

ECONOMETRIA I - LISTA 4

MODELO DE REGRESSÃO LINEAR MÚLTIPLA: INFERÊNCIA

Mateus Cardoso

28/06/2021

1) (Baseado em Wooldridge, Cap. 4) Quais são as hipóteses do modelo linear clássico (MRLC)?

2) (Wooldridge, Cap. 4) Quais dos seguintes itens podem fazer com que as estatísticas t de MQO não sejam válidas (isto é, que não tenham distribuições t sob H_0)?

- (i) Heteroscedasticidade.
- (ii) Um coeficiente de correlação de 0,95 entre duas variáveis independentes que estão no modelo.
- (iii) Omitir uma variável explicativa importante.

3) 4. As taxas de aluguel são influenciadas pela população de estudantes em uma cidade onde há universidades? Seja $rent$ o aluguel médio mensal pago pela unidade alugada em uma determinada cidade, onde há universidades. Seja pop o total da população da cidade, $avginc$, a renda média da cidade e $pctstu$, a população de estudantes como um percentual da população total. Um modelo para testar uma relação é

$$\log(rent) = \beta_0 + \beta_1 \log(pop) + \beta_2 \log(avginc) + \beta_3 pctstu + u.$$

- (i) Formule a hipótese nula de que o tamanho da população estudantil relativo à população das cidades não tem efeito *ceteris paribus* sobre os aluguéis mensais. Formule a alternativa de que há um efeito.
- (ii) Quais sinais você espera para β_1 e β_2 ?
- (iii) A equação estimada, usando dados de 1990 de 64 cidades com universidades do arquivo RENTAL, é

$$\widehat{\log(rent)} = 0,043 + 0,066 \log(pop) + 0,507 \log(avginc) + 0,0056 pctstu$$

$$\begin{array}{cccc}(0,844) & (0,039) & (0,081) & (0,0017) \\ n = 64, & R^2 = 0,458.\end{array}$$

que está errado com a seguinte afirmação: “Um aumento de 10% na população está associado a um aumento de cerca de 6,6% no aluguel”?

(iv) Teste a hipótese formulada no item (i) no nível de 1%.

3) (Wooldridge, Cap. 4) Considere a equação estimada do Exemplo 4.3 que poderia também ser usada para estudar os efeitos de faltar às aulas sobre a GPA em curso superior:

$$\begin{array}{cccc}\widehat{colGPA} = 1,39 + 0,412hsGPA + 0,15ACT - 0,083skipped \\ (0,33) & (0,94) & (0,011) & (0,026) \\ n = 141, & R^2 = 0,234.\end{array}$$

(i) Usando a aproximação normal padronizada, encontre o intervalo de confiança de 95% para β_{hsGPA} .

(ii) Você pode rejeitar a hipótese $H_0 : \beta_{hsGPA} = 0,4$ contra a hipótese alternativa bilateral no nível de 5%?

(iii) Você pode rejeitar a hipótese $H_0 : \beta_{hsGPA} = 1$ contra a hipótese alternativa bilateral no nível de 5%?

4) (Wooldridge, Cap. 4) Considere o modelo de regressão múltipla com três variáveis independentes, sob as hipóteses do modelo linear clássico RLM.1 a RLM.6:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

Você deseja testar a hipótese nula $H_0 : \beta_1 - 3\beta_2 = 1$.

- (i) Sejam $\hat{\beta}_1$ e $\hat{\beta}_2$ os estimadores de MQO de β_1 e β_2 . Encontre $Var(\beta_1 - 3\beta_2)$ em termos das variâncias de $\hat{\beta}_1$ e $\hat{\beta}_2$ e a covariância entre eles. Qual é o erro padrão de $\beta_1 - 3\beta_2$.
- (ii) Escreva a estatística t para testar $H_0 : \beta_1 - 3\beta_2 = 1$.
- (iii) Defina $\theta_1 = \beta_1 - 3\beta_2$ e $\hat{\theta}_1 = \hat{\beta}_1 - 3\hat{\beta}_2$. Escreva uma equação de regressão que envolva β_0 , θ_1 , β_2 e β_3 que permita que você obtenha diretamente $\hat{\theta}_0$ e seu erro padrão.

