

## ANÁLISE ESTATÍSTICA DO ERRO PADRÃO DA MÉDIA PONDERADA NA INTERPOLAÇÃO DE ANOMALIAS GRAVIMÉTRICAS

NELSI CÔGO DE SÁ E DENIZAR BLITZKOW

*Instituto Astronômico e Geofísico – USP  
Caixa Postal 30627  
C.P. 30627, 01051 – São Paulo, SP, Brasil*

The weighted mean is a mathematical model widely used in two-dimensional interpolation. A statistical analysis of the standard errors produced by this model when interpolating gravity anomalies is the subject of this work. The sampling was carried out by random search on a population of gravity stations. Subsequently the Bouguer anomalies of the sampled points were interpolated. The differences between the observed and the interpolated anomalies constitute a random sample for which it is postulated the normal probability density function. The statistical validity of this postulate is submitted to parametric and non-parametric hypotheses testing. The results show that there is no systematic errors in the sample and that the variance depends strongly on the number of points used in the interpolation.

### INTRODUÇÃO

Um conjunto de dados gravimétricos resultante da medição normalmente apresenta uma distribuição geográfica irregular em seus elementos. Mas, por razões de ordem prática, nas aplicações do método gravimétrico é preferível uma distribuição regular, já que esta:

- permite a identificação e o acesso direto a qualquer elemento,
- simplifica a representação do campo no plano e no espaço, e
- torna-se necessária na utilização de técnicas numéricas de análise e interpretação.

A obtenção de um conjunto com essa distribuição regular, a partir do original, consiste na interpolação dos seus elementos.

Por constituir um modelo matemático relativamente simples, a média ponderada é muito utilizada nessa interpolação. Pouco se sabe, entretanto, sobre o comportamento estatístico do erro padrão produzido por esse modelo.

Uma análise estatística dos erros padrões associados aos resíduos, obtidos na comparação de anomalias interpoladas com as originais, constitui o objetivo deste trabalho.

### MODELO MATEMÁTICO

A média ponderada, como modelo determinístico, foi sugerida por Bjerhammar (1973) e tem sido usada na interpolação de alturas geoidais (Kaersley, 1977; Balmino et al., 1979), altitudes topográficas (Davis, 1973) e anomalias gravimétricas (Kassim, 1980; Sünkel, 1980).

A interpolação com este modelo consiste em calcular as anomalias  $\hat{\Delta}g_i$ , nos pontos  $P_i$  ( $i = 1, \dots, n$ ), utilizando as anomalias conhecidas  $\Delta g_j$ , nos pontos mais próximos  $P_j$  ( $j = 1, \dots, m$ ), com pesos proporcionais às funções das distâncias ( $d_{ij}$ ) entre  $P_i$  e  $P_j$  (Fig.).

Simbolicamente:

$$\hat{\Delta}g_i = \frac{\sum_{j=1}^m p_{ij} \Delta g_j}{\sum_{j=1}^m p_{ij}} \quad (1)$$

com os pesos dados por

$$p_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2]^{-\frac{\nu}{2}} \quad (2)$$

O modelo expresso em (1) envolve duas constantes arbitrárias:  $m$ , que representa o número de anomalias conhecidas utilizadas na interpolação de cada ponto, e  $\nu$ , que é o expoente vinculado aos pesos.

A variância associada à anomalia interpolada ( $\sigma^2_{\hat{\Delta}g_i}$ ) provém de duas fontes independentes, isto é, da representação ( $\sigma^2_{r_i}$ ) e da observação ( $\sigma^2_{o_i}$ ). Então, usando a lei de propagação das variâncias, resulta

