

O ENSINO DE NOÇÕES BÁSICAS DE ESTATÍSTICA ATRAVÉS DA GEOMETRIA N-DIMENSIONAL

CLEUSA ROCHA ASANOME^a
Dra. MARIA APARECIDA PERRE^a
MARIE-CLAIRE RIBEIRO PÓLA^b

RESUMO

Nosso propósito é desenvolver um método geométrico que permita que técnicas estatísticas básicas sejam apresentadas de forma rigorosa matematicamente mas de fácil compreensão. Neste trabalho apresentaremos apenas a média e a variância sob um enfoque geométrico.

PALAVRAS-CHAVE: *Projeção ortogonal; Geometria; Média; Variância.*

1 - INTRODUÇÃO

O enfoque geométrico da Estatística, tem como principal benefício o fato de que através de poucos princípios básicos é possível desenvolver os conceitos estatísticos principalmente em modelos lineares. Isto é evidenciado em trabalhos científicos de FISHER (1915), BARTLETT (1933-1934), DURBIN e KENDALL (1951), KRUSKAL (1961), ZYSKIND (1967) e WATSON (1967). Mais recentemente, trabalho como o de BOX, HUNTER e HUNTER (1978) que aplicam a geometria vetorial em seus cursos de planejamento de experimentos; MARGOLIS (1979) que apresenta vários exemplos do uso da Geometria Vetorial na derivação de resultados estatísticos elementares e defende a Geometria para a apresentação de conceitos básicos de Estatística; HERR (1980) revisa trabalhos de Estatística com inclinação geométrica; BRYANT (1984) dá um esquema de um método para ensinar estatística usando a Geometria Elementar de Vetores, e SAVILLE e WOOD (1986) fazem duas experiências de ensino de Estatística usando a Geometria, sendo uma na Universidade de Canterbury na Nova Zelândia, e outra na Universidade da Califórnia, em Davids, nos EUA.

SAVILLE e WOOD (1986) afirmaram que os cursos foram bem recebidos e os resultados obtidos garantem um substancial melhoramento sobre o tradicional "livro de receitas".

Neste trabalho apresentar-se-á um método geométrico para a média e a variância, dando subsídios para desenvolver técnicas básicas de Estatística, tais como a Análise de Variância em delincamentos experimentais.

As idéias básicas de Geometria serão apresentadas e o seu uso será ilustrado através de exemplos gráficos.

O método consiste em trabalhar com dois objetos: o vetor de observação e o espaço do modelo e dois processos: projeção do vetor de observação no espaço do modelo e Teorema de Pitágoras.

2 - MÉDIA

De posse de um conjunto de dados buscamos através deles obter informações com o objetivo de obter inferências sobre o fenômeno em estudo. Para tal devemos encontrar a melhor forma de resumí-los e de, assim, extrair o desejado.

Utilizaremos para descrever o fenômeno o modelo linear

$$Y = X\beta + e, \quad (2.1)$$

onde

$Y = [Y_1 \ Y_2 \ \dots \ Y_n]^T$ é o vetor $n \times 1$ de observações,

$X = [1 \ 1 \ \dots \ 1]^T$ é o vetor $n \times 1$ de uns,
 β representa o vetor de parâmetros populacional,

$e = [e_1 \ e_2 \ \dots \ e_n]^T$ é o vetor $n \times 1$ dos erros independentes, com $e \sim N(0, \sigma^2)$.

Esse modelo pode ser relacionado ao modelo

$$Y = \mu + e \text{ ou}$$

$$Y = \mu_Y + (Y - \mu_Y) \quad (2.2)$$

a. Departamento de Matemática Aplicada/CCE — Universidade Estadual de Londrina

b. Departamento de Matemática/CCE — Universidade Estadual de Londrina

onde

μ_Y é a média populacional
 $(Y - \mu_Y)$ é o desvio de Y em relação à sua média.

Nosso objetivo é obter uma estimativa para Y e, portanto, uma estimativa do parâmetro β .

Para tal, utilizaremos o Método dos Mínimos Quadrados, que consiste em determinar valores para β , tais que

$$S = \| Y - X \hat{\beta} \|^2 \text{ seja mínima.}$$

Geometricamente, então, o Método dos Mínimos Quadrados para o caso diz que devemos considerar todos os possíveis modelos $X\beta$ e escolher o mais próximo a Y, como podemos ver na Figura 2.1.

