

Econometria - Lista 5

Professores: Hedibert Lopes, Priscila Ribeiro e Sérgio Martins
Monitores: Gustavo Amarante e João Marcos Nusdeo

Exercício 1

Utilizando a base de dados disponível em **TEM COPROD.wtf1**, estimamos o seguinte modelo que explica o logaritmo do salário em função da escolaridade e da experiência prévia dentro da empresa:

$$\log(\widehat{\text{salário}}) = 10,16 + 0,04educ_i + 0,01anosemp_i$$

(0,04) (0,008) (0,003)

$$n = 46 \quad R^2 = 0,7501$$

Entretanto, sabe-se que o salário pode depender de uma série de outras variáveis, muitas delas de difícil mensuração, como habilidade de cada indivíduo. Logo, pergunta-se:

- a) Qual teste poderia ser utilizado para verificar se a especificação adotada é a correta para explicar o logaritmo do salário.
- b) Explícite a hipótese nula e a equação a ser estimada, assim como a distribuição da estatística de teste.
- c) Conduza o teste proposto. Conclua-o utilizando $\alpha = 0,05$

Exercício 2

Utilizando os dados contidos no arquivo **CEOSAL.wf1**, estime o modelo de regressão a seguir:

$$\log(\text{salary}) = \beta_1 + \beta_2 \log(\text{sales}) + \beta_3 \log(\text{mktval}) + \beta_4 \text{profmarg} + \beta_5 \text{ceoten} + \beta_6 \text{comten} + \varepsilon$$

- a) Após estimar o modelo proposto conduza um teste RESET e comente os resultados obtidos.
- b) Após incluirmos os regressores ceoten^2 e comten^2 no modelo original, o que acontece com o coeficiente de determinação? Tais inclusões são estatisticamente relevantes?
- c) Levando-se em conta os resultados discutidos no item anterior, conduza um teste RESET e comente os resultados obtidos.

Exercício 3

O objetivo deste problema é investigar os determinantes da performance média dos alunos de diferentes escolas ($n = 420$) da Califórnia em um teste padronizado aplicado aos alunos de quinta série (TESTSCR). As informações de interesse se encontram no banco de dados **caliescom.xls**.

Para tanto, o seguinte modelo de regressão linear múltipla foi proposto:

$$TESTSCR_i = \beta_0 + \beta_1 STR_i + \beta_2 EL_PCT_i + \beta_3 MEAL_PCT_i + e_i \quad (1)$$

em que

STR – número de alunos / número de professores;

EL_PCT – porcentagem de alunos cuja língua inglesa é a segunda língua falada;

MEAL_PCT – porcentagem de alunos que se qualificam para o lanche subsidiado.

- a) Utilize o teste LM para testar a hipótese de que nenhuma das variáveis explicativas é importante para explicar a variável resposta. Ainda, para esse caso de interesse escreva o modelo restrito. Adote um nível de significância de 10%.
- b) Também utilizando o teste LM, verifique se $\beta_2 = \beta_3 = 0$. Adote um nível de significância de 5%.

Exercício 4

Um econometrista está interessado em modelar a produção de firmas de um determinado setor. Ele tem uma amostra no arquivo *simulados_producao* que contém:

Y_i - produção da i -ésima firma

K_i - estoque de capital físico da i -ésima firma

L_i - número de trabalhadores contratados pela i -ésima firma.

O primeiro modelo que o econometrista tem em mente é o modelo de produção linear:

$$Y_i = \beta_0 + \beta_1 K_i + \beta_2 L_i + \varepsilon_i$$

- a) Estime o modelo acima e conduza o teste RESET com termo quadrático e cúbico. Adote um nível de significância de 10%. O que se pode concluir sobre a especificação do modelo de produção linear?

O econometrista agora quer fazer o mesmo com a função de produção Cobb-Douglas:

$$Y_i = e^{\beta_0} K_i^{\beta_1} L_i^{\beta_2} e^{\varepsilon_i}$$

- b) Claramente a função Cobb-Douglas não é linear nos parâmetros e por isso não pode ser estimado por MQO. Faça uma transformação apropriada para encontrar um modelo que possa ser estimado por MQO, estime o modelo e conduza o teste RESET com termo quadrático e cúbico. Adote um nível de significância de 10%. O que se pode concluir sobre a especificação do modelo Cobb-Douglas?
- c) Conduza um teste LM para testar se as firmas deste setor operam com retornos constantes de escala. Explícite as hipóteses do teste, a estatística do teste e o passo a passo do teste até sua conclusão. Adote um nível de significância de 10%.