$$\sigma^2_{\hat{\Delta}g_i} = \sigma^2_{r_i} + \sigma^2_{o_i} \quad (3)$$

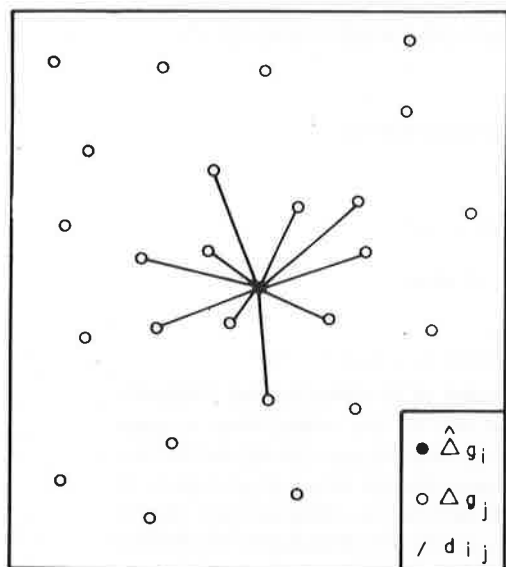


Figura 1 — Esquema de interpolação.

A variância da representação é expressa por

$$\sigma_{r_i}^2 = \frac{\sum_{j=1}^m p_{ij} (\Delta g_j - \hat{\Delta}g_i)^2}{(m-1) \sum_{j=1}^m p_{ij}} \quad (4)$$

a variância da observação é obtida com a propagação das variâncias dos valores observados

$$\sigma_{o_i}^2 = \left( \frac{\partial \hat{\Delta}g_i}{\partial \Delta g_j} \right)^2 \sigma_{\Delta g_j}^2$$

que, com (1), toma a forma

$$\sigma_{o_i}^2 = \frac{\sum_{j=1}^m p_{ij}^2 \sigma_{\Delta g_j}^2}{\left( \sum_{j=1}^m p_{ij} \right)^2} \quad (5)$$

com

$$\sigma_{\Delta g_j}^2 = \sigma_{g_j}^2 + (c \sigma_{h_j})^2, \quad (6)$$

onde  $\sigma_{g_h}^2$  e  $\sigma_{h_j}^2$  são as variâncias associadas à gravidade observada e à altitude respectivamente. O fator  $c$  depende do tipo de anomalia. Para a anomalia Bouguer ( $\rho = 2,67 \text{ g cm}^{-3}$ )

$$c_b = \frac{\partial g}{\partial h} - 2\pi k \rho \approx 0,1967 \text{ mGal m}^{-1},$$

onde  $g$  é a gravidade observada,  $k$  é a constante gravitacional e  $\rho$  é a densidade adotada.

## AMOSTRAGEM

Um conjunto com cerca de 750 estações gravimétricas, distribuídas nas regiões da Grande São Paulo (Sá & Blitzkow, 1984) e Baixada Santista, constitui a população para a análise estatística (Fig. 2).

Essas regiões têm características topográficas distintas. O relevo dominante na primeira é o planalto e na segunda é o litoral. Mas ambas chegam à Serra do Mar, onde a variação em altitude ultrapassa 1000m. A variação em anomalia Bouguer, incluindo as duas regiões, é de aproximadamente 50 mGal.

Para amostragem, as estações gravimétricas são armazenadas em disco de computador e, por busca aleatória, é obtida uma amostra com 150 registros, contendo os seguintes dados: coordenadas da estação ( $x_i, y_i$ ), anomalia Bouguer ( $\Delta g_i$ ) e variância ( $\sigma_{\Delta g_i}^2$ ). A anomalia interpolada ( $\hat{\Delta}g_i$ ) é obtida com a expressão (1) e a respectiva variância ( $\sigma_{\hat{\Delta}g_i}^2$ ) com (3), usando os dados da mesma população (Fig. 2). O resíduo é expresso por

$$v_i = \Delta g_i - \hat{\Delta}g_i \quad (7)$$

e a propagação das variâncias proporciona o erro padrão.

$$\hat{\sigma}_{v_i} = \sqrt{\sigma_{\Delta g_i}^2 + \sigma_{\hat{\Delta}g_i}^2} \quad (8)$$

Portanto, o conjunto dos resíduos

$$v = v_1, v_2, \dots, v_n \quad (9)$$

aos quais correspondem os erros padrões

$$\hat{\sigma}_v = \{\hat{\sigma}_{v_1}, \hat{\sigma}_{v_2}, \dots, \hat{\sigma}_{v_n}\}, \quad (10)$$

representa a amostra para a análise estatística neste trabalho.

