

JAIR MENDES MARQUES

O MÉTODO DA ANÁLISE DE COMPONENTES PRINCIPAIS  
NA DETECÇÃO E IDENTIFICAÇÃO DE OUTLIERS  
MÚLTIPLOS EM FOTOTRIANGULAÇÃO

Tese apresentada ao Curso de Pós-  
Graduação em Ciências Geodésicas  
da Universidade Federal do Paraná,  
para obtenção do Grau de Doutor  
em Ciências.

Orientador: prof. Dr. Quintino Dalmolin  
Co-orientador: prof. Dr. Anselmo Chaves Neto

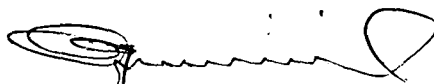
CURITIBA  
1994

O MÉTODO DA ANÁLISE DE COMPONENTES PRINCIPAIS  
NA DETECÇÃO E IDENTIFICAÇÃO DE *OUTLIERS* MÚLTIPLOS  
EM FOTOTRIANGULAÇÃO

por


JAIR MENDES MARQUES

Tese aprovada como requisito parcial  
para a obtenção do grau de Doutor em  
Ciências Geodésicas da Universidade  
Federal do Paraná, pela Banca Exami-  
nadora formada pelos professores:




---

Prof. Dr. QUINTINO DALMOLIN - Presidente - UFPR



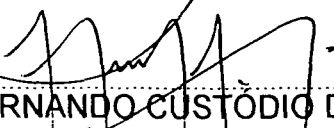
---

Prof. Dr. ANSELMO CHAVES NETO - UFPR



---

Prof. Dr. CAMIL GEMAEL - UFPR



---

Prof. Dr. JOÃO FERNANDO CUSTÓDIO DA SILVA - UNESP



---

Prof. Dr. FRANCISCO HUMBERTO SIMÕES MAGRO - UFRS

Curitiba, 25 de novembro de 1994.

## DEDICATÓRIA

Aos meus pais

Julio e Zenaide (*in memoriam*)

À minha esposa

Erly

E meus filhos

Marcos e Jaqueline

## AGRADECIMENTOS

O autor externa seus sinceros agradecimentos

- ao prof. Dr. Quintino Dalmolin pela orientação e incentivo, sem o qual esse trabalho dificilmente seria realizado;

- ao prof. Dr. Anselmo Chaves Neto, co-orientador e colega de vários anos de UFPr, pelas críticas e sugestões;

- ao prof. Dr. Camil Gemael, pela competência e dedicação que sempre demonstrou no decorrer do curso;

- ao prof. Dr. Milton de Azevedo Campos, pela dedicação e boa qualidade da sua gestão na coordenação do Curso de Pós-Graduação em Ciências Geodésicas;

- ao prof. Dr. José Bittencourt de Andrade, pela competência demonstrada nas disciplinas que lecionou;

- ao prof. Jorge Festa;

- à profa. Eva Dalmolin;

- à Bel. em Biblioteconomia Maria Eliane Stroparo, pela revisão nas referências bibliográficas;

- aos colegas e funcionários do Departamento de Estatística;

- aos colegas e funcionários do Curso de Pós-Graduação em Ciências Geodésicas;

- ao meu filho Marcos Augusto, pela ajuda na digitação deste trabalho;

- a todos que direta ou indiretamente contribuíram para a realização deste trabalho.

## SUMÁRIO

<b>TERMO DE APROVAÇÃO .....</b>	<b>II</b>
<b>DEDICATÓRIA .....</b>	<b>III</b>
<b>AGRADECIMENTOS.....</b>	<b>IV</b>
<b>SUMÁRIO .....</b>	<b>V</b>
<b>LISTA DOS PRINCIPAIS SÍMBOLOS.....</b>	<b>VIII</b>
<b>RESUMO .....</b>	<b>IX</b>
<b>ABSTRACT .....</b>	<b>X</b>
<b>1 INTRODUÇÃO.....</b>	<b>1</b>
1.1 Considerações gerais .....	1
1.2 A proposta do trabalho.....	2
1.3 Objetivos do trabalho .....	3
1.4 Desenvolvimento do trabalho .....	3
<b>2 ANÁLISE DE COMPONENTES PRINCIPAIS.....</b>	<b>5</b>
2.1 Introdução .....	5
2.2 Componentes principais populacionais .....	6
2.3 ACP para dados normais multivariados.....	12
2.4 Componentes principais obtidas de variáveis padronizadas.....	15
2.5 Componentes principais de matrizes covariâncias especiais.....	18
2.6 Componentes principais amostrais.....	21
2.7 Distribuições assintóticas para dados normais.....	27

2.8	Testes de hipóteses acerca das componentes principais .....	28
<b>3</b>	<b>FUNDAMENTOS TEÓRICOS DA ANÁLISE DE COMPONENTES PRINCIPAIS NA DETECÇÃO E IDENTIFICAÇÃO DE <i>OUTLIERS</i>.....</b>	<b>30</b>
3.1	Introdução .....	30
3.2	O modelo de Gauss-Markov .....	33
3.3	Ajustamento pelo método dos mínimos quadrados .....	35
3.4	Componentes principais dos resíduos .....	37
3.5	Resíduos padronizados .....	38
3.6	Componentes principais dos resíduos padronizados .....	41
3.7	Componentes principais padronizadas .....	42
3.8	A estatística qui-quadrado e a detecção de outliers .....	45
3.9	A análise da matriz G .....	49
3.10	O critério de rotação ortomax generalizado .....	51
3.11	Teste de discordância de Ferguson .....	58
3.12	Estimação dos vícios .....	59
3.13	Erro mínimo detectável .....	61
<b>4</b>	<b>APLICAÇÃO DO MÉTODO MULTIVARIADO EM FOTOTRIANGULAÇÃO.....</b>	<b>64</b>
4.1	Modelo matemático funcional .....	64
4.2	Modelo matemático funcional linearizado .....	64
4.3	Ajustamento livre .....	65
4.4	Componentes principais padronizadas .....	66
4.5	Estimação dos vícios das observações suspeitas .....	66
4.6	Teste de discordância de Ferguson .....	67
4.7	Aplicação do teste t de Student .....	68
4.8	Estimação final dos vícios .....	68

<b>5</b>	<b>METODOLOGIA EXPERIMENTAL EM FOTOTRIANGULAÇÃO .....</b>	<b>69</b>
5.1	Introdução .....	69
5.2	Etapas de cada experimento .....	70
5.3	Resultados de cada experimento.....	72
<b>6</b>	<b>ANÁLISE DOS RESULTADOS, CONCLUSÕES E RECOMENDAÇÕES.</b>	<b>112</b>
6.1	Análise dos resultados.....	112
6.2	Conclusões .....	114
6.3	Recomendações .....	114
	<b>ANEXO A.....</b>	<b>116</b>
	<b>ANEXO B.....</b>	<b>119</b>
	<b>ANEXO C.....</b>	<b>120</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>132</b>

## LISTA DOS PRINCIPAIS SÍMBOLOS

- $\hat{\underline{c}}$  - vetor das componentes principais dos resíduos estimados.
- $\text{cov} ( )$  - operador covariância.
- $\underline{e}_i$  - i-ésimo autovetor.
- $E$  - matriz dos autovetores.
- $E( )$  - operador esperança matemática.
- $\hat{\underline{s}}$  - vetor das componentes principais padronizadas.
- $V( )$  - operador variância.
- $\hat{\underline{V}}$  - vetor dos resíduos estimados.
- $\underline{X}$  - vetor X (todo símbolo sublinhado indica um vetor).
- $\underline{Y}$  - vetor de observações.
- $\hat{\underline{z}}$  - vetor dos resíduos estimados padronizados.
- $\hat{\underline{y}}$  - vetor das componentes principais do vetor dos resíduos estimados padronizados.
- $\lambda$  - autovalor.
- $\Lambda$  - matriz diagonal dos autovalores.
- $\underline{\mu}$  - vetor de médias.
- $\hat{\underline{\mu}}$  - vetor de médias estimadas.
- $\rho$  - correlação.
- $\sigma$  - desvio padrão.
- $\sigma^2$  - variância.
- $\Sigma$  - matriz covariância.
- $\tau_k$  - parâmetro de não centralidade.



## RESUMO

O problema da detecção e identificação de *outliers* em fototriangulação tem sido objeto de diversas pesquisas.

Esse trabalho propõe um Método Multivariado para detecção e identificação de múltiplos *outliers* nas fotocoordenadas de uma Fototriangulação por feixe de raios (*bundlemethod*). Particularmente utilizou-se o Método Multivariado conhecido como Análise de Componentes Principais (ACP) e o denominado Critério Ortomax Generalizado da Análise Fatorial (AF) em conjunto com alguns testes estatísticos como o teste de discordância de *outliers* de Ferguson e o teste t de Student.

Para testar o método proposto foi simulado um bloco de 6 fotografias, constituído de 34 pontos triangulados e 150 pontos imagens. Vinte experimentos são apresentados detalhadamente envolvendo 2 a 5 *outliers*. Os resultados e a análise de todos os experimentos realizados, bem como a comparação com os obtidos através do *Danish Method* são apresentados.

## ABSTRACT

Detecting and identifying outliers in Phototriangulation has been a constant problem in several kinds of research.

This work proposes a Multivariate Method for the detection and identification of multiple outliers in photocoordinates in Phototriangulation by bundlemethod. The Multivariate Method was particularly used, it is known as The Analysis of Principal Components (ACP, and also the so-called Generalized Orthomax Criterion of Factor Analysis (AF) together with some statistical tests such as the Ferguson's discordancy test and Student' t test.

In order to test the proposed method a simulated group of 6 photographs was used, which made up 34 triangulated points and 150 image points. Twenty experiments are presented in detail, by using from 2 to 5 outliers. The results and an analysis of every experiment which was carried out and a comparison with the obtained results when using the Danish Method are presented.

# 1 INTRODUÇÃO

## 1.1 Considerações gerais

O problema da detecção e localização de erros grosseiros em Fototriangulação tem sido objeto de diversas pesquisas (EL-HAKIM [19], [20], FIRKOWSKI [21], JUHL [27], MITISHITA [37], POPE [44], SILVA [47]). Uma técnica recomendada para a detecção e localização de erros grosseiros em Fototriangulação é conhecida como filtragem progressiva (SILVA [47]). Na aplicação da filtragem progressiva, são eliminados os erros de maior magnitude, permanecendo apenas os erros de pequena magnitude, cuja eliminação depende de métodos estatísticos. Entre os métodos estatísticos mais utilizados pode-se citar o *Data Snooping* de Baarda (BERBERAN [06]) e o *Danish Method* (BERBERAN [06], JUHL [27]), também conhecido como método de Krarup. Outras técnicas e comparação de métodos podem ser encontradas nas referências (CLERICI & HARRIS [13], DAVIES & GATHER [18], KOCH [30], KUBIK & WANG [31], OU [42], PATTERSON [43], XU [50]).

O método de Baarda é baseado na análise dos resíduos padronizados. Nesse método, a suposição fundamental é que as observações contaminadas por erros grosseiros resultam em resíduos padronizados de magnitudes significativamente altas. Dependendo da magnitude do erro grosseiro, da geometria onde ocorre e do número de graus de liberdade, o seu efeito é diluído entre os resíduos e parâmetros, no fenômeno conhecido em análise estatística como efeito *smearing*. Como consequência desse fenômeno, fica comprometida a sensibilidade dos resíduos para a análise dos erros grosseiros nas observações. A utilização de resíduos na detecção e identificação de observações contaminadas por erros grosseiros torna-se ainda mais complicada e menos confiável quando existe a presença de múltiplas observações contaminadas.

O *Danish Method* é eficiente para os problemas que envolvem múltiplas observações contaminadas por erros grosseiros. Esse método é constituído de uma função de peso que tem por objetivo diminuir, de modo iterativo, o peso das observações suspeitas de contaminação por erros grosseiros. Porém, pode ocorrer falhas do método quando existem erros de pequena magnitude, o qual identifica observações isentas de erros juntamente com observações contaminadas.

Segundo (MITISHITA [37]), em Fototriangulação os *outliers* constituem as observações contaminadas por erros grosseiros de magnitude entre  $3\sigma$  e  $100\sigma$ , sendo  $\sigma$  o desvio padrão das observações.

Com a evolução dos computadores, principalmente no que diz respeito ao aumento da velocidade de processamento e da capacidade de armazenamento de dados, o Método Multivariado (ADUOL [01], CROSILLA [14], JACKSON & MUDHOLKAR [25], surge como uma alternativa interessante na detecção e identificação de *outliers* em Fototriangulação. Principalmente porque este método, conhecido como Análise de Componentes Principais (ACP) (ANDERSON [03], CHAVES NETO & OLANDOSKI [12], CROSS & PRICE [16], JOHNSON & WICHERN [26], KENDALL [29], MARDIA; KENT & BIBBY [33], MORRISON [39], PRESS [45]) permite a otimização através do Critério de Rotação Ortomax Generalizado (CROSILLA [14], HARRIS & KAISER [24]) da Análise Fatorial (AF).

## 1.2 A proposta do trabalho

Neste trabalho propõe-se o desenvolvimento e aplicação de um Método Multivariado, para detecção e localização de *outliers* nas fotocoordenadas de uma Fototriangulação por feixe de raios (*bundlemethod*), bem como os resultados e análise de diversos experimentos.

O método será testado para detecção e identificação de *outliers* com erros grosseiros inferiores a  $30\sigma$ . Essa escolha foi feita em função das dificuldades

apresentadas pelos métodos conhecidos para detecção e identificação de erros de pequena magnitude, principalmente no caso de múltiplos *outliers*.

A aplicação do método, permitirá inicialmente, detectar a existência de *outliers* e numa segunda etapa a identificação de um certo número de observações suspeitas. De posse das observações suspeitas, entre as quais estarão incluídas as observações contaminadas por erros grosseiros, a localização dos *outliers* será feita com a aplicação de um teste t de Student em relação aos vícios estimados dessas observações, precedido da aplicação do teste de discordância de Ferguson para diminuir o número de vícios a ser estimado.

### **1.3 Objetivos do trabalho**

1.3.1 Propor um método multivariado para detecção e identificação de múltiplos *outliers* nas fotocoordenadas de uma Fototriangulação por feixe de raios (*bundlemethod*).

1.3.2 Efetuar um grande número de experimentos mostrando os resultados obtidos e as análises detalhadas dos mesmos.

1.3.3 Comparar os resultados obtidos com o *Danish Method*, mostrando a sua eficiência.

### **1.4 Desenvolvimento do trabalho**

O capítulo 2, trata da revisão bibliográfica do método multivariado conhecido como Análise de Componentes Principais. No capítulo 3 é apresentado o desenvolvimento teórico da Análise de Componentes Principais na detecção e

identificação de *outliers*. O capítulo 4 envolve o desenvolvimento da aplicação da Análise de Componentes Principais na detecção e identificação de *outliers* em Fototriangulação. No capítulo 5 é feita a aplicação da metodologia apresentada na simulação de um bloco envolvendo 6 fotografias, com a apresentação de cada etapa dos experimentos realizados. O capítulo 6 envolve a análise dos resultados obtidos nos experimentos realizados, as conclusões e recomendações.

## 2 ANÁLISE DE COMPONENTES PRINCIPAIS

### 2.1 Introdução

Para investigar as relações entre um conjunto de  $p$  variáveis correlacionadas pode ser útil transformar o conjunto de variáveis originais em um novo conjunto de variáveis não-correlacionadas chamadas *componentes principais*, tendo propriedades especiais em termos de variâncias. Essas novas variáveis são combinações lineares das variáveis originais e são derivadas em ordem decrescente de importância tal que, por exemplo, a primeira componente principal é a combinação linear normalizada com variância máxima.

A reprodução da variabilidade total do sistema requer as  $p$  variáveis, porém, freqüentemente a maior parte dessa variabilidade pode ser explicada por um número pequeno,  $k < p$ , de componentes principais. Neste caso existe praticamente a mesma quantidade de informação nas  $k$  componentes principais que nas  $p$  variáveis originais. As  $k$  componentes principais podem então substituir as  $p$  variáveis originais.

Os objetivos principais da Análise de Componentes Principais são:

- (1) reduzir o número de variáveis;
- (2) analisar quais as variáveis ou, quais os conjuntos de variáveis explicam a maior parte da variabilidade total, revelando que tipo de relacionamento existe entre elas.

O desenvolvimento da Análise de Componentes Principais não requer a suposição inicial de que as variáveis originais tenham uma distribuição específica, embora a suposição de normalidade multivariada permita o desenvolvimento de inferências.

## 2.2 Componentes principais populacionais

Seja  $\underline{X}' = [X_1, X_2, \dots, X_p]$  um vetor aleatório p-dimensional com média  $\underline{\mu}$  e matriz de covariâncias  $\Sigma$  com autovalores  $\lambda_1, \lambda_2, \dots, \lambda_p \geq 0$ .

Considere as combinações lineares

$$\begin{aligned}
 Y_1 &= c_{11}X_1 + c_{21}X_2 + \dots + c_{p1}X_p = \underline{c}'_1 \underline{X} \\
 Y_2 &= c_{12}X_1 + c_{22}X_2 + \dots + c_{p2}X_p = \underline{c}'_2 \underline{X} \\
 &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
 Y_p &= c_{1p}X_1 + c_{2p}X_2 + \dots + c_{pp}X_p = \underline{c}'_p \underline{X}
 \end{aligned} \tag{2.1}$$

ou

$$\underline{Y} = \underline{C}' \underline{X} \tag{2.2}$$

onde

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_p \end{bmatrix} \quad e \quad \underline{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ c_{p1} & c_{p2} & \dots & c_{pp} \end{bmatrix} \tag{2.3}$$

com

$$E(Y_j) = E(\underline{c}'_j \underline{X}) = \underline{c}'_j \underline{\mu} \quad j = 1, 2, \dots, p \tag{2.4}$$



$$V(Y_j) = V(\underline{c}'_j \underline{X}) = \underline{c}'_j V(\underline{X}) \underline{c}_j = \underline{c}'_j \Sigma \underline{c}_j \quad j = 1, 2, \dots, p \quad (2.5)$$

e

$$\text{cov}(Y_k, Y_j) = \text{cov}(\underline{c}'_k \underline{X}, \underline{c}'_j \underline{X}) = \underline{c}'_k \Sigma \underline{c}_j \quad k, j = 1, 2, \dots, p \quad (2.6)$$

Verifica-se que cada equação em (2.1) contém um fator de escala arbitrário. Portanto, para um determinado  $Y_j$ , ou seja,

$$Y_j = c_{1j} X_1 + c_{2j} X_2 + \dots + c_{pj} X_p = \underline{c}'_j \underline{X} \quad (2.7)$$

impõe-se a condição

$$\underline{c}'_j \underline{c}_j = \sum_{k=1}^p c_{kj}^2 = 1 \quad (2.8)$$

obtendo-se uma solução normalizada.

Pode-se então definir:

- (1) a primeira componente principal como a combinação linear  $\underline{c}'_1 \underline{X}$  que maximiza  $V(\underline{c}'_1 \underline{X})$  sujeita à restrição  $\underline{c}'_1 \underline{c}_1 = 1$ ;
- (2) a segunda componente principal como a combinação linear  $\underline{c}'_2 \underline{X}$  que maximiza  $V(\underline{c}'_2 \underline{X})$  sujeita às restrições  $\underline{c}'_2 \underline{c}_2 = 1$  e  $\text{cov}(\underline{c}'_1 \underline{X}, \underline{c}'_2 \underline{X}) = 0$ ;
- (3) a  $j$ -ésima componente principal como a combinação linear  $\underline{c}'_j \underline{X}$  que maximiza  $V(\underline{c}'_j \underline{X})$  sujeita às restrições  $\underline{c}'_j \underline{c}_j = 1$  e  $\text{cov}(\underline{c}'_j \underline{X}, \underline{c}'_k \underline{X}) = 0$  para todo  $k < j$ .

**Resultado 2.1** Seja o vetor aleatório  $\underline{X}' = [X_1, X_2, \dots, X_p]$  tendo matriz de covariância  $\Sigma$  com pares de autovalor-autovetor  $(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), \dots, (\lambda_p, \underline{e}_p)$  onde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . A  $j$ -ésima componente principal é dada por

$$Y_j = \underline{e}_j' \underline{X} = e_{1j}X_1 + e_{2j}X_2 + \dots + e_{pj}X_p \quad j = 1, 2, \dots, p \quad (2.9)$$

onde

$$V(Y_j) = \underline{e}_j' \Sigma \underline{e}_j = \lambda_j \quad j = 1, 2, \dots, p \quad (2.10)$$

e

$$\text{cov}(Y_k, Y_j) = \underline{e}_k' \Sigma \underline{e}_j = 0 \quad k \neq j \quad (2.11)$$

Se existir valores idênticos para  $\lambda_j$ , a determinação dos correspondentes  $\underline{e}_j$ , e portanto  $Y_j$ , não é única.

### *Prova*

Seja  $P: p \times p$  uma matriz ortogonal cujas colunas são os autovetores  $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$  e  $\Lambda$  uma matriz diagonal cujos elementos diagonais são os autovalores  $\lambda_1, \lambda_2, \dots, \lambda_p$ . Sejam

$$\Sigma^{1/2} = P \Lambda^{1/2} P' \quad \text{e} \quad \underline{y} = \underset{(p \times 1)}{P'} \underset{(p \times p)}{c} \underset{(p \times 1)}{c}$$

conseqüentemente,  $\underline{c} \neq \underline{0}$  implica em  $\underline{y} \neq \underline{0}$ . Portanto

$$\begin{aligned} \frac{\underline{c}' \Sigma \underline{c}}{\underline{c}' \underline{c}} &= \frac{\underline{c}' \Sigma^{1/2} \Sigma^{1/2} \underline{c}}{\underline{c}' P P' \underline{c}} = \frac{\underline{c}' P \Lambda^{1/2} P' P \Lambda^{1/2} P' \underline{c}}{\underline{y}' \underline{y}} \\ &= \frac{\underline{y}' \Lambda \underline{y}}{\underline{y}' \underline{y}} = \frac{\sum_{j=1}^p \lambda_j y_j^2}{\sum_{j=1}^p y_j^2} \leq \lambda_1 \frac{\sum_{j=1}^p y_j^2}{\sum_{j=1}^p y_j^2} = \lambda_1 \end{aligned} \quad (2.12)$$

Fazendo  $\underline{c} = \underline{e}_1$  tem-se que  $\underline{y} = P' \underline{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ , enquanto

$$\underline{e}'_k \underline{e}_1 = \begin{cases} 1, & k = 1 \\ 0, & k \neq 1 \end{cases}$$

Para essa escolha de  $\underline{c}$ ,  $\underline{y}' \Lambda \underline{y} / \underline{y}' \underline{y} = \lambda_1 / 1 = \lambda_1$ , ou

$$\frac{\underline{e}'_1 \Sigma \underline{e}_1}{\underline{e}'_1 \underline{e}_1} = \underline{e}'_1 \Sigma \underline{e}_1 = \lambda_1 = V(Y_1)$$

De forma semelhante pode ser provada a segunda parte do resultado.

Seja agora

$$\underline{c} = P\underline{y} = y_1 \underline{e}_1 + y_2 \underline{e}_2 + \dots + y_p \underline{e}_p$$

tal que

$$\underline{c} \perp \underline{e}_1, \underline{e}_2, \dots, \underline{e}_k \quad \text{implica em}$$

$$0 = \underline{e}'_j \underline{c} = y_1 \underline{e}'_j \underline{e}_1 + y_2 \underline{e}'_j \underline{e}_2 + \dots + y_p \underline{e}'_j \underline{e}_p = y_j, \quad j \leq k.$$

Portanto, para  $\underline{c}$  perpendicular aos  $k$  primeiros autovetores  $\underline{e}_j$ , tem-se da (2.12) que

$$\max_{\underline{c} \perp \underline{e}_1, \dots, \underline{e}_k} \frac{\underline{c}' \Sigma \underline{c}}{\underline{c}' \underline{c}} = \frac{\sum_{j=k+1}^p \lambda_j y_j^2}{\sum_{j=k+1}^p y_j^2} = \lambda_{k+1}, \quad k = 1, 2, \dots, p-1$$

Para a escolha  $\underline{c} = \underline{e}_{k+1}$ , onde  $\underline{e}'_{k+1} \underline{e}_k = 0$ ,  $k = 1, 2, \dots, p-1$ , obtém-se

$$\frac{\underline{e}'_{k+1} \Sigma \underline{e}_{k+1}}{\underline{e}'_{k+1} \underline{e}_{k+1}} = \underline{e}'_{k+1} \Sigma \underline{e}_{k+1} = V(Y_{k+1})$$

Resta agora mostrar que  $\underline{e}_k$  perpendicular a  $\underline{e}_j$  (isto é,  $\underline{e}'_j \underline{e}_k = 0$ ,  $j \neq k$ ) fornece  $\text{cov}(Y_j, Y_k) = 0$ . Os autovetores de  $\Sigma$  são ortogonais se todos os autovalores  $\lambda_1, \lambda_2, \dots, \lambda_p$  são distintos. Quando os autovalores não são todos distintos, os autovetores correspondentes aos autovalores comuns podem ser escolhidos como sendo ortogonais. Portanto, para quaisquer autovetores

$$\underline{e}_k \text{ e } \underline{e}_j, \quad \underline{e}'_k \underline{e}_j = 0, \quad k \neq j.$$

Como  $\Sigma \underline{e}_k = \lambda_k \underline{e}_k$ , pré-multiplicando por  $\underline{e}'_j$  fica

$$\text{cov}(Y_j, Y_k) = \underline{e}'_j \Sigma \underline{e}_k = \underline{e}'_j \lambda_k \underline{e}_k = \lambda_k \underline{e}'_j \underline{e}_k = 0$$

para qualquer  $k \neq j$ . □

**Resultado 2.2** Seja o vetor aleatório  $\underline{X}' = [X_1, X_2, \dots, X_p]$  tendo matriz de covariâncias  $\Sigma$ , com pares de autovalor-autovetor  $(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), \dots, (\lambda_p, \underline{e}_p)$  onde  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Sejam  $Y_1 = \underline{e}'_1 \underline{X}$ ,  $Y_2 = \underline{e}'_2 \underline{X}$ ,  $\dots$ ,  $Y_p = \underline{e}'_p \underline{X}$  as componentes principais. Então

$$\sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2 = \sum_{k=1}^p V(X_k) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{j=1}^p V(Y_j).$$

*Prova*

Por definição  $\sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2 = \text{tr}(\Sigma)$ . De acordo com a decomposição espectral para  $\Sigma$ , pode-se escrever  $\Sigma = P \Lambda P'$  onde  $\Lambda$  é a matriz diagonal dos autovalores e  $P = [\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p]$  a matriz dos autovetores tal que  $PP' = P'P = I$ , logo

$$\text{tr}(\Sigma) = \text{tr}(P \Lambda P') = \text{tr}(\Lambda) = \lambda_1 + \lambda_2 + \dots + \lambda_p$$

Portanto

$$\sum_{k=1}^p V(X_k) = \text{tr}(\Sigma) = \text{tr}(\Lambda) = \sum_{j=1}^p V(Y_j) \quad \square$$

De acordo com o resultado 2.2 a variância populacional total é igual a

$$\sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2 = \lambda_1 + \lambda_2 + \dots + \lambda_p \quad (2.13)$$

e conseqüentemente, a proporção da variância total devida a j-ésima componente principal é

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_p}, \quad j = 1, 2, \dots, p \quad (2.14)$$

Cada componente dos autovetores  $\underline{c}_j' = [c_{1j}, c_{2j}, \dots, c_{pj}]$  pode auxiliar na interpretação da componente principal. A magnitude de  $c_{kj}$  mede a importância da k-ésima variável  $X_k$  para a j-ésima componente principal  $Y_j$ . Na realidade,  $c_{kj}$  é proporcional ao coeficiente de correlação entre  $Y_j$  e  $X_k$ .

