

2

Regressão

O termo “regressão” foi proposto pela primeira vez por *Sir Francis Galton* (1885) num estudo onde demonstrou que a altura dos filhos não tende a refletir a altura dos pais, mas tende sim a regredir para a média da população. Atualmente, o termo “Análise de Regressão” define um conjunto vasto de técnicas estatísticas usadas para modelar relações entre variáveis e predizer o valor de uma ou mais variáveis dependentes (ou de resposta) a partir de um conjunto de variáveis independentes (ou preditoras). A relação entre as variáveis pode ser de dependência funcional¹ (*i.e.* a magnitude da variável dependente é função da magnitude da(s) variável(is) independente(s), mas o contrário já não se aplica) ou de mera associação (*i.e.* nenhuma das variáveis pode ser tida como dependente da outra, mas apenas que elas variam em conjunto). O termo variável dependente implica geralmente uma relação do tipo causa-e-efeito. Porém, a análise de regressão pode ser usada para modelar a relação funcional entre duas ou mais variáveis – *i.e.* uma relação que pode ser expressa através de uma função matemática – independente de existir ou não uma relação de tipo causa-e-efeito, o que não é o caso.

2.1

Regressão Linear

2.1.1

O Modelo de Regressão Linear

Na regressão linear a relação entre a variável dependente e as variáveis independente é uma reta – no caso de existir apenas uma variável independente – ou por extensão da reta para múltiplas dimensões – plano hipergeométrico – no caso de duas ou mais variáveis independentes.

A forma geral de um modelo de regressão linear para uma amostra de tamanho n e p variáveis é apresentada a seguir.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (2.1.1.1)$$

onde $(i = 1, \dots, n)$, Y é a variável dependente, X_j ($j = 1, \dots, p$) são as variáveis independentes, β_j ($j = 0, 1, \dots, p$) são os chamados coeficientes de regressão e ε_i representa os erros ou resíduos do modelo.

Os coeficiente de regressão β_j ($j = 1, \dots, p$) representam os declives parciais, isto é, representam uma medida de influência de X_j em Y (variação de Y por unidade de variação de X_j) enquanto que β_0 é a ordenada na origem (ou seja, é o valor de y_i – i -ésima observação da variável dependente – quando $x_{ji} = 0$, isto é, quando todas as observações de todas as variáveis independentes são nulas).

Caso o modelo seja composto por apenas uma variável independente ele é denominado **MODELO DE REGRESSÃO LINEAR SIMPLES – MRLS** e é reduzido a forma a seguir.

¹ É o caso deste estudo onde magnitude da variável dependente (consumo de energia elétrica em kWh) é função das variáveis independentes (posses dos diversos tipos de equipamentos elétricos).

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2.1.1.2)$$

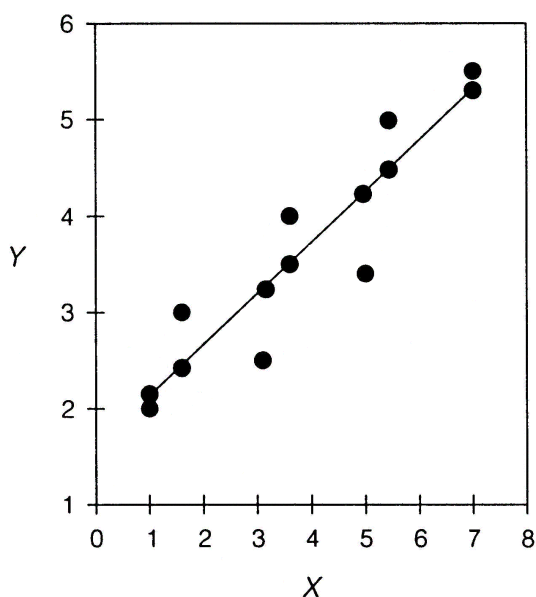


Figura 2.1.1.1 – Exemplo de Regressão Linear Simples²

Na Figura 2.1.1.1 percebe-se que os pontos distribuem-se aleatoriamente em torno de uma reta onde β_0 é o intercepto (ou seja o ponto onde a reta “corta” o eixo y), β_1 é o coeficiente de inclinação da reta (coeficiente angular).