Usando produto interno Euclidiano, "Mínimos Quadrados" significa menor distância:

$$0 \quad Y \quad Y_1 = X\hat{\beta} \quad Y_1 \quad Y_1^* \quad L(X) = L(1,1,\dots,1)$$

$$e = Y - X\hat{\beta}$$

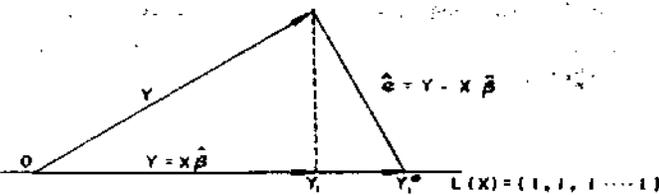


Figura 2.1 – Método de Mínimos Quadrados

Y_1 = projeção ortogonal de Y em $L(X)$

Y_1^* = qualquer outra projeção de Y em $L(X)$.

Conforme MARGOLIS (1979) o ponto mais próximo de Y é a sua projeção ortogonal Y_1 no espaço L determinado por X.

E conseqüentemente temos que :

$$Y_1 = P_X Y \tag{2.3}$$

Logo $Y_1 = Y\hat{\beta}$ para algum $\hat{\beta}$, e, pela definição de ortogonalidade (MARGOLIS, 1979) se $(Y - P_X Y) \perp X$ então

$$\langle Y - P_X Y, X \rangle = 0,$$

$$\langle Y - X\hat{\beta}, X \rangle = 0.$$

E aplicando as propriedades de produto interno na expressão acima temos

$$\langle Y, X \rangle - \hat{\beta} \langle X, X \rangle = 0 \tag{2.5}$$

Como $\langle Y, X \rangle = \sum_{i=1}^n Y_i$ e $\langle X, X \rangle = n$ então (2.5) fica.

$$\sum_{i=1}^n Y_i - \hat{\beta} n = 0,$$

$$\hat{\beta} = \frac{\sum_{i=1}^n Y_i}{n} = \bar{Y}, \tag{2.6}$$

que é a média aritmética de Y.

A média amostral $X\hat{\beta} = \bar{Y}$ é a constante que resume a amostra. A solução está ilustrada na Figura 2.2 e através dela verificamos que Y está decomposto em dois componentes ortogonais:

$PY = X\hat{\beta}$ que pertence ao espaço L_1 , gerado por X, que denominaremos espaço de estimação;

$e = Y - X\hat{\beta}$, que são os resíduos e pertencem ao espaço L_2 . Interpretaremos L_2 como subespaço resíduo, ortogonal ao espaço de estimação. Também pode ser dito que e é a projeção ortogonal complementar de Y sobre o espaço resíduo ao longo do espaço de estimação.

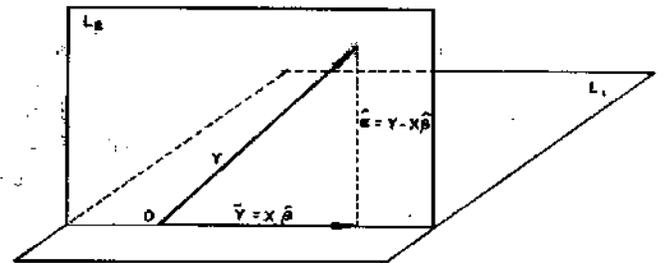


Figura 2.2 — Média amostral como projeção do vetor Y

Conforme a Figura 2.2 podemos então dizer que a projeção ortogonal do vetor de observações Y, situado no espaço amostral, sobre o espaço de estimação L_1 transforma Y linearmente, ou seja, o vetor Y é a soma de dois componentes:

$$Y = PY + (Y - PY) \text{ ou}$$

$$Y = PY + (Y - P)Y \tag{2.7}$$

$$Y = X\hat{B} + e$$

O operador P remove ou anula o vetor erro, componente de Y e simultaneamente determina a posição única de $X\hat{B}$ em L_1 .

