Notas de aula para o curso de Econometria I

Nota 6: inferência para regressão múltipla

Thiago Fonseca Morello

fonseca.morello@ufabc.edu.br

sala 301, Bloco Delta, SBC

1 Funções de distribuição de probabilidades dos estimadores

Até agora o comportamento dos estimadores de MQO na amostra foi descrito exclusivamente em função das estimativas pontuais e da variância amostral. Já, ao tratar das propriedades estatísticas, tomaram-se por base a expectativa e a variância populacional. Estas duas últimas estatísticas resumem a função de distribuição de probabilidades (FD) dos estimadores. Contudo, para realizar testes de hipóteses acerca dos valores populacionais dos parâmetros da FRP é preciso conhecer a FD completa do estimador, pois isso é requerido para a construção de regiões críticas e para o cálculo de p-valores.

Daí a utilidade da hipótese MCRL 7 (nota de aula 9). É possível demonstrar que dela decorre que $\hat{\beta}_{MQO}|X \sim N\left(\beta, \sigma^2\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\right)$. Ou seja, o estimador de MQO tem distribuição normal com média dada pela expectativa do estimador e variância equivalente ao valor populacional da variância do estimador. Esta última é igual a $\sigma^2\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}$ quando, por “estimador”, entende-se o vetor de estimadores de MQO, $\hat{\beta}_{MQO}|X$.

Também é comum se referir a $V[\hat{\beta}_{MQO}|X]$ como matriz de variância-covariância dos estimadores. De fato, a variância condicional do vetor de estimadores é uma matriz (K+1) x (K+1) que contém em sua diagonal principal as variâncias dos estimadores dos parâmetros e, fora da diagonal principal, há os coeficientes de correlação entre estimadores. Sob as hipóteses MCRL1-MCRL6, a matriz de variância-covariância assume uma forma simplificada, dada por $V[\hat{\beta}_{MQO}|X] = \sigma^2\left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}$.

Das observações dos parágrafos anteriores decorre que o estimador de MQO para cada parâmetro tem, individualmente, distribuição normal $\hat{\beta}_{MQO\,k}|X \sim N\left(\beta_k, \frac{\sigma^2}{\sum_{i=1}^{N}(x_{ik}-\bar{x}_k)^2(1-R_k^2)}\right)$. Desta maneira, pois,

$$Z=\frac{\left(\hat{\beta}_{MQOk}-E[\hat{\beta}_{MQOk}\vee X]\right)}{\sqrt{V[\hat{\beta}_{MQOk}\vee X]}}=\frac{\left(\hat{\beta}_{MQOk}-\beta_k\right)}{\sqrt{\sigma^2/\sum_{i=1}^{N}\left(x_{ik}-\overline{x}_k\right)^2\left(1-R_k^2\right)}}\quad N(0,1)$$

(a manipulação do denominador se vale do fato de que, sob as hipóteses do MCRL, os estimadores de MQO são não-viesados).

A utilidade deste resultado é comprometida pelo desconhecimento de σ^2, um parâmetro populacional, uma vez que se trata do valor populacional da variância dos termos de perturbação. A incorporação do estimador não viesado, $\hat{\sigma}^2$, conduz a uma expressão com distribuição t de Student:

$$t=\frac{\left(\hat{\beta}_{MQOk}-E[\hat{\beta}_{MQOk}\vee X]\right)}{\sqrt{\widehat{V[\hat{\beta}_{MQOk}\vee X]}}}=\frac{\left(\hat{\beta}_{MQOk}-\beta_k\right)}{\sqrt{\hat{\sigma}^2/\sum_{i=1}^{N}\left(x_{ik}-\overline{x}_k\right)^2\left(1-R_k^2\right)}}\quad t_{N-(K+1)}$$

Desta maneira, pois, chega-se a uma quantidade com distribuição t de Student com N-(K+1) graus de liberdade.

2 Teste de significância individual

2.1 Introdução

É possível verificar se a evidência gerada com base na amostra é favorável ou desfavorável à hipótese de nulidade dos parâmetros populacionais da FRP. A nulidade (populacional) de um coeficiente significa que a explicativa a ele associada não exerce efeito sobre a variável dependente, após a contabilização do efeito das demais explicativas. O teste de hipóteses adequado, para a nulidade do valor populacional do k-ésimo coeficiente é H0: $\beta_k = 0$ vs H1: $\beta_k \neq 0$. Uma vez que se está focando em um único parâmetro, o teste recebe o nome de teste de significância individual.

O procedimento de teste é equivalente ao visto em inferência estatística para o caso de testes para o valor populacional de uma média ou proporção. O primeiro passo está em calcular o valor observado da estatística do teste, assumindo a validade da hipótese nula. Ou seja, obtém-se:

$$\hat{t}(H_0)=\frac{\left(\hat{\beta}_{MQOk}-\beta_k^0\right)}{\left[\sqrt{\hat{\sigma}^2/\sum_{i=1}^{N}\left(x_{ik}-\overline{x}_k\right)^2\left(1-R_k^2\right)}\right]}\quad t_{N-(K+1)}$$

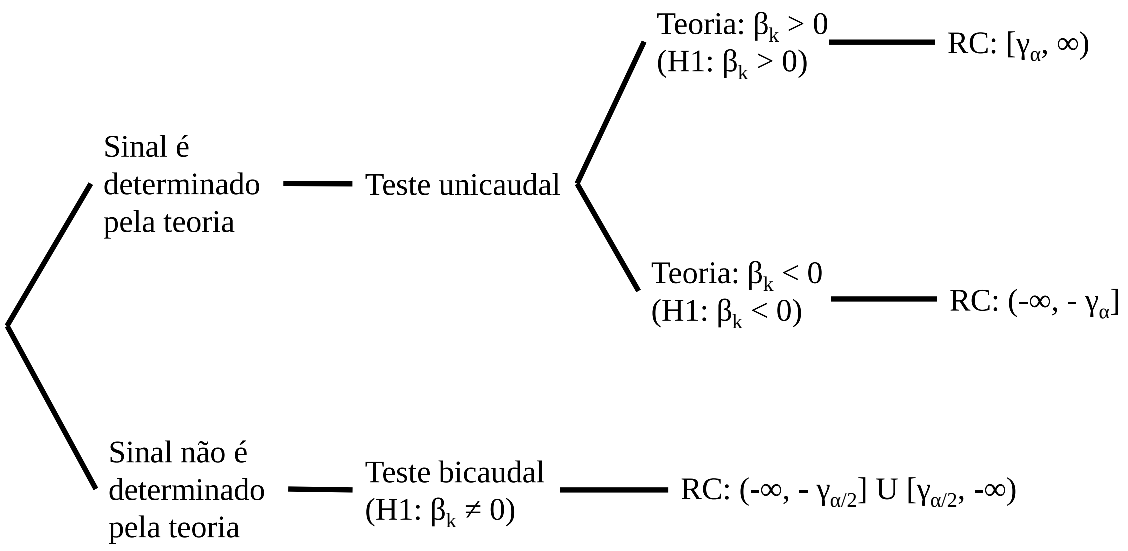
Em que β_K^0 é o valor populacional do coeficiente especificado pela hipótese nula, zero, no caso ($\beta_K^0=0$). A expressão no denominador é a estimativa pontual da variância do estimador para o k-ésimo coeficiente.

Como segundo passo, recorre-se a (pelo menos) uma das duas abordagens disponíveis para medir o grau em que a evidência é favorável à hipótese nula. Em termos gerais, a abordagem da região crítica requer que sejam identificados os valores γ_1 e γ_2 tais que $P(\gamma_1 \leq t_{N-(K+1)} \leq \gamma_2) = 0{,}95$, enquanto que a abordagem do p-valor exige a obtenção da probabilidade $2P(t_{N-(K+1)} > |\hat{t}(H_0)|)$. O símbolo "$t_{N-(K+1)}$" indica uma VA com FD t de Student com N-(K+1) graus de liberdade, uma vez que esta é a quantidade de "informação bruta" com base na qual a estatística t é calculada. Os procedimentos de implementação do teste de hipóteses específicos a cada uma das duas abordagens são apresentados de maneira detalhada no que segue.

2.2 Procedimento 1, abordagem da região crítica

Há três categorias de regiões críticas, definidas de acordo com o sinal atribuído pela teoria à influência exercida por x_k na variável dependente (Y). A figura 1 a seguir apresenta tais categorias, as quais são subdivisões das duas categorias maiores revistas no curso de inferência, as de teste unicaudal e bicaudal. Terá de ser selecionada sempre uma das três categorias e o critério para tanto é o que a teoria diz quanto ao sinal da influência de x_k sobre Y. Se ela nada disser, ou não for suficientemente clara, o mais adequado é optar pela hipótese alternativa mais completa, H1: $\beta_k \neq 0$. Ou seja, neste caso deve-se optar por um teste bicaudal. Porém, se a teoria for explícita quanto ao sinal, deve-se optar por aquela dentre as duas possibilidades para a hipótese alternativa condizente com a teoria, H1: $\beta_k > 0$ ou H1: $\beta_k < 0$.

Figura 1 Três categorias de regiões críticas e critérios para escolher entre elas



Uma vez definida a categoria de região crítica, a determinação desta se resume a encontrar o valor crítico γ_α em que o subscrito α indica o nível de significância ou probabilidade do erro tipo 1. Geralmente, $\alpha = 5\%$. Basta seguir as regras de bolso abaixo.

1. Se o teste é unicaudal, γ_α: $P(t_{N-(K+1)} \leq - \gamma_\alpha) = \alpha$ em que $t_{N-(K+1)}$ é uma VA com distribuição t com N- (K+1) graus de liberdade.

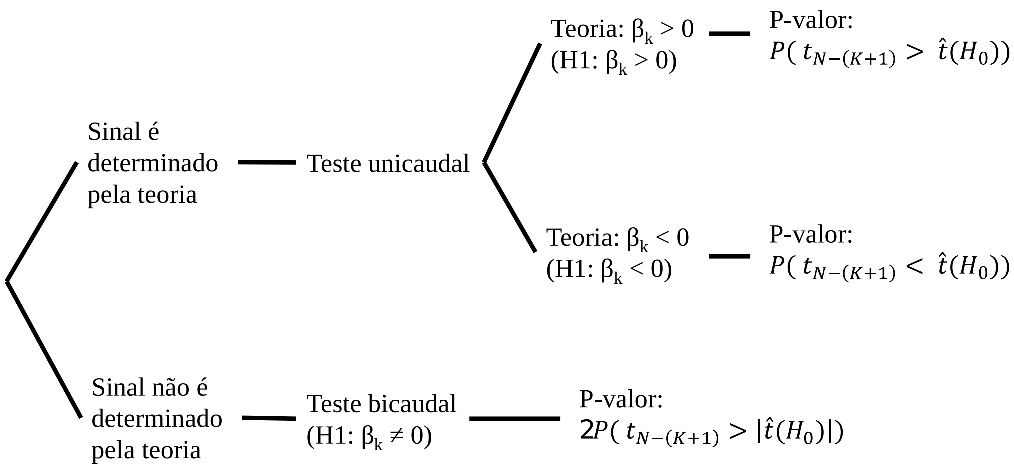
2. Se o teste é bicaudal, $\gamma_{\alpha/2}$: $P(t_{N-(K+1)} \leq -\gamma_{\alpha/2}) = \alpha/2$. Assim sendo, portanto, com α = 5%, os valores críticos do teste bicaudal, $-\gamma_{\alpha/2}$ e $\gamma_{\alpha/2}$, são tais que $P(t_{N-(K+1)} \leq -\gamma_{\alpha/2})$ = 0,05/2 = 2,5%.