**Resultado 2.3** Se  $Y_1 = \underline{e}'_1 \underline{X}$ ,  $Y_2 = \underline{e}'_2 \underline{X}$ , ...,  $Y_p = \underline{e}'_p \underline{X}$  são as componentes principais obtidas da matriz de covariância  $\Sigma$ , então

$$\rho_{Y_j X_k} = \frac{e_{kj} \sqrt{\lambda_j}}{\sigma_k} \quad k, j = 1, 2, \dots, p \quad (2.15)$$

são os coeficientes de correlação entre as componentes principais  $Y_j$  e as variáveis  $X_k$ , onde  $(\lambda_1, \underline{e}_1)$ ,  $(\lambda_2, \underline{e}_2)$ , ...,  $(\lambda_p, \underline{e}_p)$  são os pares autovalor-autovetor de  $\Sigma$ .

*Prova*

Seja  $\underline{c}'_k = [0, \dots, 0, 1, 0, \dots, 0]$  tal que

$$X_k = \underline{c}'_k \underline{X} \text{ e } \text{cov}(X_k, Y_j) = \text{cov}(\underline{c}'_k \underline{X}, \underline{e}'_j \underline{X}) = \underline{c}'_k \Sigma \underline{e}_j.$$

Visto que  $\Sigma \underline{e}_j = \lambda_j \underline{e}_j$ ,  $\text{cov}(X_k, Y_j) = \underline{c}'_k \lambda_j \underline{e}_j = \lambda_j e_{kj}$  e como  $V(Y_j) = \lambda_j$  e  $V(X_k) = \sigma_k^2$ , então

$$\rho_{Y_j, X_k} = \frac{\text{cov}(Y_j, X_k)}{\sqrt{V(Y_j)} \sqrt{V(X_k)}} = \frac{\lambda_j e_{kj}}{\sqrt{\lambda_j} \sqrt{\sigma_k^2}} = \frac{e_{kj} \sqrt{\lambda_j}}{\sigma_k} \quad k, j = 1, 2, \dots, p \quad \square$$

### 2.3 ACP para dados normais multivariados

Nos desenvolvimentos realizados anteriormente, não foi feita nenhuma suposição com relação à distribuição do vetor aleatório  $\underline{X}$ . Seja agora  $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ .

Se  $P$  denota a matriz dos autovetores de  $\Sigma$ , então as componentes principais podem ser escritas na forma

$$\underline{Y} = P'(\underline{X} - \underline{\mu}) \quad (2.16)$$

que é análoga a (2.2), exceto que, em adição à rotação ortogonal, deve-se introduzir uma mudança de origem para fazer com que  $\underline{Y}$  tenha vetor zero para as médias.

Sendo conhecida a distribuição de  $\underline{X}$  pode-se então determinar a distribuição de  $\underline{Y}$ . Como cada componente de  $\underline{Y}$  é uma combinação linear de variáveis aleatórias normalmente distribuídas, segue que cada componente de  $\underline{Y}$  também é normalmente distribuída (JOHNSON & WICHERN [26]).

A função densidade conjunta de  $\underline{X}$ ,  $f(\underline{X})$ , e a de  $\underline{Y}$ ,  $g(\underline{Y})$ , estão relacionadas por (CHATFIELD & COLLINS [11])

$$g(\underline{y}) = f(\underline{x})_y |J| \quad (2.17)$$

onde  $|J|$  = Jacobiano da transformação, ou seja,

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \dots & \frac{\partial x_1}{\partial y_p} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \dots & \frac{\partial x_2}{\partial y_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial x_p}{\partial y_1} & \frac{\partial x_p}{\partial y_2} & \dots & \frac{\partial x_p}{\partial y_p} \end{vmatrix}$$

Da (2.16) obtém-se  $\underline{X} = P\underline{Y} + \underline{\mu}$ , onde pode-se ver que o Jacobiano deve ser igual a  $|P|$ , o qual é igual a 1, visto que  $P$  é ortogonal com  $PP' = I$  e  $|P||P'| = 1$ .

Agora na Eq. (2.17),  $f(\underline{x})_y$  significa  $f(\underline{x})$  avaliada para  $\underline{y} = P'(\underline{x} - \underline{\mu})$ .

Como

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \right] \quad (2.18)$$

pode-se determinar

$$g(\underline{y}) = \frac{1}{(2\pi)^{p/2} |\Lambda|^{1/2}} \exp\left[-\frac{1}{2} \underline{y}' \Lambda^{-1} \underline{y}\right] \quad (2.19)$$

usando  $|\Sigma| = |\Lambda|$  e  $\Sigma^{-1} = C \Lambda^{-1} C'$ .

Escrevendo a Eq. (2.19) na forma

$$g(\underline{y}) = \frac{1}{(2\pi)^{p/2} \left(\prod_{i=1}^p \lambda_i\right)^{1/2}} \exp\left[-\frac{1}{2} \sum_{i=1}^p \frac{y_i^2}{\lambda_i}\right] \quad (2.20)$$

pode-se constatar que é da mesma forma que o produto de  $p$  funções densidades normais independentes.

No caso normal multivariado, as componentes principais tem uma interpretação geométrica mais simples. Seja  $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ , sabe-se que o elipsóide centrado em  $\underline{\mu}$  (JOHNSON & WICHERN [26])

$$(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) = \text{constante} = k^2$$

de densidade constante tem eixos  $\pm k \sqrt{\lambda_j} \underline{e}_j$ ,  $j = 1, 2, \dots, p$ , onde  $(\lambda_j, \underline{e}_j)$  são os pares autovalor-autovetor de  $\Sigma$ .

Um ponto no  $j$ -ésimo eixo do elipsóide terá coordenadas proporcionais a

$$\underline{e}'_j = [e_{1j}, e_{2j}, \dots, e_{pj}]$$

no sistema de coordenadas com origem  $\underline{\mu}$  e eixos  $X_1, X_2, \dots, X_p$ .

Por conveniência, considere-se  $\underline{\mu} = \underline{0}$ , podendo-se escrever

$$k^2 = \underline{X}' \Sigma^{-1} \underline{X} = \frac{1}{\lambda_1} (\underline{e}'_1 \underline{X})^2 + \frac{1}{\lambda_2} (\underline{e}'_2 \underline{X})^2 + \dots + \frac{1}{\lambda_p} (\underline{e}'_p \underline{X})^2$$



onde  $\underline{e}'_1 \underline{X}$ ,  $\underline{e}'_2 \underline{X}$ , ...,  $\underline{e}'_p \underline{X}$  são as componentes principais de  $\underline{X}$ . Fazendo

$$y_1 = \underline{e}'_1 \underline{X}, \quad y_2 = \underline{e}'_2 \underline{X}, \quad \dots, \quad y_p = \underline{e}'_p \underline{X},$$

tem-se que

$$k^2 = \frac{1}{\lambda_1} y_1^2 + \frac{1}{\lambda_2} y_2^2 + \dots + \frac{1}{\lambda_p} y_p^2 \quad (2.21)$$

onde essa equação define um elipsóide (desde que  $\lambda_1, \lambda_2, \dots, \lambda_p$  sejam positivos) em um sistema de coordenadas com eixos  $y_1, y_2, \dots, y_p$  com as direções de  $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$ , respectivamente. Se  $\lambda_1$  é o maior autovalor, então o maior eixo tem a direção de  $\underline{e}_1$ . Os remanescentes eixos menores tem direções definidas por  $\underline{e}_2, \dots, \underline{e}_p$ .

## 2.4 Componentes principais obtidas de variáveis padronizadas

As componentes principais podem ser obtidas das variáveis padronizadas

$$\begin{aligned} Z_1 &= \frac{X_1 - \mu_1}{\sigma_1} \\ Z_2 &= \frac{X_2 - \mu_2}{\sigma_2} \\ &\vdots \\ Z_p &= \frac{X_p - \mu_p}{\sigma_p} \end{aligned} \quad (2.22)$$

Em notação matricial

$$\underline{Z} = (\mathbf{V}^{1/2})^{-1}(\underline{X} - \underline{\mu}) \quad (2.23)$$

sendo  $\mathbf{V}^{1/2}$  a matriz dos desvios padrões de  $\underline{X}$ , definida pela matriz diagonal

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \sigma_p \end{bmatrix} \quad (2.24)$$

Para  $\underline{Z}$  tem-se que

$$\begin{aligned} E(\underline{Z}) &= E[(\mathbf{V}^{1/2})^{-1}(\underline{X} - \underline{\mu})] \\ &= (\mathbf{V}^{1/2})^{-1}E(\underline{X} - \underline{\mu}) = \underline{0} \end{aligned} \quad (2.25)$$

e

$$\begin{aligned} \text{cov}(\underline{Z}) &= \text{cov}[(\mathbf{V}^{1/2})^{-1}(\underline{X} - \underline{\mu})] \\ &= (\mathbf{V}^{1/2})^{-1}\Sigma(\mathbf{V}^{1/2})^{-1} = \rho \end{aligned} \quad (2.26)$$

Como se constata pela (2.23), as componentes principais de  $\underline{Z}$  podem ser obtidas da matriz de correlação  $\rho$  de  $\underline{X}$ .

De um modo geral, as componentes principais obtidas da matriz de correlação não são as mesmas que aquelas obtidas pela matriz de covariância e, além disso, não se pode passar de uma solução para outra simplesmente por uma mudança de escala nos coeficientes.

**Resultado 2.4** A  $j$ -ésima componente principal das variáveis padronizadas

$$\underline{Z}' = [Z_1, Z_2, \dots, Z_p]$$

com  $\text{cov}(\underline{Z}) = \rho$ , é dada por

$$Y_j = \underline{e}'_j \underline{Z} = \underline{e}'_j (\mathbf{V}^{1/2})^{-1} (\underline{X} - \underline{\mu}), \quad j = 1, 2, \dots, p \quad (2.27)$$

Além dessa relação, tem-se ainda que

$$\sum_{k=1}^p V(Z_k) = p \quad (2.28)$$

e

$$\rho_{Y_j, Z_k} = e_{kj} \sqrt{\lambda_j}, \quad k, j = 1, 2, \dots, p \quad (2.29)$$

onde  $(\lambda_1, \underline{e}_1), (\lambda_2, \underline{e}_2), \dots, (\lambda_p, \underline{e}_p)$  são os pares autovalor-autovetor para  $\rho$  com  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ .

*Prova*

Esse resultado pode ser obtido a partir dos Resultados (2.1), (2.2) e (2.3), com  $Z_1, Z_2, \dots, Z_p$  substituindo  $X_1, X_2, \dots, X_p$  e  $\rho$  substituindo  $\Sigma$ .  $\square$

De acordo com a (2.28), a variância populacional total (para variáveis padronizadas) é simplesmente  $p$ , a soma dos elementos diagonais da matriz  $\rho$ . Portanto, a proporção da variância total explicada pela  $j$ -ésima componente principal de  $\underline{Z}$  é dada por

$$\frac{\lambda_j}{p} \quad j = 1, 2, \dots, p \quad (2.30)$$

onde os  $\lambda_j$ 's são os autovalores de  $\rho$ .

## 2.5 Componentes principais de matrizes covariâncias especiais

Determinadas matrizes covariâncias e correlações apresentam estruturas especiais, cujas componentes principais são determinadas de uma forma simplificada.

(1) Seja  $\Sigma$  a matriz diagonal

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_p^2 \end{bmatrix} \quad (2.31)$$

Considerando  $\underline{e}_j = [0, \dots, 0, 1, 0, \dots, 0]$ , com 1 na  $j$ -ésima linha, tem-se que

$$\begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_p^2 \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \sigma_j^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

ou

$$\Sigma \underline{e}_j = \sigma_j^2 \underline{e}_j$$

e conclui-se que  $(\sigma_j^2, \underline{e}_j)$  é o j-ésimo par autovalor-autovetor. Como  $\underline{e}_j' \underline{X} = X_j$ , o conjunto de componentes principais é o próprio conjunto de variáveis aleatórias originais não-correlacionadas.

Quando o problema envolve uma matriz covariância com a estrutura da (2.31), sendo  $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$ , os contornos de densidade constante são elipsóides cujos eixos tem direções das variações máximas, não sendo necessária a rotação do sistema de eixos coordenados.

No caso de variável padronizada, tem-se que  $\rho = 1$ , uma matriz identidade  $p \times p$ . Evidentemente,  $\rho \underline{e}_j = 1 \underline{e}_j$ , tal que o autovalor 1 tem multiplicidade  $p$  e  $\underline{e}_j = [0, \dots, 0, 1, 0, \dots, 0]$   $j = 1, 2, \dots, p$ , são escolhas convenientes para os autovetores. Conseqüentemente, as componentes principais determinadas para  $\rho$  são também as variáveis originais  $Z_1, Z_2, \dots, Z_p$ .

(2) Seja  $\Sigma$  uma matriz com a seguinte estrutura

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \dots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \dots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \dots & \sigma^2 \end{bmatrix} \quad (2.32)$$

A matriz de correlação resultante

$$\rho = \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix} \quad (2.33)$$

é também a matriz de covariâncias das variáveis padronizadas. A matriz (2.33) implica que as variáveis  $X_1, X_2, \dots, X_p$  são igualmente correlacionadas.

Os  $p$  autovalores da matriz correlação (2.33) podem ser divididos em 2 grupos. Quando  $\rho > 0$ , o maior autovalor é (JOHNSON & WICHERN [26])

$$\lambda_1 = 1 + (p - 1)\rho \quad (2.34)$$

com o correspondente autovetor

$$\underline{e}'_1 = \left[ \frac{1}{\sqrt{p}} \quad \frac{1}{\sqrt{p}} \quad \dots \quad \frac{1}{\sqrt{p}} \right] \quad (2.35)$$

Os demais  $p-1$  autovalores são

$$\lambda_2 = \lambda_3 = \dots = \lambda_p = 1 - \rho \quad (2.36)$$

com autovetores

$$\underline{e}'_2 = \left[ \frac{1}{\sqrt{1 \times 2}} \quad \frac{-1}{\sqrt{1 \times 2}} \quad 0 \quad \dots \quad 0 \right] \quad (2.37)$$

$$\underline{e}'_3 = \left[ \frac{1}{\sqrt{2 \times 3}} \quad \frac{1}{\sqrt{2 \times 3}} \quad \frac{-2}{\sqrt{2 \times 3}} \quad 0 \quad \dots \quad 0 \right] \quad (2.38)$$

$$\underline{e}'_j = \left[ \frac{1}{\sqrt{(j-1)j}} \quad \dots \quad \frac{1}{\sqrt{(j-1)j}} \quad \frac{-(j-1)}{\sqrt{(j-1)j}} \quad 0 \quad \dots \quad 0 \right] \quad (2.39)$$

$$\underline{e}'_p = \left[ \frac{1}{\sqrt{(p-1)p}} \quad \dots \quad \frac{1}{\sqrt{(p-1)p}} \quad \frac{-(p-1)}{\sqrt{(p-1)p}} \right] \quad (2.40)$$

A primeira componente principal

$$Y_1 = \underline{e}'_1 \underline{X} = \frac{1}{\sqrt{p}} \sum_{k=1}^p X_k \quad (2.41)$$

é proporcional à soma das  $p$  variáveis originais. Essa componente principal explica uma proporção

$$\frac{\lambda_1}{p} = \frac{1 + (p-1)\rho}{p} = \rho + \frac{1-\rho}{p} \quad (2.42)$$

da variação populacional total. Pode-se constatar que  $\lambda_1/p \cong \rho$  para  $\rho$  próximo de 1 ou  $p$  grande.

Se as variáveis padronizadas  $Z_1, Z_2, \dots, Z_p$  tem uma distribuição normal multivariada com matriz de covariâncias dada por (2.32), então os elipsóides de densidade constante tem forma de charuto com o eixo maior ao longo da primeira componente principal  $Y_1 = (1/\sqrt{p})[1, 1, \dots, 1] \underline{X}$ .

## 2.6 Componentes principais amostrais

Todos os resultados desenvolvidos nas seções anteriores são aplicáveis somente se a matriz de covariâncias  $\Sigma$  for conhecida. Infelizmente, na prática este parâmetro é desconhecido e precisa ser estimado. Portanto, seja o vetor aleatório  $\underline{X}' = [X_1, X_2, \dots, X_p]$  com parâmetros  $E(\underline{X}) = \underline{\mu}$  e  $V(\underline{X}) = \Sigma$ , ambos desconhecidos. Suponha que  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$ , com  $n > p$ , são vetores  $p \times 1$  de observações independentes de  $\underline{X}$ .

O vetor média amostral e a matriz de covariâncias que estimam  $\underline{\mu}$  e  $\Sigma$  são dados por (Johnson & Wichern [26])

$$\hat{\underline{\mu}} = \bar{\underline{x}} = \frac{1}{n} \sum_{i=1}^n \underline{x}_i \quad (2.43)$$

ou, em notação matricial

$$\hat{\underline{\mu}} = \bar{\underline{x}} = \frac{1}{n} \mathbf{X}' \underline{\mathbf{1}} \quad (2.44)$$

onde:  $\mathbf{X}$  = matriz de dados (ordem:  $n \times p$ ),  $\underline{\mathbf{1}}$  é o vetor  $n \times 1$

$$\underline{\mathbf{1}} = \begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix} \quad (2.45)$$

e

$$\hat{\underline{\Sigma}} = \frac{1}{n} \sum_{i=1}^n (\underline{x}_i - \bar{\underline{x}})(\underline{x}_i - \bar{\underline{x}})' \quad (2.46)$$

ou, em notação matricial

$$\hat{\underline{\Sigma}} = \frac{1}{n} \mathbf{X}' \mathbf{H} \mathbf{X} \quad (2.47)$$

onde  $\mathbf{H}$  = matriz  $n \times n$  dada por

$$\mathbf{H} = \mathbf{I} - \frac{1}{n} \underline{\mathbf{1}} \underline{\mathbf{1}}' \quad (2.48)$$

sendo  $\mathbf{I}$  = matriz identidade  $n \times n$ .



Se  $\underline{X} \sim N_p(\underline{\mu}, \Sigma)$  então  $\bar{x}$  e  $\hat{\Sigma}$  são estimadores de máxima verossimilhança de  $\underline{\mu}$  e  $\Sigma$ . O estimador não-viciado de  $\Sigma$  será dado por

$$S = \frac{n}{n-1} \hat{\Sigma} \quad (2.49)$$

Sendo  $\hat{\Sigma}$ :  $p \times p$  a matriz de covariância amostral com pares de autovalor-autovetor  $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \dots, (\hat{\lambda}_p, \hat{e}_p)$ , a  $j$ -ésima componente principal amostral será dada por

$$\hat{y}_j = \hat{e}_j \underline{x} = \hat{e}_{1j} x_1 + \hat{e}_{2j} x_2 + \dots + \hat{e}_{pj} x_p \quad j = 1, 2, \dots, p \quad (2.50)$$

onde  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$  e  $\underline{x}$  é qualquer observação das variáveis  $X_1, X_2, \dots, X_p$ .

Para a  $j$ -ésima componente amostral tem-se ainda que

$$V(\hat{y}_j) = \hat{\lambda}_j, \quad j = 1, 2, \dots, p \quad (2.51)$$

$$\text{cov}(\hat{y}_k, \hat{y}_j) = 0, \quad k \neq j \quad (2.52)$$

$$\text{variância total amostral} = \sum_{j=1}^p s_j^2 = \hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_p \quad (2.53)$$

e correlação amostral entre  $\hat{y}_j$  e  $x_k$

$$r_{\hat{y}_j x_k} = \frac{\hat{e}_{kj} \sqrt{\hat{\lambda}_j}}{s_k} \quad k, j = 1, 2, \dots, p \quad (2.54)$$

Geometricamente os dados podem ser plotados como  $n$  pontos em um espaço  $p$ -dimensional. Se  $S$  é uma matriz positiva definida, todos os vetores  $\underline{x} : p \times 1$  satisfazem a equação

$$(\underline{x} - \bar{\underline{x}})' S^{-1} (\underline{x} - \bar{\underline{x}}) = k^2 \quad (2.55)$$

definindo um hiperelipsóide centrado em  $\bar{\underline{x}}$ , cujos eixos são dados pelos autovetores de  $S$ . Os comprimentos desses eixos são proporcionais a  $\sqrt{\hat{\lambda}_j}$ ,  $j = 1, 2, \dots, p$ , onde

$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p > 0$  são os autovalores de  $S$ .

Se  $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n$  representa uma amostra proveniente de uma população normal, as componentes principais amostrais,  $\hat{y}_j = \hat{\underline{e}}_j' (\underline{x} - \bar{\underline{x}})$  estimam as componentes principais populacionais  $Y_j = \underline{e}_j' (\underline{x} - \underline{\mu})$ , as quais tem distribuição  $N_p(\underline{0}, \Lambda)$ . A matriz  $\Lambda$  é uma matriz diagonal cujos elementos são os autovalores  $\lambda_1, \lambda_2, \dots, \lambda_p$  de  $\Sigma$  e  $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_p$  os correspondentes autovetores.

As componentes principais amostrais são, em geral, invariantes com respeito à mudança de escala. As variáveis medidas em diferentes escalas ou numa escala comum com ordem de grandeza bem diferente são freqüentemente padronizadas. Para a amostra, a padronização é obtida construindo

$$\underline{z}_i = D^{-1/2} (\underline{x}_i - \bar{\underline{x}}) = \begin{bmatrix} \frac{x_{i1} - \bar{x}_1}{s_1} \\ \frac{x_{i2} - \bar{x}_2}{s_2} \\ \vdots \\ \frac{x_{ip} - \bar{x}_p}{s_p} \end{bmatrix}, \quad i = 1, 2, \dots, n \quad (2.56)$$

A matriz de dados  $n \times p$  das observações padronizadas pode ser escrita como

$$Z = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1p} \\ z_{21} & z_{22} & \dots & z_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & \dots & z_{np} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{x_{11} - \bar{x}_1}{s_1} & \frac{x_{12} - \bar{x}_2}{s_2} & \dots & \frac{x_{1p} - \bar{x}_p}{s_p} \\ \frac{x_{21} - \bar{x}_1}{s_1} & \frac{x_{22} - \bar{x}_2}{s_2} & \dots & \frac{x_{2p} - \bar{x}_p}{s_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{x_{n1} - \bar{x}_1}{s_1} & \frac{x_{n2} - \bar{x}_2}{s_2} & \dots & \frac{x_{np} - \bar{x}_p}{s_p} \end{bmatrix} \quad (2.57)$$

resultando no vetor média amostral (JOHNSON & WICHERN [26])

$$\bar{z} = \frac{1}{n} Z' \underline{1} = \frac{1}{n} \begin{bmatrix} \sum_{i=1}^n \frac{x_{i1} - \bar{x}_1}{s_1} \\ \sum_{i=1}^n \frac{x_{i2} - \bar{x}_2}{s_2} \\ \vdots \\ \sum_{i=1}^n \frac{x_{ip} - \bar{x}_p}{s_p} \end{bmatrix} = \underline{0} \quad (2.58)$$

onde  $\underline{1}' = [1, 1, \dots, 1]$ , e matriz covariância amostral

$$\begin{aligned} S_Z &= \frac{1}{n-1} (Z' - \frac{1}{n} Z' \underline{1} \underline{1}') (Z' - \frac{1}{n} Z' \underline{1} \underline{1}')' \\ &= \frac{1}{n-1} (Z' - \bar{z} \underline{1}') (Z' - \bar{z} \underline{1}')' = \frac{1}{n-1} Z' Z = \end{aligned}$$

$$= \frac{1}{n-1} \begin{bmatrix} \frac{(n-1)s_1^2}{s_1^2} & \frac{(n-1)s_{12}}{s_1s_2} & \dots & \frac{(n-1)s_{1p}}{s_1s_p} \\ \frac{(n-1)s_{21}}{s_2s_1} & \frac{(n-1)s_2^2}{s_2^2} & \dots & \frac{(n-1)s_{2p}}{s_2s_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(n-1)s_{p1}}{s_ps_1} & \frac{(n-1)s_{p2}}{s_ps_2} & \dots & \frac{(n-1)s_p^2}{s_p^2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \hat{\rho}_{12} & \dots & \hat{\rho}_{1p} \\ \hat{\rho}_{21} & 1 & \dots & \hat{\rho}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{p1} & \hat{\rho}_{p2} & \dots & 1 \end{bmatrix} = R \quad (2.59)$$

Se  $\underline{z} = [z_1, z_2, \dots, z_p]$  é um vetor de observações padronizadas com matriz covariância R, a j-ésima componente principal estimada será

$$\hat{y}_j = \underline{\hat{e}}_j' \underline{z} = \hat{e}_{1j}z_1 + \hat{e}_{2j}z_2 + \dots + \hat{e}_{pj}z_p, \quad j = 1, 2, \dots, p \quad (2.60)$$

onde  $(\hat{\lambda}_j, \hat{e}_j)$  é o j-ésimo par de autovalor-autovetor de R com  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$ .

Portanto, para a variável padronizada pode-se constatar que

$$V(\hat{y}_j) = \hat{\lambda}_j, \quad j = 1, 2, \dots, p \quad (2.61)$$

$$\text{cov}(\hat{y}_k, \hat{y}_j) = 0, \quad k \neq j \quad (2.62)$$

$$\text{variância total amostral} = \text{tr}(R) = p = \hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_p \quad (2.63)$$

e correlação entre  $\hat{y}_j$  e  $z_k$  dada por

$$r_{\hat{y}_j z_k} = \hat{e}_{kj} \sqrt{\hat{\lambda}_j}, \quad j = 1, 2, \dots, p \quad (2.64)$$

A proporção da variância total amostral explicada pela  $j$ -ésima componente será determinada por

$$\frac{\hat{\lambda}_j}{p} \quad j = 1, 2, \dots, p \quad (2.65)$$

## 2.7 Distribuições assintóticas para dados normais

Para grandes amostras os seguintes resultados assintóticos, baseados no Teorema Central do Limite (MARDIA; KENT & BIBBY [33]), proporcionam distribuições úteis para os autovalores e autovetores da matriz covariância amostral.

**Resultado 2.5** Seja  $\Sigma$  uma matriz positiva definida com autovalores distintos. Seja a matriz  $M$  com distribuição de Wishart, ou seja,  $M \sim W_p(\Sigma, m)$  e  $U = m^{-1}M$ . Considere as decomposições espectrais  $\Sigma = \Gamma \Lambda \Gamma'$  e  $U = GLG'$ , e sejam  $\underline{\lambda} = \text{diag}(\Lambda)$  e  $\underline{l} = \text{diag}(L)$ . Então as seguintes distribuições assintóticas são válidas para  $m \rightarrow \infty$ :

(a)  $\underline{l} \sim N_p(\underline{\lambda}, 2\Lambda^2/m)$ , isto é, os autovalores de  $U$  são assintoticamente normais, não-viciados, e independentes, tendo cada um variância  $2\lambda_i^2/m$ .

