

## Respostas da Lista 07

1) (Barrow, Prob. 7.2 adaptado). Os dados apresentados abaixo fornecem informações sobre o consumo de margarina (em onças por pessoa por semana) e seu preço real no Reino Unido. Usando os dados do problema 7.2 do Barrow, obtemos os resultados abaixo.

Source	SS	df	MS	Number of obs =	19
Model	3.7308	1	3.7308	F( 1, 53) =	45.9798
Residual	1.3794	17	0.0811	Prob > F =	0.0000
Total	5.1102	18	0.2839	R-squared =	0.7301
				Adj R-squared =	0.7142
				Root MSE =	0.2849

  

Cons	Coef.	Std. Error	t	P> t
Preco	-0.0221	0.0033	-6.7808	0.0000
_cons	5.9936	0.3657	18.3888	0.0000

Responda:

a) Escreva o modelo de regressão estimada (o modelo teórico, somente com as letras e o modelo empírico, com os valores dos coeficientes estimados).

$$\widehat{\text{consumo}} = \widehat{\beta}_0 - \widehat{\beta}_1 * \text{preço}$$

b) Qual sinal você espera para o coeficiente da variável preço? Por quê?

$\beta_1 < 0$ , pois à medida que o preço aumenta, espera-se redução no consumo.

c) Interprete o coeficiente obtido e teste sua significância. Quantas observações há na amostra?

Há 19 observações na amostra. O coeficiente estimado do preço é -0,0221. Isto significa que se o preço aumentar em \$1, reduz-se o consumo de margarina em 0,0221 onça por pessoa por semana.

Para  $\beta_1$ , temos que as hipóteses são:  $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$ . Adotando um nível de significância de 5%, observa-se o p-valor calculado pelo Excel. Como o p-valor de  $\widehat{\beta}_1$  (0,000) é menor que 5%, rejeita-se  $H_0$ , e, portanto, preço da margarina é uma variável importante para explicar a quantidade consumida.

d) Preveja qual será o consumo para 2011, se o índice de preço da margarina nesse ano for 122.

Como temos valores observados fora da amostra (pois a amostra vai até 1988), precisa-se prever o consumo para 2011. Da regressão estimada, sabemos que

$$\widehat{\text{consumo}} = \widehat{\beta}_0 + \widehat{\beta}_1 * \text{preço} \rightarrow \widehat{\text{consumo}} = 5,9936 - 0,0221 * (122).$$

Para um preço de 122, o consumo estimado para 2011 é de 3,30.

2) Com dados das Nações Unidas para vários países sobre as variáveis poluição (quantidade de emissão de CO<sub>2</sub>) como dependente e PIB como independente, obteve-se um R<sup>2</sup> de 0.786. Adicionando expectativa de vida como outra variável independente do modelo, o R<sup>2</sup> ajustado resultou em 0.787.

a) Escreva as regressões estimadas.

$$1^\circ \text{ modelo: } \widehat{\text{poluição}} = \widehat{\beta}_0 + \widehat{\beta}_1 * \text{PIB}$$

$$2^\circ \text{ modelo: } \widehat{\text{poluição}} = \widehat{\beta}_0 + \widehat{\beta}_1 * \text{PIB} + \widehat{\beta}_2 * \text{expect}$$

b) Para prever a emissão de CO<sub>2</sub>, ajudou incluir a variável expectativa de vida no modelo? Isto significa o quê sobre o poder da variável expectativa de vida em explicar a emissão de CO<sub>2</sub>?

Não ajudou. Apesar dos dois valores não poderem ser comparados diretamente porque são modos diferentes de serem calculados, a proximidade dos resultados indica que incluir a variável expectativa de vida não foi significativa para explicar a variação da emissão de poluição.