O critério de decisão consiste em rejeitar a hipótese sempre que o valor observado da estatística do teste, $\hat{t}(H_0)$, pertencer à região crítica e não rejeitar caso contrário.

2.3 Procedimento 2, abordagem do p-valor

O p-valor é a probabilidade de obter-se um valor mais extremo para a estatística do que o valor observado. O que define se por "valor mais extremo" se entende um valor positivo ou negativo é, tal como no caso da abordagem da região crítica, a teoria, conforme a figura 2 aponta.

Figura 2 Três categorias de p-valores e critérios para escolher entre elas



Para comparar o nível de significância do teste, α = 5%, geralmente, com o p-valor, pode-se seguir as regras de bolso abaixo:

1. Se o teste é unicaudal
   a. E positivo, $P(t_{N-(K+1)} > \hat{t}(H_0))$ deve ser comparado com 5%;
   b. E negativo, $P(t_{N-(K+1)} < \hat{t}(H_0))$ deve ser comparado com 5%;
2. Se o teste é bicaudal, $P(t_{N-(K+1)} > |\hat{t}(H_0)|)$ deve ser comparado com 2,5%. Ou, de maneira equivalente, $2P(t_{N-(K+1)} > |\hat{t}(H_0)|)$ deve ser comparado com 5%, em que $|\hat{t}(H_0)|$ é o valor absoluto da estatística.

O critério de decisão consiste em rejeitar a hipótese sempre que o p-valor for inferior ao nível de significância do teste (5%, geralmente) e não rejeitar caso contrário.

2.4 Exemplo 1: Kassouf, 1994, explicando a desnutrição

A tabela da próxima página apresenta os resultados obtidos por Kassouf (1994). A autora estimou uma FRP explicando o escore Z para a altura de crianças de zero a cinco anos. As principais explicativas são a remuneração por hora trabalhada recebida pelos pais das crianças, a educação das mães e medidas de acesso a serviços de utilidade pública com implicações relevantes para a saúde de crianças, como o abastecimento de água e o saneamento básico. Como controles, a autora selecionou características socioeconômicas das crianças e das famílias a que as crianças pertencem como etnia, gênero e idade das crianças, idade dos pais e escore Z para a altura dos pais. A descrição precisa das variáveis pode ser encontrada após a tabela.

A autora optou por estimar a FRP separadamente para cada uma das quatro regiões brasileiras cobertas pela pesquisa que gerou os dados estudados.

Tabela 1 Resultados obtidos por Kassouf (1994). Coeficientes e respectivas estatísticas t entre parêntesis, dados: Pesquisa Nacional sobre saúde e nutrição, INAN/IBGE/IPEA, 1988

Variáveis	Regiões			
	Nordeste	Sudeste	Sul	Central
Constante	-3,260	-2,482	-2,826	-0,791
	(-3,93)***	(-2,82)***	(-3,32)***	(-0,86)***
PLUMB	0,261	0,135	-0,076	0,303
	(2,24)**	(1,35)	(-0,78)	(3,11)***
SEWERAGE	0,049	0,120	0,226	0,151
	(0,37)	(1,00)	(2,48)**	(1,30)
ELECTRIC	0,027	0,342	0,403	0,091
	(0,25)	(3,24)***	(3,85)***	(0,82)
STRPAVEM	0,124	0,176	-0,031	0,091
	(1,00)	(1,61)*	(-0,28)	(0,78)
URBAN	-0,167	-0,261	-0,117	-0,222
	(-1,35)	(-2,05)**	(-1,06)	(-1,98)**
ZHAMOT	0,295	0,360	0,319	0,298
	(8,48)***	(9,57)***	(9,18)***	(7,55)***
ZHAFAT	0,297	0,227	0,224	0,319
	(8,89)***	(6,21)***	(6,12)***	(8,44)***
MOTAGE21	0,323	0,380	0,179	0,011
	(2,06)**	(2,20)**	(1,10)	(0,07)
FATAGE21	-0,028	-0,98	-0,847	-0,323
	(-0,06)	(-0,23)	(2,11)**	(-0,86)
FATLITER	0,135	0,082	0,077	-0,016
	(1,62)*	(0,74)	(0,69)	(-0,15)
EDUFAT5	-0,206	-0,009	0,109	-0,174
	(-1,43)	(-0,07)	(1,03)	(-1,42)
MOTLITER	0,215	0,110	0,251	0,118
	(2,61)***	(0,97)	(2,13)**	(1,06)
EDUMOT5	0,253	-0,005	-0,075	0,204
	(1,81)*	(-0,04)	(-0,72)	(1,73)
KIDAGE3	-0,312	-0,271	-0,312	-0,210
	(-3,30)***	(-2,51)***	(-3,14)***	(-2,02)**
KIDAGE4	-0,220	-0,304	-0,386	-0,362
	(-2,31)**	(-2,82)***	(-3,86)***	(-3,30)***
KIDAGE5	-0,226	-0,404	-0,365	-0,350
	(-2,36)**	(-3,81)***	(-3,65)***	(-3,17)***
WHITE	-0,053	0,175	-0,051	0,094
	(-0,60)	(2,04)**	(-0,49)	(1,16)
BLACK	-0,105	0,101	0,444	0,310
	(-0,50)	(0,69)	(1,76)*	(1,19)
ASIAN	0,790	-0,334	0,242	1,037
	(1,80)*	(-0,58)	(0,45)	(0,94)
MALE	0,002	0,012	0,029	-0,014
	(0,02)	(0,16)	(0,42)	(-0,19)
LFLINCPC	0,530	0,410	0,318	0,237
	(5,02)***	(3,16)***	(2,68)***	(1,73)*
LWAGEMOT	-0,141	-0,008	-0,011	-0,180
	(1,66)*	(-0,08)	(-0,12)	(-1,85)*
LWAGEFAT	-0,113	-0,242	-0,227	0,163
	(-0,87)	(-1,67)*	(-1,68)*	(1,16)
R^2	0,31	0,30	0,30	0,30
Teste *F*	23,72***	16,70***	16,17***	15,57***
Observações	1.213	918	905	878

OBS.: Os testes *t* são dados entre parênteses abaixo dos coeficientes.
* Indica nível de significância de 10%.
** Indica nível de significância de 5%.
*** Indica nível de significância de 1%.

Definições das explicativas da FRP estimada por Kassouf (1994)

1 Características das crianças

1.a) Medidas de etnia: WHITE, BLACK, ASIAN; tratam-se de dummies que assumem valor 1 para a etnia indicada pelos nomes das variáveis, branca, negra e asiática

1.b) Gênero: MALE, indica com valor 1 se a criança é do sexo masculino

1.c) Idade: KIDAGE3, KIDAGE4, KIDAGE5, variáveis binárias indicando, respectivamente, se a criança tem 3, 4 ou 5 anos

2 Características dos pais

2.a) saúde: ZHAMOT, ZHAFAT, escore Z para a altura da mãe (MOT) e do pai (FAT)

2.b) Idade: MOTAGE21, FATAGE21, binárias indicando com valor unitário se pai e mãe, respectivamente, são maiores de 21 anos

2.c) Educação: FATLITER, MOTLITER, binárias que indicam com valor 1 se, respectivamente, pai e mãe são alfabetizados. E há também duas binárias que indicam se pai e mãe têm pelo menos 5 anos de estudo, EDUFAT5 e EDUMOT5.

2.d) logaritmo da renda da família, captada por LFLINCPC

2.e) salário por hora recebido por pai e mãe, LWAGEMOT e LWAGEFAT

3 Características do local

3.a) acesso à serviços públicos: há uma série de binárias indicando com valor unitário se há acesso a abastecimento público de água, PLUMB, tratamento de esgoto, SEWERAGE, energia elétrica, ELECTRIC e ruas pavimentadas, STRPAVEM.

3.b) o pertencimento do domicílio a áreas urbanas é indicado por valor unitário para a binária URB.

É um exercício frutífero o de identificar, com base na tabela acima, as explicativas que possuem influência individual (efeito parcial) relevante sobre Y. E isso pois assim pode-se compreender um dos objetivos do teste de significância individual que é exatamente o de determinar quais explicativas têm influência nula, na população, e quais têm influência relevante.

A estatística do teste de significância individual para o k-ésimo coeficiente é:

$$\hat{t}(H_0) = \frac{\hat{\beta}_{MQOk}}{\left[\sqrt{\hat{\sigma}^2 / \sum_{i=1}^{N} \left(x_{ik} - \overline{x}_k\right)^2 \left(1 - R_k^2\right)}\right]} \quad t_{N-(K+1)}$$

Os valores desta estatística para as explicativas em todas as regiões estão disponíveis na tabela acima. Resta calcular os valores críticos e os p-valores.

Uma vez que há 23 explicativas nas FRPs estimadas para todas as quatro regiões, K = 23. Já o número de observações varia com a região, sendo N_{NE} = 1213, N_{SE} = 918, N_S = 905 e N_C = 878. É possível obter os valores críticos para o teste de significância bicaudal a 5% partindo de $P(t_{N-(K+1)} \leq -\gamma_{\alpha/2}) = 0{,}025$. Os valores críticos variam apenas em função da FD, ou melhor, dos parâmetros dela, no caso, graus de liberdade. É preciso, pois, identificar um par de valores críticos para cada uma das quatro regiões. Para isso, basta utilizar o comando do software R dado por qt(0.025,df = N_i – 23 - 1), i = NE, SE, S, C. A tabela a seguir apresenta, na terceira coluna, os resultados.

Tabela 2 Valores críticos para o teste de significância individual, exemplo Kassouf (1994)

Região	K, N e N-K-1	Valores críticos, bicaudal, sig = 95%
NE	K: 23; N = 1213, N-K-1 = 1189	-1,961961; +1,961961
SE	K: 23, N = 918 , N-K-1 = 894	-1,962621; +1,962621
SE	K: 23; N = 905, N-K-1 = 881	-1,962660; + 1,962660
C	K: 23; N = 878, N-K-1 =854	-1,962660; + 1,962660

Como se pode ver, os valores críticos diferem em magnitude desprezível ao longo das quatro regiões, podendo-se tomar por base, para todas, a região crítica (-∞;-1,96] U [1,96; ∞).

Desta maneira, pode-se ver que a binária que indica com valor unitário o acesso a recursos hídricos é significativa apenas nas regiões SE e C, pois é apenas nelas em que a estatística t para a estimativa pontual do coeficiente da variável é superior, em módulo, a 1,96.

É interessante, como exercício de fixação, calcular alguns p-valores. Primeiramente, o valor da estatística t para a binária que indica com valor unitário se a criança reside em uma área urbana é equivalente a -1,35 para a região NE. O sinal negativo significa que crianças de áreas urbanas tendem a ter menor valor para o indicador de desnutrição, e, portanto, menor nível de saúde. O p-valor para o teste bicaudal é 2P(t>|-1,35|) e pode ser obtido a partir do comando 2*pt(-1.35,df=1189) ou, alternativamente, 1 – pt(1.35,df=1189). Cabe recordar que o comando pt() do R reporta a probabilidade acumulada à esquerda do valor entrado como argumento, i.e., pt(t^O,df=N-(K+1)) = P($t_{N-(K+1)} < t^O$). Resulta que 2*pt(-1.35,df=1189) = 2*0,089 = 0,18 > 0,05 e, portanto, não se pode rejeitar a hipótese de nulidade para o coeficiente da binária “URB” na regressão para a região NE.

