Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Estatística

Relatório Técnico – RTP 03/2013

UMA ABORDAGEM ALTERNATIVA PARA O ENSINO DO MÉTODO DE MÍNIMOS QUADRADOS NO NÍVEL MÉDIO E INÍCIO DO CURSO SUPERIOR

Edna A. Reis Emílio Suyama Roberto C. Quinino
Dr. Lupércio F. Bessegato

Juiz de Fora
2013

Uma Abordagem Alternativa para o Ensino do Método dos Mínimos Quadrados no Nível Médio e Início do Curso Superior

Roberto C. Quinino
Edna A. Reis
Emílio Suyama
Departamento de Estatística – ICEx – UFMG - Brasil

Lupércio F. Bessegato
Departamento de Estatística – ICE – UFJF - Brasil

1. Introdução

O método dos mínimos quadrados é o procedimento de estimação dos parâmetros de um modelo de regressão por meio da minimização da soma dos quadrados das diferenças entre os valores observados da variável resposta em uma amostra e seus valores preditos pelo modelo. Possui aplicações em áreas como biologia, engenharia, estatística, física matemática, entre outras, principalmente aquelas que objetivam relacionar uma variável dependente (Y) em função de variáveis explicativas ($X_1,...,X_k$). O método foi proposto independentemente pelos matemáticos Carl Friedrich Gauss por volta de 1795 e Adrien Marie Legendre em torno de 1805. Constitui-se num conteúdo programático usualmente apresentado no início de um curso superior para alunos que já estudaram o conceito de derivadas parciais com foco em métodos de otimização.

O método de mínimos quadrados não é, em geral, apresentado no ensino médio, pois demandaria o uso de derivada parcial, que é, normalmente, um assunto destinado ao curso superior. Neste sentido, Barbosa&Breitschaft (2006) sugerem que para o ensino médio o método de mínimos quadrados seja substituído pelo ajuste visual de uma reta aos pontos experimentais o que possibilitaria, também, a determinação dos parâmetros desta reta. Concordamos, mas entendemos que podemos melhorar o procedimento, no caso em que Y depende linearmente apenas de uma variável independente e um erro aleatório ε com média zero, ($Y=a+bX+\varepsilon$). Tal modelo permitiria apresentar o método de mínimos quadrados no ensino médio e início do ensino superior sem demandar conteúdo programático mais avançado. Nosso objetivo nessa nota é descrever como isso pode ser feito.

A abordagem aqui descrita foi baseada em Theil (1950) e Birkes&Dodge (1993) e demanda apenas noções de combinação, média ponderada e solução de sistemas de duas equações lineares a duas incógnitas, conteúdo normalmente pertencente ao ensino fundamental e médio. Essencialmente o procedimento resume-se em encontrar todas as retas que passam por, pelo menos dois pontos e, utilizando-se de uma média ponderada de todos interceptos e de todas inclinações, calcular, respectivamente, a estimativa do intercepto e da inclinação da reta final. O objetivo desta metodologia seria que o aluno chegasse ao ensino superior com uma postura mais crítica e sugestiva, inclusive em relação a outras possibilidades além do método de mínimos quadrados.

2. Abordagem Alternativa para Explicar o Método de Mínimos Quadrados

Sem perda de generalidade, demonstraremos a sugestão dessa nota com a utilização de um exemplo. Considere que desejamos obter uma estimativa da reta $Y=a+bX+\varepsilon$, representada por $\hat{Y} = \hat{a} + \hat{b}X$, que melhor se ajuste aos dados descritos e ilustrados na Figura 1.

Figura 1: Gráfico de dispersão de X versus Y.

Em sala de aula, a primeira motivação seria que os alunos descrevessem todas as retas que poderiam ser obtidas por meio da seleção de dois pares ordenados. A Figura 2 ilustra as seis retas possíveis. Por exemplo, $Y=a_1+b_1X$ é relativo aos pares ordenados (4;7) e (5;12). De maneira geral, se temos *n* pontos, o número de retas é igual a $\binom{n}{2}$, salvo casos em que os dois pares ordenados possuem o mesmo valor de X, situação que não deverá ser considerada, uma vez que gerará sistemas sem solução ou com infinitas soluções.

Figura 2: Todas as seis retas considerando conjuntos de dois pares ordenados.

Observe que os valores dos interceptos ($a_1,...,a_6$) e das inclinações ($b_1,...,b_6$) podem ser obtidos simplesmente resolvendo sistemas de duas equações lineares a duas incógnitas. Por exemplo, para o pares ordenados (4;7) e (5;12), temos o seguinte sistema:

$$\begin{cases} 7 = a_1 + 4b_1 \\ 12 = a_1 + 5b_1 \end{cases}$$

Resolvendo-o por comparação, temos que $a_1=-13$ e $b_1=5$. A Tabela 1 mostra todos os interceptos e inclinações para as seis retas.

Tabela 1: Descrição das seis retas da Figura 1.

Pontos (X,Y)	**Reta**	**Intercepto a_i**	**Inclinação b_i**	**Peso**
(4,7) e (5,12)	$Y=a_1+b_1X$	-13	5	$(5-4)^2 = 1$
(1,2) e (4,7)	$Y=a_2+b_2X$	0,33	1,67	$(1-4)^2 = 9$
(2,6) e (5,12)	$Y=a_3+b_3X$	2	2	$(2-5)^2 = 9$
(1,2) e (2,6)	$Y=a_4+b_4X$	-2	4	$(1-2)^2 = 1$
(1,2) e (5,12)	$Y=a_5+b_5X$	-0,5	2,5	$(1-5)^2 = 16$
(2,6) e (4,6)	$Y=a_6+b_6X$	5	0,5	$(2-4)^2 = 4$

Nosso interesse está em saber se alguma das seis retas poderia ser candidata à reta estimada $\hat{Y} = \hat{a} + \hat{b}X$. Cada uma das seis retas é determinística para os dois pares ordenados que a gerou, mas pode ser a "mais próxima", ou não, ao restante dos pontos. Pela Figura 2, podemos mostrar aos alunos que, quanto mais afastados entre si estão os dois valores de X que formam a reta, melhor parece ficar a reta em relação a todos os pontos. Neste sentido, a reta $Y=a_5+b_5X$ se destaca no sentido de parecer estar mais próxima de todos os pontos e candidata a se tornar a reta procurada $\hat{Y} = \hat{a} + \hat{b}X$.