Exercício 5

Neste exercício usaremos os dados da eleição presidencial americana do ano 2000. O arquivo de dados *florida2000.xls* contém o número de votos para diversos candidatos em cada um dos 67 municípios do estado da Flórida, antes da recontagem. O município que teve problemas de contagem é o de Palm Beach, que é a observação número 50 na base de dados. A variável DUMPALM é uma variável dummy que assume valor 1 para o município de Palm Beach e 0 para os demais. A variável TOTAL é o número total de votos naquele município. A recontagem de votos na Flórida foi motivada, em parte, por possíveis erros dos eleitores que queriam votar em Gore (segundo candidato, mas colocado em terceiro na cédula de voto) e acabaram votando em Buchanan (terceiro candidato, mas colocado em segundo na cédula de voto). Esse erro gerou várias cédulas em que ambos os candidatos estavam marcados. A diferença de votos entre Bush e Gore, antes da recontagem, era de 975 votos a favor de Bush.

- a) Faça uma regressão do número de votos de Buchanan em uma constante e no número de votos de Gore. Investigue a presença de *outliers*.
- b) Estime o número de votos acidentais para Buchanan (v) no município de Palm Beach incluindo a variável DUMPALM no modelo do item (a). Teste a hipótese $H_0: v \leq 975$ contra $H_A: v > 975$.
- c) Como cada município tem um tamanho de população diferente, isso pode gerar termos de erro heterocedásticos no modelo do item (a). Conduza um teste Breusch-Pagan assumindo que $\sigma_i^2 = \delta_0 + \delta_1 TOTAL_i$.
- d) Responda os itens (b) e (c) novamente, só que agora substitua o número de votos de cada candidato pela proporção de votos que ele teve no município. Para o teste BP assumindo que $\sigma_i^2 = \delta_0 + \delta_1 \frac{1}{TOTAL_i}$.
- e) Baseado nos itens anteriores, formule um modelo intuitivamente plausível para a variância do erro do modelo do item (a). Existe alguma transformação que possa nas variáveis para que tenhamos um modelo capaz de testar nossa hipótese de interesse (Bush não deveria ter ganho a eleição na Flórida) que tenha erros homocedásticos?
- f) Discuta e investigue se as suposições necessárias para a conclusão (politicamente importante) do item (e) são possíveis para esses dados.

Exercício 6

Um econometrista está interessado em estimar a equação de demanda por um bem normal:

$$\ln(q_i) = \beta_1 + \beta_2 \ln(p_i) + \beta_3 + \varepsilon_i, \quad i = 1, 2, \dots, N$$

Em que N é o número de empresas; q , p são, respectivamente, a quantidade e o preço de equilíbrio e d é uma dummy de localização igual a 1 se afirma localiza-se na região central da cidade e 0 caso contrário. Usando MQO, ora incluindo a dummy ora excluindo-a, ele obteve os resultados da Tabela 1.

Tabela 1 - Estimação da Demanda por MQO

Variável Explicativa	Variável Dependente: $\ln(q)$			
	Modelo 1		Modelo 2	
	Coefficiente	Erro-Padrão	Coefficiente	Erro-Padrão
Constante	2	0,5	1,8	0,4
$\ln(p)$	1,5	0,3	1,6	0,4
d			0,05	0,01
N	60		60	
R^2	0,65		0,95	
SQR	12,25		7,50	
Teste de Heterocedasticidade	15,98		16,89	

Nota: O teste de Heterocedasticidade de White apresenta o resultado da estatística F deste teste, sem a inclusão de termos cruzados.