Porém, como se pode observar a maioria dos pontos marcados na figura 2.1.1.1 não pertencem à reta. Nas equações 2.1.1.1 e 2.1.1.2 o termo de erro (resíduo) ε serve justamente para considerar os desvios dos pontos a reta que é a diferença entre o valor observado y e a reta.

O erro pode ser encarado como a soma de duas componentes: o erro de medição e o erro amostral/aleatório/estocástico. O erro de medição deve-se ao fato de que tanto a medição de y como de x pode ser deficiente e o erro amostral, aleatório ou estocástico deve-se a irreplicabilidade dos fenômenos em estudo, onde este último pode ser interpretado como a representação da influência, em y, de muitas variáveis omissas no modelo.

2.1.2

Os Pressupostos do Modelo de Regressão Linear

À semelhança do que acontece com todos os outros modelos estatísticos, o modelo de regressão é baseado em alguns pressupostos. Os principais pressupostos recaem sobre a componente ε de erro do modelo. Os pressupostos são os seguintes:

- i. A média da distribuição de probabilidade dos resíduos ε é nula.
- ii. A variância da distribuição de probabilidade de ε é constante, ou seja, o erro é homocedástico.
- iii. A distribuição de probabilidade do erro ε é normal.
- iv. O valores do erro ε associados a dois valores quaisquer observados de y são independentes, isto é, o valor de ε associado a um valor de y não tem qualquer relação com o valor de ε associado com qualquer outro valor de y .
- v. Não existe associação entre as variáveis independentes.

Resumindo: o erro é $\varepsilon \sim N(0, \sigma^2)$ e iid (independentes e identicamente distribuídos).

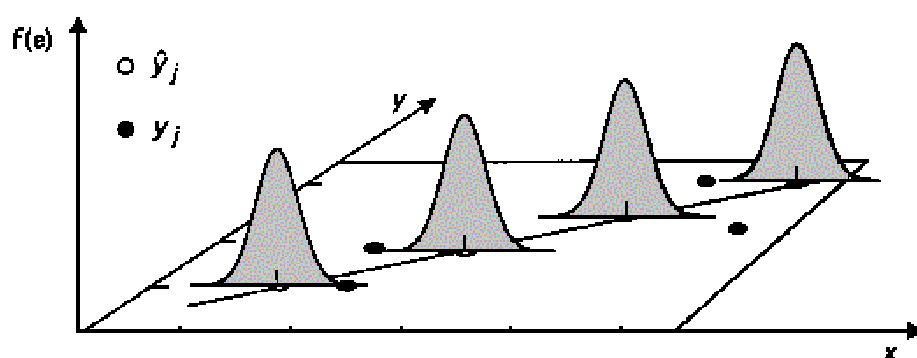


Figura 2.1.2.1 – Distribuição dos Resíduos³

² Figura adaptada de Maroco (2003).

³ Figura retirada de Maroco (2003).

2.2

O Método dos Mínimos Quadrados Ordinários (MQO)⁴

Este método utiliza como modelo a curva⁵ cuja soma dos quadrados da distâncias entre os dados e a curva seja a menor possível.

Vamos supor num primeiro momento o MRLS. Como a população não é conhecida deve-se primeiro estimar os coeficientes do modelo de regressão a partir de uma amostra representativa da população sob estudo usando os estimadores apropriados:

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1}$ onde $\hat{\beta}_0$ é a estimativa e b_0 o estimador de β_0 ; e ainda $\hat{\beta}_1$ é a estimativa e b_1 o estimador de β_1 .

Seja o modelo de regressão linear simples a seguir,

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \quad (2.2.1)$$

Seja ainda n o número de pontos observados e SQE a soma dos quadrados das distâncias verticais entre os pontos observados e a reta (veja figura 2.2.1 a seguir).

⁴ O MQO está aqui descrito de forma resumida, para um maior detalhamento existe ampla bibliografia sobre o assunto.

⁵ Reta no caso de duas variáveis. Plano no caso de três variáveis. Hiperplano no caso de mais variáveis.

