Notas de aula para o curso de Econometria I

Nota 3: propriedades algébricas, coeficiente de determinação e propriedades estatísticas

Thiago Fonseca Morello

fonseca.morello@ufabc.edu.br

sala 301, Bloco Delta, SBC

1 Propriedades algébricas

Há três propriedades algébricas essenciais do estimador de MQO. Tais propriedades dizem respeito à estrutura matemática do estimador, sendo, portanto, sempre válidas, sem que seja necessário assumir qualquer hipótese que as garanta.

(A) A soma dos resíduos da regressão é nula. A primeira condição de primeira ordem a partir da qual se obtém o estimador de MQO é -2E[(Y_i – β_0 – $\beta_1 X_i$)] = 0, ou seja, E[(Y_i – β_0 – $\beta_1 X_i$)] = 0. Aplicando o princípio de analogia, obtém-se $\sum_{i=1}^{N}\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right)=0$, ou, alternativamente, $\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)=\sum_{i=1}^{N}\hat{u}_i=0$;

Há um corolário fundamental da propriedade (A) que é a de que, na média, o modelo de regressão acerta. Para ver isso, basta tomar a última passagem, em que se afirma que $\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)=0$, e, pois, $\sum_{i=1}^{N} y_i=\sum_{i=1}^{N}\hat{y}_i$. Este corolário é importante, pois ele nos diz que, obrigatoriamente, se a regressão superestima alguns valores, ela obrigatoriamente tem de subestimar os demais, pois apenas assim os erros podem se cancelar quando somados.

(B) A covariância amostral entre a variável independente e o resíduo é nula. Esta propriedade também decorre das condições de primeira ordem, mas, neste caso, da segunda delas. Basta aplicar o princípio da analogia à equação (2) acima. Esta é tal que $\sum_{i=1}^{N} x_i\left(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i\right)=0$. O que é igual a $\sum_{i=1}^{N} x_i\hat{u}_i=0$.

(C) O ponto do plano cartesiano que corresponde aos valores médios para Y e X, ($\overline{x}$, $\overline{y}$) é parte da reta de regressão. A demonstração é simples. No primeiro passo, nota-se que, como a reta contém todos os valores de X ela também contém $\overline{X}$. No segundo passo, basta verificar que, na função da reta de regressão, o valor que corresponde a $\overline{X}$ é $\overline{Y}$.

$$\hat{\beta}_0+\hat{\beta}_1\overline{X}=N^{-1}\sum_{i=1}^{N}\left(\hat{\beta}_0+\hat{\beta}_1 x_i\right)=\overline{\hat{y}}=(1)\,N^{-1}\sum_{i=1}^{N} y_i=\overline{Y}$$

A passagem {1} se devendo à propriedade 1.

2 Coeficiente de determinação e ANOVA

Uma razão pela qual os resíduos podem assumir valores consideráveis repousa no fato de que a variável explicativa X não explica completamente a variação de Y na amostra.

Existe, porém, uma diferença entre explicar uma proporção muito baixa da variação de Y, praticamente desprezível, e uma proporção relevante, mesmo que inferior a 100%. Daí porque é relevante saber qual é, exatamente, a proporção de Y explicada pelo modelo estimado.

O coeficiente de determinação, ou r^2, é uma medida para a qualidade do ajuste do modelo estimado aos dados. Ele equivale à razão entre a variação amostral de Y explicada pelo modelo (numerador) e a variação total de Y na amostra (denominador).

Como medida para a variação explicada pelo modelo, é tomada a soma dos quadrados da diferença entre o valor de Y previsto pelo modelo e a média (amostral) de Y. O que é denominado por soma dos quadrados explicada (SQE). E isso pois, sem recorrer à regressão linear, a maneira mais simples de prever o valor de Y para cada observação é tomando a média de Y. Esta previsão "primitiva" é uma base a partir da qual o conteúdo informacional trazido pela regressão linear tem de ser julgado: se tal técnica não explica uma proporção da variação de Y consideravelmente superior à explicada pela média, não vale a pena recorrer a ela.

De modo coerente, a variação total a ser explicada é medida pela soma dos desvios de Y em torno de sua média, ou soma dos quadrados total (SQT), uma medida quase equivalente à variância de Y.