Como segundo exercício para fixação do p-valor, a atenção repousará sobre o coeficiente para o salário da mãe, LWAGEMOT, na região Central. O valor da estatística t observado é de -1,85 e, portanto, mães com maior custo de oportunidade para o tempo dedicado ao cuidado das crianças tendem a alocar menos tempo para esta atividade, o que repercute em um nível de saúde inferior para as crianças. O p-valor para o teste bi-caudal é obtido a partir do comando 2*pt(-1.35,df=1189), o qual reporta como resultado o número 0,064 > 0,05 e, portanto, não se pode rejeitar a hipótese de nulidade do coeficiente do salário materno para a região Central. Cabe observar que, contudo, a tabela de Kassouf (acima) traz o símbolo “*” na célula referente ao coeficiente em questão, o que indica que há significância estatística, mas porém, a um nível de significância de 10%. Esta conclusão está de acordo com o cálculo que se acabe de fazer, uma vez que 0,064 < 0,1.

2.5 Exemplo 2, Basant et al., 2006, investimento em tecnologia de informação e comunicação (TIC), amostra de empresas industriais localizadas no Brasil, 2003

Tabela 3 Resultados da estimação por MQO de quatro modelos para o investimento em informação e comunicação

	(1) %trab_com_PCs	(2) %trab_com_automação	(3) ln(Inv./trab)	(4) ln(Inv./vendas)
30-250emp	11,885	-1,325	-0,442	-0,485
	(5,627)**	(4,482)	(0,265)*	(0,340)
250+emp	24,076	8,265	-0,324	-0,359
	(6,638)***	(6,022)	(0,314)	(0,399)
mne	13,850	16,646	-0,528	0,804
	(9,697)	(14,888)	(0,365)	(1,106)
so	-18,481	-9,116	0,631	0,367
	(12,116)	(8,151)	(0,706)	(0,802)
listed	-9,740	4,984	0,776	0,660
	(7,042)	(7,122)	(0,542)	(0,809)
jvd	8,445	1,591	-0,324	-0,567
	(4,100)**	(3,738)	(0,232)	(0,303)*
jvf	-0,069	5,031	-0,200	-0,654
	(5,867)	(6,079)	(0,277)	(0,465)
union	-0,238	1,882	-0,086	-0,341
	(4,067)	(3,643)	(0,217)	(0,293)
Admin	0,427	0,320	0,021	0,008
	(0,137)***	(0,160)**	(0,008)**	(0,013)
Managers	-0,239	0,371	0,029	0,034
	(0,471)	(0,343)	(0,016)*	(0,019)*
Other	0,063	0,197	-0,006	-0,004
	(0,141)	(0,155)	(0,010)	(0,014)
Prod:US	0,047	0,245	0,003	-0,001
	(0,067)	(0,059)***	(0,004)	(0,004)
Prod:Coll	0,113	0,554	0,019	0,012
	(0,136)	(0,239)**	(0,008)**	(0,011)
Admin:US	0,247	-0,067	-0,000	0,002
	(0,099)**	(0,074)	(0,005)	(0,006)
Admin:Coll	0,281	-0,073	0,005	0,005
	(0,108)***	(0,077)	(0,005)	(0,006)
Age<10	3,311	2,809	0,494	0,325
	(4,390)	(3,863)	(0,232)**	(0,297)
Age11-20	4,763	4,486	0,476	0,710
	(4,497)	(4,264)	(0,240)**	(0,313)**
Observations	358	358	256	226
R-squared	0,23	0,24	0,21	0,16

Nota: desvio padrão entre parêntesis.

Definições das explicativas da FRP estimada por Basant et al (2006)

1 Variáveis dependentes

- %trab_com_PCs ≡ porcentagem de trabalhadores administrativos (colarinho branco) que utilizam computadores pessoais (PCs);
- %trab_com_automação ≡ porcentagem de trabalhadores de chão de fábrica (colarinho azul) que operam máquinas controladas por TIC;
- ln(Inv./trab) ≡ logaritmo natural da razão investimento em TIC / número de trabalhadores;
- ln(Inv./vendas) ≡ logaritmo natural da razão investimento em TIC / vendas.

2 Variáveis explicativas

O estudo não define as explicativas, e apenas foi possível deduzir da breve descrição (p.9 e p.21) as explicativas abaixo.

- 30-250emp: tamanho da empresa entre 30-250 empregados? (dummy)
- 250+emp: tamanho da empresa entre 30-250 empregados? (dummy)
- mne: multinacional? (dummy)
- union: porcentagem de trabalhadores sindicalizados
- Admin: porcentagem de trabalhadores administrativos
- Managers: porcentagem de diretores
- Other: porcentagem de trabalhadores que não realizam atividades de produção, administração e direção;
- Prod:US, proporção de trabalhadores empregados na produção que concluíram o ensino médio;
- Prod:Coll, proporção de trabalhadores empregados na produção que concluíram o ensino superior;
- Admin:US, proporção de trabalhadores administrativos que concluíram o ensino médio;
- Admin:Coll, proporção de trabalhadores administrativos que concluíram o ensino superior;
- Age<10: idade da empresa
- Age11-20: idade da empresa

A tabela acima contém os resultados de regressões que procuram explicar o investimento em tecnologias de informação e comunicação (TIC), realizados por empresas brasileiras do setor de indústria da transformação. Trata-se, pois, do investimento em computadores, automação, aparatos de comunicação, etc. E as variáveis explicativas contém características das empresas, tais como tamanho medido em número de funcionários, nacionalidade dos proprietários da empresa, nível educacional dos funcionários, entre outros determinantes do investimento em TIC sugeridos pela teoria. Os resultados provém de Basant et al (2006)[1].

Cabe realizar testes de hipóteses para os coeficientes de algumas explicativas. Seja considerada a dummy que indica com valor unitário se a empresa é multinacional, “mne”. O valor da estimativa pontual na equação (1) é de 13,85 e o desvio padrão de 9,697, o que implica em um valor para a estatística t de aproximadamente 1,43. O número de explicativas pode ser calculado como 27, apesar de não estar explicitamente indicado no artigo, mas é dito que além das 17 explicativas da tabela acima há 10 explicativas adicionais captando regiões do país e setor produtivo. Uma vez que o número de observações da equação (1) é de 358, a estatística t é calculada a partir de 358-27 = 331 graus de liberdade. Para este número de graus de liberdade, os valores críticos a um nível de significância de 5% são de -1,97 e 1,97. Estes valores compõem um intervalo que contém o valor da estatística t, dado por 1,43. Deste modo, pois, não se deve rejeitar a hipótese de que o fato de a empresa ser multinacional tem influência desprezível sobre a taxa de uso de computadores pelos trabalhadores administrativos.

Para a equação (3), a estimativa pontual para a dummy “multinacional” foi de -0,442 com desvio padrão de 0,265. A estatística t tem valor aproximadamente de -1,45. O número de graus de liberdade é de 256 – 27 = 229 e os valores críticos para o nível de significância de 5% são de -1,97 e 1,97. Conclusivamente, pois, o fato da empresa ser ou não multinacional tem influência desprezível sobre o investimento médio em TIC por trabalhador.

Seja considerada agora a explicativa binária que indica se a empresa possui mais de 250 funcionários, podendo, pois, ser considerada “grande” dentro da amostra. A estimativa pontual para o coeficiente desta variável foi de 24,076 com desvio padrão de 6,638. Consequentemente, a estatística t tem valor 3,63. Conforme visto anteriormente, tem-se 331 graus de liberdade e valores críticos de -1,97 e 1,97 e, portanto, deve-se rejeitar a hipótese de que o fato de uma empresa possuir mais de 250 funcionários tem influência desprezível sobre a proporção de trabalhadores administrativos com PCs. Conclusivamente, há evidência de que tal proporção é tão maior quanto maior for a empresa, o que talvez possa ser explicado a partir da existência de economias de escala na aquisição e manutenção de equipamentos de TIC. I.e., tais custos talvez sejam menores quanto maior o número de funcionários administrativos e, portanto, maior o número de computadores de que a empresa necessita.

2.5 Exemplo 3: Agee, 2010, explicando a desnutrição na Nigéria

Tabela 4 coeficientes e respectivas estatísticas t entre parêntesis, dados: Pesquisa demográfica e de saúde, de abrangência nacional, para a Nigéria, ano de 2003

[1] O artigo completo pode ser encontrado no link a seguir: http://ftp.iza.org/dp2294.pdf.

Estimates of child height-for-age and weight-for-age ($n = 1359$).

	Height-for-Age	Weight-for-Age
	OLS[a]	OLS[a]
	(1)	(5)
constant term	−7.384**(−6.34)	−6.09**(−7.35)
Community variables:		
Family resides in the country-side	−0.093(−0.76)	−0.195**(−2.05)
Family resides in a large city	−0.073(−0.65)	−0.346**(−3.82)
Child has vaccination card	1.481**(4.54)	1.369**(5.53)
Percent of regional women reporting big problems accessing health care due to not knowing where to go for treatment	−0.015**(−1.96)	−0.012**(−2.10)
Iodine (ppm) in household salt	0.003(0.52)	0.0002(0.11)
Family variables:		
Birth order of referent child	0.032(1.48)	0.007(0.44)
Referent child is male	−0.092(−1.07)	−0.063(−0.94)
Referent child's age	−0.051**(−4.58)	−0.045**(−4.63)
Referent child's age squared	0.0006**(3.63)	0.0006**(4.35)
Mother's age	0.018**(2.34)	0.015**(2.32)
Mother's height	0.004**(6.49)	0.003**(6.90)
Mother has a primary education	0.061(0.41)	0.206(1.56)
Mother has a secondary education	0.058(0.39)	0.097(0.82)
Mother indicates short to moderate distance to a usable health care facility	0.575(1.34)	0.093(0.55)
Household size	0.058(1.30)	0.040(1.16)
Household size squared	−0.005**(−2.14)	−0.002(−1.49)
Family's standardized wealth level	0.359**(6.05)	0.154**(3.14)
F-statistic	15.49**	14.14**

**Significant at less than 5 percent.
*Significant at less than 10 percent.
[a] Asymptotic *t*-statistics in parentheses are computed using robust standard errors corrected for region-level clustering.

A tabela acima apresenta uma parte dos resultados obtidos por Agee (2010) em um estudo empírico da desnutrição infantil na Nigéria.

A pergunta do autor é: crianças pertencentes a famílias mais ricas e cujas mães possuem mais informação acerca dos serviços de saúde disponíveis para a comunidade apresentam melhor status nutricional?

A hipótese que o estudo procura testar é a de que a riqueza familiar e a quantidade de informação acerca de serviços de saúde a que a mãe tem acesso têm efeitos positivos no nível de saúde das crianças.

Há uma diferença em relação a artigo de Kassouf. Considera-se, como variável dependente, não apenas o escore Z para a altura das crianças, mas também o escore Z para a massa corpórea das crianças. As principais variáveis cujos efeitos se deseja mensurar são as que captam a informação detida pelas mães acerca dos serviços de saúde disponíveis e o nível de riqueza da família. Quanto à primeira, ela é indicada, na tabela, pela frase “*percent of regional women reporting big problems accessing health care due to not knowing where to go for treatment*” (proporção de mulheres que declararam ter muita dificuldade para acessar serviços de saúde por não conheceram a localização das unidades que os prestam). A riqueza é captada por “*family´s standardized wealth level*” (nível de riqueza familiar padronizado).