(b)  $\underline{g}_{(k)} \sim N_p(\underline{\gamma}_k, \mathbf{V}_k/m)$ , onde  $\mathbf{V}_k = \lambda_k \sum_{j \neq k} \frac{\lambda_j}{(\lambda_j - \lambda_k)^2} \underline{\gamma}_{(j)} \underline{\gamma}'_{(j)}$ , isto é, os autovetores de  $U$  são assintoticamente normais e não-viciados.

(c) A covariância entre o  $r$ -ésimo elemento de  $\underline{g}_{(k)}$  e o  $s$ -ésimo elemento de  $\underline{g}_{(j)}$  é  $-\lambda_k \lambda_j \gamma_{rj} \gamma_{sk} / m (\lambda_k - \lambda_j)^2$ .

(d) Os elementos de  $\underline{l}$  são assintoticamente independentes dos elementos de  $G$ .

## 2.8 Testes de hipóteses acerca das componentes principais

Muitas vezes é útil ter um procedimento estatístico para decidir se as  $k$  primeiras componentes principais incluem toda variação importante de  $\underline{X}$ . Evidentemente que as demais  $(p-k)$  componentes são ignoradas se os correspondentes autovalores populacionais forem todos nulos. Contudo, isso somente acontece se  $\Sigma$  tem posto  $k$ . Essa é uma situação trivial encontrada na prática.

Uma segunda possibilidade consiste em testar a hipótese de que a proporção da variância explicada pelas últimas  $(p-k)$  componentes principais é menor que um determinado valor.

Uma outra hipótese conveniente envolve a questão da igualdade entre os últimos  $(p-k)$  autovalores. Isso implica que a variação é a mesma em todas as direções do espaço  $(p-k)$ -dimensional gerado pelos correspondentes  $(p-k)$  autovetores. Essa é uma situação de variação *isotrópica*, e implicará que se uma dessas componentes é descartada, então todas as últimas  $(p-k)$  componentes deverão ser descartadas.

No desenvolvimento de cada teste é assumido que a amostra aleatória é normal de tamanho  $n$ .

### (1) Testar a hipótese

$$H_0: (\lambda_1 + \lambda_2 + \dots + \lambda_k) / (\lambda_1 + \lambda_2 + \dots + \lambda_p) = \Psi$$

Sejam  $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p$  os autovalores de  $S$  e o valor amostral de  $\Psi$  dado por  $\hat{\Psi} = (\hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_k) / (\hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_p)$ . Do Resultado (2.5 - a) sabe-se que os elementos de  $\hat{\lambda}$  são assintoticamente normais. De acordo com (MARDIA; KENT & BIBBY [33]) tem-se que  $\hat{\Psi}$  é assintoticamente normal com média  $\Psi$  e variância

$$\begin{aligned}\tau^2 &= \frac{R}{(n-1)(\text{tr}\Sigma)^2} \left\{ (1-\Psi)^2 (\lambda_1^2 + \lambda_2^2 + \dots + \lambda_k^2)^2 + \Psi^2 (\lambda_{k+1}^2 + \dots + \lambda_p^2) \right\} \\ &= \frac{2\text{tr}\Sigma^2}{(n-1)(\text{tr}\Sigma)^2} (\Psi^2 - 2\alpha\Psi + \alpha) .\end{aligned}\quad (2.66)$$

onde  $\alpha = (\lambda_1^2 + \lambda_2^2 + \dots + \lambda_k^2) / (\lambda_1^2 + \lambda_2^2 + \dots + \lambda_p^2)$ . Note que  $\alpha$  deve ser interpretada como a proporção de variância total explicada pelas  $k$  componentes principais de uma variável cuja matriz covariância é  $\Sigma^2$ .

(2) Testar a hipótese de que os últimos  $(p-k)$  autovalores de  $\Sigma$  são iguais, ou seja

$$H_0: \lambda_{k+1} = \lambda_{k+2} = \dots = \lambda_{p-1} = \lambda_p$$

Esse teste é útil para determinar o número de componentes principais a ser usado para descrever os dados. Suponha que se tenha decidido pela inclusão de pelo menos  $k$  componentes principais, e deseja-se decidir pela inclusão ou não de qualquer outra componente. A aceitação da hipótese  $H_0$  implica que a inclusão de mais que  $k$  componentes principais deve representar na realidade a inclusão de todas as  $p$  componentes principais, pois cada uma das remanescentes componentes contém a mesma quantidade de informação.

Uma estatística usada para esse teste (MARDIA; KENT & BIBBY [33]), distribuída assintoticamente é dada por

$$\left( n - \frac{2p+11}{6} \right) (p-k) \log \left( \frac{a_0}{g_0} \right) \sim \chi_{(p-k+2)(p-k-1)/2}^2 \quad (2.67)$$

onde:  $a_0$  e  $g_0$  são, respectivamente, as médias aritmética e geométrica de  $\hat{\lambda}_{k+1}, \hat{\lambda}_{k+2}, \dots, \hat{\lambda}_p$ .

### 3 FUNDAMENTOS TEÓRICOS DA ANÁLISE DE COMPONENTES PRINCIPAIS NA DETECÇÃO E IDENTIFICAÇÃO DE *OUTLIERS*

#### 3.1 Introdução

Seja  $\underline{Y}' = (Y_1, Y_2, \dots, Y_p)$  um vetor aleatório tal que  $\underline{Y} \sim N_p(\mu_{\underline{Y}}, \Sigma_{\underline{Y}})$ . Considerando-se que foram realizadas  $n$  observações de  $\underline{Y}$ , a matriz  $Y$  dessas observações pode ser representada como

$$Y = \begin{bmatrix} Y_{11} & Y_{12} & \dots & Y_{1p} \\ Y_{21} & Y_{22} & \dots & Y_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{n1} & Y_{n2} & \dots & Y_{np} \end{bmatrix} \quad (3.1)$$

onde

$$\underline{Y}'_i = (Y_{i1}, Y_{i2}, \dots, Y_{ip}) \quad (3.2)$$

representa a  $i$ -ésima observação  $p$ -variada. De uma maneira geral a matriz  $Y$  representa um conjunto de  $n$  pontos observados em um espaço  $p$ -dimensional.

**Resultado 3.1** Para o vetor aleatório  $\underline{Y}$  com uma certa distribuição conjunta de probabilidade com vetor de médias  $\mu_{\underline{Y}}$  e matriz de covariâncias  $\Sigma_{\underline{Y}}$ , os estimadores não-viciados do vetor de médias  $\mu_{\underline{Y}}$  e da matriz de covariâncias  $\Sigma_{\underline{Y}}$  são, respectivamente,

$$\bar{\underline{Y}} = \frac{1}{n} \sum_{i=1}^n \underline{Y}_i \quad (3.3)$$



e

$$S_Y = \frac{1}{n-1} \sum_{i=1}^n (\underline{Y}_i - \bar{Y})(\underline{Y}_i - \bar{Y})' \quad (3.4)$$

*Prova*

$$E(\bar{Y}) = E\left(\frac{1}{n} \sum_{i=1}^n \underline{Y}_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n \underline{Y}_i\right) = \frac{1}{n} \sum_{i=1}^n E(\underline{Y}_i) = \frac{1}{n} n \mu_Y = \mu_Y$$

$$\begin{aligned} \text{Seja } \hat{\Sigma}_Y &= \frac{1}{n} \sum_{i=1}^n (\underline{Y}_i - \bar{Y})(\underline{Y}_i - \bar{Y})' = \frac{1}{n} \sum_{i=1}^n \underline{Y}_i \underline{Y}_i' - \bar{Y} \bar{Y}' \\ &= \frac{1}{n} \sum_{i=1}^n (\underline{Y}_i - \mu_Y)(\underline{Y}_i - \mu_Y)' - (\bar{Y} - \mu_Y)(\bar{Y} - \mu_Y)' \\ &= \left(\frac{1}{n} - \frac{1}{n^2}\right) \sum_{i=1}^n (\underline{Y}_i - \mu_Y)(\underline{Y}_i - \mu_Y)' - \frac{1}{n^2} \sum_{i \neq j} (\underline{Y}_i - \mu_Y)(\underline{Y}_j - \mu_Y)', \end{aligned}$$

como

$$E\{(\underline{Y}_i - \mu_Y)(\underline{Y}_j - \mu_Y)'\} = \underline{0} \quad \text{para } i \neq j, \text{ então}$$

$$E(\hat{\Sigma}_Y) = n \left(\frac{1}{n} - \frac{1}{n^2}\right) \Sigma_Y = \frac{n-1}{n} \Sigma_Y$$

Portanto

$$E(S_Y) = \frac{n}{n-1} E(\hat{\Sigma}_Y) = \Sigma_Y \quad \square$$

Considere a existência de um dado sub-espço  $k$ -dimensional do espaço  $p$ -dimensional no qual os  $n$  pontos tendem a se comportar conforme um modelo específico. O exame do ajustamento desses pontos ao modelo no sub-espço  $k$ -dimensional pode revelar que algumas observações são discordantes, ou seja, situam-se muito distantes do esperado. Estas observações serão denominadas de *outliers*. A inspeção dessas observações é feita através da análise das projeções ortogonais das mesmas sobre o sub-espço  $(p-k)$ -dimensional perpendicular ao sub-espço  $k$ -dimensional onde essas observações se ajustam. Como mostra (ADUOL [01]), o sub-espço  $k$ -dimensional para o melhor ajuste tem seus eixos ortogonais centrados no vetor  $\bar{Y}$  e orientados de tal forma que a soma dos quadrados das distâncias dos pontos observados a esses  $k$  eixos é um mínimo.

Para ilustrar, considere  $n=12$  observações num espaço bi-dimensional ( $p=2$ ), como mostra a Figura 3.1.

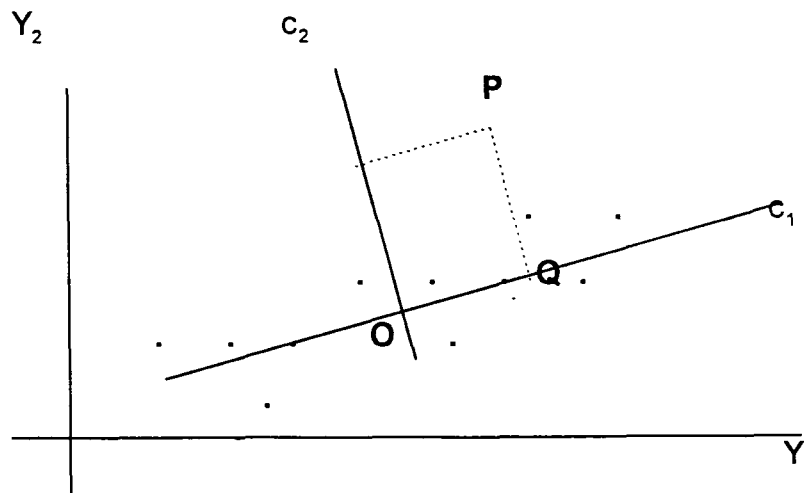


Fig. 3.1 - Observações num espaço bi-dimensional

O ponto O indica o ponto cujo vetor posição é  $\bar{Y}$  enquanto  $c_1$  e  $c_2$  são denominados eixos principais. O eixo  $c_1$  em relação ao qual se tem o melhor ajuste ( $k=1$ ) é denominado de primeiro eixo principal enquanto  $c_2$  é o segundo eixo principal.

Projetando ortogonalmente o ponto P sobre  $c_2$ , então PQ representa a coordenada do ponto P em relação ao eixo  $c_2$ . O exame das coordenadas de todos os pontos em relação a  $c_2$  indica que P é uma observação discordante das demais, ou seja um possível *outlier*. Esse procedimento pode ser naturalmente estendido para o caso p-dimensional, onde nesse caso a distância projetada (correspondente a PQ no caso bi-dimensional) será dada pela raiz quadrada da soma dos quadrados das coordenadas do ponto no sub-espço (p-k)-dimensional.

Como se pode demonstrar (ADUOL [01]), no caso p-dimensional a distância projetada para o j-ésimo ponto é determinada por

$$\|d_j\| = \sqrt{\sum_{i=k+1}^p (e_i' \hat{v}_j)^2} \quad (3.5)$$

onde:  $e_i$ , ( $i = k+1, k+2, \dots, p$ ) são os (p-k) autovetores de  $\hat{\Sigma}_Y$ ,

$$\hat{v}_j = \underline{Y}_j - \underline{\hat{Y}} \quad e$$

$e_{k+1} \hat{v}_j, e_{k+2} \hat{v}_j, \dots, e_p \hat{v}_j$  são as (p-k) componentes principais do vetor  $\hat{v}_j$ .

Na análise dos dados para a identificação de possíveis *outliers*, o interesse será concentrado naqueles pontos que apresentam altos valores para  $\|d_j\|$ .

### 3.2 O modelo de Gauss-Markov

Para o desenvolvimento do método proposto será adotado o modelo de Gauss-Markov

$$\underline{Y} = A\underline{X} + \underline{V} \quad (3.6)$$

onde:  $\underline{Y}$  = vetor das observações, de dimensão  $n$ ,

$A$  = matriz do modelo ou dos coeficientes dos parâmetros, de ordem  $n \times u$ , e posto (*rank*)  $u$ ,

$\underline{X}$  = vetor dos parâmetros, de dimensão  $u$ ,

$\underline{V}$  = vetor dos erros das observações, de dimensão  $n$ .

No modelo de Gauss-Markov (3.6), assume-se que

$$E(\underline{V}) = \underline{0}, \quad (3.7)$$

portanto

$$E(\underline{Y}) = E(A\underline{X} + \underline{V}) = A\underline{X} \quad (3.8)$$

onde

$$V(\underline{Y}) = V(\underline{V}) = \Sigma$$

com

$$\Sigma_{\underline{Y}} = \Sigma_{\underline{V}} = \sigma_0^2 P^{-1} \quad (3.9)$$

sendo:  $\sigma_0^2$  = variância da unidade de peso (variância a priori).

$P$  = matriz positiva definida de ordem  $n \times n$ , dos pesos das observações.

O modelo estimado de (3.6) pode ser escrito como

$$\underline{\hat{Y}} = A\underline{\hat{X}} \quad (3.10)$$

sendo  $\hat{\underline{X}}$  a estimativa do vetor dos parâmetros  $\underline{X}$  e  $\hat{\underline{Y}}$  de  $E(\underline{Y})$ .

### 3.3 Ajustamento pelo método dos mínimos quadrados

**Resultado 3.2** O estimador de mínimos quadrados ponderados do vetor dos parâmetros  $\underline{X}$  do modelo (3.6) é dado por

$$\hat{\underline{X}} = (\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}\underline{Y} \quad (3.11)$$

com matriz covariância

$$\Sigma_{\hat{\underline{X}}} = \hat{\sigma}_0^2(\underline{A}'\underline{P}\underline{A})^{-1} \quad (3.12)$$

e matriz covariância dos resíduos

$$\Sigma_{\hat{\underline{Y}}} = \hat{\sigma}_0^2[\underline{P}^{-1} - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'] \quad (3.13)$$

*Prova*

A determinação de  $\hat{\underline{X}}$  é feita minimizando a forma quadrática fundamental

$$\begin{aligned} \Phi &= \underline{V}'\underline{P}\underline{V} = (\underline{Y} - \underline{A}\underline{X})'\underline{P}(\underline{Y} - \underline{A}\underline{X}) \\ &= (\underline{Y}' - \underline{X}'\underline{A}')\underline{P}(\underline{Y} - \underline{A}\underline{X}) \\ &= \underline{Y}'\underline{P}\underline{Y} - \underline{Y}'\underline{P}\underline{A}\underline{X} - \underline{X}'\underline{A}'\underline{P}\underline{Y} + \underline{X}'\underline{A}'\underline{P}\underline{A}\underline{X} \\ &= \underline{Y}'\underline{P}\underline{Y} - 2\underline{X}'\underline{A}'\underline{P}\underline{Y} + \underline{X}'\underline{A}'\underline{P}\underline{A}\underline{X} \end{aligned}$$

que derivando em relação a  $\underline{X}$  e igualando a zero resulta

$$\frac{\partial \phi}{\partial \underline{X}} = -2\underline{A}'\underline{P}\underline{Y} + 2\underline{A}'\underline{P}\underline{A}\hat{\underline{X}} = 0$$

cuja solução fornece

$$\hat{\underline{X}} = (\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}\underline{Y}$$

Aplicando a lei de propagação das covariâncias para a (3.11), obtém-se

$$\begin{aligned}\Sigma_{\hat{\underline{X}}} &= (\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}\Sigma_{\underline{Y}}[(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}]' \\ &= (\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}\hat{\sigma}_0^2\underline{P}^{-1}\underline{P}\underline{A}(\underline{A}'\underline{P}\underline{A})^{-1} \\ &= \hat{\sigma}_0^2(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}\underline{A}(\underline{A}'\underline{P}\underline{A})^{-1} \\ &= \hat{\sigma}_0^2(\underline{A}'\underline{P}\underline{A})^{-1}\end{aligned}$$

Como  $\hat{\underline{V}} = \underline{Y} - \underline{A}\hat{\underline{X}}$ , substituindo  $\hat{\underline{X}}$  pela (3.11) resulta

$$\hat{\underline{V}} = [\underline{I} - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}]\underline{Y}$$

que aplicando a lei de propagação de covariâncias

$$\begin{aligned}\Sigma_{\hat{\underline{V}}} &= [\underline{I} - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}]\Sigma_{\underline{Y}}[\underline{I} - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}]' \\ &= [\underline{I} - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}]\hat{\sigma}_0^2\underline{P}^{-1}[\underline{I} - \underline{P}\underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'] \\ &= \hat{\sigma}_0^2[\underline{P}^{-1} - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}' - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}' + \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}\underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'] \\ &= \hat{\sigma}_0^2[\underline{P}^{-1} - 2\underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}' + \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'] \\ &= \hat{\sigma}_0^2[\underline{P}^{-1} - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}']\end{aligned}$$

□

### 3.4 Componentes principais dos resíduos

Seja o vetor dos resíduos estimados

$$\underline{\hat{V}}' = [\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n]' \quad (3.14)$$

sendo :  $\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n$  n variáveis correlacionadas onde

$$E(\underline{\hat{V}}) = \underline{0} \quad (3.15)$$

$$\underline{\Sigma}_{\hat{v}} = \hat{\sigma}_0^2 [P^{-1} - A(A'PA)^{-1}A'] \quad (3.16)$$

Para o modelo adotado, o número de graus de liberdade é  $k = n - u$ , que de acordo com propriedade dos autovalores (MORRISON [39]), pode-se concluir que existem k autovalores não-nulos na matriz  $\underline{\Sigma}_{\hat{v}}$ . Portanto, pode-se determinar k componentes principais dos resíduos. O vetor  $\underline{\hat{c}}$  das componentes principais dos resíduos estimados, de acordo com o Resultado 2.1, é dado por

$$\underline{\hat{c}} = E'_{\hat{v}_k} \underline{\hat{V}} \quad (3.17)$$

onde

$$\underline{\hat{c}} = \begin{bmatrix} \hat{c}_1 \\ \hat{c}_2 \\ \vdots \\ \hat{c}_k \end{bmatrix} \quad (3.18)$$

e  $E_{\hat{v}_k}$  = matriz de ordem  $n \times k$  cujas colunas são os k autovetores normalizados da matriz  $\underline{\Sigma}_{\hat{v}}$ , associados aos k autovalores não-nulos, ou seja,

$$E_{\hat{V}_k} = \begin{bmatrix} \hat{e}_{11} & \hat{e}_{21} & \dots & \hat{e}_{k1} \\ \hat{e}_{12} & \hat{e}_{22} & \dots & \hat{e}_{k2} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{e}_{1n} & \hat{e}_{2n} & \dots & \hat{e}_{kn} \end{bmatrix} \quad (3.19)$$

Para o vetor das componentes principais tem-se que

$$E(\hat{C}) = E(E'_{\hat{V}_k} \hat{V}) = \underline{0} \quad (3.20)$$

e matriz de covariâncias, de acordo com Resultado 2.1, dada por

$$\Sigma_{\hat{C}} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{bmatrix} \quad (3.21)$$

onde:  $\lambda_1, \lambda_2, \dots, \lambda_k$  são os  $k$  autovalores não-nulos da matriz  $\Sigma_{\hat{V}}$ .

### 3.5 Resíduos padronizados

**Resultado 3.3** O vetor dos resíduos  $\hat{V}$  pode ser padronizado, obtendo-se o vetor  $\hat{Z}$  tal que

$$\hat{Z} = \Sigma_{D\hat{V}}^{-1/2} \hat{V} \quad (3.22)$$

onde



$$\underline{\hat{z}} = \begin{bmatrix} \hat{z}_1 \\ \hat{z}_2 \\ \vdots \\ \hat{z}_n \end{bmatrix} \quad (3.23)$$

e

$$\Sigma_{\underline{\hat{v}}} = \text{diag}(\Sigma_{\hat{v}}) \quad (3.24)$$

com o vetor  $\underline{\hat{z}}$  tendo

$$E(\underline{\hat{z}}) = E(\Sigma_{\underline{\hat{v}}}^{-1/2} \underline{\hat{V}}) = \underline{0} \quad (3.25)$$

e matriz de covariâncias

$$\Sigma_{\underline{\hat{z}}} = \rho_{\underline{\hat{V}}} = \begin{bmatrix} 1 & \hat{\rho}_{12} & \dots & \hat{\rho}_{1n} \\ \hat{\rho}_{21} & 1 & \dots & \hat{\rho}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{n1} & \hat{\rho}_{n2} & \dots & 1 \end{bmatrix} \quad (3.26)$$

ou seja  $\Sigma_{\underline{\hat{z}}}$ , é a própria matriz de correlação de  $\underline{\hat{V}}$ , aqui denotada por  $\rho_{\underline{\hat{V}}}$ .

*Prova*

Considerando

$$\hat{z}_1 = \frac{\hat{v}_1}{\sigma_{\hat{v}_1}}$$

$$\hat{z}_2 = \frac{\hat{v}_2}{\sigma_{\hat{v}_2}}$$

$$\vdots$$

$$\hat{z}_n = \frac{\hat{v}_n}{\sigma_{\hat{v}_n}}$$

que na forma matricial fica

$$\begin{bmatrix} \hat{z}_1 \\ \hat{z}_2 \\ \vdots \\ \hat{z}_n \end{bmatrix} = \begin{bmatrix} 1/\sigma_{\hat{v}_1} & 0 & \dots & 0 \\ 0 & 1/\sigma_{\hat{v}_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sigma_{\hat{v}_n} \end{bmatrix} \begin{bmatrix} \hat{v}_1 \\ \hat{v}_2 \\ \vdots \\ \hat{v}_n \end{bmatrix}$$

ou

$$\underline{\hat{z}} = \Sigma_{\underline{\hat{v}}}^{-1/2} \underline{\hat{v}}$$

Para  $E(\underline{\hat{z}})$  tem-se que

$$E(\underline{\hat{z}}) = E(\Sigma_{\underline{\hat{v}}}^{-1/2} \underline{\hat{v}}) = \Sigma_{\underline{\hat{v}}}^{-1/2} E(\underline{\hat{v}}) = \underline{0}$$

e para a matriz de covariâncias

$$\Sigma_{\underline{\hat{z}}} = \text{Cov}(\Sigma_{\underline{\hat{v}}}^{-1/2} \underline{\hat{v}}) = \Sigma_{\underline{\hat{v}}}^{-1/2} \text{Cov}(\underline{\hat{v}}) \Sigma_{\underline{\hat{v}}}^{-1/2} = \Sigma_{\underline{\hat{v}}}^{-1/2} \Sigma_{\underline{\hat{v}}} \Sigma_{\underline{\hat{v}}}^{-1/2}$$

$$\Sigma_{\hat{\underline{z}}} = \rho_{\hat{\underline{y}}} = \begin{bmatrix} 1 & \hat{\rho}_{12} & \dots & \hat{\rho}_{1n} \\ \hat{\rho}_{21} & 1 & \dots & \hat{\rho}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{n1} & \hat{\rho}_{n2} & \dots & 1 \end{bmatrix}$$

□

### 3.6 Componentes principais dos resíduos padronizados

O vetor  $\hat{\underline{y}}$  das  $k$  componentes principais obtidas do vetor dos resíduos estimados padronizados  $\hat{\underline{z}}$ , de acordo com o Resultado 2.1, é dado por

$$\hat{\underline{y}} = E'_{\hat{\underline{z}}_k} \hat{\underline{z}} \quad (3.27)$$

onde

$$\hat{\underline{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_k \end{bmatrix} \quad (3.28)$$

e  $E_{\hat{\underline{z}}_k}$  = matriz de ordem  $n \times k$  dos  $k$  autovetores normalizados da matriz  $\Sigma_{\hat{\underline{z}}} = \rho_{\hat{\underline{y}}}$  associados aos correspondentes  $k$  autovalores não-nulos.

**Resultado 3.4** Para o vetor  $\hat{\underline{y}}$  das componentes principais dos resíduos padronizados tem-se que

$$E(\hat{\underline{y}}) = E(E'_{\hat{\underline{z}}_k} \hat{\underline{z}}) = \underline{0} \quad (3.29)$$

e

$$\Sigma_{\hat{\underline{y}}} = \Lambda_{\hat{\underline{z}}} = \begin{bmatrix} \hat{\lambda}_1 & 0 & \dots & 0 \\ 0 & \hat{\lambda}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \hat{\lambda}_k \end{bmatrix} \quad (3.30)$$

onde  $\Lambda_{\hat{\underline{z}}}$  = matriz diagonal dos k autovalores não-nulos da matriz  $\rho_{\hat{\underline{V}}}$ .

