



UNIVERSIDADE FEDERAL DE MINAS GERAIS

**MÍNIMOS QUADRADOS ORDINÁRIOS: UMA APLICAÇÃO NA ANÁLISE DAS
QUESTÕES INSTITUCIONAIS DE MUNICÍPIOS BRASILEIROS**

Flávio Hugo Pangrácio Silva 
flaviopangracio@cedeplar.ufmg.br
Cedeplar - UFMG

Guilherme Gomes Ferreira 
guilhermegf2019@cedeplar.ufmg.br
Cedeplar - UFMG

DOCENTE: Ana Hermeto.

Belo Horizonte - MG

Abril - 2024

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

INTRODUÇÃO	1
O MODELO CLÁSSICO DE REGRESSÃO LINEAR	1
Linearidade do modelo:	2
Posto Completo:	2
Exogeneidade	2
Homocedasticidade e não autocorrelação residual	2
Processo Gerador dos dados para a regressão	3
Normalidade dos erros	3
REGRESSÃO POR MÍNIMOS QUADRADOS	3
REFERÊNCIAS	5

INTRODUÇÃO

O presente trabalho se propõe a explorar de maneira detalhada o método de mínimos quadrados ordinários (MQO), apresentando uma aplicação na análise das questões institucionais presentes nos municípios brasileiros. Este método estatístico é amplamente utilizado na análise econômica, sendo fundamental para compreender as relações entre variáveis e realizar previsões.

A escolha desse enfoque se justifica pela relevância crescente do estudo das instituições no contexto municipal brasileiro, visto que as políticas públicas e a gestão eficiente dessas instituições desempenham um papel fundamental no desenvolvimento socioeconômico local. Nesse sentido, compreender como diferentes variáveis institucionais estão relacionadas entre si e como influenciam indicadores de crescimento e desenvolvimento municipal torna-se uma questão de interesse.

Por meio deste trabalho, pretendemos não apenas apresentar a aplicação prática do modelo de MQO, mas também fornecer uma base sólida de compreensão teórica, destacando os fundamentos matemáticos e estatísticos subjacentes a esse método. Para isso, organizaremos o conteúdo em várias seções, nas quais abordaremos desde os princípios básicos da regressão linear até aspectos mais avançados, passando pela discussão sobre a formulação teórica do modelo de MQO.

Inicialmente, abordaremos os principais conceitos e definições relacionados à regressão linear, discutindo os pressupostos e as limitações desse modelo estatístico. Posteriormente, dedicaremos atenção especial à formulação teórica do modelo de MQO, descrevendo o processo de estimativa dos parâmetros e apresentando as principais propriedades estatísticas dos estimadores obtidos por esse método. Além disso, discutiremos técnicas de diagnóstico e avaliação da qualidade do modelo, destacando a importância da interpretação correta dos resultados obtidos.

Por fim, demonstraremos a aplicação do modelo de MQO na análise das questões institucionais de municípios brasileiros, utilizando dados da REGIC para ilustrar o processo de formulação, estimação e interpretação do modelo. Espera-se que este trabalho contribua para ampliar o entendimento sobre o método de MQO e sua aplicação.

O MODELO CLÁSSICO DE REGRESSÃO LINEAR

A priori, antes de adentrar em detalhes do estimador de MQO, é preciso explicar o modelo clássico de regressão linear, bem como suas hipóteses subjacentes. Nesse sentido, deve se salientar que o modelo clássico de regressão linear admite a forma simples e a forma múltipla. No modelo simples, também conhecido como modelo de regressão bivariada, temos apenas uma variável explicada e uma variável explicativa, além de um intercepto e dos resíduos do modelo.

Um problema fundamental do modelo de regressão simples, no entanto, é a dificuldade de fazer uma análise parcial com apenas uma variável explicativa, ignorando todas outras variáveis que afetam a variável explicada, Y , e são não correlacionadas com a variável independente, X . É nesse sentido que existe o modelo de regressão linear múltipla, o qual permite explicar uma variável através de uma junção de mais variáveis independentes e não correlacionadas uma com a outra. Doravante, este trabalho focará no modelo de regressão linear múltipla, com a justificativa de que os pressupostos são análogos aos pressupostos do modelo simples e que com mais variáveis, o que só é permitido neste modelo, é possível fazer uma análise mais robusta.

Nesta perspectiva, para a definição do modelo clássico de regressão linear, são necessárias algumas hipóteses:

Linearidade do modelo:

A primeira hipótese implica que o modelo deve ser linear nos parâmetros estimados. Disso decorre que as variáveis explicativas podem ser não lineares. Essa hipótese basicamente indica que a relação das variáveis independentes com o parâmetro estimado é linear, ou seja, uma variação marginal nas variáveis independentes resultará em uma variação constante na variável explicada.