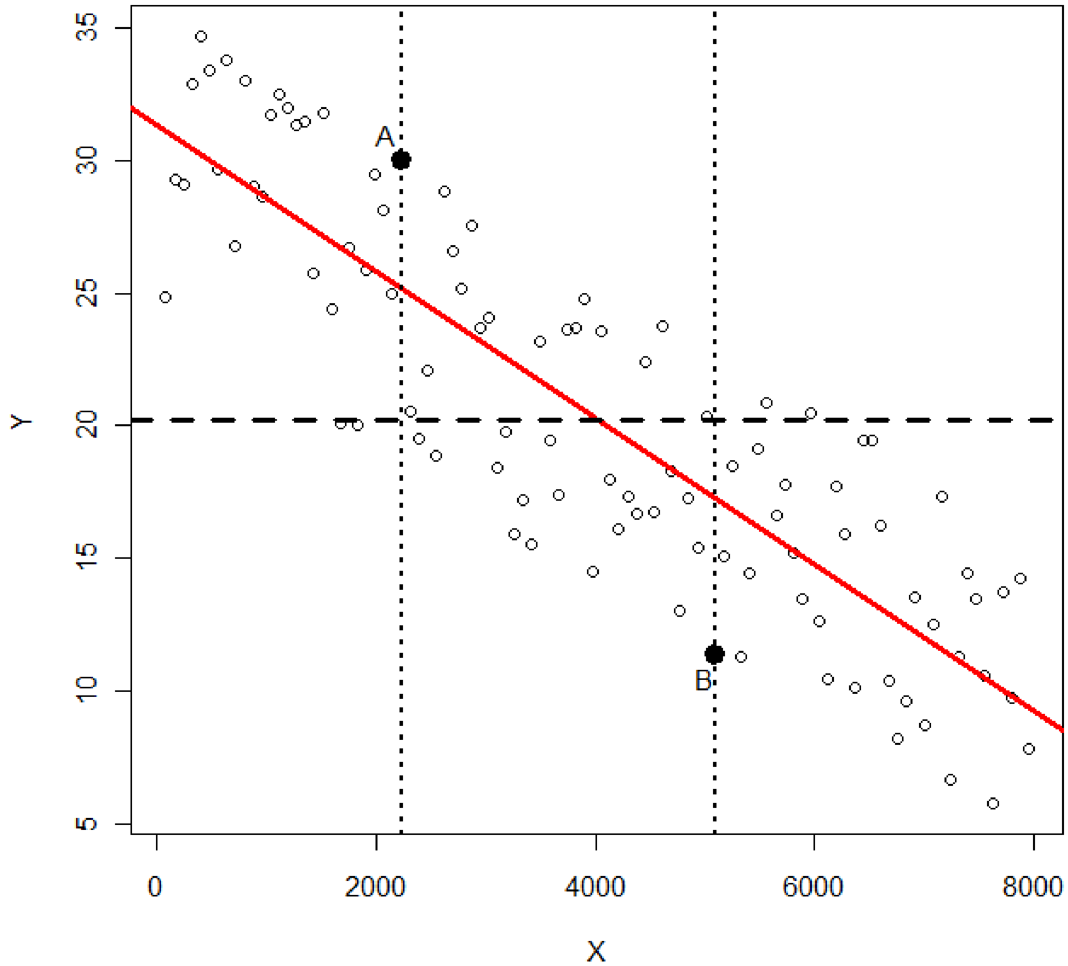
Formalmente, tem-se:

$$r^2 = \frac{SQE}{SQT} = \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{N} \hat{u}_i^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

Em que $\sum_{i=1}^{N} \hat{u}_i^2$ é a soma dos quadrados dos resíduos (SQR), medida para a proporção da variação que permanece não explicada.

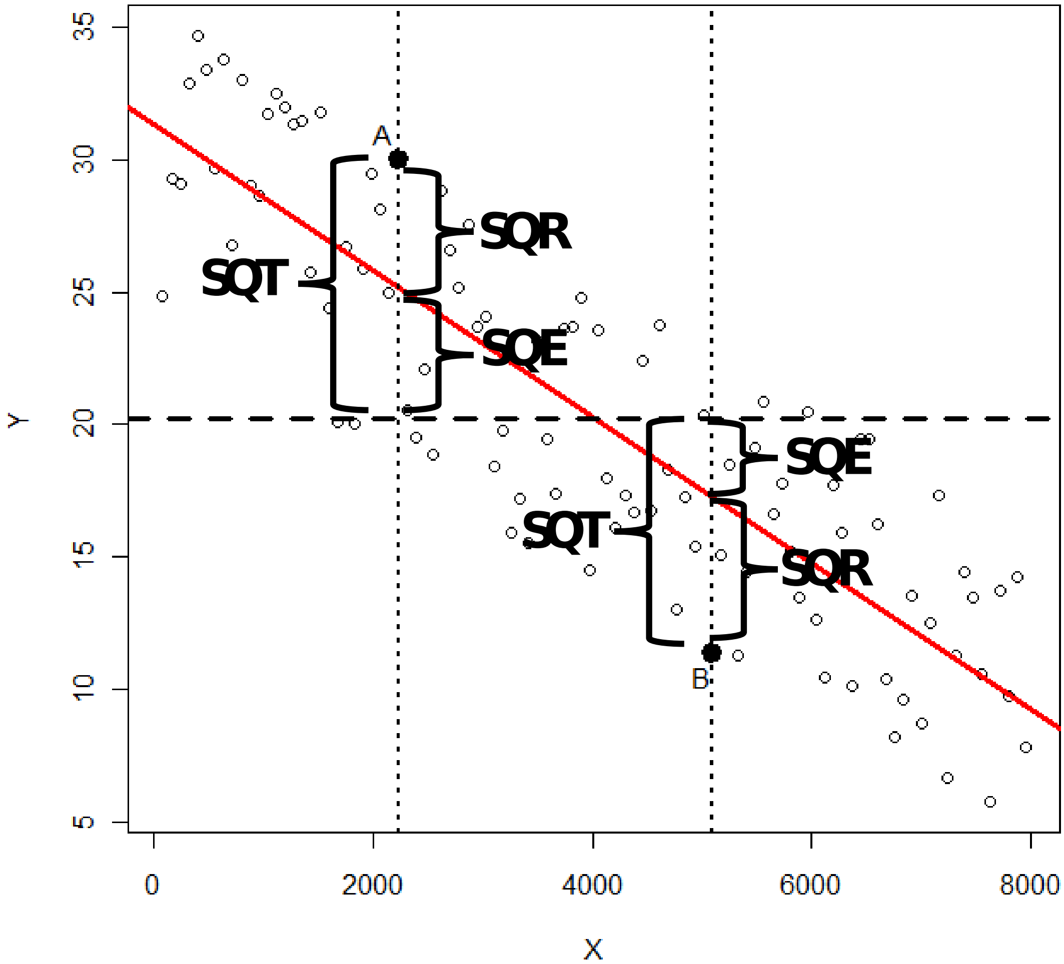
Uma interpretação visual do coeficiente de determinação é provida pela figura abaixo.

