ANÁLISE DOS MÍNIMOS QUADRADOS.

Flávio Felipe Kirchner*

SUMMARY

In this paper the derivation of the set of simultaneous normal equations by least squares method is presented.

This set of normal equations can be centered or non-cenetred to its mean and the method to obtain these two sets is mathematically demonstrated.

1. INTRODUÇÃO

Muitos pesquisadores e administradores tem descoberto a utilidade dos métodos de regressão para derivar e testar relações empíricas entre vários fenômenos observados. No campo da engenharia florestal, por exemplo, o volume das árvores tem sido expresso como uma função do diâmetro, altura comercial e classe de forma; as propriedades da madeira tem sido relacionadas às características como peso específico, idade, e taxa média de crescimento radial; estudos tem sido feitos de como os custos de exploração são afetados pelo tamanho médio da árvore, volume total e distância das estradas asfaltadas; e índice de sítio para várias espécies tem sido relacionado a certas propriedades do solo e topografia.

Análise de regressão proporciona uma rotina objetiva e amplamente aceitável para ajustar modelos matemáticos que envolvem várias variáveis. Adicionalmente, existem procedimentos que podem ser sempre usados para avaliar a equação ajustada e com o desenvolvimento dos computadores, a maior parte do trabalho penoso foi eliminado.

Infelizmente, o valor óbvio e o aumento da disponibilidade de métodos de regressão tem resultado em uso por pessoas que têm pouco conhecimento do mecanismo e suas limitações. Isto não é necessariamente uma catástrofe estatística — muitas pessoas dirigem um carro sem ter a mínima noção do que o faz ir. Mas, o usuário de regressão, como o motorista de carro, irá realizar um trabalho melhor se ele aprendeu os melhores procedimentos de operação e sabe alguma coisa sobre o que o maquinário pode ou não fazer.

O propósito deste trabalho é proporcionar algum conhecimento sobre a obtenção das equações normais pelo método dos mínimos quadrados, em termos relativamente simples.

2. CONCEITUAÇÃO

Apesar da terminologia especial usada nos livros de receita estatísticos para disfarçar similaridades, a maioria das técnicas estatísticas comuns são casos especiais do método geral da análise dos mínimos quadrados dos modelos lineares.

As apresentações destes livros de receitas são projetados para oferecer caminhos curtos (atalhos) nos cálculos. Contudo, o preço destes atalhos é a proliferação da confusão das regras para o conhecimento de casos especiais nos quais os caminhos curtos são aplicados.

Com a disponibilidade de grandes computadores, estes atalhos são mais um peso ou responsabilidade do que um benefício. Uma pessoa leva mais tempo situando o problema do que o computador exige para realizar os cálculos.

Tais técnicas aparentemente divergentes como análise de regressão, análise de variância, análise de covariância e análise de funções discriminantes, são todas casos especiais da análise dos mínimos quadrados dos modelos lineares.

Em reconhecendo-se as bases comuns destas técnicas, uma pessoa pode usar um único programa de computação para mínimos quadrados lineares, com uma subroutina de transformação, para realizar o volume de seu trabalho.

* Professor do Curso de Engenharia Florestal — UFPR.

3. O MODELO LINEAR GERAL

Sempre que nós desejarmos "explicar" uma variável particular em termos de várias outras variáveis observadas ou medidas nos mesmos indivíduos, o modelo linear é um candidato potencial. De fato, é o modelo mais simples e, conseqüentemente, o modelo preferido se ele parecer de todo realístico.

Para ser mais específico, assuma que nós observamos ou medimos a variável Y_i $(i = 1, \ldots, n)$ em cada um dos "n" indivíduos. Nos mesmos "n" indivíduos, nós também medimos ou observamos outras "m" variáveis, as quais são representadas por X_{i1} $(i=1,\ldots,n)$, X_{i2} $(i=1,\ldots,n)$, $\ldots$, X_{im} $(i=1,\ldots,n)$.

O nosso propósito é de "explicar" a variável-Y em termos das "m" variáveis-X. Isto é, o nosso modelo toma a forma de

$$Y_i = b_1X_{i1} + b_2X_{i2} + \ldots + b_mX_{im},$$

onde os coeficientes "b" são constantes a serem determinadas baseando-se nos dados. Na terminologia usual, nós chamamos "Y" a variável dependente e os "X's" as variáveis independentes.

A linearidade deste modelo é em termos dos "b's" em vez dos "X's". Nós somos livres para escolher os "X's", contudo, cada termo no modelo pode conter somente um único "b", e o "b" não deve estar elevado à potência. Desta forma, representando de forma genérica, nós temos:

Primeira observação →

$$\begin{array}{c|cccc|}
Y_i & X_{i1} & X_{i2} & X_{i3} \ldots & X_{im} \\
Y_1 & X_{11} & X_{12} & X_{13} \ldots & X_{1m} \\
Y_2 & X_{21} & X_{22} & X_{23} \ldots & X_{2m} \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
Y_n & X_{n1} & X_{n2} & X_{n3} \ldots & X_{nm}
\end{array}$$

→ Matriz das variáveis independentes

Se as variáveis -X são variáveis categóricas (dummy), o modelo é usualmente chamado de "Análise de Variância". Por exemplo, a variável a qual somente pode assumir os valores -1, 0, 1 é uma variável categórica.

Se as variáveis -X são variáveis medidas, o modelo é chamado de "Análise de Regressão".

Se algumas das variáveis -X são variáveis categóricas e o restante são variáveis medidas, o modelo é usualmente chamado de modelo de "Análise de Covariância".

4. AJUSTANDO O MODELO GERAL DOS MÍNIMOS QUADRADOS PARA OS DADOS

Até agora nós não falamos nada sobre a determinação das valores dos coeficientes "b", exceto que nós iremos obtê-los baseando-se nos dados disponíveis. Naturalmente, nós desejamos que estes valores sejam obtidos da melhor forma possível, no sentido da palavra "melhor".

Isto significa que nós desejamos "maximizar" ou "minimizar" alguma coisa no processo do cálculo dos "b's".

Para isto, vamos re-examinar o modelo linear:

$$Y_i = b_1X_{i1} + b_2X_{i2} + \ldots + b_mX_{im}$$

Quando os "b's" forem determinados e as variáveis-X para a observação i-th introduzidas no modelo, o resultado irá ser um valor "prognosticado" (ou estimado) para a variável-Y para a observação i-th.

Para diferenciar o **valor estimado** do **valor atual** (ou verdadeiro), nós iremos daqui para frente por um acento circumflexo (ou "chapéu") sobre Y_i quando referir ao valor estimado. Então, agora nós escreveremos o modelo como:

$$\hat{Y}_i = b_1X_{i1} + b_2X_{i2} + \ldots + b_mX_{im}$$

Na maioria dos casos, o valor estimado não irá ser exatamente igual ao valor atual, e a diferença é:

$$Y_i - \hat{Y}_i = Y_i - b_1X_{i1} - b_2X_{i2} - \ldots - b_mX_{im}$$

Para o critério de "melhor", nós iremos exigir que os nossos "b's" sejam tais que a soma dos quadrados dos desvios dos **valores atuais** para os **valores estimados** seja a menor possível, então o nome de **"mínimos quadrados"**.