Para uma interpolação perfeita, a população dos resíduos tem média conhecida, isto é,  $\mu_v = 0$ . Na prática, entretanto, ela representa uma população para a qual se postula a função densidade de probabilidade normal. Assim, a função que se postula para a amostra também é a normal  $N(\mu_v, s_v^2)$ , onde  $\mu_v$  é a média da população e  $s_v^2$  é a variância da amostra. A validade estatística desse postulado, entretanto, deve ser testada e, para isso, é necessário calcular os parâmetros amostrais. A média tem a expressão

$$\bar{v} = \frac{1}{n} \sum_{i=1}^n v_i \quad (11)$$

e, com esta, se calcula a variância

$$s_v^2 = \frac{1}{(n-1)} \sum_{i=1}^n (v_i - \bar{v})^2, \quad (12)$$

onde  $n$  é o número de elementos da amostra.

## TESTES ESTATÍSTICOS

Os testes estatísticos permitem verificar se:

- a função densidade de probabilidade postulada para a amostra é apropriada,
- os valores estimados dos parâmetros populacionais são confiáveis, e
- os valores estimados desses parâmetros são consistentes com seus valores conhecidos (a priori) se estes forem disponíveis.

Uma sentença quantitativa sobre a função densidade de probabilidade postulada e um parâmetro populacional constituem uma hipótese estatística. Para um valor particular desse parâmetro tem-se a hipótese nula ( $H_0$ ) à qual se associa a hipótese alternativa ( $H_a$ ), tal que  $H_a \neq H_0$ .

## Teste para rejeição de resíduos anômalos

Este teste examina separadamente os elementos da amostra para identificar e excluir aqueles que são estatisticamente incompatíveis com os demais.

A função densidade de probabilidade postulada para  $v$  é a normal  $N(\mu_v, s_v^2)$ . Então, usa-se a hipótese nula

$$H_0 : v_i \in N(\mu_0, s_v^2) \quad (\mu_0 = 0)$$

O teste consiste em calcular a estatística (Vaníček & Kravský, 1982).

$$t = \frac{v - \mu_0}{s_v} \quad (13)$$

e compará-la com a distribuição de Student, com  $n-1$  graus de liberdade e ao nível de significância  $\alpha$ . O intervalo de confiança é expresso por

$$\mu_0 - s_v \cdot t_{(n-1, 1 - \frac{\alpha}{2})} < v_i < \mu_0 + s_v \cdot t_{(n-1, 1 - \frac{\alpha}{2})}, \quad (14)$$

