

Econometria - Lista 4 – Teste F-Parcial

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

Exercício 1

Uma agência imobiliária dispõe de informações sobre o preço de venda (y , em milhares de reais), a área (x_2 , em metros quadrados), o número de quartos (x_3) e o número de banheiros (x_4) de 14 residências. Com base nessas informações, foram estimados os parâmetros dos seguintes modelos de regressão:

A. $y_i = 47,624 + \hat{\varepsilon}_i$	$SQR = 101,82$	
B. $y_i = 7,853 + 0,191x_{2i} + \hat{\varepsilon}_i$	$SQR = 18,27$	$R^2 = 0,821$
C. $y_i = 18,177 + 0,204x_{2i} - 3,587x_{3i} + \hat{\varepsilon}_i$	$\hat{\sigma}^2 = 1,530$	$R^2 = 0,835$
D. $y_i = 19,359 + 0,213x_{2i} - 3,238x_{3i} - 1,829x_{4i} + \hat{\varepsilon}_i$		$R^2 = 0,839$

em que

SQR é a Soma dos Quadrados dos Resíduos (SSR);

$\hat{\sigma}^2$ é estimador não viesado da variância do erro.

- Interprete, quando possível, as estimativas, obtidas através do método dos mínimos quadrados ordinários, para os parâmetros associados às variáveis explicativas do modelo de regressão proposto em (D). Ainda, os sinais obtidos para as estimativas desses parâmetros estão em conformidade com aqueles que você esperaria obter? Por quê?
- Explique detalhadamente como você conduziria um teste de hipótese cuja intenção fosse verificar se o impacto de 1 quarto a mais no imóvel é igual ao dobro do impacto de um banheiro a mais no preço do imóvel, fixada a área. Adote um nível de significância de 5%. Não se esqueça de descrever, deixando bem claro: a hipótese nula e alternativa do teste; a estatística de teste e sua distribuição; a região crítica; e o(s) modelo(s) a ser(em) estimado(s) para auxiliar suas conclusões.
- Conduza, caso julgue possível, um teste de hipóteses que verifique se a inclusão de x_3 traz alguma informação relevante para um modelo que já apresenta x_2 como variável explicativa, além da constante. Adote um nível de significância de 5%. Com base nos resultados deste teste, e levando em consideração as propriedades dos estimadores de MQO no caso de omissão de variável relevante e/ou inclusão de variável irrelevante, qual modelo você escolheria? Por quê?

Exercício 2

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) Qual o sinal esperado para β_1 , β_2 e β_3 ? Justifique a sua resposta. Ainda, interprete os parâmetros do modelo proposto.
- b) Como ficaria a interpretação dos parâmetros do modelo original se a variável resposta fosse $\log(\text{TESTSCR})$? Ainda, o que aconteceria com pessoas com $\text{TESTSCR} = 0$?
- c) Com ajuda do Eviews, encontre as estimativas para os parâmetros do modelo proposto em (1), e o escreva na forma usual.
- d) Teste 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%.
- e) Conduza um teste de hipóteses adequado para se verificar se $\beta_2 = \beta_3 = 0$. Adote um nível de significância de 5%.

Exercício 3 (PF 2011/01 Adaptado)

Um grupo de pesquisadores está interessado na proposição e estimação dos parâmetros de um modelo de regressão linear cujo objetivo principal é o de verificar se há um impacto diferenciado na nota média dos estudantes que possuem ou não computador pessoal, quando o tempo de dedicação aos estudos dos mesmos aumenta em uma hora por semana.

- a) Auxilie o grupo de pesquisadores, propondo uma equação de interesse.
- b) Deixando bem claro seu raciocínio e baseando-se no item anterior, esboce um gráfico que explique o modelo proposto.

Exercício 4 (Wooldridge)

Fernando é um pesquisador da área de economia da saúde e tem interesse em entender relações entre o peso de crianças recém-nascidas e fatores familiares. Para isso, Fernando coletou uma amostra com as variáveis:

- *Pesonasc* – Peso da criança no nascimento.
- *Cigs* – Número médio de cigarros diários que a mãe fumou durante a gravidez.
- *Ordem* – Ordem de nascimento do filho. (Primeiro filho = 1, segundo filho = 2, etc.)
- *Renda* – Renda anual da família.
- *Educmae* – Anos de escolaridade da mãe.
- *Educpai* – Anos de escolaridade do pai.
- *Dhomem* – Dummy que assume valor 1 quando a criança é homem.
- *Dbranco* – Dummy que assume valor 1 quando a criança é classificada como branca.

Fernando estimou os modelos a seguir tendo $\log(\textit{pesonasc})$ como variável dependente.

Modelo 1

Variável	Coefficiente Estimado	Erro-Padrão
Intercepto	4.66	0.22
<i>Cigs</i>	-0.0044	0.0009
Log(renda)	0.0093	0.0059
Ordem	0.016	0.006
Dhomem	0.027	0.01
Dbranco	0.055	0.013
n	1388	
R ²	0.0472	

Modelo 2

Variável	Coefficiente Estimado	Erro-Padrão
Intercepto	4.65	0.38
<i>Cigs</i>	-0.0052	0.001
Log(renda)	0.011	0.0085
Ordem	0.017	0.006
Dhomem	0.034	0.011
Dbranco	0.045	0.015
Educmae	-0.003	0.003
Educpai	0.0032	0.0026
n	1191	
R ²	0.0493	

- Para o caso do primeiro modelo, interprete o coeficiente estimado da variável *cigs*.
- Para o caso do primeiro modelo, qual a diferença esperada de peso no nascimento entre uma criança branca em relação à uma criança não-branca, mantendo os demais fatores constantes? Essa diferença é estatisticamente significativa?
- Para o caso do segundo modelo, comente o efeito estimado e a significância da variável *educmae*.
- Apenas com os dados disponibilizados neste exercício não é possível calcular a estatística do teste F parcial que testa conjuntamente a significância das variáveis *educmae* e *educpai*. Explique quais são as informações que faltam e, uma vez que tenha estas disponíveis, como conduziria o teste.