

**Diagnóstico
em
análise discriminante**

Sueli Maria Beltrame Reigada

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
para
obtenção do grau
de
Mestre em Estatística

Área de Concentração: **Estatística**

Orientadora: **Profa. Dra. Silvia Nagib Elian**

São Paulo – dezembro – 2004

Diagnóstico em análise discriminante

Este exemplar corresponde
à redação final da dissertação
devidamente corrigida e
apresentada por
Sueli Maria Beltrame Reigada
e aprovada pela comissão julgadora.

São Paulo, março de 2005.

Banca examinadora:

- | | |
|--|-----------|
| – Profa. Dra. Sílvia Nagib Elian (Orientadora) | IME –USP |
| – Profa. Dra. Lúcia Pereira Barroso | IME–USP |
| – Prof. Dr. Edwin Moisés Marcos Ortega | ESALQ–USP |

Agradeço a Deus por tudo,
pela minha saúde,
pela minha família,
e por ter me dado a oportunidade
de mais esta conquista.

Agradeço às minhas filhas,
Juliana e Letícia,
por existirem na minha vida.
A elas, dedico este trabalho.

Agradecimentos

Em primeiro lugar quero agradecer à minha orientadora e amiga Profª Drª Sílvia Nagib Elian pela orientação clara e precisa, e, principalmente pelo apoio e incentivo em todos os momentos.

Agradeço a meus pais por seu amor e dedicação, e tenho certeza que, onde quer que estejam, compartilham comigo da felicidade deste momento.

Agradeço ao meu marido, Rwitter, amigo e companheiro que sempre me apóia em todos os meus sonhos e realizações.

Ao José Otávio e Felipe, agradeço a colaboração e por tornarem mais feliz a vida das minhas filhas.

Agradeço pelos valiosos ensinamentos transmitidos no decorrer desses anos, aos professores Caio, Luiz Gustavo, Cláudia, Julio, Beti, Sílvia Ferrari, Márcia, Antonio Carlos, Lúcia, Carlinhos, Julio Stern, Oscar João, e em especial aos professores Gilberto e Sérgio.

Aos meus amigos Vaudeluci, Mayra, Cléber, Elias, Isabel, Elisa, Livia, Lourdes, Diana, Rosângela, Conrad e Levi agradeço por todos os bons momentos vividos entre listas e mais listas de exercícios, e também pela atenção e carinho nos momentos difíceis, e, um agradecimento especial ao amigo João, sempre disposto a resolver as minhas dúvidas.

Finalmente, agradeço a todos que, de uma forma ou de outra contribuíram para a realização deste trabalho.

Resumo

Neste trabalho foram estudadas as medidas de diagnóstico em análise discriminante em duas situações principais: para duas populações com matriz de covariância constante e para duas populações com matrizes de covariâncias diferentes (análise discriminante quadrática). É estudada ainda a abordagem Bayesiana para a detecção de observações influentes. No final do trabalho examina-se também as medidas de diagnóstico para a análise discriminante múltipla, ou seja, quando se tem mais que duas populações. O estudo é complementado com a apresentação das principais medidas estudadas aplicadas a um conjunto de dados reais.

Abstract

In this work the measures of diagnostics in discriminant analysis in two main situations had been studied: for two populations with matrix of covariance constant and for two populations with different matrices of covariances (quadratic discriminant analysis). The Bayesian approach is studied still for the detection of influential observations. In the end of the work it is examined the measures of diagnostics for the multiple discriminant analysis when there is more than two populations. The study is complemented with the presentation of the main measures studied applied to a real data set.

Sumário

1	Introdução	1
2	Diagnóstico em análise discriminante com duas populações e matriz de covariância constante	
2.1	Introdução	4
2.2	Função de influência em análise discriminante	
2.2.1	Função de Influência	5
2.2.2	Distribuição aproximada para a função de influência $I(\mathbf{x}; \Delta^2)$	16
2.2.3	Gráficos de probabilidade para detectar <i>outliers</i>	18
2.3	Principais estatísticas no diagnóstico em análise discriminante	23
2.3.1	Distribuições assintóticas	25
2.4	Medidas relacionadas com a probabilidade de classificação incorreta	
2.4.1	Medidas baseadas na curva de influência	30
2.4.2	Medidas alternativas	40
2.5	Outras medidas de diagnóstico	
2.5.1	Medidas baseadas na média quadrática	45
2.5.2	Medidas de influência para regressão logística – uma abordagem para análise discriminante	48
3	Detecção de observações influentes no contexto Bayesiano	
3.1	Densidades preditivas e divergência de Kullback-Leibler	52
3.2	Definição dos objetivos inferenciais	55
3.3	Medida de influência global, local e individual	58
3.4	Exemplo	62
4	Comparações entre algumas medidas de diagnóstico	68
5	Observações influentes em análise discriminante quadrática	76

6	Observações influentes em análise discriminante para g grupos	
6.1	Observações influentes na probabilidade estimada de uma observação pertencer a determinado grupo	82
6.2	Medidas baseadas na probabilidade de classificação incorreta	87
6.3	Uso de funções de influência para as medidas de probabilidade de classificação incorreta	91
7	Verificação das medidas em um conjunto de dados	99
8	Considerações finais	120
A Apêndice		
A-1	Conceitos básicos em análise discriminante	
A-1.1	Função discriminante de Fisher para duas populações	122
A-1.2	Probabilidade de classificação incorreta	126
A-1.3	Dois populações normais multivariadas	129
A-1.4	Discriminação entre g populações	133
A-2	Conceitos básicos sobre diagnóstico em regressão	138
B Apêndice		
	Definições, resultados e cálculos	143
C Apêndice		
	Programas no S-plus	162
D Apêndice		
	Referências Bibliográficas	177

Capítulo 1

Introdução

Análise discriminante é uma técnica estatística multivariada que, além de separar populações em grupos com base em suas características mais influentes, também estabelece uma regra para a alocação de novos elementos (de origem desconhecida) nos grupos anteriormente determinados ou em grupos já pré-existentes.