Para ler os resultados é preciso ter em conta que estimativas estatisticamente significativas a um nível de significância de 10% estão indicadas com um asterisco, “*” e estimativas significativas a um nível de 5% com dois asteriscos, “**”.

Pode-se observar que tanto a informação quanto a serviços de saúde como a educação têm influência não nula, em termos populacionais, sobre a severidade de desnutrição. Os sinais das estimativas pontuais revelam que mães informadas e famílias mais ricas reduzem a severidade da desnutrição.

2.6 Exemplo 4: discriminação de residentes de favelas pelo mercado de trabalho

Os pesquisadores Leandro Rocha, Marcelo Pessoa e Danielle Machado, do Centro de Estudos sobre Desigualdade e Desenvolvimento (CEDE), da Universidade Federal Fluminense (UFF), Rio de Janeiro, estimaram uma FRP explicando a remuneração recebida por trabalhadores brasileiros em função de características pessoais[2]. Entre elas, há uma variável binária indicando, com valor unitário, se o indivíduo reside em uma favela. Como controles, há medidas de etnia, gênero, nível educacional, situação no mercado de trabalho e atividade exercida. Os resultados são reproduzidos na tabela na próxima página.

[2] O artigo completo pode ser encontrado em http://www.proac.uff.br/cede/sites/default/files/TD61.pdf.

Tabela 5 Resultados da regressão Minceriana, variável dependente: logaritmo do rendimento mensal. Dados: PNAD 2009

Explicativa	Estimativa pontual (coeficiente)	Efeito parcial percentual na variável dependente	P-valor
Idade	0,0485	4,97%	<0,01%
Idade ao quadrado	-0,0004	-0,04%	<0,01%
Branco?	0,1392	14,94%	<0,01%
Homem?	0,331	39,24%	<0,01%
1 ano de escolaridade ?	-0,0123	-1,22%	0,9363
2 anos de escolaridade?	0,0259	2,62%	0,7985
3 anos de escolaridade ?	-0,0145	-1,44%	0,8644
4 anos de escolaridade?	0,0004	0,04%	0,995
5 anos de escolaridade?	0,074	7,68%	0,3859
6 anos de escolaridade ?	-0,0564	-5,48%	0,5112
7 anos de escolaridade?	0,1557	16,85%	0,0634
8 anos de escolaridade?	0,1902	20,95%	0,0016
9 anos de escolaridade?	0,2064	22,92%	0,0183
10 anos de escolaridade?	0,212	23,61%	0,0028
11 anos de escolaridade?	0,435	54,50%	<0,01%
12 anos de escolaridade?	0,6202	85,93%	<0,01%
13 anos de escolaridade?	0,7071	102,81%	<0,01%
14 anos de escolaridade?	0,9336	154,36%	<0,01%
15 anos de escolaridade?	1,2112	235,75%	<0,01%
16 anos de escolaridade?	1,4493	326,01%	<0,01%
17 anos ou mais de escolaridade?	1,9923	633,24%	<0,01%
Funcionário público?	0,189	20,80%	<0,01%
Empregado sem carteira assinada?	-0,2459	-21,80%	<0,01%
Autônomo?	-0,2021	-18,30%	<0,01%
Horas de Trabalho	0,013	1,30%	<0,01%
Residente de favela?	-0,0704	-6,80%	0,0237
Observações		4280	
R^2 ajustado		0,52	
Teste de significância global	176,26		<0,01%

OBS: o efeito parcial percentual de variáveis binárias é equivalente a $\exp(\hat{\beta}_k)$ - 1, em que exp() é a função exponencial cuja base é o número neperiano. Para a variável contínua “horas de trabalho”, o coeficiente é equivalente ao efeito parcial percentual.

Há pelo menos três pontos que chamam atenção nos resultados. Em primeiro lugar, o efeito parcial de morar em uma favela é negativo e significativo, uma evidência de que, mesmo controlando por etnia, gênero e nível educacional, residentes de favelas são negativamente discriminados pelo mercado de trabalho, recebendo menores salários. Em segundo lugar, o mercado de trabalho discrimina negativamente não apenas por residência em favelas, mas também por etnia e também por gênero. Em terceiro lugar, apenas níveis educacionais acima de 8 anos de escolaridade fazem diferença, positiva, no que tange à remuneração.

3 Intervalos de confiança

O teste de significância individual se restringe a avaliar se a estimativa pontual é uma evidência favorável ou desfavorável à hipótese de nulidade populacional do k-ésimo coeficiente. Cabe refletir quanto ao que esta informação nos diz e quanto ao que ela não nos diz. Ela nos diz que o parâmetro populacional (o coeficiente da FRP) não é nulo, o que é crucial: a variável explicativa associada tem, portanto, efeito parcial relevante sobre Y na população. Mas há algo importante que a informação em questão não nos diz. Ela não nos diz quais valores, dentre aqueles que sobram após o descarte do valor zero, estão mais próximos do valor efetivo do parâmetro populacional. É esta informação que o intervalo de confiança pode nos dar – não necessariamente descartando, para isso, o valor zero.

Para construir intervalos de confiança que contenham, com uma probabilidade satisfatória, α, o valor populacional do k-ésimo coeficiente da FRP, basta tomar dois limites, I_1 e I_2, tais que $P(I_1 < \beta_k < I_2) = \alpha$ (1). Para isso, parte-se do fato de que $P(-\gamma_{\alpha/2} < t < \gamma_{\alpha/2}) = \alpha$, i.e., sempre é possível identificar dois valores de uma variável aleatória com FD t de Student entre os quais há um intervalo associado ao nível satisfatório de probabilidade, ou nível de confiança. Como segundo passo, considera-se que $\frac{\left(\hat{\beta}_{MQOk}-\beta_k\right)}{\sqrt{\hat{\sigma}^2/\sum_{i=1}^{N}\left(x_{ik}-\bar{x}_k\right)^2\left(1-R_k^2\right)}}$ $t_{N-(K+1)}$ e, pois, $P\left(-\gamma_{\alpha/2} < \frac{\left(\hat{\beta}_{MQOk}-\beta_k\right)}{\sqrt{\hat{\sigma}^2/\sum_{i=1}^{N}\left(x_{ik}-\bar{x}_k\right)^2\left(1-R_k^2\right)}} < \gamma_{\alpha/2}\right) = \alpha$ (2). A partir da manipulação que transforma (2) em (1) chega-se à especificação a seguir para o intervalo de confiança.

$$IC(\beta_k;\alpha) = \{\hat{\beta}_{MQOk} - \gamma_{\alpha/2}sd(\hat{\beta}_{MQOk});\hat{\beta}_{MQOk} + \gamma_{\alpha/2}sd(\hat{\beta}_{MQOk})\}$$

Em que $sd(\hat{\beta}_{MQOk}) = \sqrt{\hat{\sigma}^2/\sum_{i=1}^{N}\left(x_{ik}-\bar{x}_k\right)^2\left(1-R_k^2\right)}$.

É interessante calcular o intervalo de confiança para um exemplo. Retomando os resultados de Basant et al. (2006), seja considerado o coeficiente da dummy que indica com valor unitário se o tamanho da empresa é superior a 250 empregados (250+emp). Na equação (4), em que é explicado o logaritmo da razão entre o investimento em TIC (tecnologia de informação e comunicação) e as vendas, a estimativa pontual para a explicativa em questão foi de -0,359 com desvio padrão de 0,399. O número de graus de liberdade para a FD t de Student é de 226 – 27 = 199, de maneira que os coeficientes de confiança para o nível de 95% sejam de -1,97 e 1,97. Com isso, o intervalo de confiança é dado por {-0,359 – 1,97*0,399; -0,359 + 1,97*0,399} = {-0.427; 1.145}.

4 Teste de restrições de exclusão

4.1 Apresentação do teste

É possível testar a hipótese de que um subgrupo de explicativas exerce efeito conjunto nulo sobre a variável dependente, uma vez contabilizado o efeito das demais explicativas. Trata-se, portanto, de aferir a contribuição conjunta de algumas explicativas e não a contribuição individual, específica, de cada uma.

Seja assumido que as explicativas cuja significância conjunta se deseja testar são aquelas referentes aos coeficientes β_k, β_{k+1},..., β_{k+s}. Trata-se, pois, de “s” variáveis. Neste caso, o enunciado para o teste de hipóteses é tal como segue:

$$H_0: \beta_k = 0 \text{ e } \beta_{k+1} = 0 \text{ e ... e } \beta_{k+s} = 0 \text{ vs } H_1: \beta_k \neq 0 \text{ ou } \beta_{k+1} \neq 0 \text{ ou...ou } \beta_{k+s} \neq 0$$

O teste, portanto, tem como hipótese nula a possibilidade de que s das K explicativas proporcionem, quando tomadas em conjunto, contribuição nula para a explicação da variação de Y. Após, é claro, ter sido contabilizada a contribuição das demais explicativas. As condições de nulidade aplicadas a cada coeficiente são denominadas por “restrições de exclusão”.

É preciso seguir um procedimento específico para este teste, diferente do adotado para o teste de significância individual. Ele consiste nos passos enumerados a seguir.

1. Obter as estimativas pontuais para os parâmetros da FRP, empregando, para isso, uma regressão linear com todas as variáveis explicativas;
2. A partir das estimativas pontuais obtidas no passo anterior, calcular a soma dos quadrados dos resíduos da regressão, $SQR_{IR}=\sum_{i=1}^{N}\hat{u}_i^2$, em que o subscrito “IR”, abreviando “irrestrito”, indica que se trata da regressão completa;
3. Obter as estimativas pontuais empregando uma regressão linear que contenha apenas as explicativas cujos coeficientes não são definidos como nulos pela hipótese nula. Este modelo será denominado por “modelo restrito”, dado que corresponde à FRP original reduzida pelas restrições de exclusão que compõem a hipótese nula;
4. A partir das estimativas pontuais do passo 3 deve-se calcular a soma dos quadrados da regressão restrita, $SQR_R=\sum_{i=1}^{N}\tilde{u}_i^2$, com o subscrito “R” abreviando o termo “restrito”;
5. A estatística do teste é $\hat{F}=\frac{(SQR_R-SQR_{IR})/s}{SQR_{IR}/(N(K+1))}\;F_{s,N-(K+1)}$, em que $F_{s,N-(K+1)}$ indica a função de distribuição de probabilidades F de Snedecor com s graus de liberdade no numerador e N-(K+1) graus de liberdade no denominador. Cabe recordar que s é o número de restrições (de exclusão) impostas ao modelo restrito (está-se testando, aqui, a hipótese de que s das K explicativas são nulas);

6. Uma estatística com distribuição F é sempre positiva, deste modo, pois, o teste de hipóteses é sempre unicaudal.
    a. A região crítica é dada por $RC(\alpha) = [F_\alpha;\infty)$, com F_α: $P(F_{s,N-(K+1)} > F_\alpha) = 5\%$
    b. O p-valor é equivalente a $P(F_{s,N-(K+1)} > \hat{F})$.