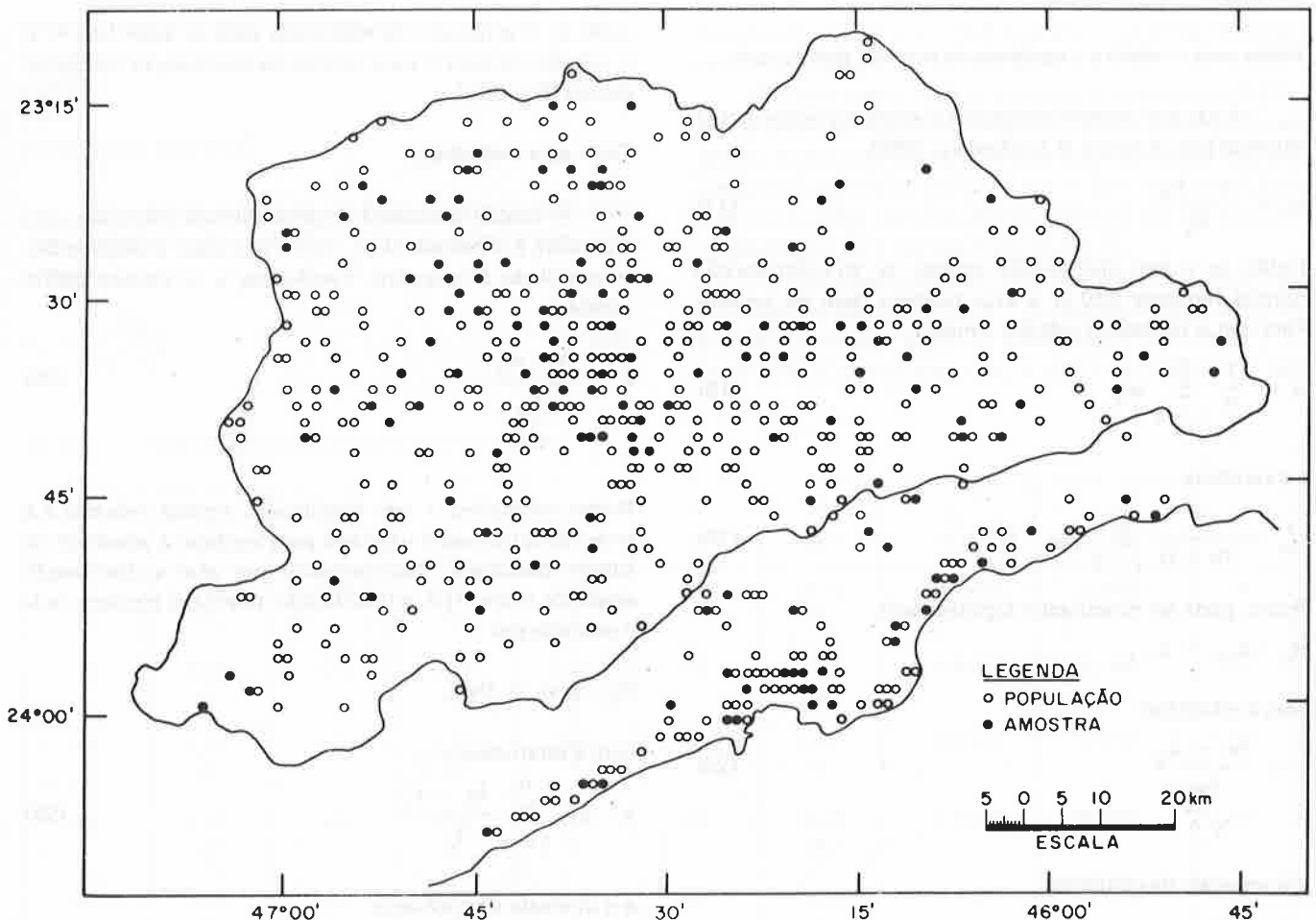


Figura 2 — Distribuição geográfica das estações.

onde  $c$  é um fator de escala. A rejeição de  $H_0$  indica continuação do teste.

### Teste para a média da amostra original

O teste da média permite verificar a existência de erros sistemáticos na amostra. Neste caso, a hipótese nula é expressa por (Mikhail & Gracie, 1981).

$$H_0 : \mu_v = \mu_0$$

e usa-se a estatística

$$t = \frac{\bar{v} - \mu_0}{\frac{s_v}{\sqrt{n}}} \quad (15)$$

com o intervalo de confiança

$$\mu_0 - \frac{s_v}{\sqrt{n}} t_{(n-1, 1 - \frac{\alpha}{2})} < \bar{v} < \mu_0 + \frac{s_v}{\sqrt{n}} t_{(n-1, 1 - \frac{\alpha}{2})} \quad (16)$$

Portanto, a rejeição de  $H_0$  ao nível de significância  $\alpha$ , significa que  $\bar{v}$  é estatisticamente diferente de  $\mu_v$ .

### Testes para a média e a variância da amostra padronizada

A variável aleatória associada à amostra padronizada é expressa por (Vanícek & Krakiwsky, 1982)

$$w = \frac{v - \mu_0}{\hat{\sigma}_v} \quad (17)$$

Então, se  $v$  tem distribuição normal,  $w$  terá distribuição normal reduzida  $N(0,1)$  e esta também deve ser testada. Para isso, é necessário calcular a média