Isto é, nós desejamos o conjunto de "b's" que minimize a função:

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - b_1X_{i1} - b_2X_{i2} - \ldots - b_mX_{im})^2$$

Daqui para frente, nós não iremos mais escrever o subscrito **i**, para economizar espaço; contudo, é subentendido estar presente.

Para minimizar esta função como desejada, nós precisamos:

1 — Achar as derivadas parciais da função com respeito a cada um dos "b's" em questão;

2 — Equacionar as derivadas parciais simultaneamente a zero;

3 — Demonstrar que o nosso resultado é, de fato, o mínimo possível.

Então, nós temos:

$$\Sigma (Y - \hat{Y})^2 = \Sigma [Y - (b_1X_1 + b_2X_2 + \ldots + b_mX_m)]^2$$

$$= \Sigma (Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2$$

Agora nós procedemos em encontrar as derivadas parciais:

Para b_1:

$$\frac{\partial \Sigma (Y - \hat{Y})^2}{\partial b_1} = \Sigma \left[\frac{(Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial b_1} \right]$$

$$= \Sigma[2\ (Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m) . \frac{\partial (Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)}{\partial b_1}]$$

$$= \Sigma\ [2\ (Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m).(-X_1)]$$

$$= \Sigma\ [-2X_1\ (Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)]$$

$$= \Sigma\ [-2\ (X_1Y - b_1X_1^2 - b_2X_1X_2 - \ldots - b_mX_1X_m)]$$

$$= -2\Sigma\ [X_1Y - b_1X_1^2 - b_2X_1X_2 - \ldots - b_mX_1X_m]$$

$$= -2\ [\ \Sigma\ X_1Y - b_1\Sigma\ X_1^2 - b_2\Sigma\ X_1X_2 - \ldots - b_m\Sigma\ X_1X_m]$$

Do mesmo modo para b_k:

$$\frac{\partial\ \Sigma\ (Y-\hat{Y})^2}{\partial\ b_k} = \Sigma\ [\frac{(Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial\ b_k}]$$

$$= \Sigma\ [2(Y-b_1X_1-b_2X_2-\ldots-b_mX_m) . \frac{\partial (Y-b_1X_1-b_2X_2-\ldots-b_mX_m)}{\partial\ b_k}]$$

$$= \Sigma\ [2(Y-b_1X_1-b_2X_2-\ldots-b_mX_m).(-X_k)]$$

$$= \Sigma\ [-2X_k(Y-b_1X_1-b_2X_2-\ldots-b_mX_m)]$$

$$= \Sigma\ [-2(X_kY-b_1X_1X_k-b_2X_2X_k-\ldots-b_mX_mX_k)]$$

$$= -2\Sigma\ [X_kY - b_1X_1X_k - b_2X_2X_k - \ldots - b_mK_mX_k]$$

$$= -2\ [\Sigma\ X_kY - b_1\Sigma X_1X_k - b_2\Sigma X_2X_k - \ldots - b_m\Sigma X_mX_k]$$

Fazendo estas **m** derivadas parciais igual a zero, nós obtemos **m** equações em **m** incógnitas (os b's).

$$0 = -2\ [\Sigma X_1Y - b_1\Sigma X_1^2 - b_2\Sigma X_1X_2 - \ldots - b_m\Sigma X_1X_m]$$

$$0 = -2\ [\Sigma X_2Y - b_1\Sigma X_1X_2 - b_2\Sigma X_2^2 - \ldots - b_m\Sigma X_2X_m]$$

$$\vdots$$

$$0 = -2\ [\Sigma X_mY - b_1\Sigma X_1X_m - b_2\Sigma X_2X_m - \ldots - b_m\Sigma X_m^2]$$

Dividindo as equações por 2 e multiplicando por menos 1, nós temos:

$$0 = \Sigma X_1Y - b_1\Sigma X_1^2 - b_2\Sigma X_1X_2 - \ldots - b_m\Sigma X_1X_m$$

$$0 = \Sigma X_2Y - b_1\Sigma X_1X_2 - b_2\Sigma X_2^2 - \ldots - b_m\Sigma X_2X_m$$

$$\vdots$$

$$0 = \Sigma X_mY - b_1\Sigma X_1X_m - b_2\Sigma X_2X_m - \ldots - b_m\Sigma X_m^2$$

$$-\Sigma X_1Y = -b_1\Sigma X_1^2 - b_2\Sigma X_1X_2 - \ldots - b_m\Sigma X_1X_m$$

$$-\Sigma X_2Y = -b_1\Sigma X_1X_2 - b_2\Sigma X_2^2 - \ldots - b_m\Sigma X_1X_2$$

$$\vdots$$

$$-\Sigma X_mY = -b_1\Sigma X_1X_m - b_2\Sigma X_2X_m - \ldots - b_m\Sigma X_m^2$$

$$b_1\Sigma X_1^2 + b_2\Sigma X_1X_2 + \ldots + b_m\Sigma X_1X_m = \Sigma X_1Y$$

$$b_1\Sigma X_1X_2 + b_2\Sigma X_2^2 + \ldots + b_m\Sigma X_2X_m = \Sigma X_2Y$$

$$\vdots$$

$$b_1\Sigma X_1X_m + b_2\Sigma X_2X_m + \ldots + b_m\Sigma X_m^2 = \Sigma X_mY$$

Este conjunto de "m" equações simultâneas é chamado de **"Conjunto de Equações Normais"**.

5. DERIVAÇÃO DO CONJUNTO CENTRADO DE EQUAÇÕES NORMAIS

A não ser que se deseje forçar a equação ajustada passar pela origem, nós usualmente damos o valor de 1 (um) para a variável X_1, para cada observação. Isto, de fato, faz X_1 ser uma constante em vez de uma variável. Então, o nosso modelo linear original:

$$\hat{Y} = b_1X_1 + b_2X_2 + \ldots + b_mX_m \text{ (modelo para a origem)}$$

reduz para:

$$\hat{Y} = b_1 + b_2X_2 + \ldots + b_mX_m$$

e as equações normais vem a ser:

$$\begin{array}{l} b_1n + b_2\Sigma X_2 + \ldots + b_m\Sigma X_m = \Sigma Y \\ b_1\Sigma X_2 + b_2\Sigma X_2^2 + \ldots + b_m\Sigma X_2X_m = \Sigma X_2Y \\ \quad . \qquad . \qquad\qquad . \qquad . \\ \quad . \qquad . \qquad\qquad . \qquad . \\ \quad . \qquad . \qquad\qquad . \qquad . \\ b_1\Sigma X_m + b_2\Sigma X_2X_m + \ldots + b_m\Sigma X_m^2 = \Sigma X_mY \end{array}$$