Posto Completo:

Essa hipótese é uma condição necessária do MCRL, haja vista que, se não satisfeita, é impossível estimar os parâmetros do modelo. Em termos matriciais, implica que a matriz das variáveis independentes deve ser não singular o que, por sua vez, exige que essas variáveis não sejam combinações lineares perfeitas umas das outras. Também é conhecida como condição de identificação

Exogeneidade

Tal condição garante que a média condicional do erro dadas as variáveis explicativas é igual a zero. Também conhecida como exogeneidade estrita, seu significado é de que as variáveis explicativas não possuem relação com o termo de perturbação. Além disso, é importante ressaltar que, como a média condicional do erro é zero, sua média incondicional também é zero, o que é garantido pela lei das expectativas iteradas. Essa é uma forte implicação que garante que uma estimação pelo MCRL sempre acerta na média. Ademais, o MCRL garante a aleatoriedade dos resíduos, isto é, a média condicional do erro i , dado um erro j qualquer é zero.

Homocedasticidade e não autocorrelação residual

Essa quarta hipótese define que a variância condicional do erro é constante e que a covariância condicional dos erros é zero. A variância constante é conhecida como

homocedasticidade, o que significa que para qualquer ponto da amostra, a variância sempre será a mesma. Quando isso não ocorre, dizemos que a variância é heterocedástica.

Já o fato da covariância condicional dos erros ser igual a zero define a não autocorrelação entre os termos de perturbação. Em termos matriciais, temos que a matriz de erros vezes a sua transposta é igual a matriz identidade vezes a variância dos resíduos. Vale ressaltar que isso não implica que as observações não são autocorrelacionadas.

Processo Gerador dos dados para a regressão

A quinta premissa se refere a não aleatoriedade do vetor de variáveis explicativas, em outras palavras, ele é não estocástico. Isso quer dizer que o vetor de variáveis explicativas é gerado exogenamente. No entanto, usualmente isso é de difícil aplicação, haja vista que o vetor x tende a ser aleatório, tal qual o vetor y . Desse modo, uma forma alternativa é assumir x como um vetor aleatório e tratar da distribuição conjunta de x e y . Desse modo, essa premissa firma que x pode ser fixo ou aleatório.

Normalidade dos erros

Implica que os termos de perturbação são normalmente distribuídos, possuindo média zero e variância constante. Essa premissa é bastante razoável, haja vista que o teorema do limite central garante essa normalidade, pelo menos, assintoticamente. Todavia, essa suposição geralmente não é necessária para obter a maioria dos resultados em uma regressão linear.

Finalizada esta parte, apresentou-se as premissas do MCRL, as quais servem como base para a construção de um modelo econométrico. O objetivo seguinte será descrever métodos de estimação de modelos, dentre eles, o famoso e amplamente utilizado, método de mínimos quadrados ordinários.

REGRESSÃO POR MÍNIMOS QUADRADOS

O método de mínimos quadrados ordinários consiste em minimizar a soma do quadrado dos resíduos, a fim de encontrar os parâmetros do modelo. O primeiro passo é distinguir entre as quantidades populacionais não observadas e os parâmetros amostrais. Em outras palavras, existem os parâmetros verdadeiros e os parâmetros calculados no modelo agem como uma estimativa desses parâmetros populacionais, desde que sejam satisfeitas as condições que tornem o MQO aplicável. Em termos matriciais, haja vista que estamos tratando de uma regressão linear múltipla, podemos escrever o modelo da seguinte forma: (colocar fórmulas)

Essa estimação nada mais é que as condições de primeira ordem do modelo. Nesse sentido, a partir dos valores estimados após encontrar os parâmetros amostrais, existem algumas relações importantes, sobretudo entre o termo de erro e os valores preditos da variável dependente: i) o MQO garante que a média dos resíduos é zero; ii) como não há covariância amostral entre

o termo de erro e as variáveis independentes, não há covariância amostral entre os valores estimados e os resíduos; iii) os pontos médios das variáveis estão sempre sobre a reta de regressão.

Além disso, sob a hipótese de homocedasticidade, tratada anteriormente no MCRL, existe um teorema, conhecido como teorema de Gauss-Markov, o qual garante que os estimadores de mínimos quadrados são os melhores estimadores não viesados da classe dos lineares. Todavia, esse teorema é muito restrito, haja vista as limitações impostas, como homocedasticidade e exogeneidade estrita, haja vista a ausência de viés. Nesse sentido, é mais vantajoso analisar as propriedades assintóticas dos estimadores, que tratam de convergência em probabilidade e flexibilizam mais o modelo estimado.

Conforme Wooldrige (2010), para um estimador ser consistente, são necessárias duas premissas. A primeira implica que a covariância entre o resíduo e o vetor de variáveis explicativas seja igual a zero e essa é uma versão mais fraca da exogeneidade. Por outro lado, a segunda premissa diz que a multiplicação matricial de x e sua transposta tem que ser igual a ordem de x , ou seja, implica independência linear. (Colocar formula - livro wooldrige)

Heiss (2020) Greene (2019)

REFERÊNCIAS

GREENE, W. H. **Econometric Analysis Global Edition**. 8. ed. [s.l.] Pearson-prentice Hall, 2019.

HEISS, F. **Using R for Introductory Econometrics**. 2. ed. [s.l: s.n.].