*Prova*

$$E(\hat{\underline{y}}) = E(E'_{\hat{\underline{z}}k} \hat{\underline{z}}) = E'_{\hat{\underline{z}}k} E(\hat{\underline{z}}) = \underline{0}$$

e para a matriz de covariâncias, de acordo com o Resultado 2.1, tem-se que

$$\Sigma_{\hat{\underline{y}}} = \text{Cov}(E'_{\hat{\underline{z}}k} \hat{\underline{z}}) = \begin{bmatrix} \sigma_{\hat{\gamma}_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\hat{\gamma}_2}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_{\hat{\gamma}_k}^2 \end{bmatrix} = \begin{bmatrix} \hat{\lambda}_1 & 0 & \dots & 0 \\ 0 & \hat{\lambda}_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \hat{\lambda}_k \end{bmatrix}$$

□

### 3.7 Componentes principais padronizadas

**Resultado 3.5** O vetor  $\hat{\underline{z}}$  das k componentes principais padronizadas do vetor  $\hat{\underline{V}}$  pode ser determinado por

$$\hat{\underline{z}} = \Lambda_{\hat{\underline{z}}}^{-1/2} \hat{\underline{y}} \quad (3.31)$$

onde

$$\underline{\hat{s}} = \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \vdots \\ \hat{s}_k \end{bmatrix} \quad (3.32)$$

com

$$E(\underline{\hat{s}}) = \underline{0} \quad (3.33)$$

e matriz de covariâncias igual a uma matriz identidade de ordem k, ou seja,

$$\Sigma_{\underline{\hat{s}}} = I_k \quad (3.34)$$

*Prova*

Padronizando  $\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_k$  e considerando o Resultado 3.4, obtém-se

$$\hat{s}_1 = \frac{\hat{\gamma}_1}{\sigma_{\hat{\gamma}_1}} = \frac{\hat{\gamma}_1}{\sqrt{\hat{\lambda}_1}}$$

$$\hat{s}_2 = \frac{\hat{\gamma}_2}{\sigma_{\hat{\gamma}_2}} = \frac{\hat{\gamma}_2}{\sqrt{\hat{\lambda}_2}}$$

$$\vdots$$

$$\hat{s}_k = \frac{\hat{\gamma}_k}{\sigma_{\hat{\gamma}_k}} = \frac{\hat{\gamma}_k}{\sqrt{\hat{\lambda}_k}}$$

que em notação matricial fica

$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \vdots \\ \hat{s}_k \end{bmatrix} = \begin{bmatrix} 1/\sqrt{\hat{\lambda}_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\hat{\lambda}_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1/\sqrt{\hat{\lambda}_k} \end{bmatrix} \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_k \end{bmatrix}$$

ou

$$\underline{\hat{s}} = \Lambda_{\hat{z}}^{-1/2} \underline{\hat{y}}$$

Para a média de  $\underline{\hat{s}}$  tem-se que

$$E(\underline{\hat{s}}) = E(\Lambda_{\hat{z}}^{-1/2} \underline{\hat{y}}) = \Lambda_{\hat{z}}^{-1/2} E(\underline{\hat{y}}) = \underline{0}$$

e para a matriz de covariâncias

$$\begin{aligned} \Sigma_{\hat{s}} &= V(\Lambda_{\hat{z}}^{-1/2} \underline{\hat{y}}) \\ &= \Lambda_{\hat{z}}^{-1/2} \Sigma_{\hat{y}} \Lambda_{\hat{z}}^{-1/2} \\ &= \Lambda_{\hat{z}}^{-1/2} \Lambda_{\hat{z}} \Lambda_{\hat{z}}^{-1/2} \\ &= I_k \end{aligned}$$

□

Substituindo (3.27) em (3.31) obtém-se

$$\underline{\hat{s}} = \Lambda_{\hat{z}}^{-1/2} E'_{\hat{z}k} \underline{\hat{z}} \quad (3.35)$$

ou ainda, substituindo (3.22) em (3.35) resulta

$$\underline{\hat{s}} = \Lambda_{\hat{z}}^{-1/2} E'_{\hat{z}k} \Sigma_{D\hat{y}}^{-1/2} \underline{\hat{v}} \quad (3.36)$$

### 3.8 A estatística qui-quadrado e a detecção de *outliers*

Como  $\hat{\underline{s}}$  é o vetor das componentes principais padronizadas tem-se, sob a hipótese de normalidade, a estatística qui-quadrado

$$\begin{aligned}\chi^2 &= \hat{s}_1^2 + \hat{s}_2^2 + \dots + \hat{s}_k^2 \\ &= \sum_{i=1}^k \hat{s}_i^2\end{aligned}\quad (3.37)$$

com  $v = k - 1$  graus de liberdade (ADUOL [01]).

**Resultado 3.6** O vetor  $\hat{\underline{s}}$  das componentes principais padronizadas pode ser colocado na forma

$$\hat{\underline{s}} = \underline{G}'\underline{Y} \quad (3.38)$$

onde

$$\underline{G} = \frac{1}{\sqrt{\hat{\sigma}_0^2}} \underline{P}\underline{Q}'_{\underline{Y}} \underline{W}_{\underline{D}\underline{Y}}^{-1/2} \underline{E}_{\underline{z}k} \underline{\Lambda}_{\underline{z}}^{-1/2} \quad (3.39)$$

com

$$\underline{\Sigma}_{\underline{Y}} = \hat{\sigma}_0^2 \underline{Q}_{\underline{Y}} \quad (3.40)$$

e

$$\underline{W}_{\underline{D}\underline{Y}} = \frac{1}{\hat{\sigma}_0^2} \text{diag}(\underline{\Sigma}_{\underline{Y}}) = \frac{1}{\hat{\sigma}_0^2} \underline{\Sigma}_{\underline{D}\underline{Y}} \quad (3.41)$$

*Prova*

Voltando ao modelo estimado de Gauss-Markov (3.10) e aos resíduos

$$\begin{aligned}\hat{\underline{V}} &= \underline{Y} - \underline{A}\hat{\underline{X}} \\ &= \underline{Y} - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}'\underline{P}\underline{Y}\end{aligned}\quad (3.42)$$

e como

$$\Sigma_{\hat{\underline{V}}} = \hat{\sigma}_0^2 \underline{Q}_{\hat{\underline{V}}}$$

então

$$\underline{Q}_{\hat{\underline{V}}} = \underline{P}^{-1} - \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}' \quad (3.43)$$

$$\therefore \underline{P}^{-1} - \underline{Q}_{\hat{\underline{V}}} = \underline{A}(\underline{A}'\underline{P}\underline{A})^{-1}\underline{A}' \quad (3.44)$$

Substituindo (3.44) em (3.42) obtém-se

$$\begin{aligned}\hat{\underline{V}} &= \underline{Y} - (\underline{P}^{-1} - \underline{Q}_{\hat{\underline{V}}})\underline{P}\underline{Y} \\ &= \underline{Y} - (\underline{Y} - \underline{Q}_{\hat{\underline{V}}}\underline{P}\underline{Y}) \\ &= \underline{Q}_{\hat{\underline{V}}}\underline{P}\underline{Y}\end{aligned}\quad (3.45)$$

que substituindo em (3.36) resulta

$$\hat{\underline{s}} = \underline{\Lambda}_{\hat{\underline{z}}}^{-1/2} \underline{E}'_{\hat{\underline{z}}} \Sigma_{\underline{D}\hat{\underline{V}}}^{-1/2} \underline{Q}_{\hat{\underline{V}}}\underline{P}\underline{Y} \quad (3.46)$$

Da (3.41) obtém-se



$$\Sigma_{D\hat{Y}} = \hat{\sigma}_0^2 W_{D\hat{Y}} \quad (3.47)$$

ou

$$\Sigma_{D\hat{Y}}^{-1/2} = \frac{1}{\sqrt{\hat{\sigma}_0^2}} W_{D\hat{Y}}^{-1/2} \quad (3.48)$$

que substituindo em (3.46), resulta

$$\hat{\underline{S}} = \frac{1}{\sqrt{\hat{\sigma}_0^2}} \Lambda_{\hat{z}}^{-1/2} E'_{\hat{z}k} W_{D\hat{Y}}^{-1/2} Q_{\hat{Y}} P \underline{Y} \quad (3.49)$$

Como

$$G = \frac{1}{\sqrt{\hat{\sigma}_0^2}} P Q_{\hat{Y}} W_{D\hat{Y}}^{-1/2} E_{\hat{z}k} \Lambda_{\hat{z}}^{-1/2}$$

a (3.49) pode ser escrita como

$$\hat{\underline{S}} = G' \underline{Y} \quad \square$$

A existência de erro em qualquer observação possibilita a decomposição de

$\hat{\underline{S}}$  em duas componentes, ou seja,

$$\begin{aligned} \hat{\underline{S}} &= \hat{\underline{S}}_0 + \delta \hat{\underline{S}} \\ &= G'(\underline{Y} + \delta \underline{Y}) \end{aligned} \quad (3.50)$$

sendo

$$\delta \hat{\underline{S}} = G' \delta \underline{Y} \quad (3.51)$$

a componente que contém o efeito do erro e

$$\hat{\underline{s}}_0 = G' \underline{Y} \quad (3.52)$$

a componente de  $\hat{\underline{s}}$  isenta de erro.

Não existindo *outliers* a estatística

$$\chi^2 = \sum_{i=1}^k \hat{s}_i^2 \quad (3.53)$$

tem o comportamento de uma distribuição qui-quadrado com  $v = k - 1$  graus de liberdade e parâmetros

$$E(\chi^2) = v \quad (3.54)$$

e

$$V(\chi^2) = 2v \quad (3.55)$$

Na presença de *outliers* a distribuição de  $\chi^2$  é não-centrada (ADUOL [01]) com parâmetros

$$E(\chi'^2) = v + \tau_k \quad (3.56)$$

e

$$V(\chi'^2) = 2(v + \tau_k) \quad (3.57)$$

onde  $\tau_k$  é o parâmetro de não-centralidade da distribuição.

O teste para a detecção de *outliers* pode então ser formulado com base nas hipóteses:

$H_0 : \tau_k = 0$  (a presença de *outliers* não é significativa)

$H_1 : \tau_k \neq 0$  (a presença de *outliers* é significativa).

Portanto, para um certo nível de significância  $\alpha$ , a hipótese  $H_0$  será rejeitada quando a estatística

$$\chi^2 = \sum_{i=1}^k \hat{s}_i^2 > \chi_\alpha^2 (k-1), \quad (3.58)$$

o que indica a presença de *outliers*.

### 3.9 A análise da matriz G

Se for verificada a desigualdade (3.58), num certo nível de significância  $\alpha$ , a existência de *outliers* é significativa, sendo que o alto valor dessa estatística conduz à existência de pelo menos um alto valor de  $|\hat{s}_i|$ . Portanto, o maior (ou maiores) valor (ou valores) de  $|\hat{s}_i|$  permite identificar a coluna (ou colunas) da matriz G, que contém os coeficientes de  $\hat{s}_i$ , pois como foi visto,  $\underline{\hat{s}} = G' \underline{Y}$  (ADUOL [01], DALMOLIN & MARQUES [17]).

Seja a matriz G:  $n \times k$ , dada por

$$G = \begin{bmatrix} g_{11} & g_{21} & \cdots & g_{i1} & \cdots & g_{k1} \\ g_{12} & g_{22} & \cdots & g_{i2} & \cdots & g_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_{1n} & g_{2n} & \cdots & g_{in} & \cdots & g_{kn} \end{bmatrix} \quad (3.59)$$

Sendo, por exemplo, a maior componente de  $\hat{\underline{s}}$ , em valor absoluto, a componente  $\hat{s}_i$ , então

$$\begin{aligned}\hat{s}_i &= g_{i1}Y_1 + g_{i2}Y_2 + \dots + g_{in}Y_n \\ &= \sum_{j=1}^n g_{ij}Y_j\end{aligned}\quad (3.60)$$

onde:  $g_{i1}, g_{i2}, \dots, g_{in}$  são os elementos da  $i$ -ésima coluna da matriz  $G$ . A ocorrência de altos valores de  $|\hat{s}_i|$  decorre da influência dos elementos  $g_{ij}$ . Portanto, inicialmente são localizados os elementos  $g_{ij}$  da coluna  $i$  que superam, em valor absoluto, os elementos correspondentes das demais colunas. Separados os elementos  $g_{ij}$  identifica-se as observações que correspondem a cada um deles.

Essas observações são as suspeitas de contaminação por erros grosseiros, de modo que entre elas estão os *outliers*. Se, por exemplo, os elementos separados fossem  $g_{i2}, g_{i5}$  e  $g_{i10}$ , eles indicariam que as observações suspeitas de contaminação que correspondem aqueles elementos são:  $Y_2, Y_5$  e  $Y_{10}$ .

O próximo passo consiste em calcular novamente a estatística  $\chi^2$ , agora com  $k - 1$  componentes principais, ou seja, eliminando-se a  $i$ -ésima componente  $\hat{s}_i$ . Repete-se o teste  $\chi^2$  e, se  $\chi^2 > \chi^2_{\alpha}(k-2)$ , identifica-se como foi visto anteriormente outras observações suspeitas, caso contrário o processo de separação das observações suspeitas está completo.

Como pode-se constatar o processo funciona de forma seqüencial, ou seja, ele deve ser aplicado até que a estatística  $\chi^2$  não seja mais significativa no nível de significância adotado.

### 3.10 O critério de rotação ortomax generalizado

Como foi visto, a análise da matriz  $G$  permite a separação das observações suspeitas de contaminação por erros grosseiros. Porém, a sensibilidade dessa matriz depende da distribuição geométrica e das magnitudes dos erros, além do número de graus de liberdade, de maneira que, uma forma de melhorar o desempenho dessa matriz é otimizar as condições de separação das observações suspeitas de contaminação.

Uma forma de proceder essa otimização consiste em efetuar uma transformação oblíqua das componentes principais, num método conhecido em Análise Fatorial como **Critério Ortomax Generalizado**. Basicamente, esse critério consiste em rotacionar a matriz dos autovetores da matriz covariância (ou correlação) dos resíduos de forma que apenas um elemento para cada linha e coluna da matriz transformada possa ser considerado significativamente diferente de zero. Essa condição é fundamental para que se possa identificar univocamente a observação caracterizada como *outlier*.

Seja  $A$  uma matriz  $n \times r$  a ser rotacionada. A solução pelo critério de rotação ortomax generalizado deve proporcionar uma matriz de transformação ortogonal  $T: r \times r$ , tal que a matriz transformada  $C = AT$  satisfaça às condições requeridas.

A solução do problema assegura uma solução oblíqua através da transformação ortogonal. O desenvolvimento que segue mostra a obtenção da rotação oblíqua de uma dada matriz por rotação ortogonal de matrizes adequadas.

Muitas pesquisas foram realizadas para a formulação de critérios de separação por linhas e colunas com o objetivo de simplificar as matrizes dos fatores (HARRIS & KAISER [24], KAISER [28]). Em 1958, Kaiser [28] sugeriu um critério que maximizava a função

$$Q = \sum_{i=1}^n \left[ \sum_{j=1}^r c_{ij}^4 - \frac{1}{r} \left( \sum_{j=1}^r c_{ij}^2 \right)^2 \right] \quad (3.61)$$

que no caso ortogonal corresponde ao conhecido **Cr terio Quartimax**, caracterizado por ter somente um elemento expressivamente diferente de zero em cada linha.

Um outro crit rio que simplifica a estrutura da matriz C por colunas, consiste na maximiza o da fun o

$$R = \sum_{j=1}^r \left[ \sum_{i=1}^n c_{ij}^4 - \frac{1}{n} \left( \sum_{i=1}^n c_{ij}^2 \right)^2 \right] \quad (3.62)$$

que no caso ortogonal corresponde ao conhecido **Cr terio Varimax**, caracterizado por ter somente um elemento expressivamente diferente de zero em cada coluna.

Os diferentes crit rios podem ser expressos pela equa o geral

$$Z = k_1 Q + k_2 R \quad (3.63)$$

onde: Z = crit rio de rota o satisfazendo as condi es desejadas,

$k_1, k_2$  = pesos adequados de acordo com o m todo que se pretende utilizar (por exemplo, para  $k_1 = 0$  e  $k_2 = 1$  o crit rio seria o Varimax enquanto que  $k_1 = 1$  e  $k_2 = 1$  resultaria num m todo intermedi rio do Quartimax e Varimax).

Substituindo (3.61) e (3.62) em (3.63) resulta, ap s determinadas simplifica es alg bricas, em

$$Z = d_3 \sum_{i=1}^n \sum_{j=1}^r c_{ij}^4 - d_1 \sum_{i=1}^n \left( \sum_{j=1}^r c_{ij}^2 \right)^2 - d_2 \sum_{j=1}^r \left( \sum_{i=1}^n c_{ij}^2 \right)^2 \quad (3.64)$$

$$\text{onde: } d_3 = k_1 + k_2 \quad (3.65)$$

$$d_1 = k_1/r \quad (3.66)$$

$$d_2 = k_2/n \quad (3.67)$$

Sendo  $A$  a matriz a ser rotacionada e  $T$  uma matriz de transformação ortonormal tal que  $C = AT$ , pode-se escrever

$$C = [c_{ij}] = \left[ \sum_{p=1}^r a_{ip} t_{pj} \right] \quad (3.68)$$

onde a matriz  $C$  é uma função da matriz  $T$ .

O objetivo é maximizar  $Z$  sob a restrição de que a matriz  $T = [t_{qs}]$  é uma matriz ortogonal, ou seja,

$$\mu_{jh} = \sum_{k=1}^r t_{kj} t_{kh} = \begin{cases} 1 & \text{se } j = h \\ 0 & \text{se } j \neq h \end{cases} \quad j, h = 1, 2, \dots, r \quad (3.69)$$

Como  $\mu_{jh} = \mu_{hj}$ , existem somente  $r(r+1)/2$  restrições independentes. Para cada restrição dois tipos de equações podem ser formadas. Entre essas,  $r$  equações são da forma

$$\mu_{jj} - 1 = \sum_{k=1}^r t_{kj}^2 - 1 = 0, \quad j = 1, \dots, r \quad (3.70)$$

enquanto  $r(r-1)$  equações tem a forma

$$\mu_{jh} = \sum_{k=1}^r t_{kj} t_{kh} = 0, \quad j \neq h = 1, \dots, r \quad (3.71)$$

De acordo com o método dos multiplicadores de Lagrange, a condição necessária para que  $Z$  em (3.64) apresente ponto de máximo sujeito às restrições vistas em (3.69), é que

$$\frac{\partial Z}{\partial t_{qs}} - \frac{\partial}{\partial t_{qs}} \sum_{j=1}^r \sum_{h=1}^r 2\theta_{jh} \left( \sum_{k=1}^r t_{kj} t_{kh} - \delta_{jh} \right) = 0 \quad (3.72)$$

onde:  $\delta_{jh}$  = delta de Kronecker ( $\delta_{jh} = 1$  se  $j = h$  e  $\delta_{jh} = 0$  se  $j \neq h$ ),

$\theta_{jh}$  = é o  $jh$ -ésimo elemento da matriz simétrica  $\Theta = [\theta_{jh}]$  dos multiplicadores de Lagrange.

Assim, obtém-se a derivada

$$\frac{\partial Z}{\partial t_{qs}} = 4d_3 \sum_{i=1}^n c_{is}^3 a_{iq} - 4d_1 \sum_{i=1}^n \left( \sum_{j=1}^r c_{ij}^2 \right) c_{is} a_{iq} - 4d_2 \sum_{i=1}^n c_{is}^2 \left( \sum_{i=1}^n c_{is} a_{iq} \right) \quad (3.73)$$

onde  $\frac{\partial c_{ij}}{\partial t_{qs}} = 0$  se  $j \neq s$  e para a (3.68) resulta que  $\frac{\partial c_{is}}{\partial t_{qs}} = \frac{\partial \sum_{k=1}^r a_{im} t_{ms}}{\partial t_{qs}} = a_{iq}$ .

Por outro lado,

$$\frac{\partial}{\partial t_{qs}} \sum_{j=1}^r \sum_{h=1}^r 2\theta_{jh} \left( \sum_{k=1}^r t_{kj} t_{kh} - \delta_{jh} \right) = 4 \sum_{j=1}^r \theta_{js} t_{qj} \quad (3.74)$$



Substituindo (3.73) e (3.74) em (3.72), dividindo ambos os membros por 4, resulta

$$d_3 \sum_{i=1}^n c_{is}^3 a_{iq} - d_1 \sum_{i=1}^n \left( \sum_{j=1}^r c_{ij}^2 \right) c_{is} a_{iq} - d_2 \sum_{i=1}^n c_{is}^2 \left( \sum_{j=1}^n c_{is} a_{iq} \right) - \sum_{j=1}^r \theta_{js} t_{qj} = 0 \quad (3.75)$$

Definindo as matrizes W, F e H como

$$W = [w_{ij}] = d_3 [c_{ij}^3] \quad (3.76)$$

$$F = d_2 [\text{diag}(C'C)] \quad (3.77)$$

tendo como elementos diagonais

$$f_{ij} = d_2 \left( \sum_{i=1}^n c_{ij}^2 \right) \quad (3.78)$$

e

$$H = d_1 [\text{diag}(CC')] \quad (3.79)$$

tendo como elementos diagonais

$$h_{ii} = d_1 \left( \sum_{j=1}^r c_{ij}^2 \right) \quad (3.80)$$

Substituindo (3.76), (3.78) e (3.80) em (3.75), resulta

$$\sum_{i=1}^n w_{is} a_{iq} - \sum_{i=1}^n h_{ij} c_{is} a_{iq} - f_{ss} \sum_{i=1}^n c_{is} a_{iq} - \sum_{j=1}^r \theta_{js} t_{qj} = 0 \quad (3.81)$$

que pode ser escrita na forma matricial como

$$A'W - A'HC - A'CF - T\Theta = 0 \quad (3.82)$$

Pré-multiplicando os dois membros por  $T'$ , sendo  $T$  uma matriz ortogonal, obtém-se

$$T'A'(W - HC - CF) = \Theta \quad (3.83)$$

Como  $\Theta$  é uma matriz simétrica, isto é  $\Theta = \Theta'$ , então

$$T'A'(W - HC - CF) = (W - HC - CF)'AT \quad (3.84)$$

Pré-multiplicando os dois membros da (3.84) por  $T$  e pós-multiplicando por  $T'$  obtém-se

$$A'(W - HC - CF)T' = T(W - HC - CF)'A \quad (3.85)$$

Com base na (3.85) pode-se desenvolver um algoritmo para calcular  $T$ .

Seja

$$G = A'(W - HC - CF) \quad (3.86)$$

tal que a (3.85) pode ser colocada na forma

$$GT' = TG' \quad (3.87)$$

Pós-multiplicando os dois membros da (3.87) por  $TG'$ , obtém-se

$$GG' = [TG']^2 \quad (3.88)$$

que extraindo a raiz quadrada dos dois membros resulta em

$$[GG']^{1/2} = TG' \quad (3.89)$$

Pré-multiplicando os dois membros por  $T'$  e pós-multiplicando por  $[GG']^{-1/2}$  obtém-se

$$T' = G' [GG']^{-1/2} \quad (3.90)$$

que transpondo, resulta

$$T = [GG']^{-1/2} G \quad (3.91)$$

Essa última equação pode ser usada como base para a obtenção de um algoritmo iterativo (CROSILLA [14]) resumido a seguir.

Primeiro faz-se  $T_{(0)} = I$ , onde  $I$  = matriz identidade  $rxr$ , como valor inicial da matriz de transformação ortonormal  $T_{(g)}$ , onde  $g$  indica a  $g$ -ésima iteração. O procedimento computacional pode ser resumido na seguinte seqüência:

$$C_{(g)} = AT_{(g)} \quad (3.92)$$

$$W_{(g)} = d_3 [c_{i,j}^3]_{(g)} \quad (3.93)$$

$$H_{(g)} = d_1 [\text{diag}(C_{(g)} C'_{(g)})] \quad (3.94)$$

$$F_{(g)} = d_2 [\text{diag}(C'_{(g)} C_{(g)})] \quad (3.95)$$

$$G_{(g)} = A' [W_{(g)} - H_{(g)} C_{(g)} - C_{(g)} F_{(g)}] \quad (3.96)$$

$$T_{(g, 1)} = [G_{(g)} G'_{(g)}]^{-1/2} G_{(g)} \quad (3.97)$$

### 3.11 Teste de discordância de Ferguson

Identificadas as observações suspeitas de contaminação, através da análise da matriz  $G$ , o próximo passo consiste em estimar os vícios dessas observações e os correspondentes erros padrões. De posse dos vícios estimados deve-se decidir se eles indicam a presença ou não de *outliers*. O simples fato de um certo vício estimado ser considerado “pequeno” ou “grande” em valor absoluto, não é suficiente para decidir se a observação correspondente é um *outlier*. A tomada dessa decisão deve ser feita com base num teste estatístico.

Considerando-se que os vícios estimados são normalmente distribuídos com média zero e variância desconhecida, para decidir se um dado vício estimado corresponde a um *outlier*, basta determinar o quociente entre o mesmo e seu erro padrão, sujeitando-o a um teste de hipótese envolvendo a distribuição  $t$  de Student.

Para que a estatística seja confiável é necessário que o erro padrão estimado seja também confiável. Porém, quando o número de observações suspeitas é grande, o número de vícios que devem ser estimados também é grande, o que resulta num pequeno número de graus de liberdade. Como conseqüência, os erros padrões estimados dos vícios tornam-se grandes e instáveis, não sendo confiáveis.

Uma forma de reduzir o número de observações suspeitas é a utilização de um teste de discordância, como é o caso do teste de Ferguson (BARNETT & LEWIS [04]). O teste de Ferguson é adequado para dados normais de média conhecida e variância desconhecida, como é o caso dos vícios estimados.

De uma maneira geral, o teste de discordância de Ferguson utiliza a estatística

$$T = \frac{1}{ns^4} \sum_{i=1}^n (x_i - \mu)^4 \quad (3.98)$$

onde:  $x_i$  =  $i$ -ésimo vício estimado,

$\mu$  = média dos vícios estimados,

$n$  = número de vícios estimados,

$s$  = desvio padrão dos vícios estimados.

Os valores críticos para o teste de Ferguson são encontrados na Referência (BARNETT & LEWIS [04]).

Para aplicar o teste de Ferguson, calcula-se inicialmente a estatística  $T$  que deve ser comparada, num certo nível de significância, com o respectivo valor crítico. Se a estatística  $T$  superar o valor crítico, identifica-se o maior valor absoluto de  $x_i$  como sendo discordante dos demais. Esse valor é separado dos demais e o teste é novamente aplicado para os  $n - 1$  valores restantes. O procedimento deve continuar até que a estatística  $T$  não supere o correspondente valor crítico.

### 3.12 Estimação dos vícios

Para estimar os vícios e seus respectivos erros padrões das observações suspeitas de contaminação por erros grosseiros, adotou-se o modelo matemático funcional

$$\underline{Y} = \underline{A}\underline{X} + \underline{B}\underline{W} + \underline{V} \quad (3.99)$$

que difere do modelo (3.6) em virtude da introdução da matriz  $\underline{B}$  e do vetor  $\underline{W}$ , sendo  $\underline{W}$  = vetor  $p \times 1$  dos  $p$  vícios a serem estimados, de modo que

$$\underline{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} \quad (3.100)$$

e

$B$  = matriz  $n \times p$  dos coeficientes dos vícios.

O modelo (3.99) estimado será

$$\underline{\hat{Y}} = A\underline{\hat{X}} + B\underline{\hat{W}} \quad (3.101)$$

sendo  $\underline{\hat{W}}$  o estimador de  $\underline{W}$  e  $\underline{\hat{Y}}$  de  $E(\underline{Y})$ .

Para refazer o ajustamento para o modelo (3.99), como foi feito para o modelo (3.6), o modelo (3.99) pode ser colocado na forma

$$\underline{Y} = D\underline{U} + \underline{V} \quad (3.102)$$

onde

$$D_{n \times (u+p)} = \begin{bmatrix} A_{n \times u} & B_{n \times p} \end{bmatrix} \quad (3.103)$$

e

$$\underline{U}_{(u+p) \times 1} = \begin{bmatrix} \underline{X}_{u \times 1} \\ \underline{W}_{p \times 1} \end{bmatrix} \quad (3.104)$$

A matriz das equações normais de (3.102) pode ser escrita na forma (ADUOL |01])

$$D'PD = \begin{bmatrix} A'PA & A'PB \\ B'PA & B'PB \end{bmatrix} \quad (3.105)$$

O modelo estimado de (3.102) será

$$\underline{\hat{Y}} = D\underline{\hat{U}} \quad (3.106)$$

sendo  $\underline{\hat{U}}$  o vetor estimado de  $\underline{U}$  e  $\underline{\hat{Y}}$  de  $E(\underline{Y})$ .

A aplicação do método dos mínimos quadrados ponderados conduz à solução

$$\underline{\hat{U}} = (D'PD)^{-1}D'P\underline{Y} \quad (3.107)$$

com matriz covariância

$$\Sigma_{\underline{\hat{U}}} = \sigma_0^2(D'PD)^{-1} \quad (3.108)$$

### 3.13 Erro mínimo detectável

A confiabilidade do modelo está relacionada à capacidade de detecção de *outliers*. A confiabilidade usualmente é considerada sob dois aspectos: interna e externa. A confiabilidade interna é entendida como a capacidade de detecção de erros grosseiros nas observações enquanto que a confiabilidade externa é medida pela influência de erros não detectáveis sobre as coordenadas objeto (SILVA [47]).