Da primeira equação normal, nós temos:

$$b_1n + b_2\Sigma X_2 + \ldots + b_m\Sigma X_m = \Sigma Y$$

$$b_1n = \Sigma Y - b_2\Sigma X_2 - \ldots - b_m\Sigma X_m$$

$$b_1 = \frac{\Sigma Y}{n} - b_2\frac{\Sigma X_2}{n} - \ldots - b_m\frac{\Sigma X_m}{n}$$

$$\boxed{b_1 = \bar{Y} - b_2\bar{X}_2 - \ldots - b_m\bar{X}_m}$$

Substituindo esta equação por b_1, no modelo, nós obtemos:

$$\hat{Y} = b_1 + b_2X_2 + \ldots + b_mX_m$$

$$\hat{Y} = (\bar{Y} - b_2\bar{X}_2 - \ldots - b_m\bar{X}_m) + b_2X_2 + \ldots + b_mX_m$$

$$\hat{Y} = \bar{Y} + (b_2X_2 - b_2\bar{X}_2) + \ldots + (b_mX_m - b_m\bar{X}_m)$$

$$\hat{Y} = \bar{Y} + b_2 (X_2 - \bar{X}_2) + \ldots + b_m (X_m - \bar{X}_m)$$

$$\hat{Y} - \bar{Y} = b_2 (X_2 - \bar{X}_2) + \ldots + b_m (X_m - \bar{X}_m)$$

Em efeito, se nós eliminarmos b_1 do modelo e medir as variáveis-X remanescentes como desvios das suas respectivas médias amostrais, o modelo irá estimar a variável-Y como um desvio da média amostral de Y. Além disto, em entrando-se os valores médios para as variáveis-X, nós obtemos:

$$\hat{Y} = \bar{Y} + b_2 (X_2 - \bar{X}_2) + \ldots + b_m (X_m - \bar{X}_m)$$

$$\hat{Y} = \bar{Y} + b_2 (\bar{X}_2 - \bar{X}_2) + \ldots + b_m (\bar{X}_m - \bar{X}_m)$$

$$\hat{Y} = \bar{Y} + 0 + \ldots + 0$$

$$\hat{Y} = \bar{Y}$$

Conseqüentemente, o efeito de fazer $X_1 = 1$ é de "centrar" o modelo sobre uma média comum de todas as variáveis envolvidas.

Em termos das eqüações normais centradas, usando os desvios das médias é equivalente usar a primeira equação para eliminar b_1 das equações remanescentes. Por exemplo, as duas primeiras equações normais centradas são:

$$b_1n + b_2\Sigma X_2 + b_3\Sigma X_3 + \ldots + b_m\Sigma X_m = \Sigma Y$$

$$b_1\Sigma X_2 + b_2\Sigma X_2^2 + b_3\Sigma X_2X_3 + \ldots + b_m\Sigma X_2X_m = \Sigma X_2Y$$

Divida a primeira equação por $\underline{n}$ e multiplique-a por ΣX_2, para obter:

$$b_1\Sigma X_2 + b_2 \frac{(\Sigma X_2)^2}{n} + b_3 \frac{(\Sigma X_2)(\Sigma X_3)}{n} + \ldots + b_m \frac{(\Sigma X_2)(\Sigma X_m)}{n} = \frac{(\Sigma X_2)(\Sigma Y)}{n} \qquad (5.1)$$

$$b_1\Sigma X_2 + b_2 \Sigma X_2^2 + b_3 \Sigma X_2 X_3 + \ldots + b_m \Sigma X_2 X_m = \Sigma X_2 Y \qquad (5.2)$$

Em seguida, subtraia a equação (5.1) da equação (5.2), para obter:

$$b_2[\Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n}] + b_3[\Sigma X_2 X_3 - \frac{(\Sigma X_2)(\Sigma X_3)}{n}] + \ldots +$$

$$+ b_m[\Sigma X_2 X_m - \frac{(\Sigma X_2)(\Sigma X_m)}{n}] = [\Sigma X_2 Y - \frac{(\Sigma X_2)(\Sigma Y)}{n}]$$

ou

$$b_2\Sigma(X_2 - \bar{X}_2)^2 + b_3\Sigma(X_2 - \bar{X}_2)(X_3 - \bar{X}_3) + \ldots + b_m\Sigma(X_2 - \bar{X}_2)(X_m - \bar{X}_m) =$$

$$= \Sigma(X_2 - \bar{X}_2)\ (Y - \bar{Y})$$

como a primeira do nosso conjunto centrado de equações normais. Procedendo-se da mesma maneira para as equações remanescentes, nós chegamos no seguinte conjunto de equações centradas:

$$b_2\Sigma(X_2-\bar{X}_2)^2 + b_3\Sigma(X_2-\bar{X}_2)(X_3-\bar{X}_3) + \ldots + b_m\Sigma(X_2-\bar{X}_2)(X_m-\bar{X}_m) = \Sigma(X_2-\bar{X}_2)(Y-\bar{Y})$$

$$b_2\Sigma(X_3-\bar{X}_3)(X_2-\bar{X}_2) + b_3\Sigma(X_3-\bar{X}_3)^2 + \ldots + b_m\Sigma(X_3-\bar{X}_3)(X_m-\bar{X}_m) = \Sigma(X_3-\bar{X}_3)(Y-\bar{Y})$$

$$\begin{array}{cccc} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array}$$

$$b_2\Sigma(X_m-\bar{X}_m)(X_2-\bar{X}_2) + b_3\Sigma(X_m-\bar{X}_m)(X_3-\bar{X}_3) + \ldots + b_m\Sigma(X_m-\bar{X}_m)^2 = \Sigma(X_n-\bar{X}_m)(Y-\bar{Y})$$

que suplementamos com:

$$b_1 = \bar{Y} - b_2\bar{X}_2 - b_3\bar{X}_3 - \ldots - b_m\bar{X}_m$$

para completar a solução.

Nós podemos fazer o conjunto centrado de equações normais mais simples e compacto em revisando um pouco a nossa notação.

Primeiro, nós subtraímos 1 (um) do índice de cada variável e coeficiente (fazendo $\ell = m-1$), sendo que o nosso modelo fica:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_\ell X_\ell$$

Fazendo em seguida:

$$y = Y - \bar{Y}$$

$$x_1 = X_1 - \bar{X}_1$$

$$\vdots$$

$$x_\ell = X_\ell - \bar{X}_\ell$$

Com esta notação, o nosso conjunto centrado de equações fica:

$$\begin{array}{l}
b_1 \, \Sigma x_1{}^2 + b_2 \, \Sigma x_1 x_2 + \ldots + b_\ell \, \Sigma x_1 x_\ell = \Sigma x_1 y \\
b_1 \, \Sigma x_1 x_2 + b_2 \, \Sigma x_2^2 + \ldots + b_\ell \, \Sigma x_2 x_\ell = \Sigma x_2 y \\
\quad \vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \\
b_1 \, \Sigma x_1 x_\ell + b_2 \, \Sigma x_2 x_\ell + \ldots + b_\ell \, \Sigma x_\ell^2 = \Sigma x_\ell y
\end{array}$$

com

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 - \ldots - b_\ell \bar{X}_\ell$$

Note que com o nosso conjunto centrado de equações,

[1] $$\Sigma \hat{y}^2 = b_1 \Sigma x_1 y + b_2 \Sigma x_2 y + \ldots + b_\ell \Sigma x_\ell y$$

nos dá a seguinte repartição de variação

[2] $$\Sigma y^2 = \Sigma \hat{y}^2 + \Sigma \, (y - \hat{y})^2$$

Isto é, nós agora estamos repartindo a variação nos desvios da variável-Y da sua média. Nós precisamos mostrar somente que:

$$\Sigma Y^2 = \Sigma \bar{Y}^2 + \Sigma(Y - \bar{Y})^2$$
$$= n\bar{Y}^2 + \Sigma(Y - \bar{Y})^2$$

para estabelecer o fato de que o conjunto centrado, nos dá uma repartição em 3 maneiras da variação total.