Figura 1 Visualizando o coeficiente de determinação: reta de regressão amostral para X e Y (linha vermelha), média de Y (linha pontilha horizontal) e duas observações (pontos A e B)



A distância vertical entre o ponto A e a linha de regressão, em vermelho, corresponde à o que a SQR capta, i.e., à porção da variação de Y não explicada pelo modelo. Já a distância vertical entre a linha vermelha e a média de Y, indicada pela linha pontilhada, corresponde à porção explicada da variação, medida pela SQE. É nestas duas parcelas que se desdobra o desvio em Y em relação à sua média, incorporada à SQT, equivalente à distância entre a coordenada vertical do ponto A e a média de Y. O gráfico abaixo deixa mais claro esta repartição da variação total.

Figura 2 Visualizando a repartição da SQT em SQR e SQE



Estimado o modelo a partir dos dados disponíveis, as somas dos quadrados podem ser dispostas em uma tabela de Análise de Variância (ANOVA, na sigla em inglês), cujo formato geral, para a regressão simples, é apresentado na tabela 1 abaixo.

Tabela 1 Tabela ANOVA

Fonte de variação	**Soma dos quadrados**	**Graus de liberdade**	**Soma dos quadrados média**
Devido à regressão	SQE	1	SQEM =SQE/1
Devido aos resíduos	SQR	N – 2	SQRM =SQR/(N-2)
Total	SQT	N – 1	SQTM= SQT/(N-1)

A terceira coluna compreende as contagens de graus de liberdade, isto é, de partículas informacionais contidas na amostra cujo valor não é fixado pelas estatísticas. Uma explicação mais detalhada pode ser encontrada na nota de aula suplementar 1. Por hora, basta compreender a contagem para cada uma das linhas da tabela. O cálculo da SQT depende do cálculo prévio da média de Y, o que elimina uma partícula de informação livre. Sobram, portanto, N – 1 partículas livres e este é o conteúdo informacional com base no qual a SQT é calculada. O cálculo da SQR, por sua vez, pressupõe a obtenção de duas estimativas pontuais, para o intercepto e para o coeficiente, duas estatísticas, de modo que restam, N – 2 partículas informacionais livres para, com base nelas, obter a SQR. A contagem para a SQE é feito de maneira diferente. Leva-se em conta o fato de que SQE = SQT – SQR, ou seja, a SQE pode ser calculada diretamente a partir desta diferença. Os graus de liberdade associados correspondem, analogamente, à diferença dos graus de liberdade correspondentes à SQT e à SQR, i.e., N – 1 – (N – 2) = 1.

3 Propriedades estatísticas

Há três propriedades estatísticas cuja verificação garante confiabilidade a um estimador, quais sejam, ausência de viés, eficiência e consistência. É crucial saber se os estimadores de MQO possuem essas propriedades. Por hora, apenas a primeira propriedade será estudada, deixando-se a consideração completa das duas outras para o tópico de regressão múltipla. Como as últimas se referem às variâncias dos estimadores, esta seção se restringirá a calcular estas estatísticas, dando início a um tópico que será concluído apenas mais a frente no curso.

3.1 Ausência de viés: comentários preliminares

Será apresentada uma entre as múltiplas demonstrações possíveis. Ela estabelece, em primeiro lugar, a ausência de viés para o estimador para o coeficiente e depois para o estimador do intercepto.

3.2 Ausência de viés do estimador para o coeficiente ($\hat{\beta}_1$)

O objetivo aqui é demonstrar que o estimador para o coeficiente é não viesado, ou seja:

$$B\left[\hat{\beta}_1 \middle| X\right] = E\left[\hat{\beta}_1 \middle| X\right] - \beta_1 = 0$$

O termo à esquerda da equação é o viés do estimador, i.e., a diferença entre seu valor esperado e o valor populacional do parâmetro-alvo.

O objetivo da análise econométrica não é explicar a distribuição completa de Y, mas sim sua distribuição condicional à variável independente, X. Daí porque toma-se, para definir o viés, a expectativa condicional do estimador em relação a X e não a expectativa incondicional. Este é um expediente que será adotado em toda a análise de propriedades estatísticas.

O primeiro passo para estabelecer a propriedade de ausência de viés para o estimador do coeficiente consiste em simplificar a fórmula de tal estimador.

3.2.1 Passo 1, simplificação de $\hat{\beta}_1$

Segundo a definição da FRP, $y_i = \beta_0 + \beta_1 x_i + u_i$. Do que decorre que:

$$\bar{y} = N^{-1}\sum_{i=1}^{N} y_i = N^{-1}\sum_{i=1}^{N}\left(\beta_0 + \beta_1 x_i + u_i\right) = N^{-1}\sum_{i=1}^{N}\beta_0 + N^{-1}\beta_1\sum_{i=1}^{N} x_i + N^{-1}\sum_{i=1}^{N} u_i \rightarrow \bar{y} = \beta_0 + \beta_1\bar{x} + \bar{u}.$$

Consequentemente, $y_i - \bar{y} = \beta_0 + \beta_1 x_i + u_i - \left(\beta_0 + \beta_1\bar{x} + \bar{u}\right) = \beta_1\left(x_i - \bar{x}\right) + u_i - \bar{u}$

Finalmente, portanto: $y_i - \bar{y} = \beta_1\left(x_i - \bar{x}\right) + \tilde{u}_i (1)$, em que $\tilde{u}_i = u_i - \bar{u}$.

