

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

São Paulo - Ago/2017

Lista 2 - Econometria I

Professora: Paula Pereda

EAE0324

Monitores: Bernardo Ostrovski e Igor Barreto

ALUNO:_____

ENTREGA: *Dia 25.08.17, até o início da monitoria*

Esta lista de exercícios tratará brevemente sobre a discussão entre causalidade vs correlação, além de alguns tópicos iniciais sobre Regressão Simples e métodos de estimação de parâmetros. Na última parte iremos treinar um pouco mais de Stata com dois exercícios que cobrirão alguns comandos básicos e avançados do programa econométrico.

1 Causalidade versus Correlação

Questão 1 Uma discussão muito importante trata 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 apenas correlacionadas mas não possuem 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 (simultaneamente). Analise os seguintes exemplos abaixo, corrigindo as afirmações feitas (se achar necessário) e dizendo qual o tipo de relação causal você é capaz de enxergar:

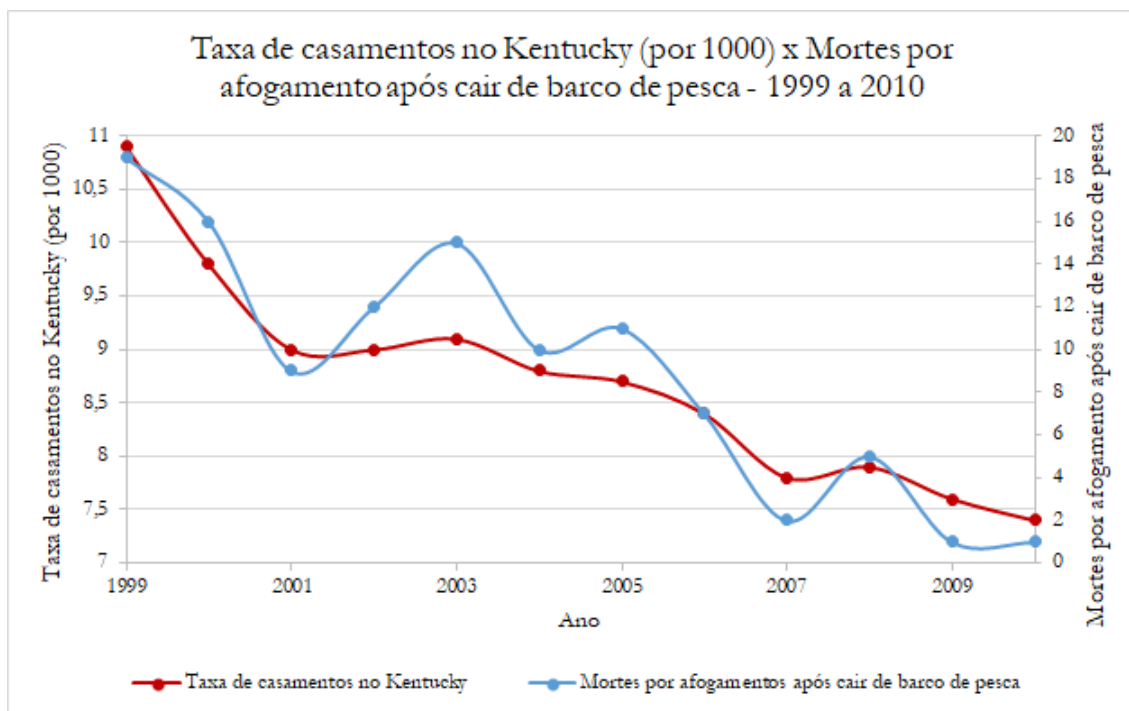
1. Um famoso exemplo usado por *Angrist e Pischke* em seu livro *Mostly harmless econometrics: An empiricist's companion*. Princeton university press (2008) nos retrata a seguinte tabela, com o resultado de uma pesquisa com indivíduos sobre seu estado de saúde (1-5, sendo 5 muito saudável) e se no último ano haviam ido ao hospital:

Grupo	Amostra	Status de Saúde Média	Desvio Padrão
Hospitalizados	7.774	3,21	0,014
Não Hospitalizados	90.049	3,93	0,003

Nota: NHIS 2005 (A&P, p.13)

Portanto, como a média de saúde dos hospitalizados é menor, pode-se concluir que os hospitais prejudicam a saúde.