Diagnóstico é uma técnica que possibilita a identificação de pontos influentes. Tais pontos, por diferirem dos demais em suas medidas, podem alterar de forma significativa as conclusões de uma análise estatística. Em problemas que envolvem ajustes de modelos, de modo geral, afetam as estimativas dos coeficientes. Métodos de diagnóstico podem ser usados ainda para apontar colinearidade entre os dados e também identificar aspectos que não condizem com as suposições iniciais, sugerindo então ações reparadoras para a análise adequada do ajuste.

A literatura sobre análise discriminante bem como sobre diagnóstico é muito vasta, mas a pesquisa na determinação de pontos influentes em análise discriminante ainda é pouco explorada.

O estatístico frequentemente se depara com o problema de como detectar e tratar observações aparentemente atípicas, e esse problema fica mais complicado ainda quando se analisa dados multivariados.

Este trabalho dedica-se ao estudo do diagnóstico em análise discriminante, apresentando as medidas definidas pelos principais pesquisadores do assunto nos últimos 26 anos.

O Capítulo 2 constitui a base da dissertação, tratando de medidas de diagnóstico para duas populações e matriz de covariância constante. Inicialmente são abordadas as medidas baseadas na função de influência de Hampel (1974). Em seguida, são apresentadas duas estatísticas fundamentais das quais várias medidas dependem. Ainda no Capítulo 2 são analisadas medidas relacionadas com a probabilidade de classificação incorreta. Medidas de influência para regressão logística em conexão com análise discriminante também são mencionadas.

O Capítulo 3 trata da detecção de observações influentes no contexto Bayesiano, utilizando as medidas de divergência de Kullback-Leibler (1951). Sob essa abordagem, são definidas medidas de influência globais, locais e individuais.

No Capítulo 4 são comparadas algumas estatísticas que, embora obtidas através de abordagens diversas por autores distintos (Johnson (1987) e Fung (1996-b)), produzem resultados semelhantes. Além disso, são comparadas medidas obtidas por um mesmo autor, Fung (1992 e 1998), que, calculadas de formas distintas também fornecem resultados similares.

O Capítulo 5 se destina a apresentar as medidas de diagnóstico utilizadas na análise discriminante quadrática, que é utilizada geralmente quando as matrizes de covariância são diferentes, apesar de ser uma análise menos robusta.

No Capítulo 6, são abordadas as medidas que detectam observações influentes em análise discriminante múltipla, ou seja, quando existem mais que duas populações. São feitas três abordagens: a primeira usa a probabilidade estimada que uma observação pertença a determinado grupo, a segunda se utiliza da probabilidade de classificação incorreta e a terceira leva em conta a função de influência discutida no Capítulo 2.

O Capítulo 7 mostra a aplicação das medidas de diagnóstico a um conjunto de dados reais e o Capítulo 8 finaliza o trabalho apresentando algumas conclusões.

Sugere-se a leitura inicial do Apêndice A, uma vez que a maior parte da notação utilizada para análise discriminante está aí definida. Nesse apêndice são revistos, de forma geral, alguns conceitos básicos sobre análise discriminante e diagnóstico que serão utilizados no decorrer da dissertação.

Capítulo 2

Diagnóstico em análise discriminante com duas populações e matriz de covariância constante

2.1 Introdução

Uma vez que em análise discriminante os dados são multivariados, a detecção e análise de *outliers* se torna muito mais complexa do que no caso univariado. Isso porque um *outlier* multivariado não é fácil de ser caracterizado, o que torna complicada a sua identificação. Pode surgir de um erro grosseiro em um de seus componentes ou de pequenos erros em todos eles. Além disso, o *outlier* multivariado pode distorcer medidas de localização e escala como por exemplo, vetor de médias e matriz de covariâncias.

Essa complexidade faz com que os procedimentos de detecção sejam elaborados com o cuidado de se proteger contra certos tipos de problemas como por exemplo distorções da correlação, em que o *outlier* pode indicar que as variáveis são

relacionadas quando realmente não o são. Além do mais, deve-se ter em mente que uma observação que é julgada discrepante para um propósito, pode ser considerada normal para outro.

Este capítulo tem por objetivo o estudo de medidas de diagnóstico que apontam *outliers* em análise discriminante com apenas duas populações e matriz de covariância constante.

2.2 Função de influência em análise discriminante

2.2.1 Função de influência

Usualmente na análise de diagnóstico, quando o procedimento estatístico adotado envolve estimação, uma estratégia intuitiva na determinação de pontos influentes é proceder ao ajuste com e sem o ponto suspeito e comparar as estimativas obtidas. Nessa linha de estudo, Campbell (1978) sugeriu o uso da função de influência, proposta por Hampel (1974).

