ACH3657

Métodos Quantitativos para Avaliação de Políticas Públicas

Aula Teórica 04

Propriedades do Método MQO e Forma Funcional

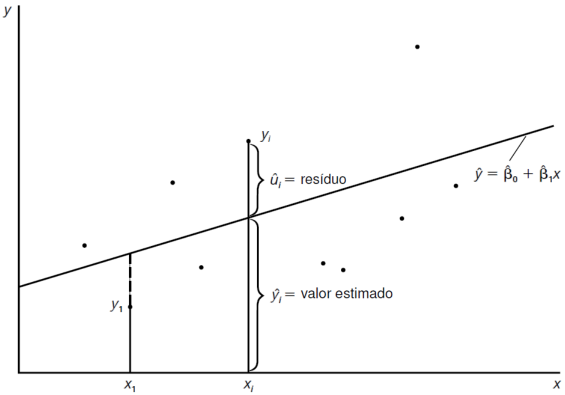
Alexandre Ribeiro Leichsenring
alexandre.leichsenring@usp.br

Mecânica do Método MQO

- Veremos algumas propriedades *algébricas* do método MQO
- São características do método para uma determinada amostra
- Propriedades Algébricas *vs* Propriedades estatísticas (estas têm a ver com a distribuição amostral dos estimadores)
- Compreender as propriedades algébricas nos ajuda a compreender o que ocorre com as estimativas de MQO e estatísticas relacionadas quando os dados são manipulados (por exemplo, mudança de unidade de medida)

Valores Estimados e Resíduos

- Assumimos que $\hat{\beta}_0$ e $\hat{\beta}_1$ foram obtidas de uma dada amostra de dados
- Dados $\hat{\beta}_0$ e $\hat{\beta}_1$ podemos obter o valor estimado $\hat{y}_i$ para cada observação



Exemplo: Salário de Diretores Executivos e Retornos de Ações

nobs	rma	salario	salario.chapeu	uchapeu
1	14.1	1095	1224.058	−129.0581
2	10.9	1001	1164.854	−163.8542
3	23.5	1122	1397.969	−275.9692
4	5.9	578	1072.348	−494.3484
5	13.8	1368	1218.508	149.4923
6	20.0	1145	1333.215	−188.2151
7	16.4	1078	1266.611	−188.6108
8	16.3	1094	1264.761	−170.7606
9	10.5	1237	1157.454	79.54626
10	26.3	833	1449.773	−616.7726
11	25.9	567	1442.372	−875.3721
12	26.8	933	1459.023	−526.0231

Propriedades algébricas das estatísticas de MQO

1. A soma, e portanto a média amostral, dos resíduos de MQO, é zero:
$$\sum_{i=1}^{n} \hat{u}_i = 0$$

2. A covariância amostral entre os regressores e os resíduos de MQO é zero:
$$\sum_{i=1}^{n} x_i \hat{u}_i = 0$$

3. O ponto $(\bar{x}, \bar{y})$ está sempre sobre a reta de regressão.

Exemplo: Salários e educação

- Para os dados do arquivo wage1.sav, temos a seguinte equação estimada:

$$\hat{salario} = -0,90 + 0,54educ$$

- Salário-hora médio da amostra $= 5,90$
- Educação formal média (medida em anos) $= 12,56$
- Inserindo $educ = 12,56$ na equação de regressão, obtemos:

$$\hat{salario} = -0,90 + 0,54(12,56) = 5,9$$

- Outro modo de interpretar uma regressão de MQO:

$$y_i = \hat{y}_i + \hat{u}_i$$

  - Da propriedade (1), a média dos resíduos é zero. Logo,

$$\bar{y} = \bar{\hat{y}}$$

- Além disso, as propriedades (1) e (2) podem ser usadas para provar que:

$$\text{Cov}(\hat{y}_i, \hat{u}_i) = 0$$

  - Ou seja, MQO decompõe cada y_i em duas partes: um valor estimado e um resíduo.
  - Os valores estimados e os resíduos são não correlacionados na amostra

Decompondo a variação em y

Definindo as seguintes quantidades:

- Soma dos Quadrados Total

$$\text{SQT} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- Soma dos Quadrados Explicada

$$\text{SQE} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

- Soma dos Quadrados dos Resíduos

$$\text{SQR} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Obtemos a seguinte igualdade:

$$\text{SQT} = \text{SQE} + \text{SQR}$$

Grau de Ajuste do Modelo

- Ainda não apresentamos uma maneira de mensurar o quão bem a variável explicativa ou independente, x, explica a variável dependente, y
- É útil ter uma medida que resume num número a “qualidade do ajuste” da reta de regressão de MQO aos dados

Coeficiente de determinação, R^2

$$R^2 = \frac{\text{SQE}}{\text{SQT}} = 1 - \frac{\text{SQR}}{\text{SQT}}$$

- R^2 é a razão entre a variação explicada e a variação total
- Ele é interpretado como a fração da variação amostral em y que é explicada por x
- $0 \leq R^2 \leq 1$

- Se todos os pontos dos dados estiverem sobre a mesma reta, MQO fornece um ajuste perfeito aos dados
  - nesse caso, $R^2 = 1$
- Um valor de R^2 próximo de zero indica um ajuste ruim da reta de MQO
  - muito pouco da variação em y_i, é capturado pela variação em $\hat{y}_i$
- Pode ser mostrado que R^2 é igual ao quadrado do coeficiente de correlação amostral entre y_i, e $\hat{y}_i$
  - Daí vem o termo “R-quadrado”.