Como o método proposto utiliza a Análise de Componentes Principais na detecção e identificação de *outliers*, pode-se então desenvolver os conceitos de confiabilidades interna e externa para a metodologia multivariada.

Foi visto que

$$\underline{s} \sim N(\underline{0}, I_k) \quad (3.109)$$

onde  $\sum_{i=1}^k s_i^2$  tem distribuição qui-quadrado com graus de liberdade  $v = n - u - 1$  ou  $v = k - 1$ .

Supondo a existência de observações contaminadas por erros grosseiros, eles irão se propagar pelas componentes principais de tal forma que a  $i$ -ésima componente principal padronizada será dada como

$$s'_i = s_i + \delta s_i \quad (3.110)$$

onde  $\delta s_i$  representa o erro propagado em  $s_i$ .

Fazendo

$$\begin{aligned} r'_k &= \sum_{i=1}^k s_i'^2 \\ &= \sum_{i=1}^k (s_i + \delta s_i)^2 \end{aligned} \quad (3.111)$$

verifica-se que  $r'_k$  tem uma distribuição qui-quadrado não-central (ADUOL [01]) com parâmetro de não-centralidade  $\tau_k$  dado por

$$\tau_k = \sum_{i=1}^k \delta s_i^2 \quad (3.112)$$

De acordo com a (3.51) pode-se escrever

$$\delta s_i = g_j \delta Y_j \quad (3.113)$$

que substituindo em (3.112) resulta



$$\begin{aligned}\tau_k &= \sum_{i=1}^k (g_{ji} \delta Y_j)^2 \\ &= \delta Y_j^2 \sum_{i=1}^k g_{ji}^2\end{aligned}\quad (3.114)$$

Isolando  $\delta Y_j$  na (3.114), obtém-se

$$\delta Y_j = \sqrt{\frac{\tau_k}{\sum_{i=1}^k g_{ji}^2}} \quad (3.115)$$

O valor de  $\delta Y_j$  é interpretado como sendo o menor erro na observação  $Y_j$  que pode ser detectado com poder  $(1 - \beta_k)$ , ao nível de significância  $\alpha$ . Os valores de  $\delta Y_j$  são tomados como medidas de **confiabilidade interna** do problema.

O vetor de erros  $\delta \underline{Y}$  propaga-se através do correspondente vetor dos parâmetros estimados  $\hat{\underline{X}}$  tal que

$$\delta \hat{\underline{X}} = (\underline{A}' \underline{P} \underline{A})^{-1} \underline{A}' \underline{P} \delta \underline{Y} \quad (3.116)$$

onde os elementos de  $\delta \hat{\underline{X}}$  indicam a **confiabilidade externa** do problema.

## 4 APLICAÇÃO DO MÉTODO MULTIVARIADO EM FOTOTRIANGULAÇÃO

### 4.1 Modelo matemático funcional

O modelo matemático funcional adotado é aquele definido pelas equações de colinearidade (LUGNANI [32], MERCHANT [35])

$$F(x) = x - x_0 - f \frac{(X - X_0)m_{11} + (Y - Y_0)m_{12} + (Z - Z_0)m_{13}}{(X - X_0)m_{31} + (Y - Y_0)m_{32} + (Z - Z_0)m_{33}} = 0 \quad (4.1)$$

$$F(y) = y - y_0 - f \frac{(X - X_0)m_{21} + (Y - Y_0)m_{22} + (Z - Z_0)m_{23}}{(X - X_0)m_{31} + (Y - Y_0)m_{32} + (Z - Z_0)m_{33}} = 0 \quad (4.2)$$

onde:  $f$  = distância focal calibrada da câmara,

$(x, y)$  = fotocoordenadas no espaço imagem,

$(X, Y, Z)$  = coordenadas dos pontos no espaço objeto,

$(X_0, Y_0, Z_0)$  = coordenadas dos centros perspectivos,

$m_{ij}$  = elementos da matriz de rotação  $R(\kappa)R(\phi)R(\omega)$ .

$x_0, y_0$  = fotocoordenadas do ponto principal em relação ao sistema fiducial.

### 4.2 Modelo matemático funcional linearizado

O modelo matemático funcional dado pelas equações de colinearidade pode ser linearizado segundo o desenvolvimento em série de Taylor (GEMAEL [22]), obtendo-se

$$\underline{AX} + \underline{L} = \underline{V} \quad (4.3)$$

onde:  $\underline{L} = \underline{L}_0 - \underline{L}_b$ , vetor de dimensão  $nx1$ ,  
 $\underline{L}_0 = F(\underline{X}_0)$ , vetor de dimensão  $nx1$ ,  
 $\underline{L}_b$  = vetor  $nx1$  das observações,  
 $\underline{X}_0$  = vetor dos valores aproximados dos parâmetros,  
 $\underline{V}$  = vetor  $nx1$  dos resíduos,  
 $\underline{X}$  = vetor  $ux1$  das correções aos parâmetros aproximados,

$$A = \left. \frac{\partial F}{\partial X_a} \right|_{X_a=X_0}, \text{ matriz de dimensão } nxu.$$

### 4.3 Ajustamento livre

Quando não se adota injunções iniciais para solucionar o sistema (4.3) pelo método dos mínimos quadrados a matriz das equações normais apresenta deficiência de *rank* (posto ou característica) ou seja, é singular. Para contornar esta situação optou-se pelo ajustamento livre, com a pseudo-inversa (MOURA [40]) na inversão das equações normais.

O vetor estimado das correções dos parâmetros, de acordo com (3.11), é dado por

$$\hat{\underline{X}} = - (A'PA)^+ A'P\underline{L} \quad (4.4)$$

As matrizes de covariâncias dos resíduos estimados e das correções dos parâmetros, de acordo com (3.13) e (3.12), são respectivamente,

$$\Sigma_{\hat{\underline{V}}} = \hat{\sigma}_0^2 [P^{-1} - A(A'PA)^+ A'] \quad (4.5)$$

$$\Sigma_{\hat{\underline{X}}} = \hat{\sigma}_0^2 (A'PA)^+ \quad (4.6)$$

#### 4.4 Componentes principais padronizadas

Como foi visto em (3.36), pode-se escrever

$$\underline{\hat{s}} = \Lambda_{\hat{z}}^{-1/2} \mathbf{E}'_{\hat{z}_k} \Sigma_{D\hat{V}}^{-1/2} \underline{\hat{V}} \quad (4.7)$$

ou

$$\underline{\hat{s}} = \frac{1}{\sqrt{\hat{\sigma}_0^2}} \Lambda_{\hat{z}}^{-1/2} \mathbf{E}'_{\hat{z}_k} W_{D\hat{V}}^{-1/2} \underline{\hat{V}} \quad (4.8)$$

ou, ainda

$$\underline{\hat{s}} = \mathbf{G}' \underline{\hat{V}} \quad (4.9)$$

com

$$\mathbf{G} = \frac{1}{\sqrt{\hat{\sigma}_0^2}} W_{D\hat{V}}^{-1/2} \mathbf{E}_{\hat{z}_k} \Lambda_{\hat{z}}^{-1/2} \quad (4.10)$$

#### 4.5 Estimação dos vícios das observações suspeitas

A análise da matriz  $\mathbf{G}$  fornece as observações suspeitas de contaminação por erros grosseiros. Os vícios dessas observações podem ser estimados, de acordo com o que foi visto na seção 3.12, adotando-se o modelo

$$\mathbf{A}\underline{X} + \mathbf{B}\underline{W} + \underline{L} = \underline{V} \quad (4.11)$$

O modelo (4.11) estimado será

$$A\hat{\underline{X}} + B\hat{\underline{W}} + \hat{\underline{L}} = \underline{0} \quad (4.12)$$

ou, de acordo com a (3.106)

$$D\hat{\underline{U}} + \hat{\underline{L}} = \underline{0} \quad (4.13)$$

cuja solução pelo método dos mínimos quadrados, resulta de acordo com a (3.107), em

$$\hat{\underline{U}} = - (D'PD)^+ D'P\underline{L} \quad (4.14)$$

com matriz covariância

$$\Sigma_{\hat{\underline{U}}} = \hat{\sigma}_0^2 (D'PD)^+ \quad (4.15)$$

#### 4.6 Teste de discordância de Ferguson

O teste de discordância de Ferguson é aplicado para os vícios estimados das observações separadas como suspeitas através da análise da matriz G.

A estatística (3.98) pode ser escrita como

$$T = \frac{1}{ps^4} \sum_{i=1}^p \hat{w}_i^4 \quad (4.16)$$

onde:  $\hat{w}_i$  = i-ésimo vício estimado,

$p$  = número de vícios estimados separados como suspeitos pela matriz G.

#### 4.7 Aplicação do teste t de Student

Com a redução do número de observações suspeitas de contaminação, pela aplicação do teste de Ferguson, estima-se novamente os vícios dessas observações e seus respectivos erros padrões, definindo-se então a estatística

$$t = \frac{\hat{w}_i}{\sigma_{\hat{w}_i}} \quad (4.17)$$

onde:  $\hat{w}_i$  = i-ésimo vício estimado,  
 $\sigma_{\hat{w}_i}$  = erro padrão do i-ésimo vício estimado.

Para  $t > t_{\alpha}(v)$ , com  $v = n - u - p + r$  graus de liberdade ( $r$  = deficiência de posto ou *rank*), o i-ésimo vício  $\hat{w}_i$  indica que a i-ésima observação correspondente constitui um *outlier*, ao nível de significância  $\alpha$ .

#### 4.8 Estimação final dos vícios

Com a identificação dos *outliers*, pode-se ainda estimar os vícios finais com a utilização dos conceitos vistos na seção 4.5. Com os vícios finais estimados pode-se efetuar as correções dos erros das observações identificadas como *outliers*.

## 5 METODOLOGIA EXPERIMENTAL EM FOTOTRIANGULAÇÃO

### 5.1 Introdução

Para testar o método multivariado proposto, foi elaborado um programa de Fototriangulação sendo o ajustamento realizado pelo método dos feixes de raios homólogos (*bundlemethod*), conjuntamente com os programas que envolvem o método multivariado e os testes estatísticos utilizados. Para elaborar os programas computacionais foi utilizado o *software* matemático MATLAB - versão 3.5 j (386-MATLAB [48]) e o estatístico MINITAB - versão 6.0 (MINITAB [36]). Os programas utilizados são apresentados no Anexo C.

Para realizar os experimentos simulou-se um bloco composto de 6 fotografias com 34 pontos triangulados totalizando 150 pontos imagens. Para esse bloco, o número de equações de observação e o número de parâmetros são dados, respectivamente, por  $n = 2 \times 150 = 300$  e  $u = 6 \times 6 + 3 \times 34 = 138$ .

As fotocoordenadas teóricas foram geradas através das equações de colinearidade (4.1) e (4.2) com a ausência de erros sistemáticos, impondo-se  $f=153,14$  mm e  $x_0 = y_0 = 0$  mm.

As fotocoordenadas observadas simuladas foram obtidas a partir das fotocoordenadas teóricas considerando-se que os erros das observações seguem a distribuição normal com desvio padrão  $\sigma = 0,004$  mm. No Anexo A encontra-se relacionado as observações teóricas, simuladas e ajustadas

Os erros grosseiros introduzidos nas observações foram considerados como tendo grandezas variando no intervalo  $(-30\sigma, +30\sigma)$ . Considerando-se esse intervalo, as grandezas dos erros em cada experimento foram determinadas por processo aleatório, eliminando-se os casos que resultaram em erro abaixo do erro mínimo detectável para o modelo adotado, conforme equação (3.115).

As observações (fotocoordenadas) que foram contaminadas por erros grosseiros, ou seja, os *outliers*, foram determinados também por processo aleatório, de maneira que todas as observações tivessem mesma probabilidade de contaminação.

O número de *outliers* em cada experimento variou de 2 a 5, de forma que em qualquer dos 20 experimentos apresentados o número mínimo de *outliers* foi 2 e o máximo 5.

Não foram aplicadas injunções, optando-se dessa forma pelo ajustamento generalizado ou livre, utilizando-se como inversa da matriz das equações normais a pseudo-inversa  $N^+$  (ALBERT [02], BEN-ISRAEL & GREVILLE [05], BJERHAMMAR [07], BOULLION [09]).

## 5.2 Etapas de cada experimento

Antes da introdução de erros grosseiros nas observações, foi aplicado o teste qui-quadrado com  $\nu = k - 1 = 169 - 1 = 168$  graus de liberdade, que resultou na estatística

$$\chi^2 = 164,150 \quad \text{com} \quad p = \Pr(\chi^2 > 164,150) = 0,5694,$$

ou seja, nos níveis de significâncias usuais (1% ou 5%), a presença de *outliers* não é detectada, como se esperava, pois não houve contaminação.

O desenvolvimento de cada experimento contendo *outliers* foi realizado em 4 etapas:

(1) Determinação (processo aleatório) das observações que serão contaminadas e os correspondentes erros (em mm) impostos a cada uma delas.

(2) Detecção dos *outliers*



- Ajustamento das observações pelo método dos mínimos quadrados ponderados (ajustamento generalizado ou livre).

- Teste qui-quadrado, sendo mostrado a estatística  $\chi^2(k-1) = \sum_{i=1}^k \hat{s}_i^2$  e o valor correspondente de  $p = \Pr[\chi^2 > \chi^2(k-1)]$ , ou seja, a probabilidade da variável  $\chi^2$  superar a estatística calculada. Portanto, num certo nível de significância  $\alpha$ , a hipótese  $H_0$  (a existência de *outliers* não é significativa) será rejeitada se  $p < \alpha$ .

### (3) Identificação dos *outliers*

- Determinação das observações suspeitas pela análise da matriz G.
- Redução do número de observações suspeitas (teste de Ferguson).
- Aplicação do teste t de Student (identificação dos *outliers*).

A terceira etapa foi desenvolvida considerando-se que inicialmente foi aplicado o método multivariado sem rotação dos autovetores da matriz covariância (ou correlações) dos resíduos, sendo o Critério de Rotação Ortomax Generalizado aplicado somente naqueles casos cuja identificação dos *outliers* não foi completa. Dessa forma verifica-se então que se trabalhou conjuntamente com as duas situações.

Quando não foi aplicada rotação os resultados foram apresentados em uma tabela contendo o número das observações selecionadas pelo teste de Ferguson, os vícios estimados  $\hat{w}$  (em mm), os erros padrões  $\sigma_{\hat{w}}$  (em mm) dos vícios estimados, a estatística t e o valor de p correspondente. Sendo aplicada rotação nos autovetores é apresentada uma nova tabela contendo os números das observações selecionadas pela matriz G após a rotação dos autovetores, juntamente com  $\hat{w}$ ,  $\sigma_{\hat{w}}$ , t e p. Quando foi utilizado o Critério de Rotação Ortomax, o teste de Ferguson somente foi aplicado nos casos em que o número de observações suspeitas foi superior a 20.

### (4) Apresentação dos *outliers* identificados pelo *Danish Method*.

### 5.3 Resultados de cada experimento

#### Experimento 5.3.1

##### (1) Observações contaminadas e erros

Quadro 5.1 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
126	0,04078	42,15600	42,19678
248	-0,03153	-81,96900	-82,00053

##### (2) Detecção dos *outliers*

$$\chi^2 = 266,532$$

$$p = 0,000002$$

##### (3) Identificação dos *outliers*

###### (a) Sem rotação dos autovetores

Observações suspeitas: 40, 42, 50, 57, 60, 69, 91, 92, 104, **126**, 132, 134, 178, 211, 231, 264.

Total = 16

Quadro 5.2 - Teste t com  $v=168$

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>126</b>	-0.04041	0.00506	<b>-7,9816</b>	<b>0,00000</b>

Como foi identificada somente a observação 126 como *outlier*, pode-se então efetuar a correção da mesma, ou seja,

$$42,19678 - 0,04041 = 42,15637$$

e após um novo ajustamento determinou-se a estatística para detecção de *outlier*, obtendo-se

$$\chi^2 = 202,826$$

$$p=0,034508$$

que indica ainda a presença de observação contaminada, tomando-se como referência um nível de significância de 5%. Portanto, o próximo passo consiste na aplicação do Método Multivariado com rotação dos autovetores para tornar completa a identificação dos *outliers*. Esse procedimento foi utilizado nos demais experimentos de uma forma automática, ou seja, sempre que a presença de *outlier* continuava significativa após a utilização do método sem rotação, imediatamente procedeu-se à aplicação do Método Multivariado com rotação dos autovetores.

(b) Com rotação dos autovetores

Quadro 5.3 - Teste t com  $v=164$

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
60	0,00554	0,00675	0,8211	0,20638
<b>126</b>	-0,03644	0,00690	<b>-5,2829</b>	<b>0,00000</b>
231	-0,00529	0,00710	-0,7447	0,22875
246	-0,00461	0,00510	-0,9041	0,18363
<b>248</b>	0,03137	0,00522	<b>6,0112</b>	<b>0,00000</b>

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: 37, **126** e **248**.

**Experimento 5.3.2**

## (1) Observações contaminadas e erros

Quadro 5.4 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
158	-0,03522	83,10200	83,06678
164	0,07668	109,12600	109,20268

(2) Detecção dos *outliers*

$$\chi^2 = 451,013$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 29, 30, 49, 57, 66, 76, 78, 87, 91, 96, **158**, **164**, 168, 189, 197, 198, 202, 205, 222, 232, 237, 238, 260, 264, 278, 286.

Total = 26

Quadro 5.5 - Teste t com  $v=167$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>158</b>	0,04102	0,00476	<b>8,6250</b>	<b>0,00000</b>
<b>164</b>	-0,07242	0,00518	<b>-13,9832</b>	<b>0,00000</b>

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **158 e 164**.

### Experimento 5.3.3

(1) Observações contaminadas e erros

Quadro 5.6 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
80	-0,03495	-10,48300	-10,51795
196	-0,03961	-105,74800	-105,78761

(2) Detecção dos *outliers*

$$\chi^2 = 242,772$$

$$p = 0,000142$$

(3) Identificação dos *outliers*

(a) Sem rotação dos autovetores

Observações suspeitas: 10, 40, 42, 50, 77, **80**, 104, 114, 211, 219, 245, 250, 251, 276.

Total = 14

Quadro 5.7 - Teste t com  $v=168$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>80</b>	0,03221	0,00492	<b>6,5442</b>	<b>0,00000</b>

(b) Com rotação dos autovetores

Quadro 5.8 - Teste t com  $v=162$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>80</b>	0,03110	0,00518	<b>6,0082</b>	<b>0,00000</b>
142	-0,01083	0,00512	-2,1142	0,01801
186	0,00041	0,00611	0,0677	0,47305
188	0,00033	0,00552	0,0589	0,47655
<b>196</b>	0,03431	0,00628	<b>5,4619</b>	<b>0,00000</b>
251	0,00026	0,00527	0,0498	0,48017
259	-0,00234	0,00489	-0,4776	0,31678

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **80** e **196**.

#### Experimento 5.3.4

(1) Observações contaminadas e erros

Quadro 5.9 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
74	0,05176	57,34100	57,39276
298	0,03428	2,75100	2,78528

(2) Detecção dos *outliers*

$$\chi^2 = 308,059$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 10, 23, 25, 40, 42, 50, 54, 72, **74**, 88, 96, 104, 120, 188, 211, 227, 245, 276, 278, 282, 286, **298**.

Total = 22

Quadro 5.10 - Teste t com  $v=167$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>74</b>	-0,02609	0,00282	<b>-9,2573</b>	<b>0,00000</b>
<b>298</b>	-0,01730	0,00293	<b>-5,8987</b>	<b>0,00000</b>

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: 37, **74** e **298**.

### Experimento 5.3.5

#### (1) Observações contaminadas e erros

Quadro 5.11 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
70	-0,09939	74,14500	74,04561
240	0,06758	-2,79500	-2,72742

#### (2) Detecção dos *outliers*

$$\chi^2 = 790,237$$

$$p = 0,000000$$

#### (3) Identificação dos *outliers*

##### (a) Sem rotação dos autovetores

Observações suspeitas: 1, 2, 3, 4, 5, 6, 24, 34, 42, 50, 52, 54, 58, 59, 62, 64, 68, 70, 72, 79, 91, 96, 98, 118, 124, 128, 130, 131, 132, 133, 134, 142, 145, 177, 182, 199, 200, 203, 207, 211, 212, 238, 239, **240**, 241, 252, 259, 264, 265, 284.

Total = 50



Quadro 5.12 - Teste t com  $v=159$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
3	-0,01371	0,01157	-1,1847	0,11894
6	-0,13023	0,16419	-0,7932	0,21441
68	0,00000	0,00622	0,0073	0,49709
<b>70</b>	-0,03541	0,16572	<b>-0,2137</b>	<b>0,41552</b>
133	-0,00014	0,01216	-0,0116	0,49538
134	-0,00228	0,00695	-0,3276	0,37182
177	0,13527	0,16199	0,8351	0,20245
239	0,00991	0,00766	1,2936	0,09883
<b>240</b>	-0,07201	0,00515	<b>-13,9874</b>	<b>0,00000</b>
241	0,13517	0,16124	0,8383	0,20155

(b) Com rotação dos autovetores

Quadro 5.13 - Teste t com  $v=161$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
5	0,00608	0,02388	0,2767	0,39118
54	0,01117	0,00486	2,2968	0,01146
67	0,00617	0,00578	1,0667	0,14384
<b>70</b>	0,09807	0,00588	<b>16,6813</b>	<b>0,00000</b>
176	0,01041	0,00592	1,7593	0,04021
177	0,00996	0,01779	0,5601	0,28809
<b>240</b>	-0,06325	0,00567	<b>-11,1541</b>	<b>0,00000</b>
267	-0,01301	0,00572	-2,2726	0,01218

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **70 e 240**.

### Experimento 5.3.6

#### (1) Observações contaminadas e erros

Quadro 5.14 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
24	0,10412	-67,83900	-67,73488
189	0,09132	82,81900	82,91032
297	-0,03638	-17,86900	-17,90538

#### (2) Detecção dos *outliers*

$$\chi^2 = 944,792$$

$$p = 0,000000$$

#### (3) Identificação dos *outliers*

##### (a) Sem rotação dos autovetores

Observações suspeitas: 2, 4, 6, 8, 16, 17, 23, **24**, 25, 26, 35, 37, 42, 49, 50, 70, **74**, 76, 78, 88, 89, 90, 91, 93, 95, 96, 107, 120, 135, 136, 138, 141, 142, 143, 144, 146, 147, 152, 157, 164, 168, 187, 188, **189**, 191, 193, 195, 202, 206, 208, 211, 216, 232, 238, 239, 240, 243, 257, 259, 262, 264, 266, 279, 288, 289, 293, 295, **297**, 298, 300.

Total = 70

Quadro 5.15 - Teste t com  $v=156$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
17	0,02902	0,01995	1,4548	0,07386
23	0,01731	0,20478	0,0845	0,46638
<b>24</b>	-0,29176	0,60849	<b>-0,4795</b>	<b>0,31612</b>
25	0,02192	0,02247	0,9754	0,16543
88	-0,08495	0,29594	-0,2871	0,38720
135	-0,01954	0,01026	-1,9045	0,02934
141	-0,00674	0,09211	-0,0732	0,47087
142	-0,09517	0,29966	-0,3176	0,37560
143	-0,00989	0,01900	-0,5203	0,30179
<b>189</b>	-0,03220	0,00732	<b>-4,3750</b>	<b>0,00001</b>
259	0,08747	0,28153	0,3107	0,37822
289	-0,01699	0,01047	-1,6224	0,05336
295	0,07891	0,21338	0,3698	0,35601

(b) Com rotação dos autovetores

Quadro 5.16 - Teste t com  $v= 160$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>24</b>	-0,11105	0,00526	<b>-21,0999</b>	<b>0,00000</b>
82	0,00155	0,00493	0,3148	0,37666
141	0,00579	0,00858	0,6743	0,25054
143	0,00159	0,00965	0,1648	0,43465
<b>189</b>	-0,08463	0,00565	<b>-14,9915</b>	<b>0,00000</b>
253	0,00632	0,00485	1,3016	0,09745
295	-0,00621	0,00754	-0,8240	0,20557
<b>297</b>	0,03342	0,00799	<b>4,1849</b>	<b>0,00002</b>
298	-0,00036	0,00589	-0,0614	0,47556

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **24**, 37, **189** e **297**.

**Experimento 5.3.7**

## (1) Observações contaminadas e erros

Quadro 5.17 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
20	0,06266	-50,59800	-50,53534
42	-0,09305	-18,03600	-18,12905
108	-0,08216	-34,78200	-34,86416

(2) Detecção dos *outliers*

$$\chi^2 = 1058,263$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 2, 17, **20**, 22, 24, 26, **42**, 45, 48, 49, 50, 84, 91, 106, **108**, 115, 116, 142, 146, 147, 149, 154, 157, 158, 211, 212, 213, 218, 221, 238, 240, 253, 255, 259, 264, 285, 295.

Total = 37

Quadro 5.18 - Teste t para  $\nu=160$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>20</b>	-0,06124	0,00514	<b>-11,9228</b>	<b>0,00000</b>
<b>42</b>	0,09095	0,00840	<b>10,8230</b>	<b>0,00000</b>
50	0,02979	0,13957	0,2134	0,41564
<b>108</b>	0,04030	0,00279	<b>14,4586</b>	<b>0,00000</b>
115	-0,00877	0,04228	-0,2075	0,41794
116	0,03438	0,14090	0,2440	0,40377
157	-0,02737	0,11186	-0,2447	0,40350
221	-0,03303	0,14628	-0,2258	0,41082
285	-0,04509	0,18465	-0,2442	0,40369

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **20, 42 e 108**.

### Experimento 5.3.8

(1) Observações contaminadas e erros

Quadro 5.19 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
102	-0,03385	-72,23100	-72,26485
140	-0,03784	-41,61900	-41,65684
278	0,09678	95,16800	95,26478

(2) Detecção dos *outliers*

$$\chi^2 = 555,628$$

$$p = 0,000000$$

### (3) Identificação dos *outliers*

#### (a) Sem rotação dos autovetores

Observações suspeitas: 2, 20, 22, 24, 29, 30, 34, 40, 42, 50, 54, 75, 88, 90, 91, 92, 98, 104, 105, 111, 115, 120, **140**, 142, 144, 184, 187, 189, 195, 197, 198, 211, 248, 255, 264, 272, 274, 276, 277, **278**, 283, 284, 285, 286.

Total = 44

Quadro 5.20 - Teste t com  $v=167$

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>140</b>	0,03548	0,00465	<b>7,6306</b>	<b>0,00000</b>
<b>278</b>	-0,09281	0,00531	<b>-17,4945</b>	<b>0,00000</b>

#### (b) Com rotação dos autovetores

Quadro 5.21 - Teste t com  $v=151$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
10	-0,00628	0,00526	-1,1934	0,11728
96	-0,01065	0,00513	-2,0777	0,01971
99	0,01803	0,02241	0,8046	0,21115
101	0,05584	0,02468	2,2627	0,01254
<b>102</b>	0,03357	0,00619	<b>5,4196</b>	<b>0,00000</b>
<b>140</b>	0,03467	0,00478	<b>7,2593</b>	<b>0,00000</b>
144	-0,01054	0,00564	-1,8687	0,03180
167	0,00504	0,00605	0,8340	0,20279
208	0,00751	0,00672	1,1177	0,13273
271	0,03212	0,02255	1,4249	0,07812
272	-0,01087	0,00639	-1,7024	0,04536
273	0,05105	0,02477	2,0612	0,02050
274	0,00841	0,00737	1,1415	0,12773
<b>278</b>	-0,08732	0,00801	<b>-10,9053</b>	<b>0,00000</b>
280	0,00054	0,00586	0,0915	0,46361
282	0,02183	0,01700	1,2841	0,10053
284	0,00809	0,00659	1,2271	0,11084
286	0,00235	0,00792	0,2962	0,38374

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **102**, **140**, **142** e **278**.