Num problema que envolve g populações com funções distribuição $F_1, F_2, \dots, F_k, \dots, F_g$, seja $\theta = T(F_1, F_2, \dots, F_k, \dots, F_g)$, um parâmetro geral, obtido a partir das funções de distribuição F_k , $k = 1, \dots, g$. Para um valor \mathbf{x} do vetor aleatório \mathbf{X} , a função de influência $I(\mathbf{x}; \theta)$ é definida como:

$$I(\mathbf{x}; \theta) = \lim_{\varepsilon \rightarrow 0} \left(\frac{\tilde{\theta} - \theta}{\varepsilon} \right)$$

em que $0 < \varepsilon < 1$, $\tilde{\theta} = T(\tilde{F})$, $\tilde{F} = (F_1, \dots, \tilde{F}_k, \dots, F_g)$

e $\tilde{F}_k = (1 - \varepsilon)F_k + \varepsilon \delta_x$ é a função distribuição F_k após sofrer perturbação no ponto \mathbf{x} , sendo δ_x uma função de distribuição que assume probabilidade 1 no ponto \mathbf{x} .

Segundo Gnanadesikan (1997), a função de influência é muito útil para avaliar o efeito das observações em uma estatística pois é usada não só como uma ferramenta para obter tipos específicos de estimadores robustos, mas também como meio de desenvolver métodos para detectar *outliers*.

O objetivo principal da função de influência teórica é calcular a influência de um particular ponto \mathbf{x} no parâmetro θ , de modo que valores altos para essa função indicariam que \mathbf{x} tem grande influência em θ . Existem três versões amostrais da função de influência.

A função de influência empírica (assim denominada por Mallows (1973) citado em Gnanadesikan (1997)) é obtida trocando F da definição pela distribuição acumulada empírica F_n .

Na segunda versão, o objetivo é estudar a diferença entre $\hat{\theta}$ e $\hat{\theta}_+$ (estimador de θ obtido com base nas n observações originais mais uma observação adicional \mathbf{x}): $I_+(\mathbf{x}; \theta) = (n+1)(\hat{\theta}_+ - \hat{\theta})$, que coincide com a curva de sensibilidade usada por Andrews et al (1972), citado em Gnanadesikan (1997), para estudar as propriedades de vários estimadores robustos para parâmetros de localização.

A terceira versão, que será aqui abordada, analisa o efeito individual das observações da amostra, sendo então adequada para calcular a influência de observações no estimador $\hat{\theta}$. É a chamada função de influência amostral de Devlin, Gnanadesikan e Kettenring (1975), definida como:

$I_-(\mathbf{x}_i; \hat{\theta}) = (n-1)(\hat{\theta} - \hat{\theta}_{-i})$ $i = 1, \dots, n$ em que $\hat{\theta}_{-i}$ é o estimador de θ calculado quando se omite a observação \mathbf{x}_i .

Tanto I_+ como I_- podem ser consideradas aproximações da função de influência empírica tomando ε como $\frac{1}{n+1}$ em I_+ e ε como $-\frac{1}{n-1}$ em I_- (Mallows (1973) citado em Gnanadesikan (1997)).

Em análise de grupos multivariados, caso em que o parâmetro estudado envolve mais que uma população, conforme será visto a seguir, a função de influência é determinada excluindo uma observação de apenas um grupo. Se \tilde{F}_k é a função de distribuição para o grupo k , eliminando-se a i -ésima observação desse grupo, avalia-se o parâmetro com base no grupo perturbado e nos restantes não perturbados, $\tilde{\theta}_k = T(F_1, F_2, \dots, \tilde{F}_k, \dots, F_g)$ e subtrai-se o parâmetro calculado para todos os grupos antes da perturbação $\theta = T(F_1, F_2, \dots, F_g)$.

No cálculo da função de influência teórica permanecem apenas os termos de *ordem* ε . Isso equivale, na versão amostral, a assumir que termos de *ordem* n^{-2} podem ser desprezados (Definição B-1 do Apêndice B).

A utilização de funções de influência para detectar *outliers* em análise discriminante foi proposta por Campbell (1978).

O autor considera o caso em que o vetor aleatório \mathbf{X} tem distribuição normal p -variada com média $\boldsymbol{\mu}_j$ e matriz de covariância $\boldsymbol{\Sigma}$, ($\mathbf{X} \sim N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$), para \mathbf{X} pertencendo a população π_j , $j = 1, 2$.

Sejam a função linear discriminante dada em (A-1.2):

$$y = \mathbf{l}'\mathbf{x} = \boldsymbol{\delta}'\boldsymbol{\Sigma}^{-1}\mathbf{x} \quad \text{sendo que } \boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2,$$

e os parâmetros de interesse :

- distância de Mahalanobis: $\Delta^2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$

- média discriminante: $\mathbf{l}'\boldsymbol{\mu}_j$
- vetor de coeficientes da função discriminante: $\mathbf{l} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$.

Uma vez que os parâmetros de interesse são funções de $\boldsymbol{\Sigma}^{-1}$ e de $\boldsymbol{\mu}_j$, para calcular a função de influência serão consideradas inicialmente perturbações em $\boldsymbol{\Sigma}^{-1}$ e $\boldsymbol{\mu}_j$.