Atenção: é preciso ter cautela ao utilizar um software como o R para obter o valor crítico e o p-valor de um teste F

Quanto ao valor crítico, o fato da distribuição F estar definida apenas para valores positivos, conforme ilustrado na figura 3 abaixo, implica que o valor crítico (sempre há apenas um valor crítico) é tal que a probabilidade acumulada pelo conjunto de valores superiores a ele é equivalente ao nível de significância, α (5%, geralmente). Ou seja, o valor crítico é sempre F_α tal que $P(F_{s,N-(K+1)} > F_\alpha) = \alpha$. É comum o erro de, não atentando para o fato de que a distribuição F é assimétrica, obter o valor crítico a partir de $P(F_{s,N-(K+1)} < F_\alpha) = \alpha$, o que tende a reportar um valor inferior à unidade e, portanto, tal erro conduz, por equívoco, à rejeição da hipótese nula (se α = 95%, a região crítica tem 95% de probabilidade de ocorrer; há, claramente, algo errado). Deve-se notar que o valor crítico equivocado corresponde, na FD da F, a um valor muito próximo de zero, conforme a figura 4 ilustra e, portanto, não pode ser tomado como valor crítico, pois este, supostamente, é um valor consideravelmente distante de zero.

Já, para o p-valor, o erro comum decorre do fato de que a maioria dos softwares reporta a função de distribuição acumulada (FDA), i.e., $P(F_{s,N-(K+1)} < \hat{F})$, mesmo sendo que o p-valor corresponde à probabilidade de obter um valor superior à $\hat{F}$ o que é equivalente a $P(F_{s,N-(K+1)} > \hat{F})$. Como, geralmente, $P(F_{s,N-(K+1)} < \hat{F}) > \alpha$, conforme sugere a figura 5, deixa-se, por equívoco, de rejeitar a hipótese nula. Uma maneira de evitar este erro está em requisitar ao software que calcule $1 - P(F_{s,N-(K+1)} < \hat{F})$.

Figura 3 FD da F de Snedecor; linha contínua: 2 graus de liberdade no numerador e 8 graus de liberdade no denominador; linha pontilhada: 6 e 8 graus de liberdade.

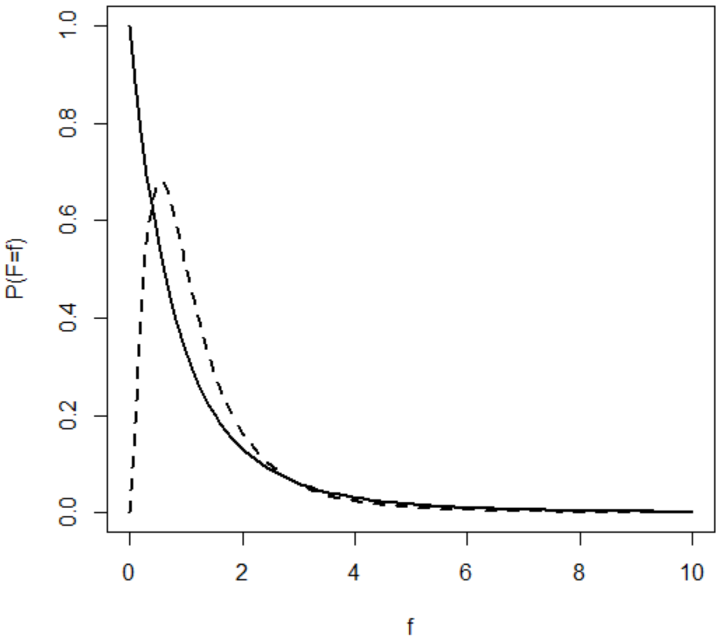


Figura 4 Valores críticos equivocado e correto para a $F_{2,8}$, α = 5%

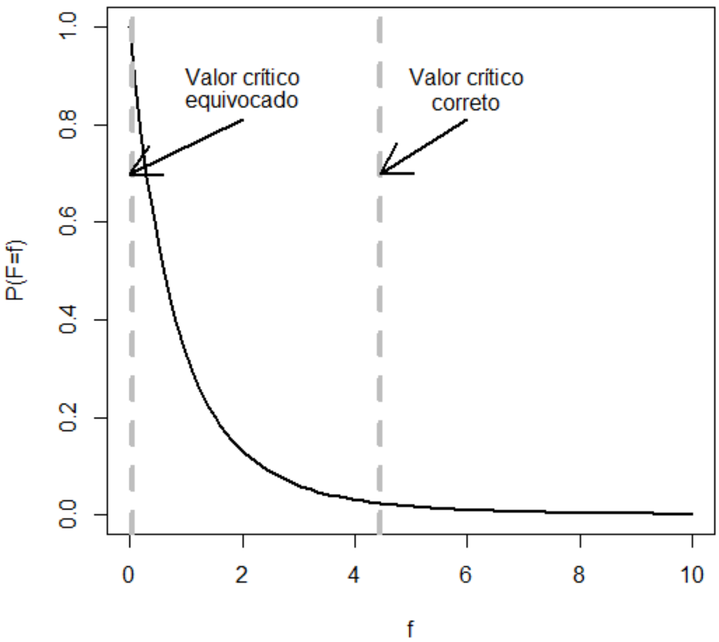
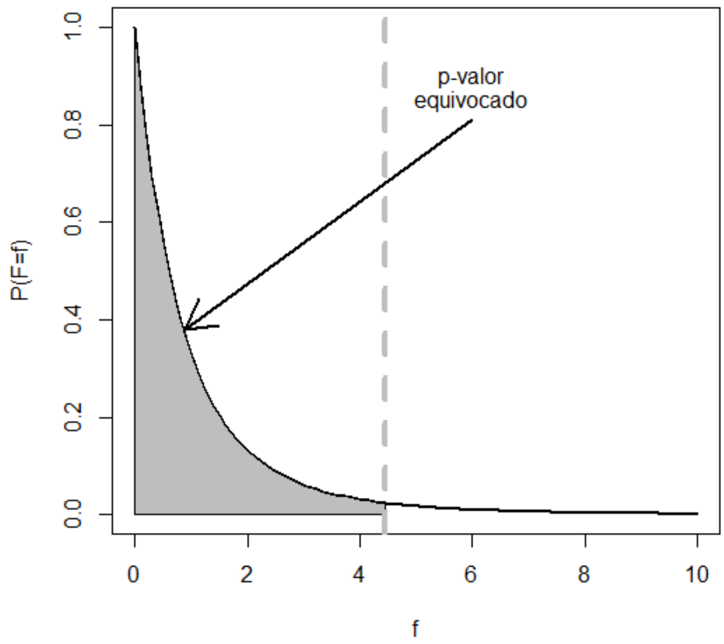


Figura 5 P-valor equivocado para a $F_{2,8}$



4.2 Significância estatística sob multicolinearidade: a vantagem do teste de restrições de exclusão

Duas variáveis explicativas entre as quais existe um coeficiente de correlação estatisticamente significativo são ditas “colineares”. No contexto da regressão múltipla é comum o uso do termo sinônimo “multicolinearidade”[3] para denotar a existência de correlações relevantes entre pelo menos duas explicativas.

Quando há multicolinearidade, as estimativas pontuais para os coeficientes das explicativas correlacionadas são menos precisas e geralmente as estatísticas t associadas se mostram próximas de zero. Ou seja, tende-se a concluir pela não rejeição da hipótese de valor nulo para os coeficientes. A razão está na dificuldade de distinguir a contribuição específica, para a explicação da variável dependente, de explicativas correlacionadas. De fato, o conteúdo informacional que duas explicativas representam é tão mais similar ou redundante quanto maior for a correlação entre elas.

O teste de significância individual para o coeficiente de uma explicativa consideravelmente correlacionada com pelo menos uma das demais explicativas tende a resultar na não-rejeição da hipótese nula. Isso ocorre mesmo quando a explicativa em questão tem, de fato, influência considerável sobre a variável dependente – um erro do tipo II. Este aparente “equívoco” quanto ao resultado do teste é, infelizmente, inevitável.

Sejam os índices das s explicativas entre as quais há correlação denotados por k, k+1,...,k+s. Testar a significância conjunta destas explicativas, i.e., se todas elas, simultaneamente, têm influência nula sobre a dependente, significa implementar o teste cujas hipóteses são:

[3] As causas e remédios factíveis para a multicolinearidade são objeto do curso de Econometria II. Uma discussão detalhada pode ser encontrada em Goldberger, A.S., A course in econometrics, cap.23.

$$H_0: \beta_k = 0 \text{ e } \beta_{k+1} = 0 \text{ e... e } \beta_{k+s} = 0 \text{ vs } H_1: \beta_k \neq 0 \text{ ou } \beta_{k+1} \neq 0 \text{ ou...ou } \beta_{k+s} \neq 0$$

Trata-se de verificar se a contribuição <u>conjunta</u> de $x_k,...,x_{k+s}$ para a explicação da variação de Y na amostra é estatisticamente relevante – daí porque o teste de múltiplas restrições de exclusão é também chamado de "teste de significância conjunta". O que não é equivalente a verificar se as contribuições individuais de cada uma das s explicativas são relevantes.

O teste de múltiplas restrições de exclusão é uma alternativa mais adequada para aferir a nulidade populacional da influência de grupos de explicativas correlacionadas entre si quando há multicolinearidade na amostra. Ao concentrar-se na contribuição conjunta, o teste de restrições de exclusão se faz, por construção, menos suscetível a distorções impostas pela multicolinearidade, uma vez que seu resultado não depende da eficácia com que a influência individual das explicativas pode ser distinguida. E é exatamente por depender desta eficácia que o teste de significância individual se mostra menos preciso sob multicolinearidade.

Uma maneira recomendável de compensar a imprecisão dos testes de significância individual quando há correlação relevante entre as explicativas (multicolinearidade) está em conduzir testes de restrições de exclusão referentes a grupos de explicativas correlacionadas entre si.

A correlação tende a ser mais comum entre explicativas que mensuram uma mesma característica socioeconômica e, portanto, os grupos para aplicar o teste de restrições de exclusão podem ser definidos com base na proximidade conceitual das características socioeconômicas medidas pelas explicativas. É esclarecedor considerar três exemplos.

Na equação de Mincer em que a remuneração dos trabalhadores é explicada em função de suas características individuais, o nível educacional, a experiência e indicadores de desempenho em exames admissionais são medidas para o nível de produtividade intrínseca geralmente correlacionadas. Um teste de restrições de exclusão pode ser aplicado às três medidas de capital humano.

É comum no estudo empírico da desnutrição infantil incorporar como explicativas binárias que indicam acesso a diversos serviços de utilidade pública, tais como abastecimento hídrico, esgotamento, suprimento de eletricidade, pavimentação, etc. Estas binárias tendem a estar correlacionadas uma vez que a infraestrutura urbana contempla, na maioria das situações, boa parte dos serviços mencionados. E, contrariamente, nos locais desprovidos de infraestrutura urbana, boa parte dos serviços em questão tendem a não ser acessíveis (i.e., geralmente, não é apenas um dos serviços que não está disponível). É adequado, pois, aplicar um teste de restrições de exclusão a todas as medidas de acesso a serviços de utilidade pública.

Um terceiro exemplo é fornecido por Wooldridge. Trata-se de estimar uma equação que explica a remuneração de *Chief Executive Officers* (CEOs) de empresas privadas, em função, por exemplo, da performance econômico-financeira da empresa. Uma vez que há diversos indicadores possíveis de performance, pode não ser claro o bastante, com base na teoria, qual indicador é mais adequado. Isso pode conduzir à inclusão de múltiplos indicadores de performance, potencialmente correlacionados (p.ex., lucro contábil, valor de mercado, competitividade no mercado internacional, etc.). Para testar a hipótese de que a performance explica a remuneração dos CEOs, um teste de restrições de exclusão aplicado aos diversos indicadores de performance é uma opção menos sujeita à imprecisões introduzidas pelas multicolinearidade do que os testes de significância individual.