$$\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i \quad (18)$$

e a variância

$$s_w^2 = \frac{1}{(n-1)} \sum_{i=1}^n (w_i - \bar{w})^2 \quad (19)$$

Assim, pode ser formulada a hipótese nula

$$H_0 : \mu_w = \mu_0,$$

com a estatística

$$t = \frac{\bar{w} - \mu_0}{\frac{s_w}{\sqrt{n}}} \quad (20)$$

e o intervalo de confiança

$$\mu_0 - \frac{s_w}{\sqrt{n}} t_{(n-1, 1 - \frac{\alpha}{2})} < \bar{w} < \mu_0 + \frac{s_w}{\sqrt{n}} t_{(n-1, 1 - \frac{\alpha}{2})}$$

$$+ \frac{s_w}{\sqrt{n}} t_{(n-1, 1 - \frac{\alpha}{2})} \quad (21)$$

No teste da hipótese nula para a variância, resulta

$$H_0 : \sigma_w^2 = \sigma_0^2 \quad (\sigma_0^2 = 1)$$

na qual se usa a estatística

$$\chi^2 = \frac{n s_w^2}{\sigma_0^2} \quad (22)$$

com o intervalo de confiança

$$\frac{\chi^2_{(n-1, \frac{\alpha}{2})} \sigma_0^2}{n} < s_w^2 < \frac{\sigma_0^2 \chi^2_{(n-1, 1 - \frac{\alpha}{2})}}{n} \quad (23)$$

onde  $\chi^2$  é a distribuição qui-quadrado.

Um teste alternativo para a variância consiste em calcular o percentual

$$P = \frac{n_1}{n} 100, \quad (24)$$

onde  $n_1$  é o número de elementos para os quais  $|w| \leq 1$ , e compará-lo com o valor teórico da distribuição normal reduzida ( $P = 68\%$ ).

### Teste para a aderência

A função densidade de probabilidade postulada para a amostra é a normal  $N(\mu_v, s_v^2)$ . Para testar a validade desse postulado é necessário transformar  $v$  na variável padronizada

$$z = \frac{v - \mu_0}{\frac{s_w}{\sqrt{n}}} \quad (25)$$

Nestas condições,  $z$  tem distribuição normal reduzida e o teste qui-quadrado é utilizado para verificar a aderência da função postulada. Representando por  $g(v)$  a distribuição empírica e por  $f(v)$  a distribuição teórica, a hipótese nula é expressa por

$$H_0 : g(v) = f(v),$$

com a estatística

$$\chi^2 = \sum_{i=1}^n \frac{(g_i - f_i)^2}{f_i} \quad (26)$$

e o intervalo de confiança

$$0 < \chi^2 < \chi^2_{(k-1, 1 - \alpha)} \quad (27)$$

onde  $g_i$  e  $f_i$  são as freqüências observada e esperada respectivamente, e  $k$  é o número de classes.

#### Teste para a curtose

A curtose representa o grau de achatamento de uma distribuição usualmente em relação à normal. O coeficiente de curtose, para  $n < 200$ , é expresso por (Snedecor & Cochran, 1973)

$$\alpha_4 = \frac{\sum_{i=1}^n |v_i - \bar{v}|}{n \sqrt{m_k}} \quad (28)$$

onde  $m_k$  é o momento de ordem  $k$  em relação à média e tem a expressão

$$m_k = \frac{\sum_{i=1}^n (v_i - \bar{v})^k}{n} \quad (29)$$

sendo  $n$  o número de elementos da amostra.

Segundo o valor numérico do coeficiente de curtose, uma distribuição pode ser: mesocúrtica, se  $\alpha_4 = 0,7979$ ; leptocúrtica, se  $\alpha_4 < 0,7979$ ; ou platicúrtica, se  $\alpha_4 > 0,7979$ .

#### Teste para a assimetria

Uma distribuição pode ainda apresentar assimetria em relação à média. O coeficiente que a caracteriza tem a expressão

$$\alpha_3 = \frac{m_3}{\sqrt{m_2^3}} \quad (30)$$

Tabela 1 — Parâmetros e testes estatísticos para  $\nu = 1,5$ .