Seja $\boldsymbol{\Sigma} = w_1\boldsymbol{\Sigma}_{F_1} + w_2\boldsymbol{\Sigma}_{F_2}$ em que $\boldsymbol{\Sigma}_{F_1}$ é a matriz de covariâncias para a população 1, $\boldsymbol{\Sigma}_{F_2}$ é a matriz de covariâncias para a população 2, $w_1 + w_2 = 1$ e $w_j > 0$. Os pesos w_1 e w_2 são adotados uma vez que as amostras podem ter tamanhos diferentes, então $w_1 = \frac{n_1}{n_1 + n_2}$ e $w_2 = \frac{n_2}{n_1 + n_2}$,

$$\boldsymbol{\Sigma}_{F_j} = \int (\mathbf{x} - \boldsymbol{\mu}_j)(\mathbf{x} - \boldsymbol{\mu}_j)' dF_j \quad \text{e}$$

$$\boldsymbol{\mu}_j = \int \mathbf{x} dF_j .$$

No caso, assume-se que $\boldsymbol{\Sigma}_{F_1} = \boldsymbol{\Sigma}_{F_2}$ ou seja, as matrizes de covariâncias são iguais. Perturbando-se a primeira população, calcula-se $\boldsymbol{\Sigma}_{F_j}$ e $\boldsymbol{\mu}_j$ no ponto $\tilde{F}_1 = (1 - \varepsilon)F_1 + \varepsilon \delta_{\mathbf{x}}$.

Usando a notação * sobrescrito para o parâmetro após perturbação, tem-se:

$$\boldsymbol{\mu}_1^* = \int \mathbf{x} d[(1 - \varepsilon)F_1 + \varepsilon \delta_{\mathbf{x}}]$$

$$= (1 - \varepsilon)\boldsymbol{\mu}_1 + \mathbf{x}\varepsilon$$

$$= \boldsymbol{\mu}_1 - \varepsilon \boldsymbol{\mu}_1 + \mathbf{x}\varepsilon ,$$

sendo que a integral aqui calculada é a integral de Riemann e Stieltjes (James (1996 p. 100)).

Portanto,

$$\mu_1^* = \mu_1 + \varepsilon(\mathbf{x} - \mu_1) = \mu_1 + \varepsilon \mathbf{z} \quad \text{com } \mathbf{z} = \mathbf{x} - \mu_1 . \quad (2.1)$$

Para $\delta = \mu_1 - \mu_2$ tem-se:

$$\begin{aligned} \delta^* &= \int \mathbf{x} d[(1-\varepsilon)F_1 + \varepsilon \delta_x] - \int \mathbf{x} dF_2 \\ &= (1-\varepsilon)\mu_1 + \varepsilon \mathbf{x} - \mu_2 \\ &= \mu_1 - \mu_2 + \varepsilon \mathbf{x} - \varepsilon \mu_1 \\ &= \mu_1 - \mu_2 + \varepsilon (\mathbf{x} - \mu_1) , \end{aligned}$$

assim,

$$\delta^* = \delta + \varepsilon \mathbf{z} . \quad (2.2)$$

Da mesma forma,

$$\begin{aligned} \Sigma_{F_1}^* &= \int (\mathbf{x} - \mu_1)(\mathbf{x} - \mu_1)' d \tilde{F}_1 \\ &= \int (\mathbf{x} - \mu_1)(\mathbf{x} - \mu_1)' d [(1-\varepsilon)F_1 + \varepsilon \delta_x] \\ &= (1-\varepsilon)\Sigma_{F_1} + \varepsilon \mathbf{z}\mathbf{z}' . \end{aligned} \quad (2.3)$$

Lembrando que

$$\Sigma = w_1 \Sigma_{F_1} + w_2 \Sigma_{F_2},$$

então, Σ após a perturbação fica

$$\Sigma^* = w_1 [(1-\varepsilon) \Sigma_{F_1} + \varepsilon \mathbf{z}\mathbf{z}'] + w_2 \Sigma_{F_2}$$

$$= w_1 \Sigma_{F_1} + w_2 \Sigma_{F_2} - \varepsilon w_1 \Sigma_{F_1} + w_1 \varepsilon \mathbf{z}\mathbf{z}' = (w_1 + w_2) \Sigma_{F_1} - \varepsilon w_1 \Sigma_{F_1} + w_1 \varepsilon \mathbf{z}\mathbf{z}',$$

pois $\Sigma_{F_1} = \Sigma_{F_2}$.

Como $\Sigma = w_1 \Sigma_{F_1} + w_2 \Sigma_{F_2}$ e $\Sigma_{F_1} = \Sigma_{F_2}$ então $\Sigma = \Sigma_{F_1} (w_1 + w_2)$,

$$\Sigma^* = \Sigma - \varepsilon w_1 \Sigma + w_1 \varepsilon \mathbf{z}\mathbf{z}' \quad \text{e assim} \quad \Sigma^* = (1 - \varepsilon w_1) \Sigma + \varepsilon w_1 \mathbf{z}\mathbf{z}' . \quad (2.4)$$

Para calcular Σ^{-1*} usa-se o seguinte resultado matricial:

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{C}^{-1} + \mathbf{D} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{D} \mathbf{A}^{-1}$$

onde \mathbf{A} , \mathbf{B} , \mathbf{C} e \mathbf{D} são matrizes de dimensões $p \times p$, $p \times n$, $n \times n$ e $n \times p$ respectivamente, sendo que no caso, $n = 1$.