4.3 Computando a estatística F a partir de valores do R^2

Há uma relação interessante entre a estatística do teste de significância conjunta e o coeficiente de determinação não ajustado, R^2, este equivalente à proporção da variação de Y explicada pelo modelo, tal como medida pela razão entre SQE e SQT. Basta partir da fórmula geral de R^2 = SQE/SQT e recordar que SQE = SQT – SQR para, com alguma manipulação chegar a SQR = (1 - R^2)SQT. Este resultado geral vale, em específico, para SQR_R e SQR_{IR} definidas acima. Incorporando isso à fórmula da estatística $\widehat{F}$, pode-se obter:

$$\widehat{F}=\frac{\left(R_{IR}^2-R_R^2\right)/s}{\left(1-R_{IR}^2\right)/\left(N-\left(K+1\right)\right)} \quad F_{s,N-(K+1)}$$

Cabe assinalar que a ordem dos modelos restrito e irrestrito no numerador é trocada na manipulação que se acaba de fazer. O que faz todo o sentido, pois o poder explicativo do modelo com menos variáveis explicativas não pode ser maior. O que tem duas implicações. Em primeiro lugar, $SQR_{IR} \leq SQR_R$, em segundo lugar, $R^2_{IR} \geq R^2_{IR}$. O que explica a ordem com que estes quatro elementos aparecem no numerador de cada uma das duas versões da estatística $\widehat{F}$. Isso é coerente com o fato de que a estatística deve ser sempre positiva, conforme exige a definição da distribuição F.

A segunda fórmula da estatística do teste sugere que é possível realizar o procedimento anteriormente descrito calculando-se, ao invés da SQT, o coeficiente de determinação para cada uma das duas versões do modelo, restrita e irrestrita.

Conforme assinalado por Wooldridge, deve-se ter por claro que o coeficiente de determinação que compõe a estatística do teste de múltiplas restrições lineares não corresponde à versão ajustada (para os graus de liberdade) da estatística, mas sim ao R^2 ordinário, não ajustado.

5 Teste de significância global

5.1 Teoria

O teste de hipóteses para coeficientes específicos verifica a significância estatística da influência individual de variáveis explicativas, tomadas uma a uma. Apesar da significância de pelo menos um coeficiente ser condição suficiente para que a fração da variação de Y explicada pelo modelo se mostre significativa, ela não é condição necessária. O que significa que a porção da variável dependente explicada pela regressão linear pode ser relevante, estatisticamente, mesmo que todas as explicativas sejam individualmente não significativas. De fato, mesmo que todas as variáveis explicativas exerçam influências estatisticamente não-significativas, quando tomadas individualmente, a composição de tais influências, quando tomadas em conjunto, pode ser estatisticamente significativa.

Desta maneira, testar a significância da porção da variação de Y explicada, ou "significância global" da regressão – pois se trata da influência conjunta de todas as variáveis explicativas - não é redundante com os testes de significância aplicados a coeficientes individuais.

O teste de significância global é definido como:

$$H_0: \beta_1 = 0 \text{ e } \beta_2 = 0 \text{ e... e } \beta_K = 0 \text{ vs } H_1: \beta_1 \neq 0 \text{ ou } \beta_2 \neq 0 \text{ ou...ou } \beta_K \neq 0$$

Assinale-se que o intercepto, β_0, é desconsiderado pelas hipóteses. Isso ocorre pois o intercepto não capta a contribuição de uma explicativa para a explicação da variação de Y, este o aspecto que se deseja colocar em xeque com o teste.

O teste que se acaba de definir é uma extensão do teste de múltiplas restrições de exclusão. Trata-se de considerar todas as restrições de exclusão possíveis, referentes, pois, a todas as explicativas. É intuitivo, portanto, que o teste de significância global se baseie na mesma estatística do teste de restrições de exclusão. De fato, a hipótese nula do teste de significância global é equivalente à afirmação de que as explicativas, tomadas conjuntamente, explicam uma proporção irrelevante da variação de Y, i.e., SQE = 0 e, pois, o valor <u>populacional</u> do coeficiente de determinação, R^2, é nulo. Para fins estritamente didáticos, uma formulação alternativa para o teste de significância global seria:

$$H_0: R^2 = 0 \text{ vs } H_1: R^2 \neq 0$$

O procedimento de teste é mais simples do que é o caso para o teste de restrições de exclusão. Vale a pena entender a razão para isso antes de prosseguir. Como já observado, a estatística do segundo teste é equivalente a:

$$\hat{F} = \frac{\left(R_{IR}^2 - R_R^2\right)/s}{\left(1 - R_{IR}^2\right)/\left(N - \left(K+1\right)\right)} \quad F_{s,N-(K+1)}$$

A que corresponde exatamente R_R^2 quando todas as restrições de exclusão possíveis se aplicam à FRP? A FRP reduzida por todas as restrições de exclusão possíveis se degenera para $Y = \beta_0 + u$, ou seja, o modelo contém apenas o intercepto. É um fato importante, ainda que simples, mostrar que explicar Y exclusivamente em função do intercepto, i.e., de uma constante, é equivalente a explicar Y em função de sua média. É preciso, para isso, obter o estimador de MQO para a FRP restrita Y = β_0 + u, conforme os passos a seguir.

$$min_{|\beta_0|} E\left[\left(Y-\beta_0\right)^2\right]$$

A condição de primeira ordem é[4]:

$$\frac{\partial}{\partial \beta_0} E\left[\left(Y-\beta_0\right)^2\right]=0 \rightarrow E\left[\frac{\partial}{\partial \beta_0}\left(Y-\beta_0\right)^2\right]=0 \rightarrow E\left[Y-\beta_0\right]=0$$

(notar que a expectativa não é condicional pois não há variáveis explicativas no modelo)

Aplicando o princípio da analogia, tem-se:

$$\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\widehat{\beta_0}\right)=0 \rightarrow \sum_{i=1}^{N} y_i - N\widehat{\beta_0}=0 \rightarrow \widehat{\beta_0}=\frac{1}{N}\sum_{i=1}^{N} y_i$$

O estimador para o intercepto no modelo restrito é, portanto, a média amostral de Y. A implicação mais importante, no momento, é a de que a soma dos quadrados explicada degenera para o valor zero. Basta notar que o valor previsto da variável dependente pelo modelo restrito é $\hat{y}_i=\widehat{\beta_0}$, e, conforme demonstrado, $\widehat{\beta_0}=\bar{y}$. Conclui-se, pois, que, $\hat{y}_i=\bar{y}$, i.e., a variável dependente é prevista como sendo equivalente à sua média amostral. Uma vez que $SQE_R=\sum_{i=1}^{N}\left(\hat{y}_i-\bar{y}\right)^2$, de modo que $SQE_R=\sum_{i=1}^{N}\left(\bar{y}-\bar{y}\right)^2=0$. Do que decorre que R^2_R = SQE_R / SQT =0.

Desta maneira, pois, não é preciso calcular SQE_R ou R^2_R, pois seus valores são conhecidos a priori. E, além disso, como tais valores são nulos, a estatística do teste passa a:

$$\hat{F}=\frac{R^2_{IR}/s}{\left(1-R^2_{IR}\right)/\left(N-\left(K+1\right)\right)} \quad F_{s,N-(K+1)}$$

Há duas maneiras de contar o número de graus de liberdade do numerador da estatística do teste, $\hat{F}$. A primeira, mais rápida, toma por base o fato de testar a significância conjunta de todas as explicativas é equivalente a impor um número de restrições de exclusão igual ao número total de coeficientes, K. Desta maneira, pois, s = K. A segunda maneira considera a versão original da estatística do teste, reproduzida abaixo:

$$\hat{F}=\left(SQR_R-SQR_{IR}\right)/¿¿$$

Para calcular SQR_R é preciso estimar um parâmetro, o intercepto, de modo que restam N – 1 graus de liberdade. E, para calcular SQR_{IR} é preciso estimar K+1 parâmetros. Desta maneira, pois, o número de graus de liberdade do numerador é N- 1 – (N – (K +1)) = K. A estatística do teste de significância global, portanto, é, finalmente:

[4] Épossível utilizar o cálculo escalar (e não matricial) pois não há explicativas e β_0 é escalar.

$$\hat{F} = \frac{R^2_{IR}/K}{(1-R^2_{IR})/\left(N-(K+1)\right)} \quad F_{K,N-(K+1)}$$

Agora, passando ao procedimento do teste.

1. Estatística do teste: $\frac{R^2/K}{(1-R^2)/N-(K+1)} \quad F_{K,N-K-1}$
2. Abordagem da região crítica: para um nível de significância α, a região crítica é: $\{\hat{F} : \hat{F} > F_{\alpha}\}$;
3. Abordagem do p-valor: para um nível de significância α, rejeitar a hipótese nula se $P¿$) < α, em que F é uma variável aleatória com função de distribuição de probabilidade (FD) dada por $F_{K,\ N-\ K-1}$ e $\hat{F}$ é o valor observado da estatística do teste. Geralmente toma-se $\alpha = 5\%$.

Cabe assinalar que o R^2 que integra a fórmula da estatística do teste de significância global corresponde à versão <u>ordinária, não ajustada</u>, da estatística.

5.2 Exemplos

A tabela abaixo contém insumos para calcular a estatística do teste de significância global para os exemplos utilizados para o teste de significância individual. Vejamos como fazer o cálculo. A fórmula da estatística é F = $\frac{R^2/K}{(1-R^2)/N-(K+1)}$. De acordo com a tabela, a regressão de Kassouf (1994), em que o indicador de desnutrição infantil é explicado na região Nordeste, $R^2 = 0{,}31$, K = 23, N-K- 1 = 1189. Desta maneira, pois, F = $\frac{0{,}31/23}{0{,}69/1189}$=23,23, conforme consta na quarta coluna da tabela 6. O p-valor pode ser obtido com o software R a partir do comando pf(23.23,df1=23,df2=1189,lower.tail=FALSE), o que resulta no número na última coluna da tabela 5. Uma vez que este é consideravelmente inferior a 5%, rejeita-se a hipótese de ausência de significância global da regressão para a região Nordeste.

Cabe destacar que a instrução “lower.tail=FALSE” informa ao R que se deseja obter $P¿$) e não $P¿$), evitando, assim, que seja reportado um p-valor equivocado (segundo discussão na subseção 4.1 acima)[5].

O mesmo exercício pode ser repetido para a regressão de Mincer, estimada pelo CEDE com o intuito de gerar evidência que permita concluir acerca da discriminação de residentes de favelas pelo mercado de trabalho. Neste caso, R^2= 0,52, K = 26 e N- K- 1

[5] Segundo se pode ler no link a seguir, que explica os comandos do R para a função de distribuição F, “lower.tail: logical; if TRUE (default), probabilities are P[X ≤ x], otherwise, P[X > x]”. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/Fdist.html

= 4253. A estatística do teste é F = $\frac{0{,}52/26}{0{,}48/4253}$=177,21. O p-valor, calculado a partir do comando pf(177.21,df1=26,df2=4253,lower.tail=FALSE) é tão pequeno que, conforme a última coluna da tabela indica, o R reporta zero. Desta maneira, pois, pode-se afirmar que a regressão do CEDE é globalmente significativa.