$m$	$\bar{v}$ Teste	$s_v$	$\bar{w}$ Teste	$s_w^2$ Teste	$P$ 68	$\chi^2$ Teste	$\alpha_3$ 0,0	$\alpha_4$ 0,7979	$T$ (s)
4	0,17 A	1,69	0,12 A	2,96 R	53	15,72 R	0,1176	0,7322	11
8	0,19 A	1,76	0,16 A	3,31 R	45	7,83 A	-0,0479	0,7455	13
12	0,22 A	1,77	0,21 A	3,90 R	43	9,41 A	-0,0397	0,7382	16
16	0,18 A	1,88	0,18 A	4,96 R	42	12,65 A	-0,2039	0,7241	20
20	0,18 A	1,91	0,21 A	5,29 R	40	14,96 R	-0,1987	0,7248	25

A = aceita  $H_0$  e R = rejeita  $H_0$  para  $n = 150$  e  $\alpha = 0,05$ .

Neste caso, a distribuição pode ser: simétrica, se  $\alpha_3 = 0$ ; assimétrica à esquerda, de  $\alpha_3 < 0$ ; ou assimétrica à direita, se  $\alpha_3 > 0$ .

Os valores de  $\alpha_3$  e  $\alpha_4$ , para  $n < 200$ , em geral são pouco representativos.

## RESULTADOS

O modelo matemático expresso em (1) envolve duas constantes arbitrárias, isto é, o número de anomalias conhecidas, utilizadas na interpolação de cada elemento da amostra ( $m$ ), e o expoente vinculado aos pesos ( $\nu$ ). A análise estatística, neste caso, tem dois objetivos: indicar os valores numéricos de tais constantes e verificar se os resultados obtidos com esses valores são estatisticamente confiáveis.

Para diversos valores de  $m$  e  $\nu$ , utilizados na interpolação dos elementos da mesma amostra o teste para a aderência ao nível de significância de 5%, indica um valor inicial para o expoente  $\nu \cong 1,5$ . Com este, é obtida a Tabela 1, incluindo diversos valores de  $m$ .

O mesmo teste, aplicado ao resultado da Tabela 1, indica o valor numérico  $m \cong 10$ , que é utilizado na obtenção da Tabela 2, para verificar o valor inicial de  $\nu$ .

A Fig. 3 mostra os valores observados e o teórico da distribuição  $\chi^2$  para  $\nu = 1,5$  (Tabela 1), e a Fig. 4 mostra esses valores para  $m = 10$  (Tabela 2).

Os parâmetros obtidos e os testes estatísticos realizados ao nível de significância de 5%, que correspondem aos valores numéricos estatisticamente mais adequados para as constantes arbitrárias ( $m = 10$  e  $\nu = 1,5$ ), estão na Tabela 3. O histograma das freqüências relativas dos resíduos padronizados ( $z$ ) está representado na Fig. 5.

O processamento foi realizado com o computador Burroughs B-6900 do Centro de Computação Eletrônica da USP. O tempo gasto na amostragem, ordenação e interpolação ( $T$ ), em segundos, é dado na Tabela 1 e está

Tabela 2 — Parâmetros e testes estatísticos para  $m = 10$ .

$\nu$	$\bar{v}$ Teste	$s_v$	$\bar{w}$ Teste	$s_w^2$ Teste	P 68	$\chi^2$ Teste	$\alpha_3$ 0,0	$\alpha_4$ 0,7979
1	0,16 A	1,90	0,13 A	4,39 R	43	6,54 A	-0,1712	0,7386
2	0,19 A	1,74	0,17 A	3,72 R	43	5,98 A	-0,0252	0,7323
3	0,19 A	1,73	0,15 A	4,17 R	48	14,92 R	-0,0001	0,7317
4	0,18 A	1,84	0,15 A	4,59 R	48	22,81 R	0,0667	0,7286
5	0,18 A	1,76	0,17 A	4,89 R	45	20,26 R	0,1425	0,7264

A = aceita  $H_0$  e R = rejeita  $H_0$  para  $n = 150$  e  $\alpha = 0,05$ .

Tabela 3 — Parâmetros e testes estatísticos para  $\nu = 1,5$  e  $m = 10$ .