A provar exigida é a seguinte:

$$\Sigma Y^2 = \Sigma(Y - \bar{Y} + \bar{Y})^2$$
$$= \Sigma[(Y - \bar{Y}) + \bar{Y}]^2$$

$$= \Sigma[(Y - \bar{Y})^2 + 2\bar{Y}(Y - \bar{Y}) + \bar{Y}^2]$$
$$= \Sigma(Y - \bar{Y})^2 + \boxed{2\ \Sigma\bar{Y}\ (Y - \bar{Y})} + \Sigma\bar{Y}^2$$

Mas,a expressão dentro do retângulo achureado fica,

$$2\Sigma\bar{Y}\ (Y - \bar{Y}) = 2\Sigma(\bar{Y}Y - \bar{Y}^2)$$
$$= 2\ \bar{Y}\Sigma Y - 2\Sigma\bar{Y}^2 = 2\ \frac{(\Sigma Y)^2}{n} - 2n\frac{(\Sigma Y)^2}{n} =$$

$$= 2\ \frac{(\Sigma Y)^2}{n} - 2n\ \frac{(\Sigma Y)^2}{n^2} = 2\ \frac{(\Sigma Y)^2}{n} - 2\ \frac{(\Sigma Y)^2}{n} = 0$$

---

[1] $\Sigma\hat{y}^2$: soma de quadrados dos desvios estimados ou soma de quadrados de regressão ou soma de quadrados do modelo.

[2] Σy^2: soma de quadrados totais dos desvios da média

Desta forma,

$$\Sigma Y^2 = \Sigma(Y - \bar{Y})^2 + \Sigma\bar{Y}^2 = \Sigma\ (Y - \bar{Y})^2 + n\bar{Y}^2 =$$
$$= \Sigma\ (Y - \bar{Y})^2 + \frac{(\Sigma Y)^2}{n}$$

Então, o conjunto centrado de equações normais nos dá a seguinte repartição em 3 caminhos da variação total na variável-Y:

Fonte	Soma de Quadrados	G.L.
Média	$n\bar{Y}^2 = (\Sigma Y)^2/n$	1
Modelo	$\Sigma \hat{y}^2 = \sum_{j=1}^{\ell} bj \sum_{i=1}^{n} x_{ij} y_i$	$\neq b'_s - 1$
Resíduo	$\Sigma (y - \hat{y})^2$	$n - \neq b'_s - 1$
Total	ΣY^2	$n-1$

Obs.: $j = i-1$

6. DERIVAÇÃO DIRETA DO CONJUNTO CENTRADO DE EQUAÇÕES NORMAIS, EM VEZ DE SECUNDARIAMENTE COMO FEITO NO ITEM 5

Partindo-se do modelo,

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \ldots + b_mX_m$$

e nós desejamos minimizar $\Sigma(Y - \hat{Y})^2$

Sendo,

$$(Y - \hat{Y}) = Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m$$

$$\Sigma(Y - \hat{Y})^2 = \Sigma[Y - (b_0 + b_1X_1 + b_2X_2 + \ldots + b_mX_m)]^2$$

$$= \Sigma(Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2$$

$$\frac{\partial \Sigma(Y - \hat{Y})^2}{\partial b_0} = \frac{\partial \Sigma(Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial b_0}$$

$$= \Sigma \left[\frac{\partial (Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial b_0} \right]$$

$$= \Sigma\ [2(Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m)\ .\frac{\partial(Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m)}{\partial b_0}]$$

$$= \Sigma\ [2\ (Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m).\ (-1)]$$

$$= \Sigma\ [-2\ (Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m)]$$

$$= -2\ \Sigma[Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m]$$

$$= -2\ [\Sigma Y - b_0 - b_1\Sigma X_1 - b_2\Sigma X_2 - \ldots - b_m\Sigma X_m]$$

$$(\div 2) \qquad 0 = -2\ [\Sigma Y - b_0 - b_1\Sigma X_1 - b_2\Sigma X_2 - \ldots - b_m\Sigma X_m]$$

$$x\ (-1) \qquad 0 = -\ \Sigma Y + b_0 + b_1\Sigma X_1 + b_2\Sigma X_2 + \ldots + b_m\Sigma X_m$$

$$0 = \Sigma Y - b_0 - b_1\Sigma X_1 - b_2\Sigma X_2 - \ldots - b_m\Sigma X_m$$

$$\boxed{nb_0 + b_1\Sigma X_1 + b_2\Sigma X_2 + \ldots + b_m\Sigma X_m = \Sigma Y}$$

$$\frac{\partial\ \Sigma(Y - \hat{Y})^2}{\partial\ b_1} = \frac{\partial\ \Sigma(Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial\ b_1}$$

$$= \Sigma\ [\ \frac{\partial(Y - b_0 - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial\ b_1}\ ]$$

$$= \Sigma\ [2\ (Y-b_0-b_1X_1-b_2X_2-\ldots-b_mX_m)\ .\frac{\partial(Y-b_0-b_1X_1-b_2X_2-\ldots-b_mX_m)}{\partial\ b_1}]$$

$$= \Sigma\ [2\ (Y-b_0-b_1X_1-b_2X_2 - \ldots - b_mX_m).(-X_1)]$$

$$= \Sigma\ [-2(X_1Y-b_0X_1-b_1X_1^2-b_2X_1X_2 - \ldots - b_mX_mX_1)]$$

$$=-2\ [\Sigma X_1Y-b_0\Sigma X_1-b_1\Sigma X_1^2-b_2\Sigma X_1X_2-\ldots-b_m\Sigma X_mX_1]$$

$$(\div 2)\quad 0 = -2\ [\Sigma X_1 Y - b_0 \Sigma X_1 - b_1 \Sigma X_1^2 - b_2 \Sigma X_1 X_2 - \ldots - b_m \Sigma X_m X_1]$$

$$x(-1)\quad 0 = -\ \Sigma X_1 Y + b_0 \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + \ldots + b_m \Sigma X_m X_1]$$

$$0 = \Sigma\ X_1 Y - b_0 \Sigma X_1 - b_1 \Sigma X_1^2 - b_2 \Sigma X_1 X_2 - \ldots - b_m \Sigma X_m X_1]$$

$$\boxed{b_0 \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + \ldots + b_m \Sigma X_m X_1 = \Sigma X_1 Y}$$

$$\frac{\partial \Sigma (Y - \hat{Y})^2}{\partial b_2} = \frac{\partial \Sigma (Y - b_0 - b_1 X_1 - b_2 X_2 - \ldots - b_m X_m)^2}{\partial b_2}$$

$$= \Sigma\ [\ \frac{\partial (Y - b_0 - b_1 X_1 - b_2 X_2 - \ldots - b_m X_m)^2}{\partial b_2}]$$