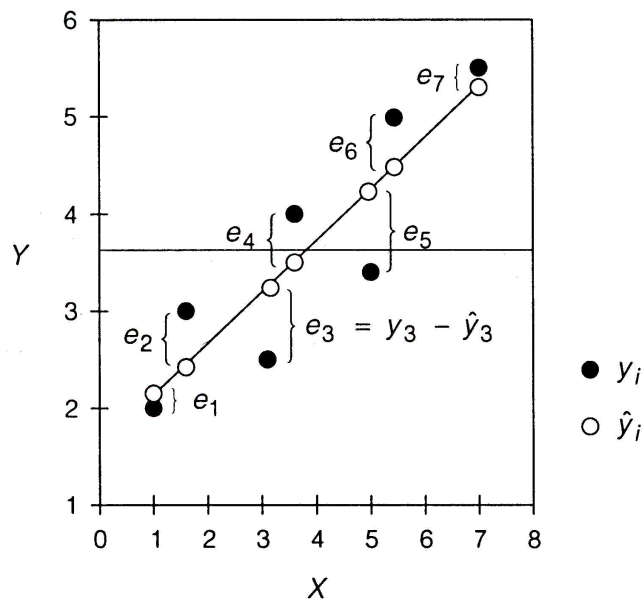


Figura 2.2.1 – Resíduos na regressão linear⁶

Isto é, sendo y_i a i -ésima observação da variável dependente, \hat{y}_i o valor predito e $\hat{e}_i = y_i - \hat{y}_i$, tem-se que o método consiste na minimização da soma dos quadrados deste último.

$$SQE = \sum_{i=1}^n \hat{e}_i^2 \quad (2.2.2)$$

Substituindo (2.2.2) em (2.2.1), tem-se:

$$SQE = \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 \quad (2.2.3)$$

Para minimizar Q com relação a β_0 e β_1 faz-se:

$$\frac{\partial SQE}{\partial \beta_0} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \quad (2.2.4)$$

e

⁶ Figura retirada de Maroco 2003.

$$\frac{\partial SQE}{\partial \beta_1} = -2 \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) x_i \quad (2.2.5)$$

Igualando a zero as derivadas acima temos

$$\beta_0 n + \beta_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \quad (2.2.6)$$

$$\beta_0 \sum_{i=1}^n x_i + \beta_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \quad (2.2.7)$$

Para se obter b_0 (estimador de β_0) e b_1 (estimador de β_1), define-se:

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{e} \quad \bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i$$

Resolvendo as equações (2.2.6) e (2.2.7) para β_0 e β_1 obtêm-se:

$$b_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x}_n \bar{y}_n}{\sum_{i=1}^n x_i^2 - n \bar{x}_n^2} \quad (2.2.8)$$

$$b_0 = \bar{y}_n - b_1 \bar{x}_n \quad (2.2.9)$$

Estendendo para o caso multivariado (modelo com duas ou mais variáveis independentes), temos:

Define-se função de regressão (Mardia, Kent e Bibby 1979) de Y em X_1, \dots, X_p como sendo a esperança condicional de Y dados quaisquer valores de X_1, \dots, X_p ou seja $E(Y|x_1, x_2, \dots, x_p)$. Sendo assim considera-se o modelo geral da regressão linear:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i \quad (2.2.10)$$

Reescrevendo a equação (2.2.10) sob a forma matricial temos:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E} \quad (2.2.11)$$

Onde:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

ou seja, \mathbf{Y} é o vetor (n x 1) de observações, \mathbf{X} é a matriz [n x (p+1)] de variáveis independentes, $\boldsymbol{\beta}$ é o vetor [(p+1) x 1] dos coeficientes de regressão, \mathbf{E} é o vetor (n x 1) dos erros.