Portanto, tomando-se $\mathbf{A} = (1 - \varepsilon w_1) \Sigma$, $\mathbf{B} = \mathbf{z}$, $\mathbf{C} = \varepsilon w_1$ e $\mathbf{D} = \mathbf{z}'$, segue que

$$\begin{aligned} \Sigma^{-1*} &= (1 - \varepsilon w_1)^{-1} \Sigma^{-1} - \frac{(1 - \varepsilon w_1)^{-1} \Sigma^{-1} \mathbf{z}\mathbf{z}' (1 - \varepsilon w_1)^{-1} \Sigma^{-1}}{(\varepsilon w_1)^{-1} + \mathbf{z}' (1 - \varepsilon w_1)^{-1} \Sigma^{-1} \mathbf{z}} \\ &= (1 - \varepsilon w_1)^{-1} \left[\Sigma^{-1} - \frac{\Sigma^{-1} \mathbf{z}\mathbf{z}' (1 - \varepsilon w_1)^{-1} \Sigma^{-1}}{(\varepsilon w_1)^{-1} (1 - \varepsilon w_1)^{-1} [(1 - \varepsilon w_1) + \varepsilon w_1 \mathbf{z}' \Sigma^{-1} \mathbf{z}]} \right] \end{aligned}$$

$$\begin{aligned}
&= (1 - \varepsilon w_1)^{-1} \left(\Sigma^{-1} - \frac{\varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1}}{1 - \varepsilon w_1 + \varepsilon w_1 \mathbf{z}' \Sigma^{-1} \mathbf{z}} \right) \\
&= \frac{\Sigma^{-1}}{(1 - \varepsilon w_1)} - \frac{\varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1}}{(1 - \varepsilon w_1) [1 - \varepsilon w_1 (1 - \mathbf{z}' \Sigma^{-1} \mathbf{z})]}.
\end{aligned}$$

Multiplicando e dividindo essa expressão por $(1 + \varepsilon w_1)$ e mantendo apenas os termos de ordem ε obtém-se :

$$\begin{aligned}
\Sigma^{-1*} &= \frac{(1 + \varepsilon w_1) \Sigma^{-1}}{(1 + \varepsilon w_1)(1 - \varepsilon w_1)} - \frac{(1 + \varepsilon w_1) \varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1}}{(1 + \varepsilon w_1)(1 - \varepsilon w_1) [1 - \varepsilon w_1 (1 - \mathbf{z}' \Sigma^{-1} \mathbf{z})]} \\
&\cong (1 + \varepsilon w_1) \Sigma^{-1} - \frac{\varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1}}{[1 - \varepsilon w_1 (1 - \mathbf{z}' \Sigma^{-1} \mathbf{z})]}.
\end{aligned}$$

Do mesmo modo, multiplica-se e divide-se o segundo termo por $[1 + \varepsilon w_1 (1 - \mathbf{z}' \Sigma^{-1} \mathbf{z})]$ e assim, segue que

$$\Sigma^{-1*} \cong (1 + \varepsilon w_1) \Sigma^{-1} - \varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1} \quad . \quad (2.5)$$

São apresentados a seguir os cálculos da função de influência para os parâmetros de interesse da análise discriminante.

a) **Função de influência para a distância de Mahalanobis** $\Delta^2 = \delta' \Sigma^{-1} \delta$

Usando (2.2) e (2.5) tem-se:

$$\Delta^{2*} = (\delta + \varepsilon \mathbf{z})' [(1 + \varepsilon w_1) \Sigma^{-1} - \varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1}] (\delta + \varepsilon \mathbf{z})$$

e efetuadas as multiplicações, esta expressão reduz-se a

$$\begin{aligned} \Delta^{2*} = & \delta' \Sigma^{-1} \delta + \delta' \varepsilon w_1 \Sigma^{-1} \delta + \varepsilon \mathbf{z}' \Sigma^{-1} \delta + \varepsilon \mathbf{z}' \varepsilon w_1 \Sigma^{-1} \delta + \delta' \Sigma^{-1} \varepsilon \mathbf{z} + \delta' \varepsilon w_1 \Sigma^{-1} \varepsilon \mathbf{z} + \\ & + \varepsilon \mathbf{z}' \Sigma^{-1} \varepsilon \mathbf{z} + \varepsilon \mathbf{z}' \varepsilon w_1 \Sigma^{-1} \varepsilon \mathbf{z} - \delta' \varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1} \delta - \varepsilon \mathbf{z}' \varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1} \delta - \\ & - \delta' \varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1} \varepsilon \mathbf{z} - \varepsilon \mathbf{z}' \varepsilon w_1 \Sigma^{-1} \mathbf{z} \mathbf{z}' \Sigma^{-1} \varepsilon \mathbf{z}. \end{aligned}$$

$$\text{Tomando } \phi = \delta' \Sigma^{-1} \mathbf{z}, \quad (2.6)$$

$$\begin{aligned} \Delta^{2*} = & \delta' \Sigma^{-1} \delta (1 + \varepsilon w_1) + \varepsilon \phi' + \varepsilon^2 w_1 \phi' + \varepsilon \phi + \varepsilon^2 w_1 \phi + \varepsilon^2 \mathbf{z}' \Sigma^{-1} \mathbf{z} + \varepsilon^3 w_1 \mathbf{z}' \Sigma^{-1} \mathbf{z} - \\ & - \varepsilon w_1 \phi \phi' - \varepsilon^2 w_1 \mathbf{z}' \Sigma^{-1} \mathbf{z} \phi' - \varepsilon^2 w_1 \phi \mathbf{z}' \Sigma^{-1} \mathbf{z} - \varepsilon^3 w_1 (\mathbf{z}' \Sigma^{-1} \mathbf{z})^2. \end{aligned}$$