3) Para os exemplos abaixo indique:

- o formato da base de dados (cross-section, séries de tempo, painel e qual é cada observação)

- a regressão estimada

- a expectativa sobre os efeitos das variáveis explicativas sobre Y. Justifique sua opinião.

a) Estudar os efeitos que o número de horas dedicadas ao estudo, a nota esperada pelo aluno e a quantidade de matérias que o aluno se matriculou no semestre sobre a nota final do segundo semestre de 2010 dos alunos das universidades federais do Brasil. Como destacar o efeito que as universidades de São Paulo obtiveram sobre as notas?

i) é uma cross-section, pois cada observação é um aluno de uma universidade federal em 2010.

ii)  $\widehat{notafinal} = \widehat{\beta}_0 + \widehat{\beta}_1 * horas + \widehat{\beta}_2 * notaesperada + \widehat{\beta}_3 * matérias$

iii)  $\beta_1 > 0$ , pois se espera que mais horas dedicadas ao estudo, aumentem a nota final do aluno.  $\beta_2 > 0$  se considerarmos um aluno honesto, pois se ele espera uma nota alta, deve ser porque domina a matéria.  $\beta_3 < 0$ , pois quanto mais matérias o aluno se matricula, menos tempo tem para estudar para cada uma.

iv) Para destacar o efeito das universidades de São Paulo, é necessário construir uma variável dummy que assume valor um se a observação é um aluno de São Paulo e zero caso contrário.

b) Emprego nos estados brasileiros ocorre em função dos valores estaduais da taxa de crescimento da produção industrial, das exportações e das importações. Como destacar o efeito das exportações dos estados da região Sudeste?

i) pode ser uma cross-section se apenas um ano for observado, em que cada observação é um estado brasileiro; ou um painel se observamos mais de um ano, em que cada observação é um estado brasileiro em determinado ano.

ii)  $\widehat{emprego} = \widehat{\beta}_0 + \widehat{\beta}_1 * cresc.prod.ind. + \widehat{\beta}_2 * exportações + \widehat{\beta}_3 * importações$

iii)  $\beta_1 > 0$ , pois se espera que um aumento da taxa de crescimento da produção industrial induza a contratação de mais trabalhadores.  $\beta_2 > 0$ , pois se as exportações aumentam, mais produção nacional é demandada, aumentando o emprego.  $\beta_3$  pode ter um efeito ambíguo:  $\beta_3 < 0$ , indicaria que os produtos importados estão tomando o lugar dos nacionais, o que reduz a produção brasileira e aumenta o desemprego; por outro lado, se  $\beta_3 > 0$ , indicaria que a produção nacional é bastante dependente da importada, de modo que ao aumentar as importações, a produção nacional se torna viável e aumenta, demandando mais trabalhadores.

iv) Para destacar o efeito das exportações dos estados da região Sudeste é necessário construir uma variável dummy que assume valor um caso o estado pertença à região Sudeste e zero caso contrário. Como se quer verificar o impacto de ser região Sudeste sobre as exportações, deve-se incluir no modelo uma variável de interação, dada por (dummy da região Sudeste)\*exportações.

4) Foram coletadas informações sobre uma amostra de prisioneiros brasileiros e argentinos para o ano de 2000 para as seguintes variáveis:

Y = duração da sentença (anos)

X1 = número de condenações anteriores

X2 = renda familiar (em 100 dólares PPP)

X3 = tempo desempregado (meses)

Considerando a duração da sentença uma variável que busca medir a gravidade do crime, responda:

a) Escreva a regressão estimada.

Com as informações acima, a regressão estimada é:

$\widehat{sentença} = \widehat{\beta}_0 + \widehat{\beta}_1 * condenações + \widehat{\beta}_2 * renda + \widehat{\beta}_3 * tempodesemp$

b) Quais os efeitos esperados das variáveis explicativas sobre Y?

O tempo de sentença do criminoso é uma variável que pode ser usada para medir a gravidade do crime.

$\beta_1 > 0$ , pois quanto mais condenações tem o prisioneiro, mais crimes deve ter cometido (ou mais violento é o criminoso).  $\beta_2 > 0$ , pois quanto maior a renda, menor a necessidade de se cometer crimes e menor seria a duração da sentença (ou os crimes seriam mais leves).  $\beta_3 > 0$ , pois segue o mesmo raciocínio da renda: quanto maior tempo desempregado, menor a renda e maior deve ser a sentença do condenado.

c) Como incluímos as informações sobre o estado civil e o gênero dos presos na regressão? Reescreva a equação estimada considerando estas informações.