Minimizando a equação (2.2.11) temos:

$$Q = \sum_{i=1}^n \varepsilon_i^2 = \mathbf{E}'\mathbf{E} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \quad (2.2.12)$$

Derivando (2.2.12) em função de $\boldsymbol{\beta}$ e igualando a zero, tem-se o estimador de mínimos quadrados

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (2.2.13)$$

2.3

Regressão Robusta

Conforme explicado no capítulo sobre regressão linear, o método mais utilizado de regressão linear é o Método dos Mínimos Quadrados Ordinários (MQO), porém este método estatístico possui pressupostos que devem ser observados na análise e que na prática, não raramente, são violados. Surge daí então a necessidade de métodos “robustos” em relação aos desvios destes pressupostos. Os dados obtidos para este estudo conforme é mostrado mais a frente não fogem a este problema. Neste trabalho então foram aplicados os dois métodos para a realização da Análise de Demanda Condicional para comparação entre ambos.

Segundo S-Plus (1997), as técnicas de regressão robusta são um importante complemento às técnicas clássicas de mínimos quadrados, uma vez que fornecem respostas similares a regressão por mínimos quadrados quando existe relação linear entre as variáveis com os erros normalmente distribuídos, porém diferem significativamente dos ajustes de mínimos quadrados quando os erros não são normalmente distribuídos ou quando os dados contêm *outliers* significantes.

O uso da regressão robusta justifica-se por ser esta considerada uma técnica robusta não somente com respeito aos *outliers*, mas também com relação aos pontos extremos, que são pontos no modelo matricial com excessiva influência sobre o resultado, e porque quanto maior o número de variáveis de um modelo, mais difícil se torna a identificação de *outliers* com o uso das técnicas de regressão clássicas (Cunha, Machado e Filho 2002).

A seguir descreve-se resumidamente alguns conceitos necessários à realização da análise de regressão via Estimadores Robustos.

2.3.1

Outliers

Barnett & Lewis (1995) definiram *outlier* em um conjunto de dados como sendo uma observação que parece ser inconsistente com o conjunto de dados remanescentes. Os *outliers* podem indicar algumas características importantes sobre um modelo, como modelo incompatível com os dados e omissão de variáveis importantes. Para Draper & Smith (1981), a rejeição automática de *outliers* não é um procedimento correto e as regras propostas para rejeição de *outliers* devem incluir a reanálise sem essas observações, que, dependendo das circunstâncias, podem ser portadoras de informações vitais dos indivíduos de uma população.

Uma observação é chamada de *outlier* de regressão, ou *outlier* na direção y , se ela se afasta do padrão linear definido pelas outras (ou pela maioria das outras) observações, este tipo de *outlier* em geral dá origem à grandes resíduos. Por outro lado, quando uma observação se destaca das demais no espaço das variáveis explanatórias (independentes), é chamada de ponto de alavanca. Cabe ressaltar que um ponto de alavanca não tem que ser necessariamente um *outlier* de regressão. Quando um ponto de alavanca é também um *outlier* de regressão, ou seja, quando um ponto é um valor discrepante tanto dentre os valores da variável dependente quanto dentre os valores das variáveis independentes, terá grande influência sobre os ajustes clássicos (Mendes 1999).

Esta grande influência reforça a idéia da busca de um método que seja “robusto” em relação a um conjunto de dados que seja contaminado, isto é, que possua valores aberrantes.

2.3.2

Equivariância

Existem 3 tipos de equivariância, regressão, escala e afim.

Um estimador de regressão T é dito equivariante de regressão se:

$$T(((x_i; y_i + x_i v); i = 1, \dots, n)) = T(((x_i; y_i); i = 1, \dots, n)) + v \quad (2.3.2.1)$$

onde v é um vetor qualquer.

Um estimador T é dito afim equivariante se:

$$T(((x_i A; y_i); i = 1, \dots, n)) = A^{-1} T(((x_i; y_i); i = 1, \dots, n)) \quad (2.3.2.2)$$

onde A é uma matriz quadrada, não singular.