A fórmula do estimador do coeficiente é:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} (2)$$

Se (1) for incorporado em (2), chega-se a:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}\left[\beta_1\left(x_i - \bar{x}\right) + \tilde{u}_i\right]\left(x_i - \bar{x}\right)}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} = \frac{\sum_{i=1}^{N}\left[\beta_1\left(x_i - \bar{x}\right)^2 + \tilde{u}_i\left(x_i - \bar{x}\right)\right]}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} = \frac{\sum_{i=1}^{N}\beta_1\left(x_i - \bar{x}\right)^2}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} + \frac{\sum_{i=1}^{N}\tilde{u}_i\left(x_i - \bar{x}\right)}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} = \beta_1\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2} + \frac{\sum_{i=1}^{N}}{\sum_{i=1}^{N}}$$

Em suma:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N}\tilde{u}_i\left(x_i - \bar{x}\right)}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}$$

E com isso se encerra a primeira parte da demonstração de ausência de viés.

3.2.2 Passo 2, aplicando a definição de viés de um estimador

É necessário retomar a definição do viés do estimador, aplicando a ela a expressão (ii) obtida como resultado da subseção anterior.

$$B\left[\hat{\beta}_1 \middle| X\right] = E\left[\beta_1 + \frac{\sum_{i=1}^{N}\tilde{u}_i\left(x_i - \bar{x}\right)}{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}\middle| X\right] - \beta_1 = ¿$$

$$E\left[\beta_1 \middle| X\right] + E\left[\frac{\sum_{i=1}^{N} \tilde{u}_i\left(x_i - \overline{x}\right)}{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2} \middle| X\right] - \beta_1 = \text{¿}$$

$$\beta_1 + \frac{\sum_{i=1}^{N} E\left[\tilde{u}_i\left(x_i - \overline{x}\right) \vee X\right]}{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2} - \beta_1 \rightarrow B\left[\hat{\beta}_1 \middle| X\right] = \frac{\sum_{i=1}^{N} E\left[\tilde{u}_i\left(x_i - \overline{x}\right) \vee X\right]}{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2}$$

As últimas duas passagens fazem uso do fato de que a expectativa de uma soma é a soma das expectativas. E também de que a expectativa está condicionada a X. Empregando este segundo fato mais uma vez, tem-se:

$$B\left[\hat{\beta}_1 \middle| X\right] = \frac{\sum_{i=1}^{N} E\left[\tilde{u}_i \vee X\right]\left(x_i - \overline{x}\right)}{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2} = \frac{\sum_{i=1}^{N} E\left[u_i - \overline{u} \vee X\right]\left(x_i - \overline{x}\right)}{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2} (3)$$

Esta expressão é crucial pois ela estabelece a importância de uma das principais hipóteses assumidas pelo modelo de regressão linear. Trata-se da hipótese de expectativa condicional zero para o termo de perturbação[1], i.e., $E\left[u_i \middle| X\right] = 0$. É importante ressaltar que esta nulidade está sendo assumida e não demonstrada. De fato, esta e outras hipóteses que aparecerão mais a frente são um componente tão importante do modelo de regressão linear como o método de estimação e a estratégia de inferência inerente ao princípio da analogia. Este assunto será retomado em regressão múltipla.

De qualquer maneira, assumindo $E\left[u_i \middle| X\right] = 0$, chega-se ao resultado que se queria demonstrar, uma vez que:

$$E\left[u_i - \overline{u} \middle| X\right] = E\left[u_i \middle| X\right] - E\left[\overline{u} \middle| X\right] = E\left[u_i \middle| X\right] - E\left[N^{-1}\sum_{i=1}^{N} u_i \middle| X\right] = E\left[u_i \middle| X\right] - N^{-1}\sum_{i=1}^{N} E\left[u_i \middle| X\right] \rightarrow$$

$$E\left[u_i - \overline{u} \middle| X\right] = E\left[u_i \middle| X\right] - N^{-1}\sum_{i=1}^{N} E\left[u_i \middle| X\right]$$

Com $E\left[u_i \middle| X\right] = 0$, tem-se, pois, que $E\left[u_i - \overline{u} \middle| X\right] = 0 (4)$

Levando o resultado (4) em (3), chega-se a:

$$B\left[\hat{\beta}_1 \middle| X\right] = \frac{\sum_{i=1}^{N} 0\left(x_i - \overline{x}\right)}{\sum_{i=1}^{N}\left(x_i - \overline{x}\right)^2} = 0$$

[1] Wooldridge 2°ed em inglês, hipótese MRL 3.