Mas $\phi = \phi'$ porque são quantidades unidimensionais, e mantendo apenas os termos de ordem ε ,

$$\Delta^{2*} \cong (1 + \varepsilon w_1) \Delta^2 + 2\varepsilon \phi - \varepsilon w_1 \phi^2. \quad (2.7)$$

Calcula-se agora a função de influência :

$$I(\mathbf{x}; \Delta^2) = \lim_{\varepsilon \rightarrow 0} \frac{(1 + \varepsilon w_1) \Delta^2 + 2\varepsilon \phi - \varepsilon w_1 \phi^2 - \Delta^2}{\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon w_1 \Delta^2 + 2\varepsilon \phi - \varepsilon w_1 \phi^2}{\varepsilon},$$

$$I(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + 2\phi - w_1 \phi^2. \quad (2.8)$$

Sendo $\mathbf{l} = \Sigma^{-1}\boldsymbol{\delta}$ o vetor de coeficientes discriminantes, observa-se que ϕ pode ser escrito como $\mathbf{l}'\mathbf{z}$ que é o mesmo que a diferença entre o escore discriminante $\boldsymbol{\delta}'\Sigma^{-1}\mathbf{x}$ e sua média discriminante para a primeira população, $\mathbf{l}'\boldsymbol{\mu}_1$:

$$\phi = \mathbf{l}'\mathbf{z} = \boldsymbol{\delta}'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) = \boldsymbol{\delta}'\Sigma^{-1}\mathbf{x} - \boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\mu}_1 = \boldsymbol{\delta}'\Sigma^{-1}\mathbf{x} - \mathbf{l}'\boldsymbol{\mu}_1.$$

Calculando a esperança e variância de ϕ , para \mathbf{x} pertencendo a população π_1 que foi perturbada, tem-se:

$$E(\phi) = E(\boldsymbol{\delta}'\Sigma^{-1}\mathbf{z}) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} E(\mathbf{X} - \boldsymbol{\mu}_1) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_1) = 0$$

$$\begin{aligned} Var(\phi) &= Var\left[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}_1)\right] = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} Var(\mathbf{X} - \boldsymbol{\mu}_1) \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \\ &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} \Sigma \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \Delta^2. \end{aligned}$$

Uma vez que $\phi \sim N(0, \Delta^2)$, então a variável aleatória padronizada $\phi_p = \frac{\phi}{\Delta}$ tem distribuição normal padrão, e tomando $\phi = \phi_p \Delta$, a função de influência para Δ^2 pode ser escrita em termos de ϕ_p como:

$$I_p(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + 2\Delta \phi_p - w_1 \Delta^2 \phi_p^2. \quad (2.9)$$

Com base nisso, uma conclusão importante de Campbell (1978) é que a função de influência para Δ^2 é uma função quadrática do desvio do escore discriminante com relação à média discriminante do grupo correspondente.

b) Função de influência para as médias discriminantes $I' \mu_j$

Foi calculada a função de influência para a média discriminante do grupo 1 uma vez que foi esse o grupo perturbado, no entanto, valeria um cálculo análogo caso a perturbação fosse feita no grupo 2.

$$(I' \mu_1)' = (\delta' \Sigma^{-1} \mu_1)'$$

Usando (2.2), (2.5) e (2.1):

$$\begin{aligned} (I' \mu_1)' &= (\delta + \varepsilon z)' [(1 + \varepsilon w_1) \Sigma^{-1} - \varepsilon w_1 \Sigma^{-1} z z' \Sigma^{-1}] (\mu_1 + \varepsilon z) \\ &= [\delta' (1 + \varepsilon w_1) \Sigma^{-1} - \delta' \varepsilon w_1 \Sigma^{-1} z z' \Sigma^{-1} + \varepsilon z' (1 + \varepsilon w_1) \Sigma^{-1} - \\ &\quad - \varepsilon z' \varepsilon w_1 \Sigma^{-1} z z' \Sigma^{-1}] (\mu_1 + \varepsilon z) \end{aligned}$$

e considerando apenas os termos de ordem ε , segue que

$$(I' \mu_1)' \cong \delta' (1 + \varepsilon w_1) \Sigma^{-1} \mu_1 - \delta' \varepsilon w_1 \Sigma^{-1} z z' \Sigma^{-1} \mu_1 + \varepsilon z' \Sigma^{-1} \mu_1 + \delta' \Sigma^{-1} \varepsilon z.$$

$$\text{Finalmente, devido a (2.6) e tomando } \eta_1 = z' \Sigma^{-1} \mu_1 \quad (2.10)$$

$$\text{então } (I' \mu_1)' \cong (1 + \varepsilon w_1) I' \mu_1 - \varepsilon w_1 \phi \eta_1 + \varepsilon \eta_1 + \varepsilon \phi.$$

Calcula-se agora a função de influência:

$$\begin{aligned} I(x; I' \mu_1) &= \lim_{\varepsilon \rightarrow 0} \frac{(1 + \varepsilon w_1) I' \mu_1 + \varepsilon \phi - \varepsilon w_1 \phi \eta_1 + \varepsilon \eta_1 - I' \mu_1}{\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\varepsilon (w_1 I' \mu_1 + \phi - w_1 \phi \eta_1 + \eta_1)}{\varepsilon} \\ &= w_1 I' \mu_1 + \phi - w_1 \phi \eta_1 + \eta_1. \end{aligned}$$

c) Função de influência para os coeficientes da função discriminante $l = \Sigma^{-1}\delta$