Vamos especificar a forma geral desse operador de projeção ortogonal P.

Como vimos em (2.4)

$$\begin{aligned} \langle Y, X \rangle - \hat{\beta} \langle X, X \rangle &= 0 \text{ ou} \\ X'Y &= X'X\hat{\beta} \end{aligned} \quad (2.8)$$

Se $X'X$ tem posto completo então

$$\hat{\beta} = (X'X)^{-1}X'Y, \quad (2.9)$$

e como $PY = X\hat{\beta}$ então

$$P = X(X'X)^{-1}X' \quad (2.10)$$

Conforme foi visto na Figura 2.2, se todos os Y 's fossem idênticos, Y deveria estar exatamente em $X\hat{\beta}$ e então $\|\hat{e}\|^2 = 0$. Assim podemos pensar em $X\hat{\beta}$ como sendo a parte de Y que pode ser satisfatoriamente explicada por uma constante. Quando isso acontece, $\|\hat{\beta}\|^2 > 0$ e essa explicação é imperfeita porque os dados variam. É preciso determinar uma medida dessa variação e é o que faremos a seguir.

3 - VARIÂNCIA

Como já dissemos, determinar simplesmente a média amostral não basta. É preciso saber qual a variabilidade existente entre os dados ou, em outras palavras, precisamos saber quão próximo está Y de $X\hat{\beta}$.

Através de MARCOLIS (1979) vemos que $\|\hat{e}\|^2$ pode ser usado para isso.

Geometricamente $\|\hat{e}\|^2 = \|Y - \bar{Y}\|^2$, a soma dos mínimos quadrados dos erros é o quadrado do comprimento do vetor erro amostral é no subespaço resíduo. Em outras palavras, é o quadrado da distância entre os vetores Y e $X\hat{\beta}$ sobre a linha definida por $L_1(X)$.

Podemos interpretar a variância amostral de Y , em termos geométricos, como a média do quadrado das normas do vetor erro estimado, \hat{e} .

Para calcular isto devemos dividir a Soma de Quadrados do Resíduo (SQR) pela dimensão do subespaço relevante, obtendo assim o Quadrado Médio do Resíduo (QMR) por dimensão.

Desde que \hat{e} está restrito a um subespaço $(n-1)$ dimensional ortogonal a $L_1(X)$, dividimos a SQR por $n-1$, obtendo:

$$QMR = \frac{SQR}{n-1} = \frac{\|\hat{e}\|^2}{n-1} = \frac{\sum(Y_i - \bar{Y})^2}{n-1} = S^2, \quad (3.1)$$

que é a usual variância amostral.

A dimensão do subespaço no qual o vetor está restrito é chamado de Graus de Liberdade associado àquele componente.

Podemos desenvolver uma fórmula para a variância amostral usando produto interno.

Seja:

$$\hat{e} = Y - X\hat{\beta}$$

a soma de quadrados dos erros ou resíduos (SQR) forma o numerador da variância amostral.

Através da definição da norma de um vetor temos que

$$\|\hat{e}\|^2 = \|Y - X\hat{\beta}\|^2 \quad (3.2)$$

$$= \langle Y - X\hat{\beta}, Y - X\hat{\beta} \rangle$$

$$= \langle Y, Y \rangle - 2 \langle Y, X\hat{\beta} \rangle + \langle X\hat{\beta}, X\hat{\beta} \rangle.$$

$$\begin{aligned} \|\hat{e}\|^2 &= \sum_{i=1}^n y_i^2 - \frac{2 \left(\sum_{i=1}^n y_i\right)^2}{n} + \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \\ \|\hat{e}\|^2 &= \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \end{aligned}$$

A variância amostral ou o quadrado médio do resíduo \hat{e} , substituindo (3.2) em (2.7)

$$S^2 = QMR = \frac{\|\hat{e}\|^2}{n-1} = \frac{1}{n-1} \left[\sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i\right)^2}{n} \right] \quad (3.3)$$

A raiz quadrada positiva da Variância Amostral é chamada de DESVIO PADRÃO. Geometricamente, é uma medida da distância de cada valor Y_i de Y até Y .