Um estimador T é dito equivariante de escala se:

$$T(((x_i; c y_i); i = 1, \dots, n)) = c T(((x_i; y_i); i = 1, \dots, n)) \quad (2.3.2.3)$$

2.3.3

Ponto de Ruptura

Devido ao fato de que uma simples observação não condizente com o resto das observações poder distorcer significativamente os resultados obtidos com o método clássico, o uso de métodos robustos que utilizem estimadores que “resistam” a um certo percentual de dados contaminados poderá fornecer resultados significativamente mais confiáveis. De modo a formalizar o quão resistente a dados contaminados é um estimador, surge o conceito de ponto de ruptura. Neste trabalho será descrito apenas o conceito de ponto de ruptura para amostra finita introduzida por DONOHO e HUBER (1983).

Ponto de ruptura é uma medida global de robustez. Ele mede qual seria a maior porcentagem de contaminação que um estimador poderia suportar e ainda assim fornecer informação confiável sobre o parâmetro considerado. No contexto de regressão, a versão amostral (finita) do ponto de ruptura significa a menor fração de *outliers* com o potencial de distorcer por completo as estimativas. Um estimador robusto adequado então, é aquele que possui ponto de ruptura igual a 0,5, ou seja, é o estimador que “resiste” a um percentual de até 50% dos dados contaminantes. Um valor de ponto de ruptura maior que 0,5 não faria sentido pois se mais da metade das observações fosse de valores discrepantes, não seria mais possível diferenciar quais seriam os dados “ruins”.

Seja um conjunto de dados com n observações de p variáveis independentes (variáveis explicativas) e uma variável dependente (variável resposta), de acordo com a equação

$$Z = \{(X_{i1}, X_{i2}, \dots, X_{ip}, Y_i); 1 \leq i \leq n\} \quad (2.3.3.1)$$

Com estes dados seguindo o modelo de regressão

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (2.3.3.2)$$

Seja também T um estimador dos parâmetros de regressão β , donde aplicando T ao conjunto de dados Z , tem-se

$$T(Z) = (\hat{\beta}_0, \dots, \hat{\beta}_p) \quad (2.3.3.3)$$

Suponha que o conjunto tenha m ($m \leq n$) observações originais substituídas por m valores arbitrários (aberrantes) produzindo assim uma amostra contaminada Z' . Logo esta amostra tem uma fração $\varepsilon = \frac{m}{n}$ de valores contaminantes.

O ponto de ruptura, versão finita de $T(\cdot)$ em Z é então definido como sendo ε^* a menor fração de *outliers* que pode distorcer por completo a estimativa.

Seja ainda o vício máximo causado pela fração ε de contaminação definido como

$$v(m; (T, Z)) = \sup \|T(Z') - T(Z)\| \quad (2.3.3.4)$$

onde o supremo é obtido examinando-se todas as amostras corrompidas possíveis de Z' . O ponto de ruptura, versão finita pode ser então definido como

$$r_n^*(T, Z) = \min \left\{ \frac{m}{n}; v(m; (T, Z)) \right\} \quad (2.3.3.5)$$

Para o caso clássico, estimado de mínimos quadrados ordinários, o ponto de ruptura é dado por

$$r_n^*(T, Z) = \frac{1}{n} \quad (2.3.3.6)$$

Donde pode-se concluir que quanto maior o tamanho da amostra menor o ponto de ruptura do estimador clássico, ainda, se o tamanho da amostra tende ao infinito o ponto de ruptura do estimador clássico tende a zero.

Se $n \rightarrow \infty$ então $r_n^*(T, Z) \rightarrow 0$

Para ilustrar o comportamento de um estimador com alto ponto de ruptura, utiliza-se como exemplo o diagrama de dispersão dos dados Rio *Kootenay* e também ilustra a comparação entre um estimador de alto ponto de ruptura o MMQ, e o estimador clássico MQO. Observa-se que a reta de mínimos quadrados foi “atraída” pelo ponto de alavanca e não representa bem o padrão linear definido pela maioria dos pontos.