Essas informações são qualitativas, então devem ser incluídas por variáveis dummy. Construímos uma dummy para estado civil (DCAS assume valor um caso o prisioneiro seja casado e zero caso contrário) e outra dummy para o gênero (DFEM assume valor um caso o prisioneiro seja do sexo feminino e zero caso contrário).

$$\widehat{sentença} = \widehat{\beta}_0 + \widehat{\beta}_1 * \text{condenações} + \widehat{\beta}_2 * \text{renda} + \widehat{\beta}_3 * \text{tempodesemp} + \widehat{\beta}_4 * \text{DFEM} + \widehat{\beta}_5 * \text{DCAS}$$

d) É possível, com o modelo da letra c, diferenciar entre os efeitos do Brasil e da Argentina? Como proceder para que isso também seja incluído no modelo?

Não, com o modelo anterior não é possível fazer nenhuma distinção entre efeitos do Brasil e da Argentina sobre a gravidade dos crimes cometidos nos países.

Para que isso ocorra, deve ser incluída uma dummy de país no modelo, como DBRA que assume valor um caso o prisioneiro seja brasileiro e zero caso contrário (seja argentino). Isto possibilita diferenciar os efeitos dos dois países. Dependendo como a dummy é inserida no modelo, são obtidas conclusões diferentes sobre os efeitos investigados (pode ser uma dummy de intercepto ou uma dummy de interação).

e) A partir da regressão estimada descrita na letra c, o pesquisador conclui seu trabalho com os seguintes resultados dispostos na tabela abaixo, sendo DBRA: dummy construída que assume valor 1 caso o indivíduo seja brasileiro e 0 caso contrário, DFEM: dummy para preso feminino e DCAS: dummy para preso casado. Interprete os efeitos de cada variável e verifique se elas são significantes. O que se pode concluir sobre a diferença da gravidade dos crimes entre os brasileiros e argentinos?

Cons	Coef.	Std. Error	t
Núm. Conden.	1.5714	0.2401	6.5448
Renda Familiar	-0.6069	0.1909	-3.1792
tempo desemp.	0.4285	0.0731	5.8618
DFEM	0.0841	0.7694	0.1093
DCAS	-1.8749	1.0683	-1.7550
DBRA*(Renda Fam.)	-0.4072	0.0947	-4.2999
DBRA*(Tempo desemp.)	-0.0194	0.0032	-6.1006
_cons	4.5639	12.460	0.3663

A regressão estimada é dada por:

$$\widehat{sentença} = \widehat{\beta}_0 + \widehat{\beta}_1 * \widehat{condenações} + \widehat{\beta}_2 * \widehat{renda} + \widehat{\beta}_3 * \widehat{tempodesemp} + \widehat{\beta}_4 * \widehat{DFEM} + \widehat{\beta}_5 * \widehat{DCAS} + \widehat{\beta}_6 * (DBRA * Renda) + \widehat{\beta}_7 * (DBRA * tempodesemp)$$

Os efeitos já foram interpretados até  $\beta_3$  na letra c e devem permanecer os mesmos com as demais variáveis incluídas neste exercício.

Para as demais variáveis temos:

$\beta_4$  não teria à princípio nenhuma preferência por sinal.  $\beta_5 > 0$ , pode ter sinal dúbio: se considerarmos que pessoas casadas pensam no bem estar da família, pode-se cometer um crime por necessidade que a família esteja passando ( $\beta_5 > 0$ ) ou afastar-se da criminalidade para proteger a família ( $\beta_5 < 0$ ).  $\beta_6$  considera o efeito que a renda no Brasil tem sobre a gravidade do crime, em comparação com o mesmo efeito na Argentina.  $\beta_7$  considera o efeito que o desemprego tem sobre a gravidade de um crime cometido no Brasil em comparação com a Argentina.  $\beta_6$  e  $\beta_7$  são variáveis de investigação das diferenças que existem entre Brasil e Argentina.

As duas variáveis de interação com dummy Brasil têm coeficientes negativos e significantes. No caso de  $\beta_5$  indica que se aumentarmos a renda familiar em 100 dólares, no Brasil a duração da sentença do crime (a gravidade) se reduz em  $(0,6469 - 0,4072) = 1,05$  anos, enquanto que para os argentinos, o mesmo aumento de 100 dólares de sua renda reduz a sentença em 0,6 anos.