É preciso assinalar que uma das implicações fundamentais da hipótese de que $E[u_i|X]=0$ é a de covariância nula entre X, a variável explicativa, e o termo de perturbação, u. Para verificar isso, é necessário percorrer os passos que seguem. De início, seja retomada a definição de covariância; $cov(x_i,u_i) = E[(x_i - E[x_i])(u_i - E[u_i])] = E[x_iu_i] - E[x_i]E[u_i]$. O segundo passo consiste na aplicação da lei das expectativas iteradas, a qual é oriunda da teoria estatística. Segundo ela, $E[W] = E_Z[E[W|Z]]$, sendo W e Z duas variáveis aleatórias quaisquer. Tomando $W = x_iu_i$ e $Z = x_i$, tem-se, pois, que $E[x_iu_i] = E[E[x_iu_i|x_i]]$ e $E[u_i] = E[E[u_i|X]]$. E, desta maneira, $cov(x_i,u_i) = E[E[x_iu_i|x_i]]$, pois $E[u_i|X] = 0 \rightarrow E[E[u_i|X]] = 0$. Consequentemente, $cov(x_i,u_i) = E[x_iE[u_i|x_i]]$, com a última passagem levando em conta o fato de que se está condicionando em X. Como, pois, $cov(x_i,u_i) = E[x_iE[u_i|x_i]]$, a hipótese de que $E[u_i|x_i] = 0$ resulta em $cov(x_i,u_i) = E[x_i0] = 0$.

3.3 Ausência de viés do estimador para o intercepto ($\hat{\beta}_0$)

Esta subseção objetiva demonstrar que:

$$B[\hat{\beta}_0|X]=E[\hat{\beta}_0|X]-\beta_0=0$$

A fórmula do estimador para o intercepto é:

$$\hat{\beta}_0=\overline{y}-\hat{\beta}_1\overline{x}$$

O que é equivalente a:

$$\hat{\beta}_0=N^{-1}\sum_{i=1}^{N}y_i-\hat{\beta}_1\left(N^{-1}\sum_{i=1}^{N}x_i\right)=N^{-1}\left(\sum_{i=1}^{N}y_i-\hat{\beta}_1\sum_{i=1}^{N}x_i\right)$$

Aplicando a definição da FRP para substituir y_i na expressão acima tem-se:

$$\hat{\beta}_0=N^{-1}\left(\sum_{i=1}^{N}\beta_0+\beta_1x_i+u_i-\hat{\beta}_1\sum_{i=1}^{N}x_i\right)\leftrightarrow$$

$$\hat{\beta}_0=N^{-1}\left(N\beta_0+\beta_1\sum_{i=1}^{N}x_i+\sum_{i=1}^{N}u_i-\hat{\beta}_1\sum_{i=1}^{N}x_i\right)$$

$$\hat{\beta}_0=\beta_0+N^{-1}\left[(\beta_1-\hat{\beta}_1)\sum_{i=1}^{N}x_i+\sum_{i=1}^{N}u_i\right](iii)$$

Retomando a definição de viés para o estimador do intercepto e nela aplicando a expressão (iii):

$$B[\hat{\beta}_0|X]=E\left\{\beta_0+N^{-1}\left[(\beta_1-\hat{\beta}_1)\sum_{i=1}^{N}x_i+\sum_{i=1}^{N}u_i\right]\Big|X\right\}-\beta_0\leftrightarrow$$

$$B\left[\hat{\beta}_0 \middle| X\right]=N^{-1}E\left\{\left[\left(\beta_1-\hat{\beta}_1\right)\sum_{i=1}^{N}x_i+\sum_{i=1}^{N}u_i\right]\middle| X\right\}\leftrightarrow$$

$$B\left[\hat{\beta}_0 \middle| X\right]=N^{-1}\left(E\left[\beta_1-\hat{\beta}_1 \middle| X\right]\sum_{i=1}^{N}x_i+\sum_{i=1}^{N}E\left[u_i \middle| X\right]\right)(iv)$$

Para avançar, é preciso assumir a hipótese $E\left[u_i \middle| X\right]=0$. Essa garante que o segundo termo dentro dos parênteses é nulo. Ela também garante que o primeiro termo é nulo, pois, como visto antes, implica em ausência de viés para o estimador do coeficiente, i.e., $E\left[\hat{\beta}_1 \middle| X\right]-\beta_1=0$, o que é equivalente a $E\left[\beta_1-\hat{\beta}_1 \middle| X\right]=0$.

A introdução da hipótese de que $E\left[u_i \middle| X\right]=0$ reduz a expressão (iv) a zero:

$$B\left[\hat{\beta}_0 \middle| X\right]=N^{-1}\left(0\sum_{i=1}^{N}x_i+\sum_{i=1}^{N}0\right)=0$$

Fica demonstrada, pois, a ausência de viés para o intercepto. Mais uma vez é preciso ressaltar que este resultado depende da hipótese de nulidade para a expectativa condicional do termo de perturbação, i.e., $E\left[u_i \middle| X\right]=0$.

3.4 A importância da hipótese $E\left[u_i \middle| X\right]=0$

As demonstrações apresentadas na seção anterior deixam claro que caso não seja válida a hipótese de que $E\left[u_i \middle| X\right]=0$, ambos os estimadores dos parâmetros da FRP, $\hat{\beta}_0$ e $\hat{\beta}_1$, serão viesados. Isso pois é apenas com base nesta hipótese que o vieses dos estimadores se tornam nulos.

Pode-se demonstrar que $E\left[u_i \middle| X\right]=0$ é equivalente a cov(x_i,u_i) = 0; conforme já discutido, pois, é também possível demonstrar que quando o termo de perturbação, u_i, e a variável explicativa, X, têm covariância não-nula entre si (cov(x_i,u_i) ≠ 0), então, obrigatoriamente, $E\left[u_i \middle| X\right]\neq 0$. A existência, portanto, de correlação entre os fatores que explicam Y e são omitidos da FRP, sendo captados pelo termo de perturbação, u, e X, que é o único fator explicativo incorporado na FRP, faz com que os estimadores de MQO sejam viesados.