### Experimento 5.3.9

#### (1) Observações contaminadas e erros

Quadro 5.22 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
159	-0,06578	63,92600	63,86022
243	-0,08862	-84,79000	-84,87862
252	0,03948	-80,92600	-80,88652

#### 2) Detecção dos *outliers*

$$\chi^2 = 655,538$$

$$p = 0,000000$$

#### (3) Identificação dos *outliers*

##### (a) Sem rotação dos autovetores

Observações suspeitas: 5, 7, 8, 9, 11, 31, 32, 34, 39, 42, 50, 54, 59, 79, 88, 91, 92, 98, 117, 119, 120, 129, 146, **159**, 161, 171, 177, 179, 181, 182, 183, 201, 203, 207, 210, 211, 232, 241, **243**, **252**, 264, 267, 276, 278, 286.

Total = 45



Quadro 5.23 - Teste t com  $v=162$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
5	-0,00537	0,00842	-0,6376	0,26231
31	-0,05184	0,02239	-2,3150	0,01093
119	-0,00202	0,02187	-0,0925	0,46321
<b>159</b>	0,06673	0,00582	<b>11,4719</b>	<b>0,00000</b>
161	0,00560	0,01531	0,3655	0,35760
201	-0,02968	0,01560	-1,9031	0,02940
<b>243</b>	0,08858	0,00536	<b>16,5142</b>	<b>0,00000</b>

(b) Com rotação dos autovetores

Quadro 5.24 - Teste t com  $v=159$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
7	0,01960	0,04261	0,4600	0,32307
8	0,01262	0,06545	0,1928	0,42368
72	0,01065	0,06668	0,1597	0,43666
79	0,00619	0,00594	1,0417	0,14956
117	0,02201	0,02175	1,0124	0,15643
<b>159</b>	0,05201	0,01521	<b>3,4192</b>	<b>0,00040</b>
179	0,00351	0,03872	0,0907	0,46392
188	0,00453	0,00597	0,7595	0,22433
<b>243</b>	0,07379	0,07571	<b>0,9746</b>	<b>0,16561</b>
<b>252</b>	-0,03706	0,00593	<b>-6,2481</b>	<b>0,00000</b>

(4) *Outliers* identificados pelo *Danish Method*Observações identificadas: **159, 243 e 252.**

### Experimento 5.3.10

#### (1) Observações contaminadas e erros

Quadro 5.25 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
4	-0,03355	62,76500	62,73145
167	-0,07636	25,89500	25,81864
262	0,04674	3,34300	3,38974

#### (2) Detecção dos *outliers*

$$\chi^2 = 480,106$$

$$p = 0,000000$$

#### (3) Identificação dos *outliers*

##### (a) Sem rotação dos autovetores

Observações suspeitas: 4, 29, 30, 39, 42, 50, 89, 121, 125, 126, 132, 134, 135, 145, 163, **167**, 187, 195, 197, 198, 199, 210, 211, 231, 239, **262**, 265, 289.

Total = 28

Quadro 5.26 - Teste t com  $v=165$

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>4</b>	0,03638	0,00495	<b>7,3528</b>	<b>0,00000</b>
125	-0,00141	0,02244	-0,0627	0,47500
<b>167</b>	0,08206	0,01579	<b>5,1975</b>	<b>0,00000</b>
<b>262</b>	-0,04135	0,00492	<b>-8,4012</b>	<b>0,00000</b>

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **4, 167 e 262.**

**Experimento 5.3.11**

## (1) Observações contaminadas e erros

Quadro 5.27 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
119	-0,06686	2,55600	2,49014
154	-0,06931	72,22800	72,15869
208	-0,09079	79,96900	79,87821
259	0,11845	45,75600	45,87445

(2) Detecção dos *outliers*

$$\chi^2 = 1547,529$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 18, 20, 22, 29, 30, 33, 35, 37, 41, 42, 43, 49, 50, 54, 66, 68, 76, 78, 79, 88, 89, 91, 103, 105, 106, 108, 111, 117, **119**, 120, 123, 126, 127, 132, 134, 136, 138, 148, 150, **154**, 158, 159, 161, 165, 169, 186, 196, 197, 198, 200, 202, 205,

206, 207, **208**, 210, 211, 212, 213, 214, 231, 232, 237, 238, 240, 248, 252, 255, **259**, 261, 262, 264, 268, 272, 274, 275, 276, 277, 278, 280, 283, 284, 286, 292, 296.

Total = 85

Quadro 5.28 - Teste t com  $v=164$

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
117	0,00356	0,00831	0,4278	0,33470
<b>154</b>	0,07433	0,00471	<b>15,7804</b>	<b>0,00000</b>
161	0,04688	0,00582	8,0576	0,00000
<b>208</b>	0,09346	0,00522	<b>17,9097</b>	<b>0,00000</b>
<b>259</b>	-0,12340	0,00473	<b>-26,0855</b>	<b>0,00000</b>

(b) Com rotação dos autovetores

Observações suspeitas: 24, 33, 35, 37, 39, 42, 43, 54, 84, 90, 101, **119**, **154**, 161, 204, 206, **208**, 210, 211, 214, **259**, 261, 267, 272, 273, 274, 276.

Total = 27

Quadro 5.29 - Teste t com  $v=161$

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
37	-0,06472	0,89933	-0,0720	0,47135
101	-0,00573	0,42210	-0,0136	0,49458
<b>119</b>	0,03162	0,01103	<b>2,8656</b>	<b>0,00236</b>
<b>154</b>	0,03711	0,00236	<b>15,7330</b>	<b>0,00000</b>
<b>208</b>	0,01682	0,44475	<b>0,0378</b>	<b>0,48495</b>
<b>259</b>	-0,06229	0,00247	<b>-25,2183</b>	<b>0,00000</b>
273	0,02037	0,02829	0,7199	0,23631
274	-0,02441	0,44086	-0,0554	0,47794

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: 37, **154**, 161, **208**, 225, **259**.

**Experimento 5.3.12**

## (1) Observações contaminadas e erros

Quadro 5.30 - Observações contaminadas

Número	Erro	Observações	Obs. + erro
8	-0,07020	58,93100	58,86080
175	0,07570	-2,59100	-2,51530
207	0,05680	52,69200	52,74880
238	-0,08624	1,98400	1,89776

(2) Detecção dos *outliers*

$$\chi^2 = 1089,296$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 1, 3, 6, 7, **8**, 12, 14, 16, 17, 18, 19, 21, 26, 29, 30, 32, 36, 38, 39, 42, 49, 50, 51, 52, 54, 64, 65, 67, 70, 72, 77, 79, 88, 91, 92, 100, 102, 103, 110, 118, 120, 129, 130, 131, 132, 133, 134, 136, 138, 143, 151, 171, 173, 174, **175**, 176, 179,

197, 198, 205, 207, 209, 210, 211, 212, 215, 224, 225, 232, **238**, 240, 242, 243, 244, 247, 248, 249, 250, 252, 259, 261, 264, 267, 275, 276, 291, 295.

Total = 87

Quadro 5.31 - Teste t com  $v=149$

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
1	-0,07336	0,09450	-0,7763	0,21939
7	0,01643	0,05749	0,2858	0,38771
<b>8</b>	0,08539	0,03616	<b>2,3612</b>	<b>0,00975</b>
39	0,54872	0,28760	1,9079	0,02916
65	-0,07388	0,09372	-0,7883	0,21588
72	0,01287	0,03738	0,3442	0,36559
103	0,11531	0,05877	1,9621	0,02580
129	0,00458	0,02179	0,2104	0,41682
131	-0,06770	0,09543	-0,7094	0,23958
133	-0,00854	0,01087	-0,7855	0,21669
171	-0,00143	0,01536	-0,0934	0,46285
174	-0,07474	0,09306	-0,8031	0,21159
<b>175</b>	-0,06936	0,00628	<b>-11,0427</b>	<b>0,00000</b>
<b>207</b>	-0,06453	0,00504	<b>-12,7947</b>	<b>0,00000</b>
209	-0,09364	0,05128	-1,8263	0,03490
210	0,56003	0,29774	1,8809	0,03096
<b>238</b>	0,01727	0,09198	<b>0,1878</b>	<b>0,42564</b>
243	-0,01620	0,05049	-0,3208	0,37440
275	-0,47710	0,25184	-1,8944	0,03005
276	0,57244	0,29943	1,9118	0,02891

(b) Com rotação dos autovetores

Observações suspeitas: 2, 4, 7, **8**, 19, 31, 38, 65, 72, 84, 110, 129, 136, 138, 146, 151, 171, 173, 174, **175**, 179, 189, 201, **207**, 211, **238**, 243, 253, 267, 291.

Total = 30

Quadro 5.32 - Teste t com  $\nu=163$

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>8</b>	0,09180	0,15210	<b>0,6036</b>	<b>0,27348</b>
31	-0,11377	0,07794	-1,4598	0,07313
<b>175</b>	-0,03262	0,00285	<b>-11,4343</b>	<b>0,00000</b>
201	-0,03265	0,02427	-1,3457	0,09013
<b>207</b>	-0,03201	0,00275	<b>-11,6441</b>	<b>0,00000</b>
<b>238</b>	0,04258	0,00280	<b>15,1995</b>	<b>0,00000</b>

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **8**, **175**, **207** e **238**.

### Experimento 5.3.13

(1) Observações contaminadas e erros

Quadro 5.33 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
45	0,07759	75,57600	75,65359
52	-0,08733	2,82400	2,73667
129	0,08774	-52,64400	-52,55626
131	-0,08086	-104,96100	-105,04186

(2) Detecção dos *outliers*

$$\chi^2 = 990,292$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 1, 2, 3, 5, 6, 24, 29, 30, 31, 39, 42, 43, **45**, 49, 50, **52**, 54, 60, 64, 70, 79, 80, 88, 109, 112, 113, 114, 117, 118, 119, 120, **129**, 130, **131**, 133, 142, 146, 150, 152, 154, 155, 158, 159, 160, 161, 162, 171, 173, 175, 177, 187, 195, 197, 198, 200, 201, 210, 211, 212, 214, 216, 218, 219, 221, 223, 225, 238, 251, 252, 257, 280, 281, 292, 295, 296, 297.

Total = 76



Quadro 5.34 - Teste t com  $v=155$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
5	0,00634	0,02122	0,2990	0,38267
31	-0,05147	0,02247	-2,2909	0,01166
<b>45</b>	-0,07559	0,00472	<b>-16,0055</b>	<b>0,00000</b>
<b>52</b>	0,08296	0,00516	<b>16,0626</b>	<b>0,00000</b>
113	0,01569	0,02168	0,7234	0,23525
119	-0,00053	0,02192	-0,0241	0,49040
<b>129</b>	-0,07876	0,02180	<b>- 3,6136</b>	<b>0,00020</b>
<b>131</b>	0,08838	0,00785	<b>11,2530</b>	<b>0,00000</b>
155	-0,01216	0,01531	-0,7941	0,21417
159	0,00261	0,00621	0,4205	0,33735
161	0,00460	0,01536	0,2997	0,38240
171	-0,00378	0,01537	-0,2458	0,40308
177	0,00848	0,01508	0,5621	0,28742
201	-0,03023	0,01568	-1,9273	0,02788

#### 4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **45, 52, 129 e 131.**

### Experimento 5.3.14

(1) Observações contaminadas e erros

Quadro 5.35 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
66	0,04619	58,99200	59,03819
136	-0,10599	-26,07400	-26,17999
245	0,09040	-78,71900	-78,62860
250	-0,05909	-100,98000	-101,03909

(2) Detecção dos *outliers*

$$\chi^2 = 1232,993$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 2, 7, 8, 9, 10, 11, 18, 19, 20, 22, 24, 29, 30, 32, 39, 41, 42, 49, 50, **66**, 77, 84, 87, 88, 104, 106, 108, 110, 116, 120, 132, 134, **136**, 137, 138, 139, 140, 142, 146, 148, 151, 158, 164, 179, 181, 183, 186, 196, 197, 198, 210, 211, 212, 215, 222, 230, 237, 238, 240, 243, **245**, **250**, 253, 257, 259, 260, 261, 267, 272, 274, 291, 295, 297.

Total = 73

Quadro 5.36 - Teste t com  $v=158$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
7	-0,00044	0,00817	-0,0533	0,47878
9	0,01267	0,00948	1,3359	0,09174
24	0,01217	0,15010	0,0811	0,46773
<b>66</b>	-0,04541	0,00469	<b>-9,6798</b>	<b>0,00000</b>
88	0,02388	0,14709	0,1624	0,43560
<b>136</b>	0,11167	0,00478	<b>23,3749</b>	<b>0,00000</b>
142	0,01103	0,14981	0,0737	0,47067
<b>245</b>	-0,09379	0,00628	<b>- 14,9319</b>	<b>0,00000</b>
<b>250</b>	0,06141	0,00546	<b>11,2531</b>	<b>0,00000</b>
259	-0,02413	0,15068	-0,1602	0,43646
295	-0,02824	0,15195	-0,1859	0,46638

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **66, 136, 245 e 250.**

### Experimento 5.3.15

(1) Observações contaminadas e erros

Quadro 5.37 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
228	0,03369	114,28100	114,31469
242	0,03689	-29,21000	-29,17311
247	0,07487	-68,69700	-68,62313
292	-0,11683	-4,51800	-4,63483

(2) Detecção dos *outliers*

$$\chi^2 = 1087,038$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 9, 10, 11, 12, 13, 14, 18, 23, 25, 28, 29, 30, 32, 36, 38, 40, 42, 49, 50, 55, 64, 68, 77, 88, 94, 100, 102, 104, 110, 120, 130, 135, 136, 137, 138, 139, 146, 151, 154, 164, 181, 183, 185, 197, 198, 204, 211, 215, 220, **228**, 238, **242**, 244, 245, 246, **247**, 249, 250, 263, 267, 268, 270, 278, 286, 289, **292**, 296, 298.

Total = 68

Quadro 5.38 - Teste t com  $v=164$

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
11	-0,11888	0,07372	-1,6127	0,05436
183	-0,07097	0,04695	-1,5118	0,06625
<b>242</b>	-0,03280	0,00499	<b>-6,5744</b>	<b>0,00000</b>
<b>247</b>	-0,04270	0,01949	<b>-2,1915</b>	<b>0,01491</b>
<b>292</b>	0,12301	0,00478	<b>25,7181</b>	<b>0,00000</b>

## (b) Com rotação dos autovetores

Quadro 5.39 - Teste t com  $\nu=152$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
8	0,00338	0,00565	0,5981	0,27532
12	0,00349	0,00729	0,4788	0,31638
14	0,00067	0,00760	0,0886	0,46476
55	-0,00134	0,00894	-0,1495	0,44068
76	0,00458	0,00710	0,6453	0,25984
78	0,00292	0,00723	0,4034	0,34360
83	-0,00442	0,00467	-0,9466	0,17266
92	0,00863	0,00565	1,5284	0,06424
164	0,00610	0,00869	0,7019	0,24190
187	-0,00962	0,00635	-1,5146	0,06597
<b>228</b>	-0,02658	0,00887	<b>- 2,9978</b>	<b>0,00159</b>
<b>242</b>	-0,03310	0,00502	<b>- 6,5937</b>	<b>0,00000</b>
243	0,00114	0,00582	0,1954	0,42267
<b>247</b>	-0,07571	0,00786	<b>- 9,6341</b>	<b>0,00000</b>
249	0,00233	0,00801	0,2912	0,38564
263	-0,00385	0,00580	-0,6641	0,25381
<b>292</b>	0,12232	0,00488	<b>25,0858</b>	<b>0,00000</b>

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **228, 242, 247 e 292.**

### Experimento 5.3.16

(1) Observações contaminadas e erros

Quadro 5.40 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
44	0,02998	-34,93600	-34,90602
97	0,10477	-35,54300	-35,43823
193	0,04658	97,97500	98,02158
254	-0,07743	-14,96000	-15,03743
290	-0,07501	-16,55000	-16,62501

(2) Detecção dos *outliers*

$$\chi^2 = 869,270$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 16, 17, 19, 23, 25, 26, 28, 40, 42, **44**, 46, 50, 76, 78, 81, 82, 86, 88, **97**, 104, 120, 135, 137, 139, 141, 147, 149, 154, 157, 186, 190, 191, 192, **193**, 194, 196, 202, 211, 213, 217, **254**, 255, 256, 258, 263, 269, 272, 274, 276, 288, 289, **290**, 291, 292, 293, 294, 295, 296, 297, 298.

Total = 60

Quadro 5.41 - Teste t com  $v=153$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
17	-0,30219	0,16406	-1,8420	0,03370
19	-0,06839	0,06478	-1,0558	0,14635
23	-0,00929	0,00725	-1,2813	0,10101
81	-0,28753	0,15803	-1,8195	0,03539
<b>97</b>	-0,11892	0,02141	<b>-5,5535</b>	<b>0,00000</b>
135	-0,29205	0,15700	-1,8602	0,03238
137	0,06969	0,06652	1,0476	0,14823
141	0,01257	0,00704	1,7858	0,03805
190	-0,29027	0,15886	-1,8272	0,03480
191	-0,03587	0,03982	-0,9008	0,18455
<b>193</b>	-0,04579	0,00603	<b>-7,5982</b>	<b>0,00000</b>
<b>254</b>	-0,21377	0,15935	<b>-1,3415</b>	<b>0,09086</b>
255	0,01729	0,01693	1,0213	0,15435
269	-0,02346	0,02127	-1,1026	0,13596
<b>290</b>	-0,22681	0,16176	<b>-1,4021</b>	<b>0,08145</b>
291	0,07871	0,07239	1,0872	0,13932

(b) Com rotação dos autovetores

Quadro 5.42 - Teste t com  $v=156$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
17	0,00388	0,01010	0,3837	0,35086
<b>44</b>	-0,02305	0,00506	<b>-4,5575</b>	<b>0,00001</b>
81	0,00644	0,00738	0,8729	0,19202
86	-0,00499	0,00473	-1,0534	0,14688
<b>97</b>	-0,11546	0,02140	<b>-5,3942</b>	<b>0,00000</b>
141	0,00603	0,00817	0,7389	0,23053
149	-0,00123	0,00631	-0,1958	0,42251
190	0,00596	0,00755	0,7887	0,21573
<b>193</b>	-0,04827	0,00578	<b>-8,3482</b>	<b>0,00000</b>
<b>254</b>	0,08274	0,00756	<b>10,9404</b>	<b>0,00000</b>
269	-0,02146	0,02126	-1,0092	0,15721
<b>290</b>	0,07401	0,00767	<b>9,6518</b>	<b>0,00000</b>
295	-0,00666	0,00724	-0,9196	0,17959

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **97, 193, 254 e 290.**

### Experimento 5.3.17

(1) Observações contaminadas e erros



Quadro 5.43 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
16	-0,04016	-16,45600	-16,49616
56	-0,10153	29,67600	29,57447
97	0,09224	-35,54300	-35,45076
244	0,04220	-47,66700	-47,62480
261	0,03134	69,68900	69,72034

(2) Detecção dos *outliers*

$$\chi^2 = 877,591$$

$$p = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 2, **16**, 24, 34, 42, 50, 51, 52, 54, **56**, 57, 58, 60, 61, 62, 74, 79, 91, **97**, 98, 118, 122, 124, 128, 132, 134, 142, 146, 149, 160, 162, 188, 203, 211, 223, 224, 227, 242, **244**, 246, 252, **261**, 264, 269, 272, 274, 284.

Total = 47

Quadro 5.44 - Teste t com  $v=160$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>16</b>	0,03355	0,00529	<b>6,3424</b>	<b>0,00000</b>
34	0,08637	0,58576	0,1475	0,44146
<b>56</b>	0,10052	0,00491	<b>20,4803</b>	<b>0,00000</b>
<b>97</b>	-0,13213	0,13435	<b>-0,9835</b>	<b>0,16341</b>
98	0,09806	0,59178	0,1657	0,43430
203	-0,09931	0,56058	-0,1772	0,42982
<b>244</b>	-0,04303	0,00499	<b>- 8,6202</b>	<b>0,00000</b>
<b>261</b>	-0,02894	0,00506	<b>- 5,7185</b>	<b>0,00000</b>
269	-0,14366	0,46801	-0,3070	0,37962

(b) Com rotação dos autovetores

Quadro 5.45 - Teste t com  $v=161$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>16</b>	0,03193	0,00691	<b>4,6237</b>	<b>0,00000</b>
28	-0,00327	0,00871	-0,3760	0,35370
<b>56</b>	0,10057	0,00491	<b>20,4913</b>	<b>0,00000</b>
90	0,00444	0,00565	0,7863	0,21642
<b>97</b>	-0,10215	0,02139	<b>- 4,7755</b>	<b>0,00000</b>
<b>244</b>	-0,04294	0,00499	<b>- 8,6006</b>	<b>0,00000</b>
<b>261</b>	-0,03028	0,00539	<b>- 5,6199</b>	<b>0,00000</b>
269	-0,02029	0,02124	-0,9508	0,17156

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **16, 56, 97, 244 e 261.**

### Experimento 5.3.18

#### (1) Observações contaminadas e rros

Quadro 5.46 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
41	0,08935	94,06400	94,15335
52	-0,07323	2,82400	2,75077
54	-0,07075	11,51400	11,44325
90	-0,07854	-85,92500	-86,00354
176	0,05262	-6,87900	-6,82638

#### (2) Detecção dos *outliers*

$$\chi^2 = 1284,500$$

$$p = 0,000000$$

#### (3) Identificação dos *outliers*

##### (a) Sem rotação dos autoverores

Observações suspeitas: 1, 2, 3, 4, 8, 10, 17, 24, 26, 36, 38, 40, **41**, 42, 43, 47, 48, 49, 50, **52**, **54**, 55, 56, 57, 59, 60, 61, 63, 64, 65, 67, 68, 75, 80, 87, 88, 89, **90**, 91, 93, 94, 100, 102, 104, 105, 106, 108, 109, 111, 114, 117, 118, 120, 122, 126, 130, 131, 133, 142, 144, 146, 148, 152, 159, 168, 174, **176**, 184, 189, 207, 209, 210, 211, 212, 214, 216, 219, 222, 225, 228, 231, 232, 238, 239, 240, 243, 245, 248, 251, 257, 259, 260, 262, 264, 265, 277, 278, 281, 282, 283, 284, 286, 296.

Total = 103

Quadro 5.47 - Teste t com  $v=158$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>41</b>	-0,08787	0,00481	<b>-18,2840</b>	<b>0,00000</b>
<b>52</b>	0,07191	0,00620	<b>11,5951</b>	<b>0,00000</b>
<b>54</b>	0,08163	0,00488	<b>16,7164</b>	<b>0,00000</b>
<b>90</b>	0,08314	0,00531	<b>15,6660</b>	<b>0,00000</b>
109	-0,12231	0,23536	-0,5197	0,30199
117	0,01642	0,02610	0,6291	0,26509
152	-0,11837	0,22961	-0,5155	0,30346
159	-0,00923	0,01928	-0,4786	0,31644
<b>176</b>	-0,04646	0,00498	<b>- 9,3270</b>	<b>0,00000</b>
216	-0,11362	0,23353	-0,4865	0,31364
282	-0,11291	0,23734	-0,4757	0,31747

#### 4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **41, 52, 54, 90 e 176.**

### Experimento 5.3.19

#### (1) Observações contaminadas e erros

Quadro 5.48 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
5	0,10008	71,56200	71,66208
102	-0,05401	-72,23100	-72,28501
157	0,09949	75,65700	75,75649
267	0,04098	-106,46800	-106,42702
275	-0,10600	-59,32400	-59,43000

(2) Detecção dos *outliers*

$$\chi^2 = 1031,793$$

$$P = 0,000000$$

(3) Identificação dos *outliers*

## (a) Sem rotação dos autovetores

Observações suspeitas: 5, 9, 10, 11, 29, 30, 32, 42, 50, 75, 90, 91, 97, 103, 110, 143, 144, 145, 146, 147, 149, 151, 153, **157**, 177, 181, 183, 184, 186, 196, 197, 198, 199, 200, 210, 211, 215, 245, 264, **267**, 269, **275**, 276, 278, 286, 297.

Total = 46

Quadro 5.49 - Teste t com  $v=161$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>5</b>	-0,10397	0,00839	<b>-12,3896</b>	<b>0,00000</b>
9	0,01048	0,00819	1,2797	0,10124
29	0,03227	0,01251	2,5800	0,00540
145	0,04169	0,02387	1,7468	0,04128
<b>157</b>	-0,10587	0,00564	<b>-18,7686</b>	<b>0,00000</b>
199	-0,03501	0,01692	-2,0686	0,02010
<b>267</b>	-0,05639	0,00565	<b>- 9,9794</b>	<b>0,00000</b>
<b>275</b>	0,10032	0,00685	<b>14,6528</b>	<b>0,00000</b>

(b) Com rotação dos autovetores

Observações suspeitas: **5**, 42, 50, 54, 80, 87, 88, 97, **102**, 103, 116, 117, 120, 144, 147, 149, 151, 153, **157**, 159, 177, 200, 211, 221, 235, 242, 251, 259, 260, **267**, 268, 269, **275**, 276.

Total = 34

Quadro 5.50 - Teste t com  $v=164$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>5</b>	-0,08738	0,02191	<b>- 3,9873</b>	<b>0,00005</b>
<b>102</b>	0,02467	0,00276	<b>8,9385</b>	<b>0,00000</b>
<b>157</b>	-0,05256	0,00336	<b>-15,6419</b>	<b>0,00000</b>
<b>267</b>	-0,02592	0,00285	<b>- 9,0864</b>	<b>0,00000</b>
<b>275</b>	0,03938	0,01107	<b>3,5574</b>	<b>0,00024</b>

(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: **5**, **102**, 103, **157** e **267**.

## Experimento 5.3.20

### (1) Observações contaminadas e erros

Quadro 5.51 - Observações contaminadas

Número	Erro	Observação	Obs. + erro
40	0,07810	-60,26300	-60,18490
103	0,09794	-24,50700	-24,40906
126	0,05436	42,15600	42,21036
140	0,11359	-41,61900	-41,50541
162	-0,06561	102,27500	102,20939

### (2) Detecção dos *outliers*

$$\chi^2 = 1301,938$$

$$p = 0,000000$$

### (3) Identificação dos *outliers*

#### (a) Sem rotação dos autovetores

Observações suspeitas: 2, 10, 16, 18, 20, 22, 24, 34, **40**, 42, 50, 51, 55, 57, 59, 60, 66, 68, 69, 74, 76, 78, 79, 80, 84, 88, 89, 92, 93, 94, 95, 97, 98, **103**, 104, 106, 108, 114, 120, **126**, 132, 134, 136, 138, **140**, 142, 143, 146, 160, **162**, 166, 170, 178, 187, 188, 191, 193, 195, 200, 202, 203, 205, 207, 209, 211, 219, 223, 224, 225, 228, 231, 234, 237, 245, 246, 251, 252, 253, 255, 257, 262, 266, 269, 275, 276, 284, 293, 297.