$$= \Sigma\ [2(Y-b_0-b_1X_1-b_2X_2-\ldots-b_mX_m) . \frac{\partial (Y-b_0-b_1X_1-b_2X_2-\ldots-b_mX_m)}{\partial b_2}]$$

$$= \Sigma\ [2\ (Y-b_0-b_1X_1-b_2X_2-\ldots-b_mX_m).(-X_2)]$$

$$= \Sigma\ [-2(X_2Y-b_0X_2-b_1X_1X_2-b_2X_2^2-\ldots-b_mX_mX_2]$$

$$= -2\ [\Sigma X_2 Y - b_0 \Sigma X_2 - b_1 \Sigma X_1 X_2 - b_2 \Sigma X_2^2 - \ldots - b_m \Sigma X_m X_2]$$

$$(\div 2)\quad 0 = -2\ [\Sigma X_2 Y - b_0 \Sigma X_2 - b_1 \Sigma X_1 X_2 - b_2 \Sigma X_2^2 - \ldots - b_m \Sigma X_m X_2]$$

$$x(-1) \qquad 0 = -\Sigma X_2 Y + b_0 \Sigma X_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^2 + \ldots + b_m \Sigma X_m X_2$$

$$0 = \Sigma X_2 Y - b_0 \Sigma X_2 - b_1 \Sigma X_1 X_2 - b_2 \Sigma X_2^2 - \ldots - b_m \Sigma X_m X_2$$

$$\boxed{b_0 \Sigma X_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^2 + \ldots + b_m \Sigma X_m X_2 = \Sigma X_2 Y}$$

$$\frac{\partial \Sigma (Y - \hat{Y})^2}{\partial b_j} = \frac{\partial \Sigma (Y - b_0 - b_1 X_1 - b_2 X_2 - \ldots - b_m X_m)^2}{\partial b_j}$$

$$= \Sigma \left[\frac{\partial (Y - b_0 - b_1 X_1 - b_2 X_2 - \ldots - b_m X_m)^2}{\partial b_j} \right]$$

$$= \Sigma \left[2(Y - b_0 - b_1 X_1 - b_2 X_2 - \ldots - b_m X_m) . \frac{\partial (Y - b_0 - b_1 X_1 - b_2 X_2 - \ldots - b_m X_m)}{\partial b_j} \right]$$

$$= \Sigma [2 (Y - b_0 - b_1 X_1 - b_2 X_2 - \ldots - b_m X_m) . (-X_j)]$$

$$= \Sigma [-2X_j (Y - b_0 - b_1 X_1 - b_2 X_2 - \ldots - b_m X_m)]$$

$$= -2\Sigma [X_j Y - b_0 X_j - b_1 X_1 X_j - b_2 X_2 X_j - \ldots - b_m X_m X_j]$$

$$= -2 [\Sigma X_j Y - b_0 \Sigma X_j - b_1 \Sigma X_1 X_j - b_2 \Sigma X_2 X_j - \ldots - b_m \Sigma X_m X_j]$$

$$(\div 2) \qquad 0 = -2 [\Sigma X_j Y - b_0 \Sigma X_j - b_1 \Sigma X_1 X_j - b_2 \Sigma X_2 X_j - \ldots - b_m \Sigma X_m X_j]$$

$$x(-1) \qquad 0 = -\Sigma X_j Y + b_0 \Sigma X_j + b_1 \Sigma X_1 X_j + b_2 \Sigma X_2 X_j + \ldots + b_m \Sigma X_m X_j]$$

$$0 = \Sigma X_j Y - b_0 \Sigma X_j - b_1 \Sigma X_1 X_j - b_2 \Sigma X_2 X_j - \ldots - b_m \Sigma X_m X_j]$$

$$b_0 \Sigma X_j + b_1 \Sigma X_1 X_j + b_2 \Sigma X_2 X_j + \ldots + b_m \Sigma X_m X_j = \Sigma X_j Y$$

Substituindo $X_j = X_m$, nós temos:

$$\boxed{b_0 \Sigma X_m + b_1 \Sigma X_1 X_m + b_2 \Sigma X_2 X_m + \ldots + b_m \Sigma X_m^2 = \Sigma X_m Y}$$

Assim sendo, o nosso conjunto de equações normais fica:

$$nb_0 + b_1 \Sigma X_1 + b_2 \Sigma X_2 + \ldots + b_m \Sigma X_m = \Sigma Y$$

$$b_0 \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + \ldots + b_m \Sigma X_1 X_m = \Sigma X_1 Y$$

$$b_0 \Sigma X_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^2 + \ldots + b_m \Sigma X_2 X_m = \Sigma X_2 Y$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$b_0 \Sigma X_m + b_1 \Sigma X_1 X_m + b_2 \Sigma X_2 X_m + \ldots b_m \Sigma X_m^2 = \Sigma X_m Y$$

Quando o modelo contém o termo constante (b_0), é possível simplificar as equações normais e suas soluções. A simplificação aparece pelo fato de que a solução das equações normais nos dá uma estimativa de b_0.

Nós sabemos do nosso conjunto de equações normais que:

$$nb_0 + b_1 \Sigma X_1 + b_2 \Sigma X_2 + \ldots + b_m \Sigma X_m = \Sigma Y$$

$$nb_0 = \Sigma Y - b_1 \Sigma X_1 - b_2 \Sigma X_2 - \ldots - b_m \Sigma X_m$$

$$b_0 = \frac{\Sigma Y}{n} - b_1 \frac{\Sigma X_1}{n} - b_2 \frac{\Sigma X_2}{n} - \ldots - b_m \frac{\Sigma X_m}{n}$$

mas $\bar{Y} = \frac{\Sigma Y}{n}$ e $\bar{X}_i = \frac{\Sigma X_i}{n}$, então

$$\boxed{b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 - \ldots - b_m \bar{X}_m} \qquad (6.1)$$

Novamente, do nosso conjunto de equações normais, temos:

$$nb_0 + b_1 \Sigma X_1 + b_2 \Sigma X_2 + \ldots + b_m \Sigma X_m = \Sigma Y \qquad (6.2)$$

Substituindo a equação 6.1 na equação 6.2, obtém-se:

$$n(\bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2 - \ldots - b_m\bar{X}_m) + b_1\Sigma X_1 + b_2\Sigma X_2 + \ldots + b_m\Sigma X_m = \Sigma Y$$

$$(n\bar{Y} - b_1 n\bar{X}_1 - b_2 n\bar{X}_2 - \ldots - b_m n\bar{X}_m) + b_1\Sigma X_1 + b_2\Sigma X_2 + \ldots + b_m\Sigma X_m = \Sigma Y$$

$$b_1(\Sigma X_1 - n\bar{X}) + b_2(\Sigma X_2 - n\bar{X}_2) + \ldots + b_m(\Sigma X_m - n\bar{X}_m) = \Sigma Y - n\bar{Y}$$

mas $\bar{X}_1 = \frac{\Sigma X_1}{n}$; $\bar{X}_i = \frac{\Sigma X_i}{n}$ e $\bar{Y} = \frac{\Sigma Y}{n}$,

então

$$b_1 (\Sigma X_1 - \frac{\cancel{n} \Sigma X_1}{\cancel{n}}) + b_2 (\Sigma X_2 - \frac{\cancel{n} \Sigma X_2}{\cancel{n}}) + \ldots + b_m (\Sigma X_m - \frac{\cancel{n} \Sigma X_m}{\cancel{n}}) = \Sigma Y - \frac{\cancel{n} \Sigma Y}{\cancel{n}}$$

$$b_1(\Sigma X_1 - \Sigma X_1) + b_2 (\Sigma X_2 - \Sigma X_2) + \ldots + b_m (\Sigma X_m - \Sigma X_m) = \Sigma Y - \Sigma Y$$

$$b_1 (0) + b_2 (0) + \ldots + b_m (0) = 0$$

assim sendo, a primeira linha do conjunto de equações normais desaparece.