5) (Wooldridge, Cap. 4) A análise de regressão pode ser usada para testar se o mercado usa eficientemente as informações ao avaliar ações. Seja $return$ o retorno total de possuir ações de uma empresa ao longo de um período de quatro anos, do final de 1990 até o final de 1994. A hipótese dos mercados eficientes diz que esses retornos não devem estar sistematicamente relacionados à informação conhecida em 1990. Se as características conhecidas da empresa no início do período ajudassem a prever os retornos das ações, poderíamos usar essas informações para escolher ações. Para 1990, seja dkr a relação dívida-capital de uma empresa, seja eps os ganhos por ação, seja $netinc$ a renda líquida e seja $salary$ a remuneração total dos CEOs da empresa.

(i) Usando os dados do arquivo RETURN, estimou-se a seguinte equação:

$$\widehat{return} = -14,37 + 0,321dkr + 0,043eps - 0,0051netinc + 0,0035salary$$

$$(6,89) \quad (0,201) \quad (0,078) \quad (0,0047) \quad (0,0022)$$

$$n = 142, \quad R^2 = 0,0395.$$

Teste se as variáveis explicativas são conjuntamente significantes ao nível de 5%. Alguma variável explicativa é individualmente significativa?

(ii) Agora, nova estimação do modelo usando a forma log para $netinc$ e $salary$ forneceu a seguinte equação:

$$\widehat{return} = -36,30 + 0,327dkr + 0,069eps - 4,74\log(netinc) + 7,24\log(salary)$$

$$(39,37) \quad (0,203) \quad (0,080) \quad (3,39) \quad (6,31)$$

$$n = 142, \quad R^2 = 0,0330.$$

Alguma de suas conclusões no item (i) mudou?

(iii) Nessa amostra, algumas empresas têm zero de dívidas e outras têm ganhos negativos. Deveríamos tentar usar $\log(dkr)$ ou $\log(eps)$ para vermos se isso melhorará o ajustamento? Explique.

(iv) Em geral, a evidência da previsibilidade dos retornos é forte ou fraca?

Exercícios no R

1) (Wooldridge, Cap. 4) O modelo a seguir pode ser usado para estudar se gastos de campanha afetam os resultados das eleições:

$$voteA = \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystA + u,$$

em que $voteA$ é a porcentagem de votos recebidos pelo Candidato A, $expendA$ e $expendB$ são os gastos de campanha dos Candidatos A e B, e $prtystA$ é uma medida da força do partido do Candidato A (uma porcentagem dos votos presidenciais mais recentes que foram para o partido A).

- (i) Qual é a interpretação de β_1 ?
- (ii) Em termos de parâmetros, defina a hipótese nula de que um aumento de 1% nos gastos de A seja compensado por um aumento de 1% nos gastos de B.
- (iii) Estime o modelo dado usando as informações do arquivo VOTE1 e registre os resultados na forma usual. Os gastos do Candidato A afetam os resultados? E os gastos do Candidato B? Você poderia usar esses resultados para testar a hipótese do item (ii)?
- (iv) Estime um modelo que dê diretamente a estatística t para testar a hipótese do item (ii). O que você conclui? (Use uma alternativa bilateral.)

2) (Wooldridge, Cap. 4) O conjunto de dados do arquivo 401KSUBS contém informações sobre os ativos financeiros líquidos ($nettfa$), idade dos respondentes da pesquisa (age), renda familiar anual (inc), tamanho da família ($fsize$) e participação em determinados planos de pensão para pessoas dos Estados Unidos. As variáveis de patrimônio e renda são registradas em milhares de dólares. Para essa questão, use somente os dados de residências com apenas uma pessoa (assim, $fsize = 1$).

- (i) Quantas residências com apenas uma pessoa existem no conjunto de dados?
- (ii) Use os MQO para estimar o modelo

$$nettfa = \beta_0 + \beta_1 inc + \beta_2 age + u,$$

e reporte os resultados usando o formato usual. Certifique-se de usar somente residências com apenas uma pessoa na amostra. Interprete os coeficientes de inclinação. Houve alguma surpresa em relação às estimativas de inclinação?

- (iii) O intercepto da regressão do item (ii) teve um significado interessante? Explique.
- (iv) Encontre o p-valor do teste $H_0 : \beta_2 = 1$ contra $H_1 : \beta_2 < 1$. Você rejeita H_0 ao nível de significância de 1%?
- (v) Se você fizer uma regressão simples de $nettfa$ sobre inc , o coeficiente estimado de renda seria muito diferente do estimado do item (ii)? Por que sim ou por que não?