Total = 88

Quadro 5.52 - Teste t com  $v=159$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>40</b>	-0,07727	0,01938	<b>- 3,9881</b>	<b>0,00005</b>
97	-0,04026	0,07096	-0,5674	0,28561
<b>103</b>	-0,13083	0,07588	<b>-1,7242</b>	<b>0,04330</b>
104	-0,00970	0,01834	-0,5287	0,29887
<b>126</b>	-0,05398	0,00507	<b>-10,6537</b>	<b>0,00000</b>
<b>140</b>	-0,11492	0,00466	<b>-24,6735</b>	<b>0,00000</b>
<b>162</b>	0,06456	0,00505	<b>12,7758</b>	<b>0,00000</b>
203	-0,00751	0,01730	-0,4341	0,33240
269	-0,05307	0,07560	-0,7020	0,24185
275	-0,03461	0,06323	-0,5474	0,29243

## b) Com rotação dos autovetores

Observações suspeitas: 38, **40**, 42, 53, 60, 76, 86, 102, **103**, 104, **126**, 138, **140**, 142, **162**, 207, 209, 226, 257, 275.

Total = 20

Quadro 5.53 - teste t com  $v=161$ 

Número	$\hat{w}$	$\sigma_{\hat{w}}$	t	p
<b>40</b>	-0,79727	0,44718	<b>-1,7829</b>	<b>0,03824</b>
<b>103</b>	-0,25972	0,11065	<b>-2,3473</b>	<b>0,01006</b>
104	-0,73719	0,45169	-1,6321	0,05230
<b>126</b>	-0,05383	0,00507	<b>-10,6257</b>	<b>0,00000</b>
<b>140</b>	-0,11465	0,00465	<b>-24,6377</b>	<b>0,00000</b>
<b>162</b>	-0,06469	0,00505	<b>12,8013</b>	<b>0,00000</b>
209	0,69023	0,42812	1,6122	0,05443
275	0,54343	0,36204	1,5010	0,06765



(4) *Outliers* identificados pelo *Danish Method*

Observações identificadas: 37, **40**, **103**, **126**, **140** e **162**.

## 6 ANÁLISE DOS RESULTADOS, CONCLUSÕES E RECOMENDAÇÕES

### 6.1 Análise dos resultados

No nível de significância tomado como referência para o teste  $\chi^2$ , verifica-se que nos 20 experimentos realizados a existência de *outliers* é significativa.

No Quadro 6.1 encontra-se resumido a identificação dos *outliers* pelo Método Multivariado e pelo *Danish Method*, no nível de significância tomado como referência para o teste t (1%).

Quadro 6.1 - Resumo dos resultados dos 20 experimentos

Experi- mento	Outliers (Simulados)	Método Multivariado		Danish Method
		Sem rotação	Com rotação	
5.3.1	126, 248	126	126, 248	37, 126, 248
5.3.2	158, 164	158, 164		158, 164
5.3.3	80, 196	80	80, 196	80, 196
5.3.4	74, 298	74, 298		37, 74, 298
5.3.5	70, 240	240	70, 240	70, 240
5.3.6	24, 189, 297	189	24, 189, 297	24, 37, 189, 297
5.3.7	20, 42, 108	20, 42, 108		20, 42, 108
5.3.8	102, 140, 278	140, 278	102, 140, 278	102, 140, 142, 278
5.3.9	159, 243, 252	159, 243	159, 252	159, 243, 252
5.3.10	4, 167, 262	4, 167, 262		4, 167, 262
5.3.11	119, 154, 208, 259	154, 161, 208, 259	119, 154, 259	37, 154, 161, 208, 225, 259
5.3.12	8, 175, 207, 238	175, 207	175, 207, 238	8, 175, 207, 238
5.3.13	45, 52, 129, 131	45, 52, 129, 131		45, 52, 129, 131
5.3.14	66, 136, 245, 250	66, 136, 245, 250		66, 136, 245, 250
5.3.15	228, 242, 247, 292	242, 292	228, 242, 247, 292	228, 242, 247, 292
5.3.16	44, 97, 193, 254, 290	97, 193	44, 97, 193, 254, 290	97, 193, 254, 290
5.3.17	16, 56, 97, 244, 261	16, 56, 244, 261	16, 56, 97, 244, 261	16, 56, 97, 244, 261
5.3.18	41, 52, 54, 90, 176	41, 52, 54, 90, 176		41, 52, 54, 90, 176
5.3.19	5, 102, 157, 267, 275	5, 157, 267, 275	5, 102, 157, 267, 275	5, 102, 103, 157, 267
5.3.20	40, 103, 126, 140, 162	40, 126, 140, 162	126, 140, 162	37, 40, 103, 126, 140, 162

A identificação pelo Método Multivariado somente não foi completa nos experimentos:

5.3.11 - onde a observação 161 foi identificada erradamente como *outlier*,

5.3.12 - onde a observação 8 não foi identificada como *outlier* e

5.3.20 - onde a observação 103 também não foi identificada como *outlier*.

Com relação à observação 161, pode-se constatar através do Anexo A, que a mesma corresponde a uma fotocoordenada do ponto 12, bem como as observações: 53, 54, 119, 120, 162, 225 e 226 também são fotocoordenadas desse ponto. Como nesse experimento a observação 119 também foi contaminada por erro grosseiro (ver Quadro 5.27), provavelmente ocorreu uma contaminação da observação 161 por parte da 119, o que está de acordo com (EL-HAKIM [19],[20]) que estudou os efeitos dos erros grosseiros para um mesmo ponto em outras fotos.

No experimento 5.3.12, verificou-se que a observação 8 não foi identificada como *outlier* no nível de significância de 1%, porém pode-se constatar (Quadro 5.31) que  $p=0,00975$ , ou seja, ao nível de significância de 2% por exemplo, a mesma já seria identificada como *outlier*.

No caso da não identificação da observação 103, no experimento 5.3.20, pode-se constatar que a mesma foi observada no ponto 4 (Anexo A) bem como a observação 40 que nesse experimento também foi contaminada por erro grosseiro. Dessa forma, existe a possibilidade de certa influência da observação 40 sobre a 103 neutralizando parte do efeito do erro. Pode-se constatar ainda que, de acordo com o Quadro (5.53), obteve-se para o teste t da observação 103 que  $p=0,01006$ , ou seja, a um nível de significância de pouco mais de 2% essa observação seria identificada como *outlier*.

Na aplicação do *Danish Method* verifica-se que não foram identificadas como *outliers* as observações

119 no experimento 5.3.11,

44 no experimento 5.3.16,

275 no experimento 5.3.19,

e foram erradamente identificadas como *outliers* as observações:

37 nos experimentos: 5.3.1, 5.3.4, 5.3.6, 5.3.11 e 5.3.20;

142 no experimento 5.3.8; 161 e 225 no experimento 5.3.11 e

103 no experimento 5.3.19.

## 6.2 Conclusões

Nos 20 experimentos realizados foram impostos erros grosseiros em 70 observações, ou seja, foram simulados 70 *outliers*. Da análise realizada pode-se constatar que uma observação foi identificada erradamente como *outlier* no Método Multivariado, enquanto que esse número subiu para 9 no *Danish Method*. Com relação ao número de *outliers* não identificados, a ocorrência foi de 2 ou 2,9% do total de observações contaminadas no Método Multivariado e 3 ou 4,3% no *Danish Method*. Esses resultados fornecem uma indicação clara da eficiência do Método Multivariado proposto na detecção e identificação de múltiplos *outliers* em Fototriangulação, devendo ressaltar ainda que os dois *outliers* não identificados pelo método poderiam ser identificados num nível de significância pouco acima do que foi adotado.

## 6.3 Recomendações

Com base nos estudos teóricos e experimentos realizados, pode-se fazer algumas recomendações para pesquisas futuras com o objetivo de melhorar ainda mais o método proposto. As principais recomendações seriam:

(1) pesquisar um método alternativo para o teste de Ferguson, já que o mesmo nem sempre proporcionou bons resultados nos experimentos realizados;

(2) desenvolver um método híbrido *Método Multivariado-Danish Method*, procurando explorar a eficiência do Método Multivariado e a facilidade de uso do *Danish Method*.

(3) procurar determinar o custo/benefício do método proposto para aplicação do mesmo em uma empresa que trabalha com fotogrametria;

(4) comparar o tempo de processamento computacional gasto pelo método proposto com os métodos usuais.

### ANEXO A - Fotocoordenadas (em mm)

Ponto	Obs.	Teórica	Obs. simulada	Ajustada	Ponto	Obs.	Teórica	Obs. simulada	Ajustada
18	1	105.99635	105.99300	105.99471	15	59	107.89904	107.89500	107.89651
18	2	54.61925	54.62100	54.62216	15	60	41.96421	41.96500	41.96636
19	3	100.45291	100.45400	100.45201	16	61	84.53422	84.52600	84.53049
19	4	62.76513	62.76500	62.76694	16	62	52.83266	52.83400	52.83040
20	5	71.56069	71.56200	71.56151	17	63	51.58486	51.58200	51.58452
20	6	69.01284	69.01200	69.01541	17	64	37.24639	37.24300	37.24436
21	7	53.83751	53.83000	53.82959	18	65	-2.62132	-2.62100	-2.62332
21	8	58.93352	58.93100	58.93298	18	66	58.99333	58.99200	58.99237
22	9	38.83966	38.83700	38.83945	19	67	-7.67260	-7.67900	-7.67815
22	10	51.59132	51.59200	51.58826	19	68	67.28065	67.27600	67.27910
23	11	21.76998	21.76400	21.76260	20	69	-34.56340	-34.56400	-34.56625
23	12	40.30054	40.29700	40.29716	20	70	74.14372	74.14500	74.14567
24	13	4.02046	4.02100	4.02114	21	71	-52.56873	-52.87200	-52.87096
24	14	26.17040	26.16900	26.16919	21	72	64.38980	64.39000	64.38990
25	15	26.17388	26.17100	26.17231	22	73	-68.73794	-68.73900	-68.73909
25	16	-16.46100	-16.45600	-16.45975	22	74	57.34281	57.34100	57.34141
26	17	81.72721	81.73100	81.73044	23	75	-86.39028	-86.38700	-86.39140
26	18	-30.49483	-30.49500	-30.49314	23	76	46.34869	46.34600	46.34736
27	19	94.09149	94.09600	94.09631	24	77	-104.46390	-104.46000	-104.46247
27	20	-50.59480	-50.59800	-50.59550	24	78	32.50496	32.50400	32.50561
28	21	101.44035	101.44400	101.44523	25	79	-82.27417	-82.97400	-82.97141
28	22	-45.82666	-45.82600	-45.82788	25	80	-10.48404	-10.48300	-10.48564
29	23	98.51645	98.52600	98.52399	26	81	-16.70552	-16.70700	-16.70414
29	24	-67.84361	-67.83900	-67.84437	26	82	-25.69816	-25.70300	-25.69959
30	25	102.59538	109.64300	109.63834	27	83	-4.67098	-4.66700	-4.67074
30	26	-91.59350	-91.59800	-91.59510	27	84	-45.69105	-45.68700	-45.69463
31	27	8.48340	8.48500	8.48425	28	85	2.57098	2.56800	2.57384
31	28	-28.02687	-28.03100	-28.02967	28	86	-41.10529	-41.10500	-41.10930
32	29	80.51677	80.50800	80.51097	29	87	-0.54156	-0.53800	-0.53910
32	30	109.63797	109.64300	109.63834	29	88	-62.67495	-62.68300	-62.67842
34	31	80.54985	80.55100	80.54847	30	89	2.96417	2.97100	2.96717
34	32	83.03700	83.03600	83.03795	30	90	-85.91815	-85.92500	-85.92223
1	33	69.23214	69.22700	69.22869	31	91	-101.94333	-101.94500	-101.94141
1	34	-58.73032	-58.72800	-58.73257	31	92	-21.64423	-21.65600	-21.65195
2	35	70.38534	70.38400	70.38573	33	93	78.62595	78.62800	78.62907
2	36	-79.55348	-79.55100	-79.55409	33	94	80.58318	80.59100	80.58577
3	37	78.42445	78.43400	78.42534	34	95	-25.55495	-25.55400	-25.55733
3	38	-71.91589	-71.91000	-71.91349	34	96	88.10373	88.11000	88.10317
4	39	80.56957	80.56800	80.56966	1	97	-35.53627	-35.54300	-35.53991
4	40	-60.25707	-60.26300	-60.25756	1	98	-58.89817	-58.90500	-58.90116
5	41	94.06783	94.06400	94.06507	2	99	-33.59773	-33.60200	-33.60200
5	42	-18.03590	-18.03600	-18.03323	2	100	-80.06187	-80.06900	-80.06479
6	43	78.25798	78.25700	78.25933	3	101	-26.15575	-26.15500	-26.15301
6	44	-34.93025	-34.93600	-34.93091	3	102	-72.23292	-72.23100	-72.23286
8	45	75.57437	75.57600	75.57761	4	103	-24.50604	-24.50700	-24.50615
8	46	-12.42239	-12.43100	-12.42443	4	104	-60.38028	-60.37800	-60.38229
9	47	36.58156	36.58300	36.58037	5	105	-10.89378	-10.89500	-10.89550
9	48	21.21845	21.22100	21.21706	5	106	-17.72766	-17.72500	-17.72478
10	49	86.98169	86.98000	86.98110	6	107	-26.82028	-26.82000	-26.82117
10	50	-9.09091	-9.08500	-9.08998	6	108	-34.78040	-34.78200	-34.78087
11	51	102.41940	102.41900	102.41919	7	109	-52.85312	-52.85100	-52.85383
11	52	2.82012	2.82400	2.82154	7	110	-8.69355	-8.69700	-8.69290
12	53	106.92825	106.92800	106.92897	8	111	-29.39995	-29.39500	-29.39764
12	54	11.52121	11.51400	-11.52155	8	112	-12.13121	-12.12900	-12.13222
13	55	113.84325	113.85000	113.84569	9	113	-67.31733	-67.31600	-67.31647
13	56	29.67726	29.67600	29.67534	9	114	21.52867	21.52700	21.52860
14	57	101.90265	101.90300	101.90310	10	115	-18.02939	-18.03100	-18.03133
14	58	39.07279	39.07100	39.07166	10	116	-8.76231	-8.69700	-8.69290

Ponto	Obs.	Teórica	Obs. simulada	Ajustada	Ponto	Obs.	Teórica	Obs. simulada	Ajustada
11	117	-2.08179	-2.08800	-2.08711	22	181	5.51671	5.52200	5.51828
11	118	3.17918	3.18200	3.18147	22	182	-69.08547	-69.08400	-69.08335
12	119	2.55838	2.55600	2.55724	23	183	16.50385	16.49800	16.49970
12	120	11.87181	11.88000	11.87312	23	184	-87.68333	-87.68700	-87.68181
13	121	9.76791	9.76500	9.76431	24	185	30.32983	30.32500	30.32519
13	122	29.94800	29.94700	29.94749	24	186	-107.02348	-107.02000	-107.01913
14	123	-2.32165	-2.31800	-2.31910	25	187	74.32822	74.32900	74.32449
14	124	39.29873	39.30100	39.29879	25	188	-85.63863	-85.63600	-85.63487
15	125	3.96773	3.97000	3.97142	26	189	82.82256	82.81900	82.82225
15	126	42.15291	42.15600	42.15616	26	190	-18.95701	-18.95800	-18.95582
16	127	-19.47727	-19.46900	-19.46991	27	191	102.50551	102.50700	102.50826
16	128	52.97652	52.97000	52.97471	27	192	-7.45848	-7.46000	-7.45922
17	129	-52.64464	-52.64400	-52.64337	28	193	97.97153	97.97500	97.97311
17	130	37.53759	37.53900	37.53633	28	194	-0.13539	-0.13200	-0.13359
18	131	-104.95947	-104.96100	-104.95878	31	195	86.39672	86.39500	86.39936
18	132	59.23939	59.23500	59.23656	31	196	-105.75381	-105.74800	-105.75094
19	133	-109.77704	-109.78100	-109.78314	32	197	-51.69499	-51.68800	-51.69273
19	134	67.49379	67.48900	67.48943	32	198	-22.82652	-22.82600	-22.82865
26	135	-108.46070	-108.45500	-108.45764	33	199	-12.46584	-12.46300	-12.46707
26	136	-26.07020	-26.07400	-26.06893	33	200	77.30953	77.31700	77.31662
27	137	-96.01377	-96.01500	-96.01569	34	201	-24.70654	-24.70600	-24.70477
27	138	-46.30154	-46.31000	-46.30273	34	202	-24.31417	-24.31300	-24.31251
28	139	-88.71241	-88.71300	-88.70896	1	203	39.22428	39.22400	39.22691
28	140	-41.61759	-41.61900	-41.61969	1	204	71.01373	71.01000	71.00830
29	141	-91.70213	-91.70700	-91.70292	2	205	60.48006	60.48100	60.47939
29	142	-63.55505	-63.54900	-63.55652	2	206	71.95045	71.94400	71.94911
30	143	-88.14753	-88.14500	-88.14449	3	207	52.68861	52.69200	52.68697
30	144	-87.25487	-87.25300	-87.25802	3	208	79.97243	79.96900	79.97114
33	145	-24.56318	-24.56000	-24.56161	4	209	40.85075	40.84900	40.85168
33	146	80.42062	80.42500	80.42249	4	210	82.17067	82.17400	82.16876
5	147	84.30629	84.30400	84.30220	5	211	-0.79064	-0.80000	-0.79298
5	148	89.55571	89.55900	89.55294	5	212	95.43343	95.42700	95.42921
6	149	100.71344	100.71200	100.71316	6	213	15.52201	15.51800	15.52241
6	150	74.22801	74.23100	74.22832	6	214	80.02913	80.02500	80.02882
7	151	74.68968	74.68700	74.68960	7	215	-10.54989	-10.54600	-10.54953
7	152	48.99998	49.00500	49.00094	7	216	53.75688	53.75400	53.75643
8	153	78.80109	78.80000	78.80185	8	217	-6.74803	-6.74400	-6.74616
8	154	72.22880	72.22800	72.23174	8	218	77.52728	77.53300	77.52893
9	155	44.91465	44.91500	44.91352	9	219	-40.46062	-40.45900	-40.45999
9	156	35.57112	35.57000	35.57097	9	220	39.78884	39.78400	39.78626
10	157	75.65613	75.65700	75.65481	10	221	-9.74792	-9.75000	-9.74918
10	158	83.10626	83.10200	83.10592	10	222	88.60338	88.60300	88.60147
11	159	63.92457	63.92600	63.92623	11	223	-20.83645	-20.83600	-20.83931
11	160	97.86166	97.86300	97.86258	11	224	103.45442	103.45400	103.45338
12	161	55.48889	55.48800	55.48999	12	225	-29.10917	-29.10900	-29.11029
12	162	102.27230	102.27500	102.27435	12	226	107.75458	107.75600	107.75435
13	163	37.89624	37.90400	37.90279	13	227	-46.29472	-46.29600	-46.29492
13	164	109.12357	109.12600	109.12828	13	228	114.28204	114.28100	114.28401
14	165	28.78321	28.78500	28.78332	14	229	-55.71579	-55.71500	-55.71486
14	166	98.23916	98.24000	98.24185	14	230	102.92475	102.92500	102.92425
15	167	25.90253	25.89500	25.89724	15	231	-58.19936	-58.20000	-58.20074
15	168	103.88614	103.88900	103.88616	15	232	108.60018	108.59700	108.59669
16	169	15.06241	15.05800	15.05746	16	233	-69.40011	-69.39800	-69.39630
16	170	82.25641	82.25300	82.25503	16	234	86.3831	86.38300	86.37720
17	171	29.72149	29.72100	29.72161	17	235	-55.63136	-55.62800	-55.63004
17	172	50.49219	50.49600	50.49356	17	236	54.61180	54.61100	54.61022
18	173	5.89717	5.89700	5.89836	18	237	-78.00366	-78.00700	-78.00528
18	174	-2.04983	-2.04600	-2.04905	18	238	1.98812	1.98400	1.98529
19	175	-2.59408	-2.59100	-2.58959	19	239	-86.12026	-86.12700	-86.12266
19	176	-6.87265	-6.87900	-6.87483	19	240	-2.79013	-2.79500	-2.79611
20	177	-11.19272	-11.19300	11.19140	20	241	-93.39314	-93.40200	-93.39937
20	178	-33.19272	-33.78000	-33.78010	20	242	-29.20324	-29.21000	-29.20735
21	179	-1.65259	-1.65600	-1.65490	21	243	-84.78673	-84.79000	-84.78923
21	180	-52.63409	-52.63300	-52.63322	21	244	-47.66434	-47.66700	-47.66751

Ponto	Obs.	Teórica	Obs. simulada	Ajustada	Ponto	Obs.	Teórica	Obs. simulada	Ajustada
22	245	-78.71659	-78.71900	-78.71868	3	273	-47.11686	-47.11400	-47.11419
22	246	-63.74603	-63.74500	-63.74787	3	274	78.71409	78.71000	78.71475
23	247	-68.69679	-68.69700	-68.69704	4	275	-59.32918	-59.32400	-59.32556
23	248	-81.96630	-81.96900	-81.96866	4	276	81.09205	81.08900	81.09170
24	249	-55.68128	-55.68600	-55.68430	5	277	-101.39977	-101.39700	-101.39907
24	250	-100.97936	-100.98000	-100.97868	5	278	95.17480	95.16800	95.17097
25	251	-11.98045	-11.98000	-11.98051	6	279	-84.86403	-84.86400	-84.86338
25	252	-80.92832	-80.92600	-80.92663	6	280	79.36746	79.37000	79.36750
26	253	7.80592	7.80100	7.80550	7	281	-110.84906	-110.85300	-110.84806
26	254	-14.95841	-14.96000	-14.95809	7	282	53.12028	53.11400	53.11831
27	255	28.34346	28.34600	28.34492	8	283	-107.63193	-107.63200	-107.62828
27	256	-3.38425	-3.38400	-3.38514	8	284	77.19646	77.19600	77.19739
28	257	23.82776	23.83200	23.83055	10	285	-110.72023	-110.71900	-110.72015
28	258	4.01330	4.02400	4.01482	10	286	88.46230	88.46300	88.45997
29	259	45.75098	45.75600	45.75242	25	287	-111.40909	-111.40900	-111.41025
29	260	0.35317	0.35400	0.35445	25	288	-84.78560	-84.78000	-84.78406
30	261	69.68720	69.68900	69.69005	26	289	-80.50869	-80.51300	-80.51036
30	262	3.34522	3.34300	3.34658	26	290	-16.55246	-16.55000	-16.55385
31	263	-1.45458	-1.44800	-1.44900	27	291	-59.26274	-59.26600	-59.26388
31	264	-100.89466	-100.89200	-100.89387	27	292	-4.51065	-4.51800	-4.51365
33	265	-95.71879	-95.71900	-95.72299	28	293	-63.87389	-63.86700	-63.87074
33	266	80.81478	80.81600	80.81731	28	294	2.87198	2.87200	2.87174
34	267	-106.47122	-106.46800	-106.47417	29	295	-41.61302	-41.60900	-41.61375
34	268	-19.90341	-19.91200	-19.90635	29	296	-0.52206	-0.52200	-0.52318
1	269	-60.65296	-60.64400	-60.64766	30	297	-17.87619	-17.86900	-17.86980
1	270	70.00858	70.00700	70.00381	30	298	2.75209	2.75100	2.75074
2	271	-38.85944	-38.86500	-38.86222	31	299	-101.60399	-101.60400	-101.60232
2	272	70.66143	70.66600	70.66171	31	300	-104.86845	-104.87000	-104.86799



## ANEXO B - Parâmetros ajustados

Foto	$\kappa$ (grau)	$\varphi$ (grau)	$\omega$ (grau)	$X_o$ (m)	$Y_o$ (m)	$Z_o$ (m)
1	83.302378	0.590523	-0.787120	1720.874	800.612	2767.710
2	84.308033	-0.366920	-0.291952	1874.078	1918.822	2764.730
3	84.231358	0.414711	0.330428	2005.132	2965.069	2759.050
4	-4.199785	-0.578447	2.305649	1190.241	1905.729	2770.117
5	-4.328907	1.515278	1.414111	2137.367	1820.992	2764.475
6	-5.085549	-1.737224	1.610067	3059.806	1741.021	2764.630

Ponto	X (m)	Y (m)	Z (m)
1	2551.767	2555.564	1200.173
2	2771.678	2551.109	1189.428
3	2696.066	2636.536	1198.989
4	2575.960	2666.386	1206.026
5	2161.814	2845.887	1202.231
6	2316.965	2668.457	1204.542
7	2027.955	2426.469	1186.906
8	2086.437	2664.673	1201.692
9	1703.931	2307.741	1182.754
10	2063.984	2782.914	1202.316
11	1959.292	2956.135	1193.525
12	1875.320	3012.428	1190.449
13	1696.722	3105.314	1183.594
14	1588.910	2991.047	1185.988
15	1564.480	3058.774	1180.528
16	1429.955	2829.287	1182.984
17	1556.329	2475.464	1188.257
18	1266.949	1944.563	1157.067
19	1174.004	1900.462	1153.595
20	1061.821	1622.327	1129.139
21	1148.569	1419.755	1136.670
22	1210.948	1249.362	1149.197
23	1309.684	1055.504	1156.099
24	1435.808	852.632	1158.729
25	1907.603	1030.772	1157.734
26	2167.736	1683.349	960.207
27	2419.127	1800.790	953.664
28	2373.073	1891.647	954.694
29	2626.279	1829.923	950.030
30	2907.051	1843.269	948.618
31	2003.339	832.220	1176.611
32	640.363	1769.697	1128.168
33	1134.455	2803.893	1169.246
34	926.035	1733.123	1132.318

## ANEXO C - Programas computacionais

```

% *****
%          PROGRAMA: FOTOL.M          *
% *          FOTOTRIANGULACAO ANALITICA EM BLOCO          *
% *          AJUSTAMENTO E DETERMINACAO DA MATRIZ G          *
% *          PROGRAMADOR: JAIR MENDES MARQUES          *
% *****
%          ENTRADA DOS DADOS          *
% *****
% ARQUIVO DADOS1 = NUM. FOTO E NUM. DO PONTO
% ARQUIVO DADOS2 = ELEM. DE ORIENTACAO
% ARQUIVO DADOS3 = PONTOS DE APOIO
% ARQUIVO DADOS4 = PARAMETROS APROXIMADOS
% ARQUIVO DADOS5 = FOTOCOORDENADAS
    load dados1.mat
    load dados2.mat
    load dados3.mat
    load dados4.mat
    load dados5.mat
% *****
% *          FORMACAO DO VETOR DOS PESOS E DO VETOR          *
% *          DAS OBSERVACOES          *
% *****
for i=1:NI,
    j=2*i;
    xlb(j-1)=xf(i)-xpp;
    xlb(j)=yf(i)-ypp;
    p1(j-1)=1./(dp^2);
    p1(j)=p1(j-1);
end
% MATRIZ DOS PESOS
p=diag(p1);
NPAR=6*NFO+3*NPT;
NR=NPAR;
% *****
% *          AJUSTAMENTO DAS OBSERVACOES          *
% *****
iter=0;
while iter<5
    J1=2*NI;
    for i=1:J1,
        for k=1:NPAR,
            a(i,k)=0.0;
        end
    end
    L=6*NFO;
    for i=1:NI,
        J=6*(ift(i)-1);
        c1=cos(x0(J+1));
        c2=cos(x0(J+2));
        c3=cos(x0(J+3));
        s1=sin(x0(J+1));
        s2=sin(x0(J+2));
        s3=sin(x0(J+3));
        xml1=c2*c1;
        xml2=c3*s1+s3*s2*c1;
    end
end