$\beta_6$  indica que se aumentarmos o tempo de desemprego, os brasileiros cometem menos crimes (ou crimes menos graves) do que os argentinos: se aumentarmos o tempo de desemprego em 1 mês, a duração da sentença do crime brasileiro aumenta em  $(0,4285 - 0,0194) = 0,41$  anos, enquanto que para os argentinos, o mesmo aumento de 1 mês de desemprego aumenta a sentença em 0,42 anos.

Os resultados indicam que, apesar de haver uma diferença significativa entre os países, em termos qualitativos, a diferença é observada melhor pela condição da renda dos criminosos nos dois países. Observa-se, nesse caso, que os criminosos brasileiros reagem mais positivamente à redução das penas (gravidade dos crimes) do que os argentinos, dado um aumento da renda familiar.

Sobre a significância, já que não temos computado o p-valor e nem o número de observações da amostra, devemos usar a regra de bolso  $t_{crítico} = \pm 2$ .

Com isso, vemos que as dummies DFEM e DCAS não são significantes indicando que não ser do sexo feminino ou masculino e ser ou não casado não explicam significativamente a duração da sentença dos criminosos. Por outro lado, as demais variáveis consideradas são todas significantes para explicar a gravidade dos crimes cometidos.

f) Se a duração da sentença e a renda familiar estivessem medidas em logaritmo, como seria interpretado o coeficiente de renda familiar? E se somente duração da sentença estivesse em logaritmo, como mudaria a interpretação?

Ambos medidos em logaritmo: se aumentarmos em 1% a renda familiar, a sentença se reduz em  $\beta_2\%$ .

Somente duração da sentença em log: se aumentarmos a renda em 100 dólares, a sentença se reduz em  $(100 * \beta_2)\%$ .

g) Indique todo o procedimento de teste (definição das hipóteses, regressões que são necessárias estimar, dados necessários para construir a estatística do teste, a própria estatística de teste, como encontrar o valor crítico do teste e a regra de decisão do teste) que queira verificar se há efeitos diferenciados para o Brasil em relação à Argentina no que diz respeito à duração da sentença de seus criminosos condenados.

Com o modelo estimado, os efeitos do Brasil e Argentina se diferenciam pelas variáveis de interação da dummy do Brasil com renda e com tempo de desemprego. Assim para verificarmos se há diferença nesses países, devemos testar:

$$\begin{cases} H_0: \beta_6 = \beta_7 = 0 \\ H_1: \beta_6 \neq 0 \text{ e } \beta_7 \neq 0 \end{cases}$$

Isso significa que temos dois modelos para serem estimados:

- i) modelo restrito considera a restrição da hipótese nula.
- ii) modelo não restrito não considera a restrição da hipótese nula e seria exatamente o estimado na letra f

Este teste é realizado pela estatística F de acordo com a seguinte fórmula:

$$F = \frac{(SQRR - SQRNR)/q}{SQRNR/(n-k-1)}$$
 Portanto, para cada regressão estimada, observa-se a Soma dos Quadrados dos

Resíduos do modelo restrito (SQRR) e do não restrito (SQRNR).

Deve-se informar o tamanho da amostra (n).

q é o número de restrições, neste caso igual a 2.

k é o número de variáveis explicativas do modelo não restrito, excluindo o intercepto, neste caso igual a 7.

Uma vez calculada a estatística de teste, basta compará-la com o valor crítico tabelado a 5% de significância, com 2 graus de liberdade no numerador e (n-8) graus de liberdade no denominador. Se a estatística de teste for maior do que o valor tabelado, rejeita-se  $H_0$ , o que implica que há diferenças significativas entre os países no que diz respeito à gravidade dos crimes cometidos por seus prisioneiros.

Obs.: quando não temos o nível de significância do teste e nem o tamanho da amostra, para comparar o  $t_{\text{calculado}}$  com algum  $t_{\text{crítico}}$ , podemos utilizar i) a regra de bolso mencionada acima ou ii) supor um nível de significância igual a 5% e que a amostra é grande, tendendo ao infinito, o que resulta em um  $t_{\text{crítico}} = \pm 2$ .