3.5 Variâncias dos estimadores de MQO

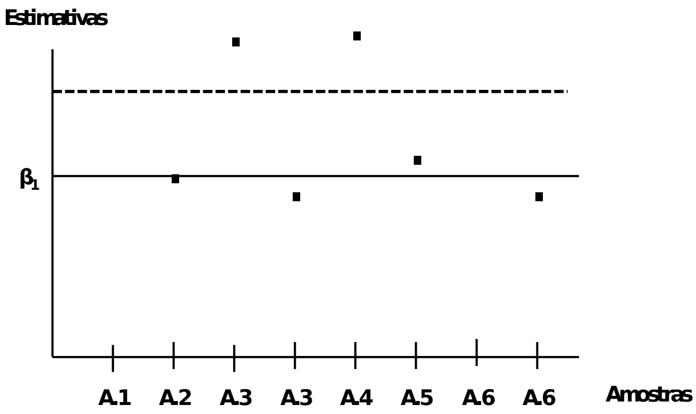
3.5.1 Variâncias populacionais

De início cabe refletir em torno do sentido do termo “variância dos estimadores”. Os estimadores geram estimativas pontuais, i.e., valores, como função da amostra. É de se esperar, pois, que tais valores sejam distintos para amostras distintas, conforme o gráfico abaixo indica. Cabe retomar o experimento mental empregado para introduzir as propriedades de ausência de viés e eficiência[2]: se for possível obter todas as amostras possíveis da população, pode-se construir a função de distribuição de probabilidades dos

[2] Ver nota de aula 1.

estimadores. E isso pois, as estimativas pontuais, assim como o valor de uma variável aleatória ordinária, também variam, de modo que é possível resumir esta variação com as estatísticas padrão, tais como média, variância, etc.

Figura 1 Valor do estimador do coeficiente em diversas amostras*



*β_1 é o valor populacional do coeficiente.

3.5.2 Variância do estimador do coeficiente[3]

A variância condicional do estimador $\hat{\beta}_1$ pode ser escrita da seguinte maneira:

$$V\left[\hat{\beta}_1 \middle| X\right] = E\left[\left(\hat{\beta}_1 - E[\hat{\beta}_1 \vee X]\right)^2 \middle| X\right]$$

Utilizando o resultado de ausência de viés:

$$V\left[\hat{\beta}_1 \middle| X\right] = E\left[\left(\hat{\beta}_1 - \beta_1\right)^2 \middle| X\right]$$

Na nota de aula 5 foi obtida uma expressão simplificada para o estimador, a expressão abaixo:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{N} u_i \left(x_i - \overline{x}\right)}{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2}$$

Esta expressão pode ser escrita como:

$$\hat{\beta}_1 - \beta_1 = \frac{\sum_{i=1}^{N} u_i \left(x_i - \overline{x}\right)}{\sum_{i=1}^{N} \left(x_i - \overline{x}\right)^2} = \sum_{i=1}^{N} u_i w_i (X)$$

[3] Esta demonstração está de acordo com a desenvolvida na página 96 do livro de Badi Baltagi, Econometrics, Springer (livro disponível para a comunidade UFABC em http://link.springer.com/.)

Em que $w_i(X)=\frac{\sum_{i=1}^{N} x_i-\overline{x}}{\sum_{i=1}^{N}(x_i-\overline{x})^2}$, uma função apenas dos dados para a variável explicativa, representados por “X”.

Introduzindo este fato na expressão da variância:

$$V\left[\hat{\beta}_1 \middle| X\right]=E\left[\left(\sum_{i=1}^{N} u_i w_i(X)\right)^2 \middle| X\right](i)$$

A última passagem leva em conta que se está tomando a expectativa condicional em relação a X. Definindo $a_i \equiv u_i(x_i-\overline{x})$, a expressão dentro da expectativa é equivalente à forma geral abaixo.

$$\left(\sum_{i=1}^{N} a_i\right)^2=(a_1+a_2+\ldots+a_N)(a_1+a_2+\ldots+a_N)=¿$$

$$a_1(a_1+a_2+\ldots+a_N)+a_2(a_1+a_2+\ldots+a_N)+\ldots+a_N(a_1+a_2+\ldots+a_N)=¿$$

$$a_1^2+a_1(a_2+a_3+\ldots+a_N)+a_2^2+a_2(a_1+a_3+\ldots+a_N)+\ldots+a_N^2+a_N(a_1+a_2+\ldots+a_{N-1})=\sum_{i=1}^{N} a_i^2+\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{i=1}^{N} a_i a_j$$

Desta maneira, pois:

$$\left(\sum_{i=1}^{N} u_i w_i(X)\right)^2=\sum_{i=1}^{N} u_i^2 w_i(X)^2+\sum_{j\neq 1}^{N}\sum_{i=1}^{N} u_i w_i(X) u_j w_j(X)(ii)$$

Levando a expressão (ii) à (i):

$$V\left[\hat{\beta}_1 \middle| X\right]=E\left[\sum_{i=1}^{N} u_i^2 w_i(X)^2+\sum_{j\neq 1}^{N}\sum_{i=1}^{N} u_i w_i(X) u_j w_j(X) \middle| X\right](i)$$

Utilizando o fato de que a expectativa das somas é a soma das expectativas:

$$V\left[\hat{\beta}_1 \middle| X\right]=\sum_{i=1}^{N} E\left[u_i^2 \vee X\right] w_i(X)^2+\sum_{j\neq 1}^{N}\sum_{i=1}^{N} E[u_i u_j \vee X] w_i(X) w_j(X)(ii)$$

Nesta passagem, mais uma vez leva-se em conta que as expectativas estão condicionadas ao valor assumido por X e, portanto, $w_i(X)$, $w_i(X)^2$ e $w_j(X)$ são constantes para a expectativa.