É a norma de \hat{e}

$$S = \sqrt{\frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1}}$$

Em particular, determinamos assim um estimador $\bar{Y} = X\hat{\beta}$ que minimiza a soma de quadrados dos desvios dada em (2.6) e a sua variância, que fornece uma estimativa da variabilidade dos dados, dada por (3.3).

4 - EXEMPLO NUMÉRICO EM R^2

a) Cálculo da média (ou obtenção de $\hat{\beta}$).

$$\text{Seja } Y = \begin{bmatrix} 8 \\ 2 \end{bmatrix}$$

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

$$\bar{Y} = X(X'X)^{-1}X'Y$$

$$\bar{Y} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 \\ 8 & 2 \end{bmatrix}$$

$$\bar{Y} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} = X\hat{\beta} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad 5$$

b) Representação geométrica.

A Figura 4.1 mostra a representação geométrica dos vetores Y , \bar{Y} e $\hat{e} = Y - \bar{Y}$.

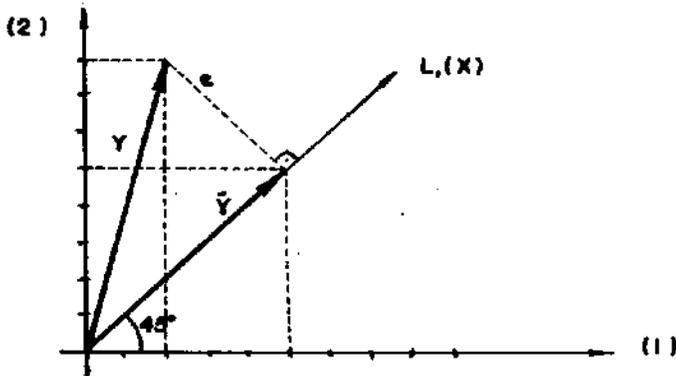


Figura 4.1 - Representação geométrica de Y , \bar{Y} e \hat{e} em R^2

Verifica-se que o espaço de estimação $L_1(X)$ é uma reta que faz 45° com os eixos cartesianos.

c) Comprimento de vetores.

O comprimento de vetores pode ser obtido calculando-se a norma dos mesmos.

$$\|Y\| = \sqrt{2^2 + 8^2} = 8,25$$

$$\|\bar{Y}\| = \sqrt{5^2 + 5^2} = 7,07$$

$$\hat{e} = Y - \bar{Y} = \begin{bmatrix} 2 \\ 8 \end{bmatrix} - \begin{bmatrix} 5 \\ 5 \end{bmatrix} = \begin{bmatrix} -3 \\ 3 \end{bmatrix}$$

$$\|\hat{e}\| = \sqrt{(-3)^2 + (3)^2} = 4,24$$

Esses valores correspondem aos do gráfico da Figura 4.1.

d) Verificação da ortogonalidade entre os espaços de estimação e resíduo.

Base de $L_1(X)$: $u = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$

Base de $L_2(X)$: $v = \left\{ \begin{bmatrix} -3 \\ 3 \end{bmatrix} \right\}$

$$\langle u, v \rangle = u^T v = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} -3 \\ 3 \end{bmatrix} = 0$$

Como o produto interno dos vetores u e v é nulo, os espaços de estimação e resíduo são ortogonais. Podemos mostrar que $\bar{Y} \perp \hat{e}$ calculando-se o produto interno de \bar{Y} e \hat{e} .

$$\langle \bar{Y}, \hat{e} \rangle = \bar{Y}^T \hat{e} = \begin{bmatrix} 5 & 5 \end{bmatrix} \begin{bmatrix} -3 \\ 3 \end{bmatrix}$$

$$\langle \bar{Y}, \hat{e} \rangle = 0$$

e) Variância e Desvio Padrão

$$s^2 = \frac{\sum (Y_i - \bar{Y})^2}{n - 1} = \frac{\|\hat{e}\|^2}{n - 1} = \frac{18}{n - 1}$$

$$s^2 = 18 = s = \sqrt{18} = 4,24$$

No caso especial em que estamos analisando um exemplo do R^2 , o desvio padrão é exatamente igual ao comprimento do vetor erro que está representado na figura 4.1.

f) Demonstração gráfica de que $\hat{\beta}$ é o estimador de mínimos quadrados.