2. No dia 19 de Junho desse ano o jornal *Gazeta do Povo* de Curitiba publicou uma notícia com a seguinte manchete: “Saiba de onde ‘surtem’ os vendedores de guarda-chuva de Curitiba”. Sempre que vemos muitos vendedores de guarda-chuva na rua, é praticamente certeza de que irá chover. Portanto pode-se concluir que mais vendedores de guarda-chuva na rua aumentam a probabilidade de chover.
Fonte: <http://www.gazetadopovo.com.br/curitiba/saiba-de-onde-surgem-os-vendedores-de-guarda-chuva-de-curitiba-7sgrukb7d3to1qlug1lmsxclv>
3. Sabe-se que quanto menor for a taxa de casamentos no estado americano de Kentucky, menor também será o número de mortes ocasionadas por afogamentos após quedas de barcos de pesca. O gráfico abaixo ilustra bem isso:



Obs: você pode entrar no site <http://tylervigen.com/spurious-correlations> para encontrar mais correlações como essa.

4. Uma professora certa vez classificou os alunos de sua classe em duas categorias: alunos que estudaram ao longo do curso (A) e alunos que só estudaram na véspera das provas (B). Além disso ela perguntou aos alunos o quanto (em litros) de cerveja que eles haviam consumido na semana anterior a prova. Após apurar todas as informações ela constatou que dentro da categoria A, um maior consumo de cerveja

na semana anterior a prova estava correlacionado a uma maior nota. Portanto, se conclui que um maior consumo de cerveja próximo a prova aumenta a nota do aluno.

5. Existe uma lenda atualmente no futebol inglês conhecida como "maldição de Ramsey". O jogador galês Aaron Ramsey, de 26 anos e que atua no Arsenal da Inglaterra é conhecido por amaldiçoar pessoas famosas: sempre que marca um gol, nos dias seguintes algum famoso morre pelo mundo. A figura abaixo mostra alguns destes casos:

Nome	Data da morte	Data do gol
Andrés Montes	16/10/2009	14/10/2009
Antonio De Nigris	16/11/2009	14/11/2009
Osama Bin Laden	02/05/2011	01/05/2011
Steve Jobs	05/10/2011	02/10/2011
Muammar Gaddafi	20/10/2011	19/10/2011
Whitney Houston	11/02/2012	11/02/2012
Chavela Vargas	05/08/2012	04/08/2012
Bebo Valdés	22/03/2013	22/03/2013
Rafael Videla	17/05/2013	14/05/2013
Ken Norton	18/09/2013	18/09/2013
Paul Walker	30/11/2013	30/11/2013
'Huracán' Carter	20/04/2014	20/04/2014
H.G.Giger	12/05/2014	11/05/2014
Robin Williams	11/08/2014	10/08/2014
Eduardo Galeano	13/04/2015	11/04/2015
Gunter Grass	13/04/2015	11/04/2015
David Bowie	11/01/2016	09/01/2016
Alan Rickman	14/01/2016	13/01/2016

Fonte: <http://www.sport.es/es/noticias/premier-league/maldicion-ramsey-3445920>

2 Métodos de Estimação

Questão 2 Suponha que você esteja interessado em estudar o efeito do aumento da velocidade máxima permitida (X) nas principais vias de sua cidade em cima do número de acidentes de trânsito (Y). Após uma leitura na bibliografia relacionada, você está confiante de que pode expressar tal relação através do seguinte modelo linear:

$$Y = a + bX + u$$

onde u capta o efeito de outras variáveis, independentes a velocidade máxima permitida, em cima do número de acidentes de trânsito, de tal forma que $\mathbb{E}[u] = 0$ e $Var[u] = \sigma^2$. Suponha que $Y|X$ possua distribuição Log-Normal (veja a dica ao final da questão).

A partir de uma base de dados com N observações, você passa a estimar os parâmetros de interesse (a , b e σ^2) por diferentes métodos.

1. Estime a e b através do Método dos Mínimos Quadrados.
2. Estime a , b e σ^2 através do Método da Máxima Verossimilhança (lembre que você é capaz de encontrar $\mathbb{E}[Y|X]$ e $Var[Y|X]$).
3. Qual(is) momento(s) populacional(is) você é capaz de identificar a partir das informações do enunciado? Estime a e b a partir do Método dos Momentos.