Exemplo: Salário de diretores executivos e retornos de Ações

- Na regressão de salários de diretores executivos, estimamos a seguinte equação:

$$\hat{salario} = 963,191 + 18,501 rma$$

$$n = 209, R^2 = 0,0132$$

- O R-quadrado indica quanto da variação no salário é, realmente, explicada pelo retorno da ação.
- Não muito! 0 retorno da ação da empresa explica somente 1,3% da variação nos salários dessa amostra de 209 diretores executivos.
- 98,7% das variações salariais desses diretores executivos são deixadas sem explicação
- Esse fato não deve ser surpreendente: muitas outras características, tanto da empresa como do diretor executivo individual, devem influenciar o salário; esses fatores ficam incluídos nos *termos de erro* de uma análise de regressão simples

- Nas ciências sociais não são incomuns R^2 baixos nas equações de regressão
- Discutiremos essa questão, de modo mais geral, sob a análise de regressão múltipla, mas...
- Vale a pena enfatizar agora que um R^2 aparentemente baixo não significa, necessariamente, que uma equação de regressão de MQO é inútil
- Ainda, é possível que a equação $salario = f(rma)$ seja uma boa estimativa da relação *ceteris paribus* entre salário e rma
- Isso não depende diretamente da magnitude do R^2...
- Porém, algumas vezes, a variável explicativa elucida uma parte substancial da variação amostral na variável dependente.

Exemplo: Resultados Eleitorais e Gastos de Campanha

- Na equação de resultados eleitorais, $R^2 = 0,856$.
- A participação dos candidatos nos gastos de campanha explica mais de 85% da variação nos resultados eleitorais nessa amostra.

  $\Rightarrow$ Explicação considerável.

Forma funcional: Incorporação de Não-Linearidades na Regressão Simples

- Relações lineares não são, em geral, suficientes para a diversidade de aplicações reais
- É fácil incorporar muitas não-linearidades na análise de regressão simples
- Trabalhos aplicados nas ciências sociais, com freqüência apresentam equações de regressão com variável dependente na forma logarítmica. Por que isso aparece?

Exemplo: Relação linear entre salário e escolaridade

$$\widehat{salario} = -0,90 + 0,54\, educ$$

- Para cada ano a mais de educação formal há aumento esperado do salário em 54 centavos de dólar por hora
- Pela natureza linear do ajuste, 54 centavos de dólar é o aumento tanto para o primeiro ano de educação quanto para o vigésimo ano
- Seria mais razoável esperar um aumento porcentual!
- O modelo com logaritmo está de acordo com essa lógica

Exemplo: Relação exponencial entre salário e educação

Suponha que o aumento percentual do salário é o mesmo dado um ano a mais na educação

- Um modelo que gera (aproximadamente) um efeito percentual constante é:

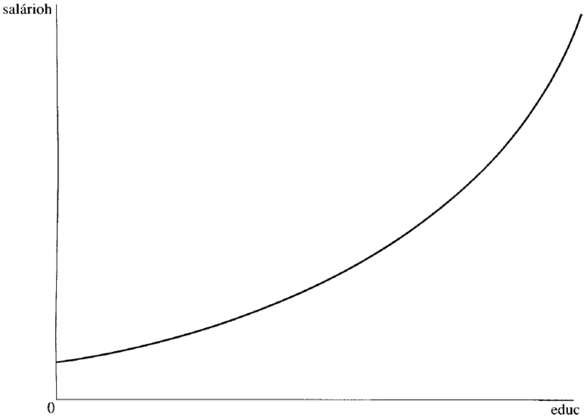
$$\log(salario) = \beta_0 + \beta_1 \; educ + u$$

  em que em que $\log(\cdot)$ é o logaritmo natural.

- Em particular, se $\Delta u = 0$, então:

$$\% \Delta salario \simeq (100 \cdot \beta_1) \Delta educ$$

$$salario = \exp(\beta_0 + \beta_1 \; educ)$$



Uma Equação do Logaritmo dos Salários-Hora

Utilizando os mesmos dados do Exemplo anterior, mas usando $\log(salarioh)$ como a variável dependente, obtemos a seguinte relação:

$$\log(salario) = 0,584 + 0,083\ educ$$

$$n = 526,\ R^2 = 0,186$$

- O coeficiente de $educ$ tem uma interpretação percentual:
  - Para cada ano adicional de educação formal, $salarioh$ aumenta 8,3%
- Uma vez obtida a equação, o $\log$ de $salarioh$ raramente é mencionado
- O importante é estabelecer o efeito de aumento percentual
- O intercepto não tem muito significado
- $educ$ explica cerca de $18,6\%$ da variação em $\log(salarioh)$
- Finalmente, a equação pode não capturar toda a não-linearidade da relação entre salário-hora e escolaridade formal
  - Se houver "efeitos-diploma", o décimo segundo ano de educação – formatura do ensino médio nos Estados Unidos – deve ser muito mais valioso que o décimo primeiro ano