Podemos agora desenvolver um racional mais objetivo. Defina e_i como a diferença entre o valor observado e o valor predito pela reta ($Y_i - \hat{Y}_i,\ i = 1,\dots,n$) e escolhamos a reta que apresentar o menor valor de $\sum |e_i|$. Por exemplo, as figuras 3 e 4 ilustram respectivamente uma comparação entre as retas $Y=a_5+b_5X$ e $Y=a_6+b_6X$. Observe que a reta $Y=a_5+b_5X$ apresenta um valor de $\sum |e_i| = 4$, menor do que a reta $Y=a_6+b_6X$, $\sum |e_i| = 8$, sendo, então, mais apropriada que esta. De fato, a reta $Y=a_5+b_5X$ apresenta o menor $\sum |e_i|$ entre as seis retas sendo, então, uma opção aceitável. Tal reta é conhecida na literatura como aquela obtida pelo *método do mínimo desvio absoluto* (em inglês, *Least Absolute Seviations - LAD*).

Figura 3: Desempenho da reta $Y=a_5+b_5X$.

Figura 4: Desempenho da reta $Y=a_6+b_6X$.

Outra alternativa, considerando a informação de todas as retas simultaneamente, seria trabalhar com a média ponderada de todas as retas, em que o peso de cada reta seria diretamente proporcional à distância ao quadrado dos valores de X que formam a reta (Tabela 1).
Com os interceptos, inclinações e pesos mostrados no Tabela 1, podemos obter estimativas para os coeficientes *a* e *b* ($\hat{a}$ e $\hat{b}$) como a média ponderada dos interceptos e inclinações, respectivamente. Temos então:

$$\begin{cases} \hat{a} = \dfrac{-13\times 1 + 0,33\times 9 + 2\times 9 - 2\times 1 - 0,5\times 16 + 5\times 4}{1+9+9+1+16+4} = 0,45 \\ \hat{b} = \dfrac{5\times 1 + 1,67\times 9 + 2\times 9 + 4\times 1 + 2,5\times 16 + 0,5\times 4}{1+9+9+1+16+4} = 2,10 \end{cases}$$

Assim, a reta Y=0,45+2,1X é a reta obtida com a participação dos coeficientes de todas as retas. Os coeficientes desta reta são aqueles obtidos pelo conhecido *método de mínimos quadrados ordinários* (em inglês, *Ordinary Least Squares - OLS*). Observe que quanto maior a distância dos valores de X para uma particular reta, maior será o seu peso na composição da estimativa.
Na abordagem tradicional de análise de regressão linear, utiliza-se a definição de que a reta OLS minimiza $\sum e_i^2$ e a reta LAD minimiza $\sum |e_i|$. Observe que a reta LAD passará sempre por pelo menos dois pares ordenados de pontos. O procedimento LAD foi proposto pelo matemático Roger Joseph Boscovich em 1757, isto é, aproximadamente 50 anos antes do método OLS. Maiores detalhes podem ser encontrados em Birkes & Dodge (1993).

3. Conclusão e Recomendações

O objetivo deste trabalho foi dividir com outros professores nossa experiência de sala de aula em introduzir o Método de Mínimos quadrados conforme descrito nesta nota. O método é simples e consiste em: obter todas as retas que passem por, pelo menos, dois pontos; utilizar uma média ponderada dos interceptos e inclinações obtidos. O exemplo apresentado na Seção 2 pode ser facilmente reproduzido pelo professor e aplicado em sala de aula. A maioria dos alunos com a qual trabalhamos com este método era do primeiro ano de graduação.

Também, em palestras, tivemos a oportunidade de apresentar a abordagem para alunos do ensino médio. Percebemos que, se motivados, muitos alunos sugerem outras possibilidades para $\hat{Y} = \hat{a} + \hat{b}X$, quase sempre sugerindo pesos alternativos para a média ponderada das retas obtidas das combinações dos pares ordenados tomados dois a dois. O melhor termômetro de que a abordagem gera grande interesse foi o fato de que vários alunos se aproximaram do quadro ao final da aula para conversar com o professor e apresentar outras sugestões.

Dessa forma, concluímos, pela nossa experiência em sala de aula, que o método é mais atrativo aos alunos numa etapa introdutória do que a equivalente e tradicional abordagem de minimização de $\sum e_i^2$. Finalmente, o método também pode ser ampliado para comtemplar estimadores alternativos para regressão linear múltipla, geralmente discutida em um estágio mais avançado da graduação.

4. Referências

1. Barbosa, V. C. ; Breitschaft, A. M. S. (2006) Um aparato experimental para o estudo do princípio de Arquimedes, *Revista Brasileira de Ensino de Física*, v.28, n.1, p.115–122.

2. Birkes, David & Yadolah Dodge. (1993) *Alternative Methods of Regression*. Wiley.

3. Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. I, II and III. Koninklijke Nederlandse. Akademie van Wetenschappen, Proceedings, ser. A, vol. 53, pp. 386–392, 521–525, 1397–1412.