Dica: Se W tem distribuição Log-Normal com média μ e variância σ^2 , então a função de densidade de probabilidade de W é dada por:

$$f(w; \mu, \sigma) = \begin{cases} \frac{1}{w\sigma\sqrt{2\pi}} \exp\left[-\frac{(\log(w) - \mu)^2}{2\sigma^2}\right], & \text{se } w > 0 \\ 0 & \text{caso contrário} \end{cases}$$

3 Introdução a Regressão Simples

Questão 3 Considere o mesmo modelo da questão anterior. Suponha que sua base de dados tenha lhe fornecido a seguinte tabela:

Observação	Nº acidentes de transito	Velocidade máxima permitida (Km/h)
1	12	55
2	15	60
3	17	70
4	24	75

Nota: cada observação representa uma rodovia em um período de um mês

1. Estime a e b pelo Método dos Mínimos Quadrados Ordinários (MQO) (Você não precisa demonstrar a fórmula se já a obteve na questão anterior).
2. Agora que você estimou os coeficientes, calcule o valor dos resíduos.
3. Verifique que $\sum_{i=1}^4 \hat{u}_i = 0$.
4. Verifique que $\sum_{i=1}^4 x_i \hat{u}_i = 0$.
5. Calcule a Soma dos Quadrados Totais (SQT) da regressão.
6. Calcule a Soma dos Quadrados dos Resíduos (SQR) da regressão.
7. Calcule a qualidade do ajuste, isto é, calcule o R-quadrado da regressão.
8. A partir do que você respondeu nos itens 3 e 4, é possível afirmar que $E[u] = 0$? E que $E[ux] = 0$?

4 *Stata*

Recomendamos que a versão utilizada seja o Stata 13, no mínimo. Se não for possível obter o programa para uso doméstico, os computadores dos laboratórios do FEA5 estão equipados com o Stata para alunos da graduação. Entregue as respostas dos exercícios abaixo num arquivo do-file (extensão .do). Nomeie o do-file com seu número USP (ex: 7600420.do). Quando a(s) resposta(s) for(em) numérica(s), coloque-a(s) como comentário após o(s) comando(s) utilizado(s), utilizando duas barras (“//”) para iniciar a sessão de comentários.

Questão 4

1. Abra a base de dados *auto.dta*, disponível no Moodle.
2. Descreva a base de dados utilizando o comando *describe*. Quantas observações há nesta base? Quantas variáveis?

3. Agora, descreva a base de dados utilizando o comando *summarize*. Qual a diferença para o comando anterior? Dentre os dois comandos, qual você diria que é mais útil se estivermos querendo informações estatísticas sobre a base de dados?
4. Assim como quase todos os comandos do Stata, o comando *summarize* permite que opções sejam adicionadas ao código para obter diferentes resultados. Suponha que nosso objetivo seja extrair o melhor detalhamento possível das variáveis pelo comando *summarize*. Que opção você adicionaria ao comando? Qual o efeito disto (isto é, o que aconteceu com os resultados)?
5. Suponha que nosso interesse, agora, seja estudar a frequência das variáveis. Qual comando adequado para isso? Qual sua abreviação? Qual porcentagem de carros importados na base de dados?
6. Faça um histograma da variável de preços. Que distribuição se assemelha à verificada?
7. Plote o gráfico da distribuição da variável de preços, utilizando o comando *kdensity*.
8. Utilizando o comando *scatter*, plote o gráfico entre preço e capacidade do motor. Baseado apenas no gráfico, você diria que motores maiores estão relacionados a maiores preços? Qual o sentido da correlação? E entre preço e eficiência (milhas por galão)?
9. Faça o mesmo para as variáveis de preço e altura da cabine. Você diria, baseado apenas no gráfico, que estas duas variáveis estão relacionadas?

Questão 5

1. Abra a base de dados *auto.dta*, disponível no Moodle.
2. Note que as unidades desta base de dados seguem o sistema imperial de medidas. Crie uma nova variável a partir da variável comprimento, transformando a unidade de polegadas para centímetros (Nota: o fator de conversão é $1 \text{ pol} = 2,54 \text{ cm}$).
3. Faça o mesmo para a variável *mpg*: crie uma nova variável, transformando a unidade de milhas por galão para quilômetros por litro (Nota: $1 \text{ milha} = 1,6 \text{ km}$ e $1 \text{ galão} = 3,79 \text{ litros}$).

4. A maioria dos comandos do Stata permite que os resultados sejam condicionais a outras opções, através da opção *if*. Utilizando o *if*, crie uma variável que tem valor 1 se o comprimento do automóvel for menor ou igual a 400 cm e 0 caso contrário. Chame esta variável de *hatch*.
5. Crie uma variável que tenha valor 1 se o automóvel for importado e valor 0 caso contrário. Note que a variável importado está no formato texto (*string*). Tire a média desta nova variável e compare-a com a frequência obtida no item 5 da questão anterior. Qual a relação entre os dois?
6. Agora, utilizando um *loop*, crie três variáveis de uma vez, tais que elas sejam os valores das variáveis peso, altura e comprimento ao quadrado.