Partindo do modelo $Y = \mu + e$ onde $\hat{\mu} = X\hat{\beta}$ podemos construir um gráfico fazendo variar os valores de $\hat{\beta}$ e obter os correspondentes valores para $\|\hat{e}\|^2$, então para:

$X\hat{\beta}$	$\hat{e} = Y - X\hat{\beta}$	$\ \hat{e}\ ^2$
[1 1]	[1 7]	50
[2 2]	[0 6]	36
[3 3]	[-1 5]	26
[4 4]	[-2 4]	20
[5 5]	[-3 3]	18
[6 6]	[-4 2]	20
[7 7]	[-5 1]	26
[8 8]	[-6 0]	36
[9 9]	[-7 -1]	50

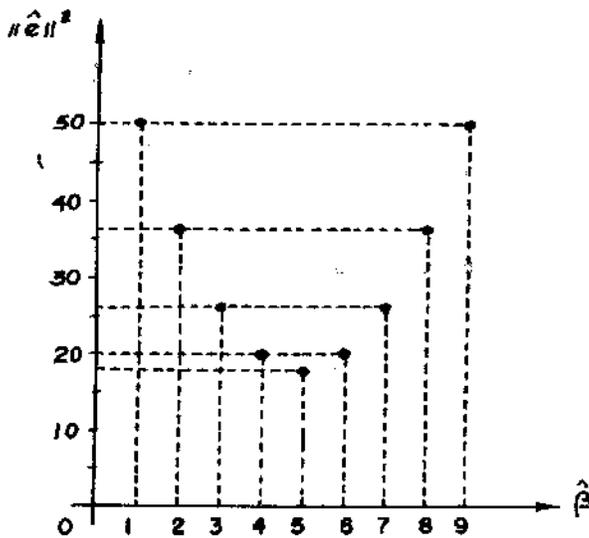
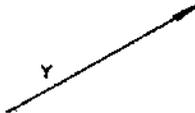
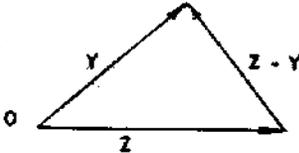
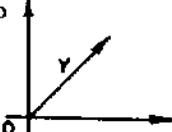
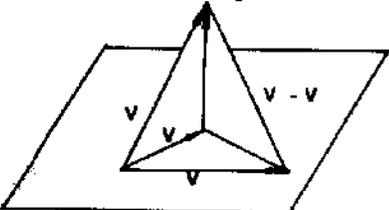


Figura 4.2 -
Representação
gráfica de \hat{e}^2
em função de $\hat{\beta}$

Observa-se que a representação gráfica desta variação nos dá uma parábola, cujo mínimo é $\hat{\beta} = 5$ que corresponde a menor Soma de Quadrados de Resíduos ou $\|\hat{e}\|^2 = 18$.

5 - SUMÁRIO

O quadro a seguir associa as idéias algébricas e geométricas à expressão estatística correspondente.

Idéia Algébrica	Idéia Geométrica	Expressão Estatística
1) Vetor Y	Vetor Y	Dados
$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$		$y_1, y_2, y_3, \dots, y_n$
2) Vetor Z - Y	Vetor diferença	Desvios, erro ou resíduo
$Z - Y = \begin{bmatrix} z_1 - y_1 \\ z_2 - y_2 \\ \vdots \\ z_n - y_n \end{bmatrix}$		$z_1 - y_1, z_2 - y_2, \dots, z_n - y_n$
3) Produto Interno de X e Y		Soma de produtos cruzados
$\langle X, Y \rangle = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$		Ex.: $\langle X, Y \rangle = \sum_{i=1}^n x_i y_i$
4) Norma ao quadrado		
a) $\ y\ ^2 = \langle X, Y \rangle$	a) Quadrado da distância de Y para zero	a) Soma de quadrados
		$\langle Y, Y \rangle = \sum_{i=1}^n y_i^2$
b) $\ Z - Y\ ^2 = \langle (Z - Y), (Z - Y) \rangle$	b) Quadrado da distância de Y para Z.	b) Soma de quadrados de resíduos (SQR) $\ \hat{e}\ ^2 = \sum (Z - Y)^2$
5) Todos vetores da forma $v = ax + by + cz + \dots$	Espaço vetorial $L(x, y, z, \dots)$ através de 0 determinado pelos vetores.	Todos modelos de forma $ax + by + cz + \dots$
6) Um vetor v em $L(v_1, v_2, \dots, v_n)$ tal que $\langle v_1 - v_2, v_3 \rangle = 0$ para todo v_3 em L.		O modelo mais adequado da forma $ax + by + cz$ (linear)