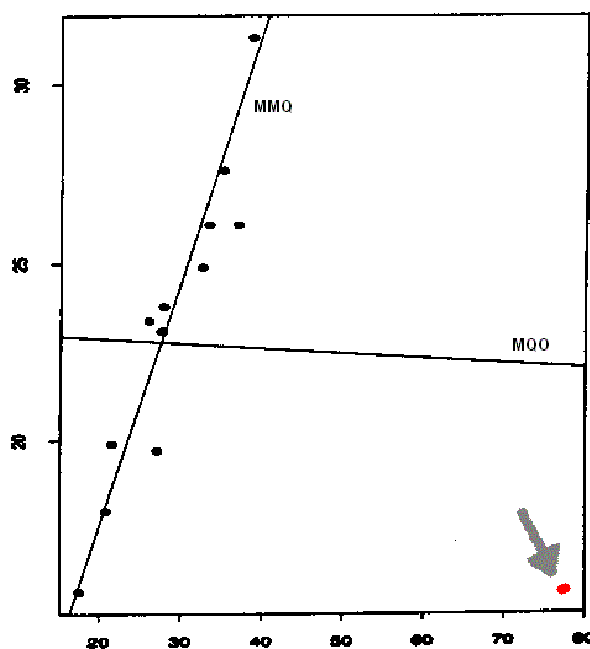


Figura 2.3.3.1 – Ponto de alavanca⁷

Muitos autores têm se baseado no critério global do alto ponto de ruptura para desenvolver propostas robustas em relação aos pontos de alavanca. Contudo, várias dessas propostas precisam de sacrificar as propriedades locais de robustez, propriedades estas determinadas pela função de influência.

⁷ Exemplo retirado de Mendes (1999).

2.3.4

Função de Influência

A função de influência investiga a robustez local de um estimador ao descrever a influência de uma contaminação local na estimativa, medindo portanto se o estimador responderia suavemente a pequenas modificações nos dados, isto é, a função de influência mede a eficácia de um estimador diante de uma contaminação infinitesimal em uma observação qualquer da amostra.

Segundo Mendes (1999), um primeiro aspecto a se considerar sobre a robustez de um estimador é relacionado com sua estabilidade local.

A função de influência (IF) de um estimador de regressão T em uma distribuição F é definida como

$$IF((x_0, y_0); T, F) = \lim_{\varepsilon \rightarrow 0} \frac{T((1-\varepsilon)F + \varepsilon\delta_0) - T(F)}{\varepsilon} \quad (2.3.4.1)$$

$\forall (x_0, y_0) \in \mathfrak{R}^{p-1}$ para os quais este limite existe e onde (x_0, y_0) é a observação contaminada. Para um maior detalhamento sobre este tópico, veja Mendes (1999) e Hampel et al. (1986).

2.3.5

Posição Geral para Regressão

Para $n \geq p$, um conjunto de dados é dito estar em posição geral para regressão se apenas um hiperplano passa através de cada um de seus subconjuntos de tamanho p , e se não existem dois destes hiperplanos iguais para diferentes subconjuntos de tamanho p , isto é, só existe um e somente um hiperplano para cada subconjunto dos dados e dois subconjuntos diferentes não possuem o mesmo hiperplano.

2.3.6

Principais Estimadores Robustos

Denominam-se estimadores robustos, os estimadores de regressão que possuem certas propriedades que os permitem estimações confiáveis mesmo quando os dados possuem um certo percentual de contaminação, isto é, são estimadores literalmente robustos a desvios dos pressupostos básicos do método de mínimos quadrados. Muitos autores vêm propondo novos métodos na tentativa de resolver o problema quase sempre encontrado na prática de desvio de pressupostos. De modo geral, estes autores propõem métodos que substituem a função a ser minimizada (que conforme já comentado, no caso do método dos mínimos quadrados ordinários é a soma dos quadrados dos resíduos) por outra que não seja tão sensível a *outliers*.

2.3.6.1

Estimadores L_1

Os estimadores L_1 , que se baseiam na minimização da soma dos valores absolutos dos resíduos, isto é, minimizam a função dos resíduos.

$$L_1 = \sum_{i=1}^n |r_i| \quad (2.3.6.1.1)$$

onde n é o número de observações e $r_i = y_i - x_i\beta$ é o i -ésimo resíduo.

Estes foram os primeiros estimadores robustos para regressão e foram propostos por Edgeworth em 1887.