O avanço requer a adoção de duas hipóteses acerca do comportamento do termo de perturbação.

Em primeiro lugar, assume-se que o valor da variância é equivalente para todas as observações, não variando entre elas, i.e.,$V\left[u_i \vee x\right]=\sigma^2$, i=1,...,N. Esta hipótese, de variância equivalente para todas as observações, é denominada por hipótese de "homocedasticidade dos resíduos".

Tal hipótese pode ser escrita de uma maneira que seja mais útil para o propósito de simplificação da variância de $\hat{\beta}_1$. Basta considerar que $V\left[u_i \vee x\right]=E\left(u_i-E[u_i \vee x]\right)^2 |x]=E\left[u_i^2 \vee x\right]-E\left[u_i \vee x\right]^2$. Sob a validade da hipótese de que $E\left[u_i | x\right]=0$, i=1,...,N, tem-se que $E\left[u_i \vee x\right]^2$ = 0 e, portanto, $V\left[u_i \vee x\right]=E\left[u_i^2 \vee x\right]$. A hipótese de homocedasticidade é, pois, equivalente a $E\left[u_i^2 \vee x\right]=\sigma^2$, i=1,...,N.

A segunda hipótese a ser feita é a de que não há correlação entre os valores que o termo de perturbação assume para observações distintas, ou seja, cov(u_i,u_j) = 0, i,j=1,...,N, hipótese esta referida como "ausência de autocorrelação". Ela também pode ser escrita de maneira mais útil, levando em conta que cov(u_i,u_j) = E[$u_i u_j$|X] – E[u_i|X]E[u_j|X], e, mais uma vez, assumindo que $E\left[u_i | x\right]=0$, i=1,...,N, tem-se cov(u_i,u_j) = E[$u_i u_j$|X] = 0, i,j=1,...,N.

Aplicando apenas a hipótese de homocedasticidade à expressão (ii), tem-se:

$$V\left[\hat{\beta}_1 | X\right]=\sum_{i=1}^{N} \sigma^2 w_i(X)^2+\sum_{j\neq 1}^{N}\sum_{i=1}^{N} E[u_i u_j \vee x] w_i(X) w_j(X)$$

Agora, aplicando também a hipótese de ausência de autocorrelação, a expressão (ii) se reduz a:

$$V\left[\hat{\beta}_1 | X\right]=\sum_{i=1}^{N} \sigma^2 w_i(X)^2$$

Reintroduzindo a definição de w_i(X):

$$V\left[\hat{\beta}_1 | X\right]=\sum_{i=1}^{N} \sigma^2 \left[\frac{\left(x_i-\bar{x}\right)}{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}\right]^2=\sum_{i=1}^{N} \sigma^2 \frac{\left(x_i-\bar{x}\right)^2}{\left[\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2\right]^2}=\frac{\sigma^2}{\left[\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2\right]^2}\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2=\frac{\sigma^2}{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2} \leftrightarrow$$

(a penúltima passagem fez uso dos fatos de que σ^2 e $\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2$ são constantes para o somatório)

Finalmente, pode-se concluir que:

$$V\left[\hat{\beta}_1 | X\right]=\frac{\sigma^2}{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}$$

A derivação da variância do estimador para o intercepto será omitida. Sua fórmula é tal como segue.

$$V\left[\hat{\beta}_0 \middle| X\right]=\frac{\sigma^2 N^{-1}\sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}$$

A precisão de um estudo econométrico depende crucialmente dos valores das variâncias dos estimadores, derivadas nesta seção. De fato, quanto menor estas variâncias, mais precisas e, pois, confiáveis são as estimativas pontuais obtidas a partir da amostra disponível.

A razão para isso está em que quanto maior a variância do estimador, mais o valor da estimativa pontual varia com a informação disponível, i.e., com a amostra. Menor é, portanto, a generalidade do resultado obtido, i.e., da estimativa pontual. O que é um problema pois a análise econométrica procura chegar a afirmações que digam respeito a relação entre duas (ou mais) variáveis tal como ela se manifesta para a população, um âmbito geral de análise.

De acordo com a fórmula de $V\left[\hat{\beta}_1 \middle| X\right]$, a variância do estimador do coeficiente é função negativa da variabilidade de X, medida esta por $\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2$, a soma dos quadrados dos desvios de X de sua média. Caso a variável X exiba variabilidade desprezível, a expressão $\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2$ será próxima de zero e, portanto, a variância do estimador será infinita, ilimitadamente grande[4]. Neste caso extremo, mesmo sendo o estimador não viesado, a estimativa pontual não pode ser tomada como uma estimativa confiável para o valor populacional do coeficiente.

É crucial, portanto, que a variável independente tenha variabilidade considerável na amostra.

3.5.3 Estimando a variância dos estimadores

A variância dos estimadores não pode ser diretamente calculada a partir dos dados amostrais uma vez que é função da variância do termo de perturbação e este nunca é observado. Consequentemente, sua variância, $E\left[u_i^2 \vee x\right]=\sigma^2$, não pode ser calculada.

De fato, as expressões obtidas na subseção anterior correspondem aos valores populacionais das variâncias dos estimadores as quais, pois, geralmente não são conhecidas. Porém, para realizar procedimentos de inferência tais como a estimação por intervalo e teste de hipóteses, é preciso conhecer a variância dos parâmetros.