$\bar{v}$ Teste	$s_v$	$\bar{w}$ Teste	$s_w^2$ Teste	P 68	$\chi^2$ Teste	$\alpha_3$ 0,0	$\alpha_4$ 0,7979
0,20 A	1,75	0,18 A	3,59 R	43	6,04 A	-0,0148	0,7387

A = aceita  $H_0$  e R = rejeita  $H_0$  para  $n = 150$  e  $\alpha = 0,05$ .

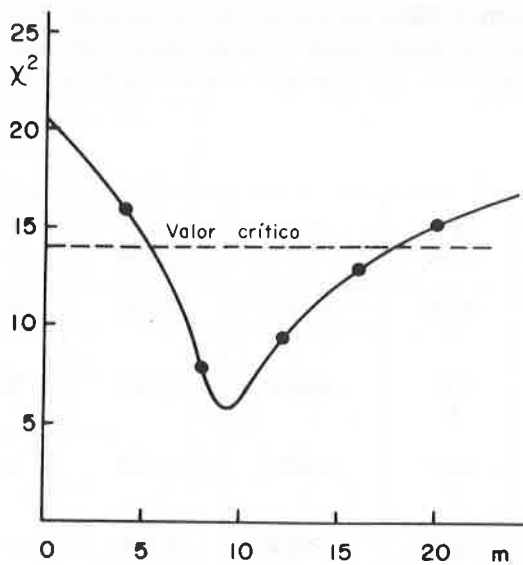


Figura 3 — Valores de  $\chi^2$  para  $\nu = 1,5$ .

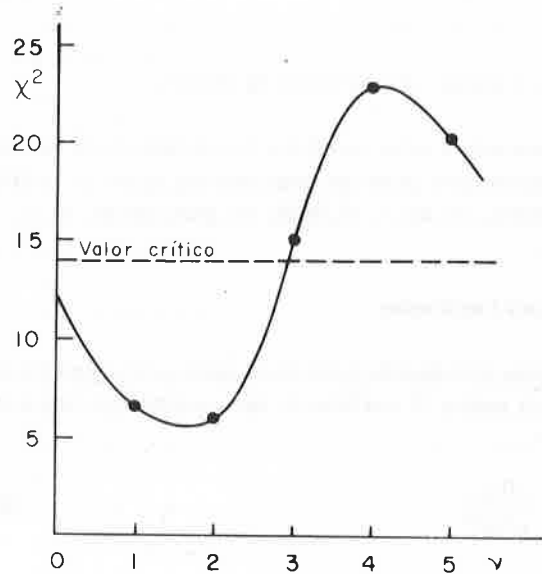


Figura 4 — Valores de  $\chi^2$  para  $m = 10$ .

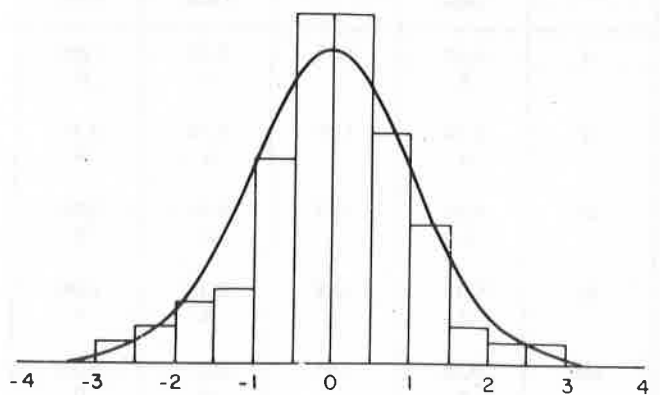


Figura 5 — Histograma dos resíduos padronizados (z).