Os estimadores L_1 são robustos em relação a *outliers* na em Y (variável dependente). Contudo, estes estimadores não são melhores que os estimadores de mínimos quadrados em relação ao ponto de ruptura visto que os mesmos não são

robustos a *outliers* em X (variáveis dependentes), ou seja, eles não são robustos s pontos de alavanca possuindo portanto ponto de ruptura igual a zero.

Os estimadores L_1 não podem então ser considerados bons pois um bom estimador deve ser robusto em relação a *outliers* em Y e em X e possuir certas propriedades de equivariância.

2.3.6.2

M-estimadores

Os M-estimadores foram propostos por HUBER (1973), generalizando o conceito de estimação de máxima verossimilhança. Estes são na verdade uma classe de estimadores muito geral e inclui, como um de seus casos particulares, o estimador de mínimos quadrados ordinários e o estimador L_1 .

Pode-se definir os M-estimadores como os valores (β_n, σ_n) que minimizam a seguinte função de dispersão dos resíduos:

$$S(\beta) = \sum_{i=1}^n \rho\left(\frac{y_i - x_i^T \beta}{\sigma}\right) = \sum_{i=1}^n \rho\left(\frac{r_i}{s}\right) \quad (2.3.6.2.1)$$

2.3.6.3

GM-estimadores ou M-estimadores Generalizados

Os Gm-estimadores, foram introduzidos por Maronna et al. (1979) e seu principal objetivo é limitar a influência dos pontos de alavanca através da aplicação de alguma função peso $\omega_i = \omega(x_i)$ em x_i . Porém como mostrou Maronna et al. (1979), o ponto de ruptura dos Gm-estimadores decresce com o aumento do número p de variáveis independentes não podendo ser maior que $1/p$.

2.3.6.4

Least Median of Squares – LMS e Least Trimmed Squares - LTS

O estimador LMS – *Least Median of Squares* (em português Menor Mediana dos Quadrados dos Resíduos – MMQ), foi proposto por Rousseeuw (1984). Rousseeuw propôs que se tomasse como estimador a menor mediana dos quadrados dos resíduos.

$$\min_b \left(\text{mediana}_i(r_i^2) \right) \quad (2.3.6.4.1)$$

onde $r_i = y_i - x_i\beta$ é o i -ésimo resíduo.

O estimador LMS é robusto em relação aos outliers em X e Y, e possui ponto de ruptura 0,5, apresentando também propriedades de equivariância. Contudo são ineficientes quando os erros seguem uma distribuição normal visto que possuem lenta taxa de convergência dada por $1/\sqrt[3]{n}$. Ainda, os estimadores LMS possuem não suave resposta a discretas alterações nos valores centrais dos dados, provocando assim grandes alterações nas estimativas (Hettmansperger & Sheather –1992 e Davies 1993).

Algum tempo depois Rousseeuw propôs o estimador LTS – *Least Trimmed Squares* (em português Mínimos Quadrados Podados – MQP) dado por:

$$\min_b \sum_{i=1}^q |y_i - x_i b|_{(i)}^2 \quad (2.3.6.4.2)$$

O estimador LTS possui a mesma robustez a *outliers* do LMS, porém o mesmo é mais eficiente pois possuem taxa de convergência dada por \sqrt{n} .

2.3.6.5

MM-estimadores

O estimador agora em questão é por ser o escolhido na execução deste trabalho. Esta nova classe de estimadores robustos pra regressão foi introduzido por Yohai (1985), e marcou um grande avanço dentre a classe de estimadores robustos.

Os MM-estimadores, possuem as seguintes propriedades:

- ponto de ruptura de 50%;
- consistência
- eficiência assintótica

Estes estimadores são definidos através de um procedimento em três estágios:

1° estágio: obtém-se estimativas para os parâmetros $\beta^{(0)}$ com alto ponto de ruptura utilizando-se o estimador LTM por exemplo, para isso não há imposição de eficiência do estimador.

2° estágio: calcula-se um M-estimador de escala s (S-estimador) a partir dos resíduos obtidos no primeiro estágio.

3° estágio: fixando-se este estimador de escala como uma escala auxiliar, calcula-se um M-estimador de regressão baseado em uma função ψ redescendente.