```

```

xm13=s3*s1-c3*s2*c1;
xm21=-c2*s1;
xm22=c3*c1-s3*s2*s1;
xm23=s3*c1+c3*s2*s1;
xm31=s2;
xm32=-s3*c2;
xm33=c3*c2;
N=L+3*(ipt(i)-1);
M=2*i;
dx=x0(N+1)-x0(J+4);
dy=x0(N+2)-x0(J+5);
dz=x0(N+3)-x0(J+6);
q=xm31*dx+xm32*dy+xm33*dz;
xm1=xm11*dx+xm12*dy+xm13*dz;
xn1=xm21*dx+xm22*dy+xm23*dz;
a(M-1,J+1)=(-df/q)*(dx*xm21+dy*xm22+dz*xm23);
a(M-1,J+2)=(df/q^2)*(q*(dx*s2*c1-dy*s3*c2*c1+dz*c3*c2*...
c1)+xm1*(dx*c2+dy*s3*s2-dz*c3*s2));
a(M-1,J+3)=(df/q^2)*(q*(dy*xm13-dz*xm12)-xm1*(dy*xm33-...
dz*xm32));
a(M-1,J+4)=(df/q^2)*(q*xm11-xm1*xm31);
a(M-1,J+5)=(df/q^2)*(q*xm12-xm1*xm32);
a(M-1,J+6)=(df/q^2)*(q*xm13-xm1*xm33);
a(M-1,N+1)=-a(M-1,J+4);
a(M-1,N+2)=-a(M-1,J+5);
a(M-1,N+3)=-a(M-1,J+6);
a(M,J+1)=(df/q)*(dx*xm11+dy*xm12+dz*xm13);
a(M,J+2)=(df/q^2)*(-q*(dx*s2*s1-dy*s3*c2*s1+dz*c3*c2*s1)...
+xn1*(dx*c2+dy*s3*s2-dz*c3*s2));
a(M,J+3)=- (df/q^2)*(q*(-dy*xm23+dz*xm22)+xn1*(dy*xm33-...
dz*xm32));
a(M,J+4)=(df/q^2)*(q*xm21-xn1*xm31);
a(M,J+5)=(df/q^2)*(q*xm22-xn1*xm32);
a(M,J+6)=(df/q^2)*(q*xm23-xn1*xm33);
a(M,N+1)=-a(M,J+4);
a(M,N+2)=-a(M,J+5);
a(M,N+3)=-a(M,J+6);
xl(M-1)=xlb(M-1)-df*(xm1/q);
xl(M)=xlb(M)-df*(xn1/q);
end
% CALCULO DA MATRIZ DAS EQ. NORMAIS
at=a';
xn1=at*p;
xn=xn1*a;
% CALCULO DO VETOR U
u=xn1*xl';
% AJUSTAMENTO COM INJUNCOES OU LIVRE?
if(NPA>2)
for i=1:NPA,
L=6*NFO;
N=L+3*(ipct(i)-1);
px=1.0/(dpx(i))^2;
py=1.0/(dpy(i))^2;
pz=1.0/(dpz(i))^2;
xn(N+1,N+1)=xn(N+1,N+1)+px;
xn(N+2,N+2)=xn(N+2,N+2)+py;
xn(N+3,N+3)+xn(N+3,N+3)+pz;
u(N+1)=u(N+1)-px*(xc(i)-x0(N+1));
u(N+2)=u(N+2)-py*(yc(i)-x0(N+2));
u(N+3)=u(N+3)-pz*(zc(i)-x0(N+3));
end
xg=inv(xn);

```

```

        else
            xg=pinv(xn);
        end
    end
    xn=xg;
% CALCULO DO VETOR DAS CORRECOES
    cx1=xn*u;
    cx=-1.0*cx1;
% CALCULO DOS PARAMETROS
    xa=x0+cx;
    x0=xa;
% PASSAGEM DE RADIANS PARA GRAUS
    for i=1:NFO,
        J=6*(i-1);
        for k=1:3,
            xa(J+k)=xa(J+k)*180.0/3.141592654;
        end
    end
    iter=iter+1
end
% SAIDA: KAPA, FI, OMEGA, X0, Y0, Z0
for i=1:6;
    par(1,i)=xa(i);
    par(2,i)=xa(i+6);
    par(3,i)=xa(i+12);
    par(4,i)=xa(i+18);
    par(5,i)=xa(i+24);
    par(6,i)=xa(i+30);
end
% SAIDA DE: X, Y, Z
for i=1:34,
    j=36+3*(i-1);
    for k=1:3,
        px(i,k)=xa(j+k);
    end
end
format long
end
save datal.mat par px
end
% *****
% * DETERMINACAO DAS MATRIZES: COVARIANCIA,          *
% *   CORRELACAO, Ezk, LAMBDA                       *
% *****
% MATRIZ COVARIANCIA DOS RESIDUOS
at=a';
an1=a*xn;
an2=an1*at;
p2=inv(p);
vv=p2-an2;
for i=1:NO,
    for j=1:NO,
        vv1(i,j)=(vv(i,j)+vv(j,i))/2.0;
    end
end
% MATRIZ DE CORRELACAO DOS RESIDUOS
for i=1:NO,
    for j=1:NO,
        if(i==j)
            r1(i,j)=1./sqrt(vv1(i,j));
        else
            r1(i,j)=0.0;
        end
    end
end

```

```

        end
    end
    rol=r1*vv1;
    ro=rol*r1;
    for i=1:NO,
    for j=1:NO,
        if (ro(i,j)>1.0)
            ro(i,j)=1.0;
        else
            ro(i,j)=ro(i,j);
        end
    end
end
end
for i=1:NO,
for j=1:NO,
    if (ro(i,j)<-1.0)
        ro(i,j)=-1.0;
    else
        ro(i,j)=ro(i,j);
    end
end
end
end
for i=1:NO,
for j=1:NO,
    ro(i,j)=(ro(i,j)+ro(j,i))/2.0;
end
end
% MATRIZ DOS AUTOVALORES E AUTOVETORES
[ez,d1]=eig(ro);
% FORMACAO DA MATRIZ Ezk
k1=0;
nc=0;
for i=1:NO;
    if (abs(d1(i,i)))>1.0
        k1=k1+1;
        d2(k1)=d1(i,i);
        for j=1:NO,
            v2(j,k1)=ez(j,i);
        end
    else
        nc=nc+1;
    end
end
end
% MONTAGEM DA MATRIZ LAMBDA zk
nd=NO-nc;
for i=1:nd,
    for j=1:nd,
        if i~=j
            v3(i,j)=0.0;
        else
            v3(i,j)=1./sqrt(d2(i));
        end
    end
end
end
% CALCULO DO VETOR DOS RESIDUOS
J=2*NI;
for i=1:J,
    v(i,1)=-x1(i);
end
end
% RESIDUOS PADRONIZADOS
z=r1*v;

```

```

% *****
% *           ARQUIVOS           SALVOS           *
% *****
save dados7.mat p vv1 r1 v2 v3 v NO NR nd
save dos11.mat a xl p z NO NR v
save auto.mat v2
end

```

```

% *****
% *           PROGRAMA: FOTO2.M           *
% *   DETERMINACAO DA MATRIZ G           *
% *   PROGRAMADOR: JAIR MENDES MARQUES   *
% *****
% *****
% *           ENTRADA DOS DADOS           *
% *****
% ARQUIVO DADOS7.MAT = GERADO PELO
% PROGRAMA FOTO1.M
load dados7.mat
% *****
% *           CALCULO DA MATRIZ G           *
% *****
% MATRIZ G
vv2=p*vv1;
r7=vv2*r1;
rv1=r7*v2;
rv=rv1*v3;
vr=rv';
% CALCULO DO VETOR DAS COMPONENTES
% PRINCIPAIS PADRONIZADAS
sr=vr*v;
% CALCULO DA ESTATISTICA QUI-QUADRADO
qq=0;
for i=1:nd,
qq=qq+(sr(i))^2;
qq=real(qq);
end
% CALCULO DO SIGMA A POSTERIORI
vt=v';
vtv=vt*v;
dpl=0.0040;
sigma=vtv*(1.0/dpl^2)/(NO-NR+7);
% *****
% *           ARQUIVOS SALVOS           *
% *****
save dos10.mat sr rv qq sigma nd
save qq2.mat qq
end

```

```

% *****
% *          PROGRAMA: FOTO3.M          *
% *  IDENTIFICACAO DAS OBSERVACOES    *
% *  SUSPEITAS E CALCULO DOS VICIOS E  *
% *  ERROS PADROES CORRESPONDENTES    *
% *          PROGRAMADOR: JAIR MENDES MARQUES    *
% *****
% *****
% *          ENTRADA DOS DADOS          *
% *****
% ARQUIVO DOS10.MAT = GERADO POR FOTO2.M
% ARQUIVO DOS11.MAT = GERADO POR FOTO1.M
% ARQUIVO CRIT.MAT = VALORES CRITICOS
% DE QUI-QUADRADO (MINITAB)
load dos10.mat
load dos11.mat
load crit.mat
% *****
% CALCULO DOS QUADRADOS DE sr
for i=1:nd;
    sq(i)=(sr(i))^2;
end
[s1,I]= sort(sq);
% DETERMINACAO DOS sr SIGNIFICATIVOS
% NO NIV. SIGN. = 5% COM G.L.=8
NK=0;
nn=nd;
sqq=qq;
j=1;
while sqq>qc(j)
    i=I(nn);
    sqq=sqq-s1(nn);
    NK=NK+1;
    J(NK)=i;
    nn=nn-1;
    j=j+1;
end
% *****
% *  IDENTIFICACAO DAS OBS. SUSP. - MATRIZ G  *
% *****
I2=0;
I3=0;
I4=0;
rm=abs(rv);
for i=1:NO,
    I1=0;
    [rn,I5]=sort(rm(i,:));
    for j=1:NK,
        if I5(nd)==J(j)
            I1=I1+1;
        else
            I2=I2+1;
        end
    end
    if I1==1
        I3=I3+1;
        sus(I3)=i;
    else
        I4=I4+1;
    end
end
end
% MONTAGEM DA MATRIZ B

```

```

for i=1:NO,
    for j=1:I3,
        if sus(j)==i
            b(i,j)=1.0;
        else
            b(i,j)=0.0;
        end
    end
end
end
% MONTAGEM DA MATRIZ C
NV=I3;
N2=NR+NV;
for i=1:NO,
    k=0;
    for j=1:N2,
        if j <= NR
            c(i,j)=a(i,j);
        else
            k=k+1;
            c(i,j)=b(i,k);
        end
    end
end
end
% *****
% *   CALCULO DOS VICIOS E ERROS PADROES   *
% *****
cp=c'*p;
xn=cp*c;
xg=pinv(xn);
u=cp*xl';
xv1=xg*u;
xv=-1.0*xv1;
for i=(NR+1):(NR+NV)
    j=i-NR;
    vic(j)=xv(i);
    dpv(j)=sqrt(xg(i,i));
    zvic(j)=vic(j)/dpv(j);
end
for i=1:NV
    susp(i,1)=sus(i);
    susp(i,2)=vic(i);
    susp(i,3)=dpv(i);
    susp(i,4)=zvic(i);
end
% *****
% *   ARQUIVOS SALVOS   *
% *****
save dat3.mat sus J NV NK vic dpv zvic susp
end

```



```

% *****
% *          PROGRAMA: FOTO4.M          *
% * ESTIMA OS VICIOS IDENTIFICADOS PELO TESTE *
% * DE FERGUSON E OS VICIOS FINAIS      *
% *          PROGRAMADOR: JAIR MENDES MARQUES      *
% *****
% *****
% *          ENTRADA DOS DADOS          *
% *****
% ARQUIVO DOS10.MAT = GERADO POR FOTO2.M
% ARQUIVO DOS11.MAT = GERADO POR FOTO2.M
% ARQUIVO DOS12.MAT = GERADO POR FERG.M
% ARQUIVO DOS32.MAT = VALORES CRITICOS DE t
% GERADOS PELO MINITAB
% *****
load dos10.mat
load dos11.mat
load dos12.mat
load dos32.mat
% *****
% MONTAGEM DA MATRIZ B
for i=1:NO,
    for j=1:NT,
        if out(j)==i
            b(i,j)=1.0;
        else
            b(i,j)=0.0;
        end
    end
end
% MONTAGEM DA MATRIZ C
N2=NR+NT;
for i=1:NO,
    k=0;
    for j=1:N2,
        if j <= NR
            c(i,j)=a(i,j);
        else
            k=k+1;
            c(i,j)=b(i,k);
        end
    end
end
% *****
% *          CALCULO DOS VICIOS E ERROS PADROES          *
% *****
cp=c'*p;
xn=cp*c;
xg=pinv(xn);
u=cp*xl';
xv1=xg*u;
xv=-1.0*xv1;
for i=(NR+1):(NR+NT)
    j=i-NR;
    vic1(j)=xv(i);
    dpv1(j)=sqrt(xg(i,i));
    zvic1(j)=vic1(j)/dpv1(j);
end
gl=169-NT;
tc=tvic(169-gl);
for i=1:NT
    vicio(i,1)=out(i);
end

```

```

        vicio(i,2)=vicl(i);
        vicio(i,3)=dpvl(i);
        vicio(i,4)=zvicl(i);
end
% *****
% *           ARQUIVOS SALVOS           *
% *****
save da4.mat vicl dpvl zvicl vicio tc gl
end

% *****
% *           PROGRAMA: FERG.M           *
% *           TESTE DE FERGUSON           *
% *           PROGRAMADOR: JAIR MENDES MARQUES *
% *****
% *****
% *           ENTRADA DOS DADOS           *
% *****
% ARQUIVO DAT13.MAT = GERADO POR FOTO3.M
% ARQUIVO FCRT.MAT = VALORES CRITICOS DE
% FERGUSON
load dat3.mat
load fcrit.mat
% *****
% ORDENANDO OS VICIOS
v2=abs(vic);
[v3,i3]=sort(v2);
% *****
% *           TESTE DE FERGUSON           *
% *****
a1=0;
b1=0;
for i=1:NV,
    a1=a1+(v3(i))^4;
end
for i=1:NV,
    b1=b1+(v3(i))^2;
end
c1=(b1/NV)^2;
t=a1/(NV*c1);
NN=NV;
k=NV-2;
j=0;
while t>vc(k);
    j=j+1;
    obs(j)=i3(NN);
    tf(j,1)=sus(obs(j));
    tf(j,2)=t;
    tf(j,3)=vc(k);
    out(j)=sus(obs(j));
    NN=NN-1;
    k=k-1;
    a1=0;
    b1=0;
    for i=1:NN
        a1=a1+(v3(i))^4;
    end
end

```

```

    for i=1:NN,
        b1=b1+(v3(i))^2;
    end
    c1=(b1/NN)^2;
    t=a1/(NN*c1);
end
NT=j;
% *****
% *           ARQUIVOS  SALVOS           *
% *****
save dat21.mat tf
save dos12.mat out NT
end

% *****
% *           PROGRAMA: ROTACAO.M           *
% *           CRITERIO ORTOMAX GENERALIZADO *
% *           PROGRAMADOR: JAIR MENDES MARQUES *
% *****
% *           ENTRADA DOS DADOS           *
% *****
% ARQUIVO AUTO.MAT = GERADO POR FOTO1.M
load auto.mat
NO=300;
nd=169;
v2=real(v2);
% *****
% ENTRADA DA MATRIZ c(0)=at(0)
c=v2;
at=v2';
% ENTRADA DOS PESOS w1 e w2
w1=1.0;
w2=1.0;
% *****
% CALCULO DAS CONSTANTES: d1, d2 e d3
d1=w1/nd;
d2=w2/NO;
d3=w1+w2;
% *****
% INICIO DO LOOP ITERATIVO
iter=0;
ite=0;
nt=NO*nd;
c5=zeros(NO,nd);
c6=ones(NO,nd);
m1=1
while ite<nt
    k1=0;
    k2=0;
    for i=1:NO,
        for j=1:nd,
            if abs(c5(i,j)-c6(i,j))<0.1
                k1=k1+1;
            else
                k2=k2+1;
            end
        end
    end
end
end

```

```

end
ite=k1;
m1=2
% CALCULO DA MATRIZ W
for i=1:NO,
for j=1:nd,
    w(i,j)=d3*(c(i,j))^3;
end
end
w=real(w);
w7=1
c5=c;
% CALCULO DA MATRIZ H
ct=c';
cct=c*ct;
cct=real(cct);
cd1=diag(cct);
cd=diag(cd1);
h=d1*cd;
h7=1
clear cct cd1 cd
% CALCULO DA MATRIZ F
cc=ct*c;
cc=real(cc);
dcl=diag(cc);
dc=diag(dcl);
f=d2*dc;
f=real(f);
f7=1
clear dcl dc
% CALCULO DA MATRIZ G
hc=h*c;
hc=real(hc);
cf=c*f;
cf=real(cf);
g1=w-hc;
g1=real(g1);
g2=g1-cf;
g2=real(g2);
g=at*g2;
g=real(g);
clear hc cf g1 g2 h w f
g7=1
% CALCULO DA MATRIZ T
gt=g';
ggt=g*gt;
ggt=real(ggt);
[ag,d]=eig(ggt);
for i=1:nd,
for j=1:nd,
    if(i==j)
        gg2(i,j)=1./sqrt(d(i,j));
    else
        gg2(i,j)=0.0;
    end
end
end
end
ag=real(ag);
ga=ag';
gg3=ag*gg2;
gg3=real(gg3);

```

```
gi=gg3*ga;
gi=real(gi);
t=gi*g;
t=real(t);
t7=1
clear ga ag gg2 gg3 gi ggt gt
% CALCULO DA NOVA MATRIZ C
c=v2*t;
c=real(c);
c6=c;
c7=1
% FIM DO LOOP ITERATIVO
iter=iter+1
end
% *****
% *           ARQUIVOS SALVOS           *
% *****
save rot7.mat c
save rot8.mat c
end
```

## REFERÊNCIAS BIBLIOGRÁFICAS

- [01] ADUOL, F.M.W. Detection of outliers in geodetic networks using principal component analysis and bias parameter estimation. **Technical Report**, Stuttgart, n. 2, p.1-104, 1987.
- [02] ALBERT, A. **Regression and the Moore-Penrose pseudoinverse**. New York:: Academic Press, 1972. 180p.
- [03] ANDERSON, T.W. **An introduction to multivariate statistical analysis**. New York : J. Wiley, 1979. 374p.
- [04] BARNETT, V.; LEWIS, T. **Outliers in statistical data**. New York: J.Wiley, 1984, 463p.
- [05] BEN-ISRAEL, A; GREVILLE, T.N.E. **Generalized inverses: theory and applications**. New York: J. Wiley, 1973. 395p.
- [06] BERBERAN, A. Outlier detection and heterogeneous observations: a simulation case study. **Australian Journal of Geodesy, Photogrammetry and Surveying**, Canberra, n. 56, p. 49-61, June 1992.
- [07] BJERHAMMAR, A. **Theory of errors and generalized matrix inverses**. Amsterdam: Elsevier, 1973. 420p.
- [08] BLAHA, G. A note on adjustment of free networks. **Bulletin Géodésique**, n.56, p.281-299, 1982.
- [09] BOULLION, T.L.; ODELL, P.L. **Generalized inverse matrices**. New York: Wiley Interscience, 1971. 103p.
- [10] CARROL, J.B. An analytical solution for aproximating simple-structure in factor analysis. **Psychometrika**, [S.I.], v. 18, n. 1, p. 343-370, march 1953.
- [11] CHATFIELD, C.; COLLINS, A.J. **Introduction to multivariate analysis**. London: Chapman & Hall, 1992. 246.
- [12] CHAVES NETO, ANSELMO; OLANDOSKI, MÁRCIA. **Tópicos de análise multivariada**. Curitiba: Universidade Federal do Paraná, 1987, 51p.
- [13] CLERICI, E.; HARRIS, M.W. A premium-protection method applied to detection and rejection of erroneous observations. **Manuscripta Geodaetica**, New York, v. 5, n. 4, p. 283-298, 1980.

- [14] CROSILLA, F. Improving the outlier separability in geodetic networks according to the generalized orthomax criterion. **Manuscripta Geodaetica**, New York, v. 11, n. 1, p. 38-47, 1986.
- [15] \_\_\_\_\_. Procrustean transformation as a tool for the construction of a criterion matrix for control networks. **Manuscripta Geodaetica**, v. 8, p. 343-370, 1983.
- [16] CROSS, P.A.; PRICE, D.R. A strategy for the distinction between single and multiple gross errors in geodetic networks. **Manuscripta Geodaetica**, New York, v. 10, n. 3, p. 172-178, 1985.
- [17] DALMOLIN, Quintino; MARQUES, Jair Mendes. O uso da análise multivariada na detecção de outlier em fototriangulação. In.: CONGRESSO BRASILEIRO DE CARTOGRAFIA, 16., Rio de Janeiro, 1993. **Anais ...** Rio de Janeiro: Sociedade Brasileira de Cartografia. v.2.
- [18] DAVIES, L.; GATHER, U. The identification of multiple outliers. **Journal of the American Statistical Association**, Alexandria, v. 423, n. 88, p.782-792, 1993.
- [19] EL-HAKIM, S.F. A practical study of gross-error detection in bundle adjustment. **The Canadian Surveyor**, Ottawa, v.35, n. 4, p. 373-386, December 1981.
- [20] \_\_\_\_\_. Data snooping with weighted observations. In.: SIMPOSIUM MATHEMATICAL MODELS, ACCURACY ASPECTS AND QUALITY CONTROL, Otaniemi, 1982. **Proceedings...** Otaniemi, Helsinki University of Technology, 1982. p.126-33.
- [21] FIRKOWSKI, Henrique. **Confiabilidade da fototriangulação: configuração e detecção de erros no apoio de campo**. Curitiba, 1986. Dissertação (Mestrado em Ciências Geodésicas) - Setor de Tecnologia, Universidade Federal do Paraná.
- [22] GEMAEL, Camil. **Introdução ao ajustamento de observações: aplicações geodésicas**. Curitiba : Diretório Acadêmico do Setor de Tecnologia, 1984. 205p.
- [23] \_\_\_\_\_. **Ajustamento livre: inversas generalizadas**. Curitiba: Universidade Federal do Paraná, Curso de Pós-Graduação em Ciências Geodésicas, 1977. 30p.
- [24] HARRIS, C.H.; KAISER, H.F. Oblique factor analytic solutions by orthogonal transformations. **Psychometrika**, [S.l.], v. 29, n. 4, p. 347-362, December 1964.
- [25] JACKSON, J.E.; MUDHOLKAR, G.S. Control procedure for residuals associated with principal component analysis. **Technometrics**, [S.l.], v. 21, n. 3, p.341-349, August 1979.

- [26] JOHNSON, R.A.; WICHERN, D.W. **Applied multivariate statistical analysis**. Englewood Cliffs : Prentice-Hall, 1988. 607p.
- [27] JUHL, J. The danish method of weight reduction for gross error detection. In.: INTERNATIONAL CONGRESS OF PHOTOGRAMMETRY AND REMOTE SENSING, 15., Rio de Janeiro, 1984. **International Archives ...** Rio de Janeiro: International Society of Photogrammetry and Remote Sensing, 1984. v. 25, t. A3a, p.468-72.
- [28] KAISER, H.F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, [S.l], v. 23, n. 3, p. 187-200, September 1958.
- [29] KENDALL, M. **Multivariate analysis**. London : C. Griffin, 1980. 210p.
- [30] KOCH, K.R. **Statistical analysis of geodetic data with applications for the deformation analysis**. Curitiba : Curso de Pós-Graduação em Ciências Geodésicas, 1985. 125p.
- [31] KUBIK, K; WANG, Y. Comparison of different principles for outlier detection. **Australian Journal of Geodesy, Photogrammetry and Surveying**, Canberra, n. 54, p.67-80, June 1991.
- [32] LUGNANI, João Bosco. **Introdução à fototriangulação**. Curitiba : Universidade Federal do Paraná, 1987. 134p.
- [33] MARDIA, K.M.; KENT, J.T.; BIBBY, J.M. **Multivariate analysis**. London : Academic, 1979. 212p.
- [34] MARQUES, Jair Mendes. **Análise estatística das discrepâncias parciais de uma rede de nivelamento de primeira ordem**. Curitiba, 1982. Dissertação (Mestrado em Ciências Geodésicas) - Setor de Tecnologia, Universidade Federal do Paraná.
- [35] MERCHANT, D.C. **Analytical photogrammetry: theory and practice**. [S.l] : The Ohio State University, 1979. 2v.
- [36] MINITAB - Reference Manual PC version. Rosemont, PA: Quickset, november 1991.
- [37] MITISHITA, Edson Aparecido. **Detecção de erros grosseiros nas aerotriangulações**. Curitiba, 1986. Dissertação (Mestrado em Ciências Geodésicas) - Setor de Tecnologia, Universidade Federal do Paraná.
- [38] MOOD, A.M.; GRAYBILL F.A.; BOES D.C. **Introduction to the theory of statistics**. Singapore: McGraw-Hill, 1974. 564p.
- [39] MORRISON, D.F. **Multivariate statistical methods**. Tokyo : McGraw-Hill, 1976. 415p.



- [40] MOURA, José Ozildo. **Aplicação do ajustamento livre à fotogrametria.** Curitiba, 1981. Dissertação (Mestrado em Ciências Geodésicas) - Setor de Tecnologia, Universidade Federal do Paraná.
- [41] ODEN, J.T. **Applied functional analysis.** New Jersey: Prentice-Hall, 1979. 426p.
- [42] OU, Z.Q. Sequential tests for outliers in the general linear model. **Australian Journal of Geodesy, Photogrammetry and Surveying.** Canberra, n. 50, p.37-49, June 1989.
- [43] PATTERSON, R.C. Outlier detection: some distribution theory. **Australian Journal of Geodesy, Photogrammetry and Surveying.** Canberra, n. 43, December 1985.
- [44] POPE, A.J. The statistics of residuals and the detection of outliers. **NOAA Technical Report NOS 65 NGS, [S.I.]**, v. 1, p. 1-133, 1976.
- [45] PRESS, J. **Applied multivariate analysis: using byesian and frequentist methods of inference.** London : R. Krieger, 1982. 600p.
- [46] SHERIN, R.J. A matrix formulation of Kaiser's varimax criterion. **Psychometrika, [S.I.]**, v. 31, p. 535-538, december 1966.
- [47] SILVA, João Fernando Custodio. **A inferência bayesiana e a detecção e localização automática erros grosseiros em fototriangulação por feixe de raios.** Curitiba, 1987. Tese (Doutorado em Ciências Geodésicas) - Setor de Tecnologia, Universidade Federal do Paraná.
- [48] 386-MATLAB user's guide. Natick, MA : The Math Works, Inc., October 1990.
- [49] UOTILA, U.A. Generalized inverse as a weight matrix. **The Canadian Surveyor**, v. 28, n.5, p. 698-701, december 1974.
- [50] XU, P. Consequences of constant parameters and confidence intervals of robust estimation. **Bollettino di Geodesia e Scienze Affini.** Firenze, v. 3, p. 231-249,