Tomando a segunda linha do conjunto de equações normais:

$$b_0 \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + \ldots + b_m \Sigma X_1 X_m = \Sigma X_1 Y \qquad (6.3)$$

Substituindo a equação 6.1 na equação 6.3, temos,

$$(\bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 - \ldots - b_m \bar{X}_m) \Sigma X_1 + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + \ldots + b_m \Sigma X_1 X_m = \Sigma X_1 Y$$

$$(\bar{Y} \Sigma X_1 - b_1 \bar{X}_1 \Sigma X_1 - b_2 \bar{X}_2 \Sigma X_1 - \ldots - b_m \bar{X}_m \Sigma X_1) + b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + \ldots + b_m \Sigma X_1 X_m = \Sigma X_1 Y$$

$$b_1(\Sigma X_1^2 - \bar{X}_1 \Sigma X_1) + b_2(\Sigma X_1 X_2 + \bar{X}_2 \Sigma X_1) + \ldots + b_m(\Sigma X_1 X_m - \bar{X}_m \Sigma X_1) = \Sigma X_1 Y - \bar{Y} \Sigma X_1$$

mas, $\bar{X}_1 = \frac{\Sigma X_1}{n}$; $\bar{X}_i = \frac{\Sigma X_i}{n}$ e $\bar{Y} = \frac{\Sigma Y}{n}$, então

$$b_1(\Sigma X_1^2 - \frac{(\Sigma X_1)(\Sigma X_1)}{n}) + b_2 (\Sigma X_1 X_2 - \frac{(\Sigma X_2)(\Sigma X_1)}{n}) + \ldots + b_m(\Sigma X_1 X_m - \frac{(\Sigma X_m)(\Sigma X_1)}{n}) =$$

$$= \Sigma X_1 Y - \frac{(\Sigma Y)(\Sigma X_1)}{n}$$

$$b_1(\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n}) + b_2 (\Sigma X_1 X_2 - \frac{(\Sigma X_2)(\Sigma X_1)}{n}) + \ldots + b_m(\Sigma X_1 X_m - \frac{(\Sigma X_m)(\Sigma X_1)}{n}) =$$

$$= \Sigma X_1 Y - \frac{(\Sigma Y)(\Sigma X_1)}{n}$$

mas, $\Sigma x_j x_k = \Sigma X_j X_k - \frac{(\Sigma X_j)(\Sigma X_k)}{n}$, então

$$\boxed{b_1 \; \Sigma x_1^2 + b_2 \; \Sigma x_1 x_2 + \ldots + b_m \; \Sigma x_1 x_m = \Sigma x_1 y}$$

O mesmo procedimento acima é aplicado para as outras partes do conjunto de equações normais.

Como percebe-se, a primeira coluna do conjunto de equações normais também desaparece, e o conjunto de equações para o modelo centrado será:

$$\begin{array}{l}
b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 + \ldots + b_m \Sigma x_1 x_m = \Sigma x_1 y \\
b_1 \Sigma x_1 x_2 + b_2 \Sigma x_2^2 + \ldots + b_m \Sigma x_2 x_m = \Sigma x_2 y \\
\quad \vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \qquad\qquad \vdots \\
b_1 \Sigma x_1 x_m + b_2 \Sigma x_2 x_m + \ldots + b_m \Sigma x_m^2 = \Sigma x_m y
\end{array}$$

com,

$$b_0 = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 - \ldots - b_m \bar{X}_m$$

7. DERIVAÇÃO DO CONJUNTO NÃO-CENTRADO DE EQUAÇÕES NORMAIS COMO UM CASO ESPECIAL DO CONJUNTO CENTRADO

Tomando-se o modelo:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_m X_m$$

Nós desejamos forçar a equação ajustada passar pela origem. Isto é obtido fazendo-se $b_0 = 0$ e o modelo não-centrado será:

$$\hat{Y} = b_1 X_1 + b_2 X_2 + \ldots + b_m X_m$$

e deseja-se minimizar

$$\Sigma \, (Y - \hat{Y})^2$$

então,

$$(Y - \hat{Y}) = Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m$$

$$\Sigma(Y - \hat{Y})^2 = \Sigma[Y - (b_1X_1 + b_2X_2 + \ldots + b_mX_m)]^2$$

$$= \Sigma(Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2$$

$$\frac{\partial \Sigma(Y - \hat{Y})^2}{\partial b_1} = \frac{\partial \Sigma(Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial b_1}$$

$$= \Sigma\left[\frac{\partial(Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial b_1}\right]$$

$$= \Sigma\left[2(Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m) . \frac{\partial(Y-b_1X_1-b_2X_2-\ldots-b_mX_m)}{\partial b_1}\right]$$

$$= \Sigma[2(Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m).(-X_1)]$$

$$= \Sigma[-2(X_1Y - b_1X_1^2-b_2X_1X_2 - \ldots - b_mX_mX_1)]$$

$$= -2[\Sigma X_1Y - b_1\Sigma X_1^2-b_2\Sigma X_1X_2 - \ldots - b_m\Sigma X_mX_1]$$

$$\div (2) \quad 0 = -2[\Sigma X_1Y - b_1\Sigma X_1^2 - b_2\Sigma X_1X_2-\ldots-b_m\Sigma X_mX_1]$$

$$x(-1) \quad 0 = -\Sigma X_1Y + b_1\Sigma X_1^2 + b_2\Sigma X_1X_2 + \ldots + b_m\Sigma X_mX_1$$

$$0 = \Sigma X_1Y - b_1\Sigma X_1^2 - b_2\Sigma X_1X_2 + \ldots + b_m\Sigma X_mX_1$$

$$\boxed{b_1\Sigma X_1^2 + b_2\Sigma X_1X_2 + \ldots + b_m\Sigma X_1X_m = \Sigma X_1Y}$$

$$\frac{\partial \Sigma(Y - \hat{Y})^2}{\partial b_2} = \frac{\partial \Sigma(Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial b_2}$$

$$= \Sigma\left[\frac{\partial(Y - b_1X_1 - b_2X_2 - \ldots - b_mX_m)^2}{\partial b_2}\right]$$

$$= \Sigma\left[2(Y - b_1X_1 - b_2X_2 - \dots - b_mX_m) . \frac{\partial(Y-b_1X_1-b_2X_2-\dots-b_mX_m)}{\partial b_2}\right]$$

$$= \Sigma[2(Y - b_1X_1 - b_2X_2 - \dots - b_mX_m).(-X_2)]$$
$$= \Sigma[-2(X_2Y - b_1X_1X_2 - b_2X_2^2 - \dots - b_mX_mX_2)]$$