[4] Basta relembrar do que foi estudado acerca do conceito matemático de limite. A divisão por um número que tende a zero é arbitrariamente grande, tendendo ao infinito.

Uma saída está em tomar os resíduos ($\hat{u}_i$,i=1,...,N) como estimativas para as perturbações (u_i,i=1,...,N). Desta maneira, recorrendo mais uma vez ao princípio da analogia, a contrapartida amostral da variância dos erros é:

$$\tilde{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{u}_i^2$$

Este estimador, contudo, é viesado, pois N não é o número de graus de liberdade[5] que sobram após calcular os resíduos. Como os resíduos são obtidos a partir de $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$, é preciso, pois, estimar os dois parâmetros para obtê-los, o que consome dois graus de liberdade da amostra, restando N – 2 graus de liberdade. O estimador correto e não viesado para a variância do termo de perturbação é:

$$\hat{\sigma}^2 = \frac{1}{N-2} \sum_{i=1}^{N} \hat{u}_i^2$$

Substituindo, na fórmula das variâncias dos estimadores, a variância populacional dos erros, σ^2, pela expressão acima, obtém-se estimadores não viesados para as variâncias populacionais dos estimadores, tal como segue.

$$\widehat{V\left[\hat{\beta}_1 \middle| X\right]} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2}$$

$$\widehat{V\left[\hat{\beta}_0 \middle| X\right]} = \frac{\hat{\sigma}^2 N^{-1} \sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N} \left(x_i - \bar{x}\right)^2}$$

Cabe assinalar a introdução do símbolo “^” acima das variâncias com o objetivo de indicar que se trata de estimadores para a as variâncias, os quais utilizam informação da amostra. De fato, as variâncias derivadas na subseção anterior são calculadas a partir dos valores assumidos pelos estimadores em todas as amostras possíveis da população. Este conjunto de dados dificilmente está disponível. Geralmente tem-se apenas uma amostra. Com base nela, pois, apenas é possível estimar as variâncias “populacionais” dos estimadores, derivadas anteriormente.

Apêndice Decomposição do coeficiente

Há uma relevante relação entre o estimador do coeficiente da FRP e o coeficiente de correlação linear de Pearson. Basta retomar a fórmula de ambos para a amostra para que tal relação fique aparente.

[5] Consultar a nota de aula suplementar 1 no repositório da disciplina no TIDIA-AE.

$$\hat{\beta}_1=\frac{\sum_{i=1}^{N}(y_i-\overline{y})(x_i-\overline{x})}{\sum_{i=1}^{N}(x_i-\overline{x})^2}\,(1)$$

$$\hat{\rho}=\frac{1/N\sum_{i=1}^{N}(y_i-\overline{y})(x_i-\overline{x})}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\overline{y})^2}}\,(2)$$

A fórmula do estimador do coeficiente, (1), pode ser escrita de maneira a explicitar melhor a relação com a fórmula (2):

$$\hat{\beta}_1=\frac{1/N\sum_{i=1}^{N}(y_i-\overline{y})(x_i-\overline{x})}{1/N\sum_{i=1}^{N}(x_i-\overline{x})^2}\,(1')$$

Uma vez que $1/N\sum_{i=1}^{N}(x_i-\overline{x})^2$ e $1/N\sum_{i=1}^{N}(y_i-\overline{y})^2$ são as variâncias amostrais de X e Y, respectivamente, elas podem ser escritas como $\widehat{V(X)}$ e $\widehat{V(Y)}$. De modo que:

$$\hat{\beta}_1=\frac{1/N\sum_{i=1}^{N}(y_i-\overline{y})(x_i-\overline{x})}{\widehat{V(X)}}\,(1')$$

$$\hat{\rho}=\frac{1/N\sum_{i=1}^{N}(y_i-\overline{y})(x_i-\overline{x})}{\sqrt{\widehat{V(X)}}\sqrt{\widehat{V(Y)}}}\,(2)$$

E, portanto:

$$\hat{\rho}=\frac{\hat{\beta}_1\widehat{V(X)}}{\sqrt{\widehat{V(X)}}\sqrt{\widehat{V(Y)}}}\rightarrow\hat{\rho}=\hat{\beta}_1\frac{\sqrt{\widehat{V(X)}}}{\sqrt{\widehat{V(Y)}}}$$

Ou, de maneira equivalente:

$$\hat{\beta}_1=\hat{\rho}\frac{\sqrt{\widehat{V(Y)}}}{\sqrt{\widehat{V(X)}}}$$

Tem-se, pois, que o valor da estimativa pontual para o coeficiente é dado pelo produto do valor do coeficiente de correlação linear de Pearson por uma medida que corresponde ao relativo de dispersão em torno da média entre a variável explicada e variável explicativa. Esta última medida, indica, pois, em que magnitude a variabilidade de Y é superior ou inferior à variabilidade de X na amostra.

Uma vez que o desvio padrão de uma variável aleatória é sempre positivo, o sinal da estimativa pontual para o coeficiente é sempre equivalente ao sinal do coeficiente de correlação.

Pode-se concluir que, quanto maior o valor absoluto da correlação entre X e Y, i.e, quanto mais próximo da unidade estiver o valor absoluto de $\hat{\rho}$, e maior for a variabilidade de Y em relação a X, maior será o valor absoluto de $\hat{\beta}$. Esta conclusão é importante uma vez que o valor absoluto da estimativa pontual corresponde ao valor absoluto do impacto de X sobre Y, tal como estimado com base na amostra.