Tabela 6 Insumos para calcular a estatística de significância global e resultados, exemplos de regressões múltiplas

Artigo/ Estatística	R^2/ R^2 ajustado	K, N e N-K-1	Valor da estatística F (calculado)	Valor da estatística F (artigo)	p-valor da estatística F (calculado)
Kassouf (1994), NE	0,31	K: 23; N = 1213, N-K-1 = 1189	23,23	23,72	9,00E-80
Kassouf (1994), SE	0,3	K: 23, N = 918 , N-K-1 = 894	16,66	16,7	1,26E-54
Kassouf (1994), S	0,3	K: 23; N = 905, N-K-1 = 881	16,42	16,17	1,10E-53
Kassouf (1994), C	0,3	K: 23; N = 878 (C), N-K-1 =854	15,91	15,57	9,93E-52
Agee (2010)	NA	K = 17 ; N = 1359, N-K-1 = 1341	NA	15,49 e 14,14	<5%
CEDES-UFF (2011)	0,52	K = 26 ; N = 4280, N-K-1 = 4253	177,21	176,26	0

Apêndice A Teste de combinações lineares para os coeficientes

As restrições de exclusão são apenas uma categoria de condições ou restrições que podem ser impostas aos parâmetros da FRP. Em algumas aplicações, é interessante saber se existem relações lineares entre os coeficientes. Um exemplo interessante, explorado por Gujarati (cap.8) é o da estimação de uma função de produção que descreva um processo produtivo.

A especificação Cobb-Douglas é uma forma funcional adotada por um número relevante de estudos empíricos. Assumindo que são utilizados apenas dois fatores de produção, trabalho, L e capital (máquinas e equipamentos), representado pela letra “G”, a produção de Y unidades é $Y = F(G, L) = G^{\alpha}L^{\beta}$.

Uma característica da especificação Cobb – Douglas está em sua generalidade. Em particular, ela permite representar as três possibilidades para os retornos de escala, quais sejam (a) crescentes, i.e., um aumento dos dois fatores em uma mesma proporção resulta em um aumento mais do que proporcional da produção; (b) constantes, com o aumento equivalente dos dois fatores resultando em um aumento exatamente proporcional da produção e (c) decrescentes, caso em que o aumento dos dois fatores na mesma proporção induz um aumento menos do que proporcional da produção. Para verificar que estas três possibilidades são a priori possíveis na especificação acima, basta realizar o exercício abaixo.

$$F(\lambda G, \lambda L) = (\lambda G)^{\alpha}(\lambda L)^{\beta} = \lambda^{\alpha+\beta} G^{\alpha}L^{\beta} = \lambda^{\alpha+\beta}F(G, L) = \lambda^{\alpha+\beta}Y$$

Tem-se, pois:

1. Se $\alpha + \beta > 1$, há retornos crescentes (caso a);
2. Se $\alpha + \beta = 1$, há retornos constantes (caso b);
3. Se $\alpha + \beta < 1$, há retornos decrescentes (caso c);

Para estimar a função de produção teórica, introduz-se um termo que sintetiza elementos não-observados que explicam o nível de produção, ε, e uma constante, “A”, a qual pode captar o nível de progresso técnico. Chega-se, com essas adições, a $Y = AG^{\alpha}L^{\beta}\varepsilon$. Para transformar este modelo em uma forma funcional linear, basta aplicar o logaritmo dos dois lados da equação

$$\log(Y) = \log(G^{\alpha}L^{\beta}u) = \log A + \alpha\log G + \beta\log L + \log\varepsilon$$

Denotando $\log A$ por η, $\log\varepsilon$ por u, e as VAs em logaritmo com letras minúsculas, chega-se ao modelo de regressão linear abaixo:

$$y = \eta + \alpha g + \beta l + u$$

Como testar, com base na estimação deste modelo, a hipótese de retornos constantes? Conforme discutido, há retornos constantes sempre que $\alpha + \beta = 1$. Basta, portanto, realizar o teste de hipóteses H0: $\alpha + \beta = 1$ vs H1: $\alpha + \beta \neq 1$. É oportuno observar que este teste pode também ser representado como H0: $\alpha + \beta - 1 = 0$ vs H1: $\alpha + \beta - 1 \neq 0$, o que permite uma interpretação de teste de significância individual em que o parâmetro cuja nulidade está em xeque é $\alpha + \beta - 1$.

A principal diferença deste teste em relação aos anteriores, no que respeita à implementação está em que a estatística deve incorporar uma combinação linear de duas estimativas pontuais.

$$\hat{t}(H_0) = \frac{\hat{\alpha} + \hat{\beta} - (\alpha_0 + \beta_0)}{\sqrt{\widehat{V[\hat{\alpha} + \hat{\beta}]}}} \quad t_{N-(K+1)}$$

Em que $\alpha_0 + \beta_0$ é o valor da combinação linear dos coeficientes definida pela hipótese nula, ou seja, $\alpha_0 + \beta_0 = 1$. E $K+1 = 3$. O numerador da estatística, pois, pode ser calculado facilmente a partir da estimação dos coeficientes. O cálculo do denominador pressupõe o conhecimento da variância da soma dos parâmetros. De partida, com base em um resultado conhecido de estatística, é de se esperar que

$\widehat{V[\hat{\alpha}+\hat{\beta}]}=\widehat{V[\hat{\alpha}]}+\widehat{V[\hat{\beta}]}+2\widehat{cov(\hat{\alpha},\hat{\beta})}$. A dificuldade está em calcular a covariância entre os parâmetros. Não é possível, neste particular, recorrer à hipótese de independência entre os estimadores de α e β, isso pois ambos são obtidos do mesmo conjunto de dados, simultaneamente, a partir de um sistema de equações lineares que os relaciona. Não podem, portanto, ser independentes. A fórmula para calcular a variância é obtida a partir da matriz de covariância dos estimadores, $\widehat{V[\hat{\beta}]}=\hat{\sigma}^2(X'X)^{-1}$, adotando-se, é claro, as hipóteses de homocedasticidade e autocorrelação (MCRL 5 e MCRL6). Entre os elementos que não pertencem à diagonal principal há a correlação procurada.

Há, contudo, duas alternativas mais diretas para realizar o teste de restrições lineares. A primeira delas, proposta no livro-texto de Wooldridge, consiste em definir α +β = θ e reescrever o teste da seguinte maneira:

$$H0: \theta = 1 \text{ vs } H1: \theta \neq 1$$

A simplificação possibilitada por esta convenção pode ser compreendida introduzindo-se o parâmetro θ no modelo a partir da relação α = θ – β. O que resulta no modelo alterado dado por y = η + θg + β(l-k) + u. É preciso, pois, criar uma nova explicativa, l-k, o que pode ser feito facilmente. Feito isso, a estimação do modelo alterado permite obter diretamente uma estimativa pontual para a soma dos coeficientes α e β e também para o desvio padrão do estimador para esta soma. A estatística do teste é como segue:

$$\hat{t}(H_0)=\frac{\hat{\theta}-1}{\widehat{V[\hat{\theta}]}} \quad t_{N-(K+1)}$$

Trata-se de uma estatística com distribuição t, cujos componentes são conhecidos pois correspondem à estimativa pontual e ao desvio padrão do coeficiente da explicativa “g”. O teste, bicaudal, no caso, pode ser implementado a partir da construção da região crítica (- ∞, - $\gamma_{\alpha/2}$] U [$\gamma_{\alpha/2}$, ∞), em que α é o nível de significância do teste (geralmente 5%) e $\gamma_{\alpha/2}$ é tal que F(t ≤ - $\gamma_{\alpha/2}$) = α/2 (geralmente 2,5%). Também há a opção de calcular o p-valor $P¿$.

Uma terceira maneira de realizar o teste é comparando-se a SQR ou o R^2 da regressão restrita pela condição α + β = 1 com a regressão irrestrita. A estatística do teste e o procedimento são equivalentes ao que se tem para o teste de múltiplas restrições lineares. Com a qualificação de que o presente teste consiste em uma única restrição, i.e., s =1, de modo que a estatística é:

$$\hat{F}=\frac{(SQR_R-SQR_{IR})/1}{SQR_{IR}/(N(K+1))} \quad F_{1,N-(K+1)}$$

Apêndice B Abordagem alternativa para o teste de restrições lineares aos parâmetros

A.1 Introdução

A estatística F vista nas subseções 4 a 6 é uma medida para a discrepância da SQR entre dois modelos, um que incorpora restrições lineares e outro que não as incorpora. Há, contudo, uma estatística F alternativa, cuja construção se apoia em uma interpretação distinta do problema de testar a validade de múltiplas restrições lineares. Ela toma por base uma notação geral que permite representar todas as possíveis restrições lineares que se pode impor ao vetor de parâmetros, β. Tal notação, quando empregada para especificar o teste de hipóteses, resulta em:

$$H0: R\beta = q \text{ vs. } H1: R\beta \neq q$$

O primeiro componente da notação é a matriz R. Esta tem dimensão s x K+1 em que s é o número de restrições lineares cuja validade está em xeque. O segundo componente é o vetor q, de dimensão s x 1.

A melhor maneira de entender a composição de R e de q é a partir de exemplos de testes de restrições lineares.

Seja considerado o teste para restrições de exclusão referentes ao intercepto e aos coeficientes das duas primeiras explicativas. Ou seja, trata-se do teste H_0: $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0$ vs $\beta_0 \neq 0$ ou $\beta_1 \neq 0$, $\beta_2 \neq 0$. Para este teste, R assume a forma a seguir:

$$R_{sx(K+1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

Já o vetor q é dado por:

$$q = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Ou seja, assumem valor unitário apenas os elementos de R correspondentes à posição dos coeficientes cuja nulidade se deseja testar. Os demais elementos assumem valor nulo (o que os permite excluir do teste).

Deste modo, pois:

$$R\beta = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \dots \\ \beta_K \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

E, consequentemente, H_0: Rβ – q é equivalente a:

$$H_0 : \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Esta versão da hipótese nula é equivalente à vista anteriormente, de que H_0: $\beta_0 = 0$, $\beta_1 = 0$, $\beta_2 = 0$.

Para o teste de significância global, R assume a forma de uma matriz com valores unitários em sua diagonal principal e zeros nas demais posições, i.e.:

$$R_{sx(K+1)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Já “q” é um vetor K x 1 de zeros, pois há restrições de exclusão para todos os coeficientes, i.e., K restrições de exclusão.

De modo que:

$$H_0 : R\beta = q \longleftrightarrow H_0 : \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \dots \\ \beta_K \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

O terceiro exemplo é o de um teste para relações lineares envolvendo quatro parâmetros no total, sendo elas: (i) $\beta_1 + \beta_2 + \beta_3 = C_0$ e (ii) $\beta_3 + \beta_4 = D_0$, em que C_0 e D_0 são constantes com valores arbitrários. Como escrever essas duas restrições lineares na forma Rβ = q? Responder a esta questão é um exercício importante para fixar a notação aqui descrita.