- Considere o modelo 2, interprete a estimativa de β_2
- Faça um teste F-parcial para escolher entre os modelos 1 e 2. Adote um nível de significância de 10%. Não se esqueça de descrever: a hipótese nula e alternativa; a estatística de teste; a distribuição da estatística de teste; a região crítica (região de rejeição) e o(s) modelo(s) a ser(em) estimado(s).
- Faça o Teste de Heterocedasticidade de White (sem termos cruzados) para o modelo escolhido no item (b). Adote um nível de significância de 5%. Não se esqueça de descrever: a hipótese nula e alternativa; a estatística de teste; a distribuição da estatística de teste; a região crítica (região de rejeição) e o(s) modelo(s) a ser(em) estimado(s).
- Considerando os resultados obtidos, escreva a matriz de variância e covariância do vetor de erros, inclusive sua dimensão. É necessário usar algum estimador robusto à heterocedasticidade? Em caso afirmativo, qual estimador robusto deve ser adotado?
- Os resultados do teste de heterocedasticidade tem alguma implicação sobre a forma utilizada para selecionar o modelo na letra (b)?
- Observe que os modelos 1 e 2 a estimativa $\hat{\beta}_2 > 0$. Qual seria uma possível explicação para esse resultado? Você alteraria o método de estimação? Explique porque em caso afirmativo, e diga qual modelo você utilizaria. Neste caso você utilizaria alguma variável adicional? Qual?

Exercício 7

O senhor Flávio tem uma equipe especializada em apartamentos de luxo em São Paulo e no Rio de Janeiro. Para ajudar a precificar os imóveis, eles utilizam um software que dá uma estimativa para o valor do imóvel quando ele atende a algumas características. A ideia por trás da programação é a seguinte regressão linear:

$$valor_i = \beta_1 + \beta_2 tam_i + \beta_3 cidade_i + \beta_4 quartos_i + \varepsilon_i$$

em que

tam_i – é o tamanho do apartamento, em m²;

$cidade_i$ – é uma *dummy* que assume valor 1 quando o apartamento está localizado no Rio de Janeiro, e 0 se está localizado em São Paulo;

$quartos_i$ – é a quantidade de quartos que existem no apartamento;

ε_i – é o termo de erro.

Com base nos resultados a seguir, verifique se a suposição acerca da heterocedasticidade dos erros é válida. Para tanto, conduza seu teste admitindo 10% de significância, não se esquecendo de explicar as hipóteses (e suas interpretações), a estatística de teste sob a hipótese nula, a região crítica, a regra de decisão, e finalmente, o resultado. Ainda, responda como a não validade dessa suposição prejudicaria as propriedades dos estimadores e a inferência estatística.

Dependent Variable: VALOR
Method: Least Squares
Date: 05/19/12 Time: 21:05
Sample: 1 172
Included observations: 172

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-39964.59	10072.24	-3.967797	0.0001
CIDADE	282596.0	6185.856	45.68422	0.0000
QUARTOS	4534.212	1824.631	2.485002	0.0139
TAM	40475.17	85.71815	472.1890	0.0000
R-squared	0.999253	Mean dependent var	4049718.	
Adjusted R-squared	0.999240	S.D. dependent var	1238016.	
S.E. of regression	34139.38	Akaike info criterion	23.73727	
Sum squared resid	1.96E+11	Schwarz criterion	23.81047	
Log likelihood	-2037.405	Hannan-Quinn criter.	23.76697	
F-statistic	74901.68	Durbin-Watson stat	2.194186	
Prob(F-statistic)	0.000000			

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/19/12 Time: 21:08
Sample: 1 172
Included observations: 172

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19234793	6.65E+08	0.028923	0.9770
CIDADE	3.69E+09	4.08E+08	9.032939	0.0000
QUARTOS	1.97E+08	1.20E+08	1.635619	0.1038
TAM	-1498919.	5659590.	-0.264846	0.7915
R-squared	0.349740	Mean dependent var	1.14E+09	
Adjusted R-squared	0.338128	S.D. dependent var	2.77E+09	
S.E. of regression	2.25E+09	Akaike info criterion	45.93287	
Sum squared resid	8.54E+20	Schwarz criterion	46.00606	
Log likelihood	-3946.226	Hannan-Quinn criter.	45.96256	
F-statistic	30.11937	Durbin-Watson stat	1.419924	
Prob(F-statistic)	0.000000			