricial de regressão múltipla. Indique quais são necessárias para a estimação por MQO, quais são necessária para mostrar que $\hat{\beta}_{MQO}$ é não-viesado e quais são necessárias para mostrar o valor de $Var[\hat{\beta}_{MQO}|X]$ obtido por você no item 2 (Note que nem toda(s) hipótese(s) utilizada(s) no Modelo de Regressão Linear Simples é(são) necessária(s), assim como alguma(s) a mais pode(m) ser necessária(s)).

6. Suponha que o vetor estimado por você tenha sido $\hat{\beta} = \begin{pmatrix} 1,0 \\ 0,004 \\ -0,04 \\ 0,008 \end{pmatrix}$.

Estime quanto deve ser a nota de um aluno que estudou 550 minutos, frequentou 850 minutos de monitoria e mora a 50km da universidade.

Recomendamos que as duas próximas questões sejam feitas com auxílio do Excel, a fim de otimizar a parte de cálculo matricial. Veja o material sobre matrizes no excel, disponível no moodle, caso não esteja familiarizado com funções que envolvem matrizes. Tome cuidado para não colocar apenas os resultados em sua resolução dos exercícios, colocando também o desenvolvimento das questões

Questão 4 Existe uma discussão na literatura de Economia da Saúde sobre o efeito de saúde na infância nos rendimentos individuais. Uma boa *proxy* de saúde na infância seria a altura da pessoa, tal como discutido em Thomas e Strauss (1997)⁴, Schultz (2002)⁵ e Persico, Postlewaite e Silverman (2004)⁶. Assim, você imagina o seguinte modelo:

$$Y = X\beta + U, \quad (7)$$

onde Y é um vetor com os rendimentos dos indivíduos; X é uma matriz com as alturas (em centímetros) dos indivíduos; e U é um vetor de erros homocedásticos e ortogonais a X . Você matou as últimas aulas e monitorias e não entendeu muito bem o conceito de **multicolinearidade**, então pediu ajuda aos monitores, que lhe disseram para imaginar um modelo análogo ao dado em (7) mas com uma variável a mais: altura dos indivíduos medida em metros. Eles lhe fornecem uma base de dados com quatro indivíduos: o primeiro com 1,6m de altura; o segundo com 1,5m de altura; o terceiro com 1,9m de altura; e o último com 1,75m de altura. Você pergunta sobre os valores dos rendimentos de cada um mas os monitores lhe dizem que isso não será preciso, pois você nem conseguirá estimar os coeficientes por MQO.

1. Explícite a matriz X a partir dos dados fornecidos pelos monitores (seguindo a lógica do modelo proposto por eles, com duas variáveis independentes).
2. Em seguida, explícite a matriz X' a partir dos dados fornecidos pelos monitores.
3. Verifique que de fato você não poderá encontrar uma estimativa para os coeficientes (Para isso, primeiro escreva a fórmula de $\hat{\beta}_{MQO}$; depois note que o termo da direita envolve o vetor Y que você não possui, enquanto que o termo da esquerda depende apenas de X e X' ; lembre-se da fórmula para obter uma matriz inversa; veja o que ocorre com o determinante). Qual a intuição por de trás disso?

⁴Thomas, D., Strauss, J. (1997). Health and wages: Evidence on men and women in urban Brazil. *Journal of econometrics*, 77(1), 159-185.

⁵<https://www.econstor.eu/bitstream/10419/98380/1/cdp841.pdf>

⁶<https://core.ac.uk/download/pdf/6330617.pdf>

Questão 5 Suponha o mesmo contexto da questão anterior. Agora os monitores lhe passam um modelo um pouco alterado, ainda com uma variável que mede a altura (em centímetros) das pessoas e outra que mede o tamanho do pé dos indivíduos. Os quatro indivíduos usados na base são os mesmos que anteriormente e o tamanho dos seus pés (em determinada unidade de medida) são, respectivamente: 40, 37.45, 47.5 e 43.75.

1. Considerando o modelo com apenas a variável de altura (em centímetros), explicita as matrizes X , X' e $X'X$. Estime $Var[\hat{\beta}_{MQO}|X]$, considerando $Var[U|X] = \sigma^2 I_4$ (Note que o valor de $(X'X)^{-1}$ aqui não será o mesmo dos itens seguintes).
2. Explicita agora a matriz X a partir dos dados fornecidos pelos monitores (seguindo a lógica do modelo proposto por eles, com duas variáveis independentes).
3. Em seguida, explicita a matriz X' a partir dos dados fornecidos pelos monitores.
4. Obtenha a matriz $(X'X)^{-1}$. Compare esse resultado com o obtido no terceiro item da questão anterior. Porque dessa vez foi possível estimar essa matriz?
5. Considerando o modelo com a variável de altura (em centímetros) e o tamanho dos pés, estime $Var[\hat{\beta}_{MQO}|X]$, considerando $Var[U|X] = \sigma^2 I_4$. Compare o segundo elemento da diagonal principal dessa matriz com o segundo elemento da diagonal principal da matriz obtida no primeiro item da questão. O que ocorreu de um para o outro? (Pense na possível correlação entre as duas variáveis independentes do modelo e em como isso afeta a matriz $(X'X)^{-1}$)