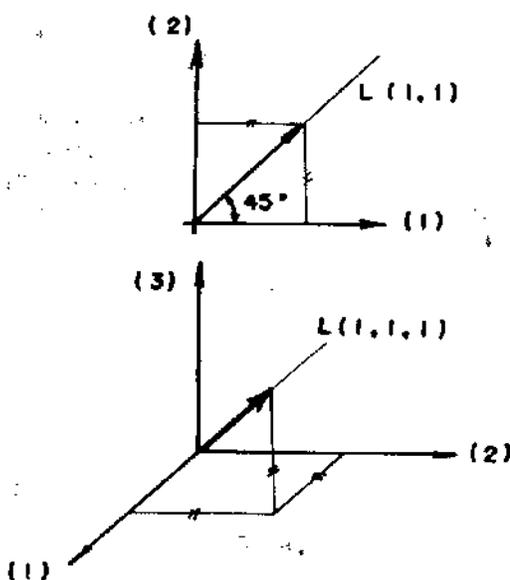
7) Vetor unitário

$$X = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Linha equiangular aos eixos coordenados $L(1, 1, \dots, 1)$
 EXEMPLO EM R^2 E R^3

Constante

$$X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$



ABSTRACT

Our objectives is to develop a geometric method wich allows basic statistic. Procedures to be presented in a mathematically rigorous way and at the same time to be easy to understand. In this paper we will present only the mean and the variance in a geometric point of view.

KEY-WORDS: Orthogonal projections; Geometric; Mean; Variance.

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 - BARTLETT, M. S. The Vector Representation of Sample. *Proceedings of the Cambridge Philosophical Society*, 30: 327-340, 1933.
- 2 - BRYANT, P. Geometry, Statistics, Probability: Variations on a Common Theme. *The American Statistician*, 40 (3): 205-214, 1984.
- 3 - BOX, G. E. P.; HUNTER, W. G. & HUNTER, J. S. *Statistics for Experimenters An Introduction to Design, Data Analysis, and Model Building*. New York, J. Wiley & Sons, 1978.
- 4 - DURBIN, J. & KENDALL, M. G. The Geometry of Estimation. *Biometrika*, 38 : 150-158, 1951.
- 5 - FISHER, R. A. Frequency Distribution of the Values of the Correlation Coefficient in Samples From an Indefinitely Large Population. *Biometrika*, 10 : 507 - 521, 1915.
- 6 - HERR, D. G. A Geometric Characterization of Connectedness in a Two-Way Design. *Biometrika*, 63 : 93-100, 1976.
- 7 - HERR, D. G. On the History of the Use of Geometry in the General Linear Model. *The American Statistician*, 34(1) : 34-47, 1980.
- 8 - KRUSKAL, W. H. The Coordinate-free Approach to Gauss-Markov Estimation and Its Application to Missing and Extra Observation. *4th Berkeley Symposium on Mathematical Statistics and Probability*, 1 : 435-451, 1961.
- 9 - MARGOLIS, M. S. Perpendicular Projections and Elementary Statistics. *The American Statistician*, 33 (3) : 131-135, 1979.
- 10 - SAVILLE, D. J. & WOOD, G. R. A Method for Teaching Statics Using N-Dimensional Geometry. *The American Statistician*, 40 (3) : 205-214, 1986.
- 11 - ZISKIND, G. On canonical Forms, Non-Negative Covariance Matrices and Best and Simple Least Linear Estimators in Linear Models. *Annals of Mathematical Statistics*, 38: 1092-1109, 1967.
- 12 - WATSON, G. S. Linear Least Squares Regression. *Annals of Mathematical Statistics*, 38: 1679-1699, 1967.

Recebido para publicação em 30/10/89