Em primeiro lugar, há duas restrições lineares, então a matriz R tem duas linhas. Cada uma dessas linhas deve ser preenchida com valores unitários e nulos de maneira a que, quando pré-multiplicada por β, resulte um vetor de duas linhas, a primeira exatamente equivalente a $\beta_1 + \beta_2 + \beta_3$ e a segunda equivalente a $\beta_3 + \beta_4$. A matriz R apropriada, portanto, é tal que:

$$R\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\\ \dots\\ \beta_K\end{bmatrix}=\begin{bmatrix}\beta_1+\beta_2+\beta_3\\ \beta_3+\beta_4\end{bmatrix}$$

Refletindo um pouco, chega-se a:

$$R=\begin{bmatrix}0 & 1 & 1 & 1 & 0 & 0 & \dots & 0\\ 0 & 0 & 0 & 1 & 1 & 0 & \dots & 0\end{bmatrix}$$

O vetor q é 2 x 1 e seus elementos são as constantes a que correspondem as formas lineares dos parâmetros, C_0 e D_0, i.e.:

$$q=\begin{bmatrix}C_0\\ D_0\end{bmatrix}$$

Reunindo as formas de R e q em duas hipóteses, tem-se:

$$H_0:\begin{bmatrix}0 & 1 & 1 & 1 & 0 & 0 & \dots & 0\\ 0 & 0 & 0 & 1 & 1 & 0 & \dots & 0\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\\ \beta_3\\ \dots\\ \beta_K\end{bmatrix}=\begin{bmatrix}C_0\\ D_0\end{bmatrix}$$

O estudo da subseção que ora se encerra é relevante para a realização de testes de restrições lineares no software R, uma vez que o comando linerHypothesis requer o uso da notação Rβ = q[6].

A.2 Estatística F alternativa

Uma vez discutida a notação que dá base ao teste F alternativo, pode-se apresentar a estatística que ele toma por base.

$$F=\frac{\left(R\hat{\beta}-q\right)'\left(R\widehat{V[\hat{\beta}\vee X]}R'\right)^{-1}\left(R\hat{\beta}-q\right)}{s}\quad F_{s,N-K-1}(1)$$

Esta estatística tem uma distribuição F com s graus de liberdade no numerador e N-K-1 graus de liberdade no denominador, em que s é o número de restrições lineares.

---

6 Sobre este comando, ver a seção 3.4 de Farnsworth, G. V, “Econometrics in R”, documento disponível no link a seguir http://cran.r-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf

Há uma implicação fundamental desta estatística alternativa para testes de restrições lineares, a qual se restringe ao teste para uma <u>única</u> restrição de exclusão referente a um coeficiente. Este pode ser genericamente enunciado como H_0: $\beta_k = 0$ vs $\beta_k \neq 0$. Trata-se, portanto, de um teste de significância individual, o qual é geralmente realizado com base na estatística t. De fato, uma restrição de exclusão que se refere a apenas uma explicativa é equivalente à hipótese nula de um teste de significância individual para o coeficiente da explicativa. Não é surpreendente, portanto, que a estatística do teste F para H_0: $\beta_k = 0$ vs $\beta_k \neq 0$ tenha relação com a estatística t para o mesmo teste. Para chegar a esta relação, basta especificar a matriz R e o vetor q correspondentes à hipótese nula. Assumindo, sem perda de generalidade que k = 3, de modo que H_0: $\beta_3 = 0$ vs $\beta_3 \neq 0$, tem-se:

$$R=\begin{bmatrix}0 & 0 & 0 & 1 & 0 & 0 & \dots & 0\end{bmatrix}, q=[0]$$

Assim, pois:

$$R\hat{\beta}-q \longleftrightarrow \begin{bmatrix}0 & 0 & 0 & 1 & 0 & 0 & \dots & 0\end{bmatrix}\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \dots\\ \hat{\beta}_K\end{bmatrix}=0 \longleftrightarrow \beta_3=0$$

E, pois, $R\hat{\beta}-q=\beta_3$. Deste modo, a estatística F passa a:

$$F_3=\frac{\hat{\beta}_3\left(\widehat{V\left[\hat{\beta}_3 \middle| X\right]}\right)^{-1}\hat{\beta}_3}{s}(1')$$

Como $\hat{\beta}_3$ e $\widehat{V[\hat{\beta}_3 \vee X]}$ são escalares e há apenas uma restrição, i.e., s = 1, é correto escrever:

$$F_3=\frac{\hat{\beta}_3^{\,2}}{\widehat{V[\hat{\beta}_3 \vee X]}} \quad F_{1,N-K-1}(1'')$$

Esta estatística é exatamente equivalente ao quadrado da estatística t para o teste de significância estatística de β_3. Basta recordar que a última é equivalente a:

$$t_3=\frac{\hat{\beta}_3}{\sqrt{\widehat{V[\hat{\beta}_3 \vee X]}}} \quad t_{N-K-1}(2)$$

Deste modo, pois, por (1´´) e (2), $F_3 = t_3^2$. Este resultado é válido, qualquer que seja a explicativa cuja nulidade de seu coeficiente é submetida a teste.

Apêndice C R^2 ajustado e explicativas não estatisticamente significativas

O impacto sobre o coeficiente de determinação ajustado ($\overline{R}^2$) causado pela inclusão de uma explicativa adicional está relacionado ao valor absoluto da estatística t desta variável e, portanto, à sua significância estatística. Analogamente, considerando a incorporação de um conjunto de explicativas adicionais, o efeito desta expansão da FRP sobre o $\overline{R}^2$ depende da significância conjunta das explicativas adicionais, tal como medida pelo valor da estatística F em um teste de múltiplas restrições de exclusão em que o modelo original e o modelo expandido são comparados.

Uma maneira de compreender o fundamento destes dois princípios está em estudar as relações existentes entre a variação no $\overline{R}^2$, de um lado, e os valores das estatísticas F e t, de outro lado, este o objetivo desta subseção. O argumento apresentado demonstra apenas o primeiro princípio, de maneira que a análise que segue se restringe à inclusão de apenas uma explicativa adicional. Contudo, a argumentação pode ser estendida para a inclusão de mais de uma explicativa adicional.

Conforme estabelecido na nota 8, dois modelos que diferem apenas em função da inclusão de uma explicativa, a K+1-ésima, i.e., $x_{K+1,}$ apresentam coeficientes de determinação cuja diferença é dada por:

$$\overline{R}_1^2 - \overline{R}_0^2 = \frac{N-1}{SQT}\left[\left(\frac{SQR_0}{N-K-1}\right) - \left(\frac{SQR_1}{N-K-2}\right)\right] (1)$$

Em que SQR_0 se refere ao modelo que não inclui x_{K+1} e SQR_1 ao modelo que a inclui.

Definindo $\frac{\delta_0}{\rho} \equiv \frac{SQR_0}{N-K-1}$ e $\frac{\delta_1}{\rho-1} \equiv \frac{SQR_1}{N-K-2}$, $\rho \equiv N - K - 1$, a expressão acima muda para:

$$\overline{R}_1^2 - \overline{R}_0^2 = \frac{N-1}{SQT}\left(\frac{\delta_0}{\rho} - \frac{\delta_1}{\rho-1}\right) (1')$$

Um pouco de manipulação algébrica permite expressar (1’) de maneira mais adequada.

$$\overline{R}_1^2 - \overline{R}_0^2 = \frac{N-1}{SQT}\left(\frac{\delta_0}{\rho} - \frac{\delta_1}{\rho} + \frac{\delta_1}{\rho} - \frac{\delta_1}{\rho-1}\right) = \frac{N-1}{SQT}\left(\frac{\delta_0 - \delta_1}{\rho} + \delta_1\left(\frac{1}{\rho} - \frac{1}{\rho-1}\right)\right) \rightarrow$$

$$\overline{R}_1^2 - \overline{R}_0^2 = \frac{N-1}{SQT}\left(\frac{\delta_0 - \delta_1}{\rho} + \delta_1\left(\frac{1}{\rho} - \frac{1}{\rho-1}\right)\right) (1'')$$

(na primeira passagem, somou-se e subtraiu-se o mesmo termo, $\frac{\delta_1}{\rho}$ dentro dos parêntesis).

A estatística F para o teste da restrição de exclusão $\beta_{K+1} = 0$ assume a forma:

$$F=\frac{\left(SQR_0-SQR_1\right)/1}{SQR_1/\left(N-K-2\right)}\left(2\right)$$

Empregando as definições de δ_0, δ_1 e a ρ à (2):

$$F=\frac{\delta_0-\delta_1}{\delta_1/\left(\rho-1\right)}=\frac{\left(\delta_0-\delta_1\right)\left(\rho-1\right)}{\delta_1}\rightarrow\left(\delta_0-\delta_1\right)=\frac{F\,\delta_1}{\left(\rho-1\right)}\left(2'\right)$$

Incorporando o resultado (2') em (1''):

$$\overline{R}_1^2-\overline{R}_0^2=\frac{N-1}{SQT}\left(\frac{F\,\delta_1}{\rho\left(\rho-1\right)}+\delta_1\left(\frac{1}{\rho}-\frac{1}{\rho-1}\right)\right)=\frac{N-1}{SQT}\left[\delta_1\left(\frac{F}{\rho\left(\rho-1\right)}-\frac{1}{\rho\left(\rho-1\right)}\right)\right]=\left(\frac{N-1}{SQT}\right)\left(\delta_1\frac{F-1}{\rho\left(\rho-1\right)}\right)$$

Desta maneira, pois:

$$\overline{R}_1^2-\overline{R}_0^2=\delta_1\left(\frac{N-1}{SQT}\right)\left(\frac{F-1}{\rho\left(\rho-1\right)}\right)$$

Uma vez que $\frac{\delta_1}{\left(\rho-1\right)}\left(\frac{N-1}{SQT}\right)=1-\overline{R}_1^2$, chega-se, finalmente, a:

$$\frac{\overline{R}_1^2-\overline{R}_0^2}{1-\overline{R}_1^2}=\left(\frac{F-1}{\rho}\right)\left(2'''\right)$$

Dado que $1-\overline{R}_1^2=\frac{SQR_1/N-K-2}{SQT/N-1}$, SQR_1 e SQT são somas de quadrados e, portanto, sempre positivos e N – K - 2 e N - 1 são sempre positivas[7], então $1-\overline{R}_1^2$ é sempre positivo. O elemento ρ = N – K – 1 também é sempre positivo. Consequentemente, pois:

$$S\left(\overline{R}_1^2-\overline{R}_0^2\right)=S\left(F-1\right)\left(2^{iv}\right)$$

Em que S(.) é a função que dá o sinal de seu argumento. A equação 2^{iv} deixa claro que uma condição suficiente para que a inclusão de uma explicativa condicional tenha impacto nulo sobre o R^2 ajustado é a de que o valor da estatística do teste F de significância individual para tal explicativa seja equivalente a 1. Outra conclusão de grande importância é a de que se a estatística F for inferior a 1, o que geralmente é o caso quando a explicativa adicional é não estatísticamente significativa, então a inclusão dela acaba por reduzir o R^2 ajustado.

Cabe retomar a relação entre a estatística alternativa para o teste F, quando aplicada à restrição de exclusão para a k-ésima explicativa, e a estatística t para o teste de significância individual para a k-ésima explicativa. Conforme discutido na última subseção, $F_k = t_k^2$. Incorporando isso à equação 2^{iv}, tem-se:

[7] Esta afirmação é violada apenas para K+2 > N e N < 1, possibilidades, que contudo, inviabilizam a estimação do modelo.

$$\left(\overline{R_1^2} - \overline{R_0^2}\right) = S\left(t^2 - 1\right) \quad (2^v)$$

Uma implicação crucial da equação (2^v) é a de que apenas quando a explicativa adicional tem estatística t em módulo superior à unidade é que sua inclusão repercute na redução do R^2 ajustado – o que pressupõe a consideração do fato de que apenas números com valor absoluto maior do que a unidade têm quadrados maiores do que a unidade. Como geralmente os valores críticos para um teste t com nível de significância pelo menos equivalente a 5% são superiores à unidade em módulo, pode-se concluir que, na maioria das vezes, que a inclusão de uma explicativa não estatísticamente significativa contribui para a redução do R^2 ajustado. Este resultado é mencionado nos livros de Wooldridge e Gujarati, entre outros.