$$= -2[\Sigma X_2Y - b_1\Sigma X_1X_2 - b_2\Sigma X_2^2 - \dots - b_m\Sigma X_mX_2]$$

$$\div(2)\;0 = -2[\Sigma X_2Y - b_1\Sigma X_1X_2 - b_2\Sigma X_2^2 - \dots - b_m\Sigma X_mX_2]$$

$$x(-1)\;0 = -\Sigma X_2Y + b_1\Sigma X_1X_2 + b_2\Sigma X_2^2 + \dots + b_m\Sigma X_mX_2$$

$$0 = \Sigma X_2Y - b_1\Sigma X_1X_2 - b_2\Sigma X_2^2 - \dots - b_m\Sigma X_mX_2$$

$$\boxed{b_1\Sigma X_1X_2 + b_2\Sigma X_2^2 + \dots + b_m\Sigma X_2X_m = \Sigma X_2Y}$$

$$\frac{\partial\Sigma(Y - \hat{Y})^2}{\partial b_j} = \frac{\partial\Sigma(Y - b_1X_1 - b_2X_2 - \dots - b_mX_m)^2}{\partial b_j}$$

$$= \Sigma\left[\frac{\partial(Y - b_1X_1 - b_2X_2 - \dots - b_mX_m)^2}{\partial b_j}\right]$$

$$= \Sigma\left[2(Y-b_1X_1-b_2X_2-\dots-b_mX_m) . \frac{\partial(Y-b_1X_1-b_2X_2-\dots-b_mX_m)}{\partial b_j}\right]$$

$$= \Sigma[2(Y - b_1X_1 - b_2X_2 - \dots - b_mX_m).(-X_j)]$$

$$= \Sigma[-2X_j(Y - b_1X_1 - b_2X_2 - \dots - b_mX_m)]$$

$$= -2\Sigma[X_jY - b_1X_1X_j - b_2X_2X_j - \dots - b_mX_mX_j]$$

$$= -2[\Sigma X_jY - b_1\Sigma X_1X_j - b_2\Sigma X_2X_j - \dots - b_m\Sigma X_mX_j]$$

$$\div (2) \quad 0 = -2[\Sigma X_j Y - b_1 \Sigma X_1 X_j - b_2 \Sigma X_2 X_j - \ldots - b_m \Sigma X_m X_j]$$

$$x(-1) \quad 0 = -\Sigma X_j Y + b_1 \Sigma X_1 X_j + b_2 \Sigma X_2 X_j + \ldots + b_m \Sigma X_m X_j$$

$$0 = \Sigma X_j Y - b_1 \Sigma X_1 X_j - b_2 \Sigma X_2 X_j - \ldots - b_m \Sigma X_m X_j$$

$$b_1 \Sigma X_1 X_j + b_2 \Sigma X_2 X_j + \ldots + b_m \Sigma X_m X_j = \Sigma X_j Y$$

Substituindo $X_j = X_m$, nós temos:

$$\boxed{b_1 \ \Sigma X_1 X_m + b_2 \Sigma X_2 X_m + \ldots + b_m \Sigma X_m^2 = \Sigma X_m Y}$$

Assim, o conjunto final de equações normais para o modelo não-centrado será:

$$b_1 \Sigma X_1^2 + b_2 \Sigma X_1 X_2 + \ldots + b_m \Sigma X_1 X_m = \Sigma X_1 Y$$

$$b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2^2 + \ldots + b_m \Sigma X_2 X_m = \Sigma X_2 Y$$

$$\vdots$$

$$b_1 \Sigma X_1 X_m + b_2 \Sigma X_2 X_m + \ldots + b_m \Sigma X_m^2 = \Sigma X_m Y$$

Nós podemos verificar que esta solução é a mínima soma de quadrados, examinando as segundas derivadas parciais.

A primeira derivada parcial i^{th} é:

$$\frac{\partial SQ}{\partial b_i} = \Sigma \ [2(Y - b_1 X_1 - b_2 X_2 - \ldots - b_m X_m).(-X_i)]$$

$$= \Sigma[-2(X_i Y - b_1 X_1 X_i - b_2 X_2 X_i - \ldots - b_i X_i^2 - \ldots - b_m X_i X_m)]$$

$$= -2 \ [\Sigma X_i Y - b_1 \Sigma X_1 X_i - b_2 \Sigma X_2 X_i - \ldots - b_i \Sigma X_i^2 - \ldots - b_m \Sigma X_i X_m]$$

e a segunda derivada parcial com respeito a b_i é

$$\frac{\partial^2 SQ}{\partial b_i^2} = \Sigma[-2\{-(X_i^2)\} \ ] = \Sigma[2X_i^2] = 2\Sigma X_i^2$$

que é sempre positiva. Sendo que as segundas ordens parciais em todas as dimensões são sempre positivas, nós temos que ter de fato um único "mínimo".

8. DIVISAO DA SOMA DE QUADRADOS TOTAIS EM PORÇÕES EXPLICADA E RESIDUAL

Uma maneira de avaliar a adequabilidade de um modelo ajustado é determinar a quantidade de variação que é explicada pelo modelo. Como medida de variação, nós iremos usar a soma de quadrados. O nosso interesse, então, reside nas seguintes fontes de variação em Y:

$$\text{SQ Total} = \Sigma\, Y^2$$

$$\text{SQ Modelo} = \Sigma\, \hat{Y}^2$$

$$\text{SQ Resíduo} = \Sigma\, (Y - \hat{Y})^2$$

Seria conveniente se as duas últimas fontes de variação fossem aditivas e totalizassem para a primeira, isto é,

$$\Sigma Y^2 = \Sigma\hat{Y}^2 + \Sigma(Y - \hat{Y})^2 \qquad (8.1)$$

Se isto fosse o caso, nós poderíamos determinar a variação residual em subtraindo a variação explicada pelo modelo da variação total.

Para isto vamos provar que as fontes de variação são aditivas, com respeito a fórmula acima, para o conjunto de mínimos quadrados dos "b's".

No processo, nós iremos obter um método conveniente de calcular a $\Sigma\hat{Y}^2$. Nós começamos com a expressão ΣY^2, na qual nós somamos e subtraímos $\hat{Y}$:

$$\Sigma Y^2 = \Sigma(Y - \hat{Y} + \hat{Y})^2$$

$$= \Sigma[(Y - \hat{Y}) + \hat{Y})]^2$$

$$= \Sigma\ [(Y - \hat{Y})^2 + 2\hat{Y}(Y - \hat{Y}) + \hat{Y}^2]$$

Mudando o termos $\hat{Y}^2$ para o começo, nós temos:

$$= \Sigma[\hat{Y}^2 + 2\hat{Y}(Y - \hat{Y}) + (Y - \hat{Y})^2]$$

Passando o sinal de somatório para dentro:

$$= \underbrace{\Sigma\hat{Y}^2}_{1^{\circ}} + \underbrace{2\Sigma\hat{Y}(Y - \hat{Y})}_{2^{\circ}} + \underbrace{\Sigma(Y - \hat{Y})^2}_{3^{\circ}}$$

O primeiro e o último termo, são aqueles que nós desejamos reter e, então, permanece para nós mostrarmos que o segundo termo é zero.