Substituindo Σ^{-1} e δ pelas as expressões (2.5) e (2.2):

$$\begin{aligned} \Gamma^* &= (\Sigma^{-1}\delta)^* = [(1 + \varepsilon w_1)\Sigma^{-1} - \varepsilon w_1 \Sigma^{-1} \mathbf{z}\mathbf{z}' \Sigma^{-1}](\delta + \varepsilon \mathbf{z}) \\ &= (1 + \varepsilon w_1)\Sigma^{-1}\delta - \varepsilon w_1 \Sigma^{-1} \mathbf{z}\mathbf{z}' \Sigma^{-1}\delta + (1 + \varepsilon w_1)\Sigma^{-1}\varepsilon \mathbf{z} - \varepsilon^2 w_1 \Sigma^{-1} \mathbf{z}\mathbf{z}' \Sigma^{-1}\mathbf{z}. \end{aligned}$$

Usando (2.6) e eliminando o termo em ε^2 segue que

$$\Gamma^* \cong (1 + \varepsilon w_1)\mathbf{1} - \varepsilon w_1 \Sigma^{-1} \mathbf{z}\phi' + \Sigma^{-1}\varepsilon \mathbf{z},$$

Lembrando que $\phi = \phi'$, a função de influência é calculada como:

$$\begin{aligned} I(\mathbf{x}; \mathbf{l}) &= \lim_{\varepsilon \rightarrow 0} \frac{(1 + \varepsilon w_1)\mathbf{1} - \varepsilon w_1 \Sigma^{-1} \mathbf{z}\phi + \Sigma^{-1}\varepsilon \mathbf{z} - \mathbf{l}}{\varepsilon} \\ &= w_1 \mathbf{l} - w_1 \Sigma^{-1} \mathbf{z}\phi + \Sigma^{-1} \mathbf{z} \\ &= w_1 \mathbf{l} + (1 - w_1 \phi) \Sigma^{-1} \mathbf{z}. \end{aligned} \tag{2.11}$$

A quantidade $I(\mathbf{x}; \mathbf{l})$ é um vetor. A fim de se obter um valor numérico como nas funções de influência anteriores, basta calcular a função de influência para $\mathbf{l} = \mathbf{1}$, dada por

$$\begin{aligned} (\delta' \Sigma^{-1})^* (\Sigma^{-1}\delta)^* &= (\delta + \varepsilon \mathbf{z})' [(1 + \varepsilon w_1)\Sigma^{-1} - \varepsilon w_1 \Sigma^{-1} \mathbf{z}\mathbf{z}' \Sigma^{-1}]^2 (\delta + \varepsilon \mathbf{z}) \\ &= (\delta' + \varepsilon \mathbf{z}') [(1 + \varepsilon w_1)^2 \Sigma^{-1} \Sigma^{-1} - 2(1 + \varepsilon w_1)\Sigma^{-1} \varepsilon w_1 \Sigma^{-1} \mathbf{z}\mathbf{z}' \Sigma^{-1} + \varepsilon^2 w_1^2 \Sigma^{-1} \mathbf{z}\mathbf{z}' \Sigma^{-1} \Sigma^{-1} \mathbf{z}\mathbf{z}' \Sigma^{-1}]. \\ & \quad (\delta + \varepsilon \mathbf{z}). \end{aligned}$$

Desprezando-se os termos em ε^2 obtém-se :

$$\begin{aligned}
 (\mathbf{l}'\mathbf{l})^* &\cong (\boldsymbol{\delta}' + \varepsilon \mathbf{z}') \left[(1 + 2\varepsilon w_1) \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} - 2\varepsilon w_1 \boldsymbol{\Sigma}^{-1} \mathbf{z} \mathbf{z}' \boldsymbol{\Sigma}^{-1} \right] (\boldsymbol{\delta} + \varepsilon \mathbf{z}) \\
 &= \boldsymbol{\delta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} + \boldsymbol{\delta}' 2\varepsilon w_1 \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} - \boldsymbol{\delta}' 2\varepsilon w_1 \boldsymbol{\Sigma}^{-1} \mathbf{z} \mathbf{z}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} + \varepsilon \mathbf{z}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} + \boldsymbol{\delta}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \varepsilon \mathbf{z} \\
 &= \mathbf{l}'\mathbf{l} + 2\varepsilon w_1 \mathbf{l}'\mathbf{l} - 2\varepsilon w_1 \mathbf{l}' \boldsymbol{\Sigma}^{-1} \mathbf{z} \phi + 2\mathbf{l}' \boldsymbol{\Sigma}^{-1} \varepsilon \mathbf{z} .
 \end{aligned}$$

Finalizando,

$$\begin{aligned}
 I(\mathbf{x}; \mathbf{l}'\mathbf{l}) &= \lim_{\varepsilon \rightarrow 0} \frac{\mathbf{l}'\mathbf{l} + 2\varepsilon w_1 \mathbf{l}'\mathbf{l} - 2\varepsilon w_1 \mathbf{l}' \boldsymbol{\Sigma}^{-1} \mathbf{z} \phi + 2\mathbf{l}' \boldsymbol{\Sigma}^{-1} \varepsilon \mathbf{z} - \mathbf{l}'\mathbf{l}}{\varepsilon} \\
 &= 2w_1 \mathbf{l}'\mathbf{l} + 2(1 - w_1 \phi) \mathbf{l}' \boldsymbol{\Sigma}^{-1} \mathbf{z} . \tag{2.12}
 \end{aligned}$$

Nas próximas seções será vista a utilização de algumas dessas medidas.