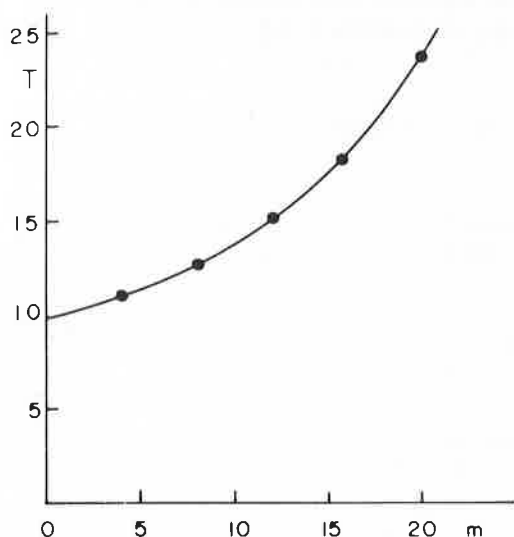


Figura 6 — Tempo de processamento.

representado na Fig. 6, onde o tempo que corresponde à interpolação se distingue dos demais.

## CONCLUSÕES

A análise estatística dos resíduos, obtidos com diversos valores pré-fixados para as constantes arbitrárias, indica os valores numéricos destas para os quais o modelo matemático é estatisticamente adequado.

O teste da hipótese nula para a média (Tabela 1), mostra que não há erro sistemático significativo nos resíduos originais ( $\mu_v = 0$ ) nem nos resíduos padronizados ( $\mu_w = 0$ ).

A rejeição da hipótese nula no teste da variância ( $\sigma_w^2 > 1$ ) indica que o cálculo do erro padrão é otimista, isto é, os valores obtidos para  $\hat{\sigma}_v$  são inferiores aos esperados, o que se observa também no teste alternativo pelo percentual (Tabela 1). As causas mais prováveis são: o efeito topográfico, cuja correção não está incluída na anomalia Bouguer, e a deficiência na distribuição das estações, existentes na Serra do Mar.

Quando  $m$  aumenta, a variância da amostra padronizada ( $s_w^2$ ) também aumenta e isto ocorre porque a variância da representação ( $\sigma_r^2$ ) diminui.

A função densidade de probabilidade da amostra é normal para os seguintes intervalos aproximados das constantes arbitrárias:  $5 < m < 18$  (Fig. 3) e  $0 < \nu < 3$  (Fig. 4), nos quais o valor mínimo de  $\chi^2$  ocorre para  $m \doteq 10$  e  $\nu \doteq 1,5$ . No exterior desses intervalos das constantes, a amostra tem distribuição assimétrica e leptocúrtica (Tabelas 1 e 2).

## REFERÊNCIAS

- BALMINO, G., BROSSIER, C., CASNAVE, A., NOUEL, F., DOMINH, K. & VALES, N. — 1979 — Geoid of the Kerguelen Island area determined from GEOS-3 altimeter data. *J. Geophys. Res.*, **84**: 3827-3831.
- BJERHAMMAR, A. — 1973 — Theory of errors and generalized matrix inverses. New York, Elsevier, 420 pp.
- DAVIS, J.C. — 1973 — Statistics and data analysis in Geology. New York, John Wiley & Sons, 550 pp.
- KAERSLEY, A.H.W. — 1977 — The prediction and mapping of geoidal undulations from GEOS-3 altimetry. Report No. 267, The Ohio State University, Columbus, 20 pp.
- KASSIM, F.A. — 1980 — An evaluation of three techniques for the prediction of gravity anomalies in Canada. Technical Report No. 73, University of New Brunswick, Fredericton, 82 pp.
- MIKHAIL, E.M. & GRACIE, G. — 1981 — Analysis and adjustment of survey measurements. New York, Van Nostrand, 340 pp.
- SÁ, N.C. & BLITZKOW, D. — 1984 — Levantamento gravimétrico na Região da Grande São Paulo. XXXIII Congresso Brasileiro de Geologia, Resumos, Rio de Janeiro, RJ, 338 pp.
- SNEDECOR, G.W. & COCHRAN, W.G. — 1973 — Statistical methods. Iowa, The Iowa State University Press, 593 pp.
- SUNKEL, H. — 1980 — A general surface representation module designed for Geodesy. Report No. 292, The Ohio State University, Columbus, 158 pp.
- VANÍČEK, P. & KRAKIWSKY, E. — 1982 — Geodesy: The Concepts. Amsterdam, North-Holland, 691 pp.