Sendo que,

$$2\Sigma\hat{Y}\ (Y - \hat{Y}) = 0$$

somente se,

$$\Sigma\hat{Y}\ (Y - \hat{Y}) = 0$$

Nós podemos eliminar o multiplicador 2, em que se segue. Então,

$$\Sigma\hat{Y}\ (Y - \hat{Y}) = \Sigma\ (\hat{Y}Y - \hat{Y}^2) = \Sigma\hat{Y}Y - \Sigma\hat{Y}^2$$

Substituindo $\hat{Y} = b_1X_1 + b_2X_2 + \dots + b_mX_m$,

nós obtemos

$$\Sigma\hat{Y}Y - \Sigma\hat{Y}^2 = \Sigma Y(b_1X_1 + b_2X_2 + \dots + b_mX_m) - \Sigma(b_1X_1 + b_2X_2 + \dots + b_mX_m)^2$$

$$= \Sigma(b_1X_1Y+b_2X_2Y+\dots+b_mX_mY)-\Sigma(b_1X_1+b_2X_2+\dots+b_mX_m)^2$$

$$= \underbrace{(b_1\Sigma X_1Y+b_2\Sigma X_2Y+\dots+b_m\Sigma X_mY)}_{1}-\underbrace{\Sigma(b_1X_1+b_2X_2+\dots+b_mX_m)^2}_{2} \quad (8.2)$$

Para completar a prova, nós precisamos somente mostrar que os dois termo acima são idênticos.

Vamos então expandir o termo no lado direito (2) da equação 8.2:

$$\Sigma(b_1X_1+b_2X_2+\dots+b_mX_m)^2=\Sigma(b_1X_1+b_2X_2+\dots+b_mX_m).(b_1X_1+b_2X_2+\dots+b_mX_m)$$

$$=\Sigma[b_1X_1(b_1X_1+b_2X_2+\dots+b_mX_m)+b_2X_2(b_1X_1+b_2X_2+\dots+b_mX_m)+\dots+$$

$$+b_mX_m(b_1X_1+b_2X_2+\dots+b_mX_m)]$$

$$=\Sigma[(b_1^2X_1^2+b_1b_2X_1X_2+\dots+b_1b_mX_1X_m)+(b_1b_2X_1X_2+b_2^2X_2^2+\dots+b_2b_mX_2X_m)+\dots+$$

$$+(b_1b_mX_1X_m+b_2b_mX_2X_m+\dots+b_m^2X_m^2)]$$

$$=\Sigma(b_1^2X_1^2+b_1b_2X_1X_2+\dots+b_1b_mX_1X_m)+\Sigma(b_1b_2X_1X_2+b_2^2X_2^2+\dots+b_2b_mX_2X_m)+\dots+$$

$$+\Sigma(b_1b_mX_1X_m+b_2b_mX_2X_m+\dots+b_m^2X_m^2)$$

$$=(b_1^2\Sigma X_1^2+b_1b_2\Sigma X_1X_2+\ldots+b_1b_m\Sigma X_1X_m)+(b_1b_2\Sigma X_1X_2+b_2^2\Sigma X_2^2+\ldots+$$

$$+b_2b_m\Sigma X_2X_m)+\ldots+(b_1b_m\Sigma X_1X_m+b_2b_m\Sigma X_2X_m+\ldots+b_m^2\Sigma X_m^2)$$

$$=\underbrace{b_1(b_1\Sigma X_1^2+b_2\Sigma X_1X_2+\ldots+b_m\Sigma X_1X_m)}_{1}+\underbrace{b_2(b_1\Sigma X_1X_2+b_2\Sigma X_2^2+\ldots+b_m\Sigma X_2X_m)}_{2}+\ldots+$$

$$\underbrace{+b_m(b_1\Sigma X_1X_m+b_2\Sigma X_2X_m+\ldots+b_m\Sigma X_m^2)}_{3}$$

Mas, as expressões nos parêntesis (1, 2 e 3) são simplesmente os lados direitos das equações normais, isto é:

$$b_1\Sigma X_1^2 + b_2\Sigma X_1X_2 +\ldots+ b_mX_1X_m = \Sigma X_1Y$$

$$b_1\Sigma X_1X_2 + b_2\Sigma X_2^2 + \ldots + b_m\Sigma X_2X_m = \Sigma X_2Y$$

$$\vdots$$

$$b_1\Sigma X_1X_m + b_2\Sigma X_2X_m + \ldots + b_m\Sigma X_m^2 = \Sigma X_mY$$

Sendo assim

$$= b_1\Sigma X_1Y + b_2\Sigma X_2Y + \ldots + b_m\Sigma X_mY \quad (8.3)$$

Desta forma, substituindo a equação 8.3 no item 2 da equação 8.2, nós temos:

$$\Sigma\hat{Y}(Y - \hat{Y})=(b_1\Sigma X_1Y+b_2\Sigma X_2Y +\ldots+ b_m\Sigma X_mY)-(b_1\Sigma X_1Y+b_2\Sigma X_2Y+\ldots+b_m\Sigma X_mY)$$

$$\Sigma\hat{Y}\ (Y - \hat{Y}) = 0,$$

e a prova está completa.

Conseqüentemente, nós comprovamos pela equação 8.1 que as fontes de variação são aditivas.

No processo de desenvolvimento desta prova, nós também mostramos que:

$$^3\Sigma\hat{Y}^2 = b_1\Sigma X_1Y + b_2\Sigma X_2Y + \ldots + b_m\Sigma X_mY$$

Isto diz que a porção da soma de quadrados total "explicada" pelo modelo, pode ser calculada multiplicando-se cada coeficiente "b" pelo lado -Y de sua equação normal correspondente e somando os resultados assim obtidos. Então, nós repartimos a variação em Y (como medido pela soma de quadrados) como segue:

TOTAL = MODELO + RESÍDUO

$$\Sigma Y^2 = \Sigma \hat{Y}^2 + \Sigma(Y - \hat{Y})^2$$

ou

$$\Sigma(Y - \hat{Y})^2 = \Sigma Y^2 - \Sigma \hat{Y}^2$$

com

$$\Sigma \hat{Y}^2 = b_1 \Sigma X_1 Y + b_2 \Sigma X_2 Y + \ldots + b_m \Sigma X_m Y$$

[3]Isto só é válido para a solução dos mínimos quadrados.

9. RESUMO

Neste trabalho é apresentada a derivação do conjunto simultâneo de equações normais pelo método dos mínimos quadrados.

Este conjunto de equações normais pode ou não ser centrado para a sua média e o método para se obter estes dois conjuntos é demonstrado matematicamente.

10. LITERATURA CONSULTADA

1. Draper, N. e Smith, H. Applied Regression Analysis. John Wiley and Sons, Inc. New York, 1966.
2. Freese, F. Linear Regression Methods for Forest Research. U.S. Department of Agriculture. Forest Service. Forest Products Laboratory Research Paper FPL 17. Madison, Wis., 1964.
3. Graybill, F.A. Theory and Applications of the Linear Model. Duxbury Press. Mass., 1967.