2.2.2 Distribuição aproximada para a função de influência $I(\mathbf{x}; \Delta^2)$

A função de influência pode ser considerada como uma variável aleatória, uma vez que ela é uma transformação do vetor aleatório \mathbf{X} .

Da expressão (2.9), sabe-se que a função de influência para a distância de Mahalanobis (Δ^2) é:

$$I(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + 2\Delta \phi_p - w_1 \Delta^2 \phi_p^2 ,$$

$$\text{com } \phi_p = \frac{\phi}{\Delta} , \quad \phi = \boldsymbol{\delta}' \boldsymbol{\Sigma}^{-1} \mathbf{z} \quad \text{e} \quad \mathbf{z} = \mathbf{x} - \boldsymbol{\mu}_1 .$$

Se \mathbf{X} tem distribuição normal multivariada, está demonstrado na Seção B-4 do Apêndice B que a distribuição de probabilidades da função de influência para a distância de Mahalanobis (Δ^2) tem assimetria negativa.

Tomando a função de influência para a distância de Mahalanobis com ϕ não padronizado, $I(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + 2\phi - w_1 \phi^2$ e derivando em relação a ϕ obtém-se o ponto de máximo para $\phi = \frac{1}{w_1}$, assim, $I_{max}(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + \frac{1}{w_1}$.

Campbell (1978) sugere o uso de $I_m(\mathbf{x}; \Delta^2) = I_{max}(\mathbf{x}; \Delta^2) - I(\mathbf{x}; \Delta^2)$, dada por

$$\begin{aligned} I_m(\mathbf{x}; \Delta^2) &= w_1 \Delta^2 + w_1^{-1} - (w_1 \Delta^2 + 2\Delta\phi_p - w_1 \Delta^2 \phi_p^2) & (2.13) \\ &= w_1 \Delta^2 + w_1^{-1} - w_1 \Delta^2 - 2\Delta\phi_p + w_1 \Delta^2 \phi_p^2 \\ &= w_1^{-1} (1 - 2w_1 \Delta\phi_p + w_1^2 \Delta^2 \phi_p^2) . \end{aligned}$$

Esta variável aleatória é sempre positiva e, como $I(\mathbf{x}; \Delta^2)$ tem assimetria negativa, segue que $I_m(\mathbf{x}; \Delta^2)$ tem distribuição positivamente assimétrica.

Multiplicando e dividindo $I_m(\mathbf{x}; \Delta^2)$ por ϕ_p^2 obtém-se

$$\begin{aligned} I_m(\mathbf{x}; \Delta^2) &= \frac{w_1^{-1}}{\phi_p^2} (\phi_p^2 - 2w_1 \Delta\phi_p^3 + w_1^2 \Delta^2 \phi_p^4) \\ &= \frac{w_1^{-1}}{\phi_p^2} (\phi_p - w_1 \Delta\phi_p^2)^2 . \end{aligned}$$

Como no ponto máximo $\phi_p = (w_1 \Delta)^{-1}$,

$$\begin{aligned} I_m(\mathbf{x}; \Delta^2) &= \frac{w_1^{-1}}{w_1^{-2} \Delta^{-2}} (\phi_p - w_1 \Delta (w_1 \Delta)^{-2})^2 \\ &= w_1 \Delta^2 (\phi_p - w_1^{-1} \Delta^{-1})^2. \end{aligned}$$

Mas, como $\phi_p \sim N(0,1)$ então $(\phi_p - w_1^{-1} \Delta^{-1}) \sim N_\phi(-w_1^{-1} \Delta^{-1}, \mathbf{1})$

e usando o Resultado B-2 do Apêndice B tem-se que

$(\phi_p - w_1^{-1} \Delta^{-1})^2 = (\phi_p - w_1^{-1} \Delta^{-1}) \mathbf{1}^{-1} (\phi_p - w_1^{-1} \Delta^{-1})$ tem distribuição quiquadrado não central com 1 grau de liberdade e parâmetro de não centralidade $(-w_1^{-1} \Delta^{-1}) \mathbf{1}^{-1} (-w_1^{-1} \Delta^{-1}) = (w_1^{-1} \Delta^{-1})^2$.

Portanto $\frac{I_m(\mathbf{x}; \Delta^2)}{w_1 \Delta^2}$ tem uma distribuição quiquadrado não central com 1 grau

de liberdade e parâmetro de não centralidade $(w_1^{-2} \Delta^2)^{-1}$.

Observa-se que quiquadrado com um grau de liberdade é a distribuição nula da medida, sendo assim a distribuição esperada na inexistência de pontos discrepantes. A utilização desse fato será discutida na próxima seção.

2.2.3 Gráficos de probabilidade para detectar *outliers*

Campbell (1978) usou a função de influência amostral para confirmar a aproximação quiquadrado por ele sugerida através de dados simulados de uma distribuição normal bivariada com variâncias unitárias e correlação 0,9. As médias

das populações foram $(0, 0)$ para o primeiro grupo e $(0, 2)$ para o segundo grupo. Foram geradas 50 observações para cada grupo.

Uma vez que a distribuição quiquadrado é um caso particular da distribuição gama, o autor usou um gráfico do tipo “QQ” dos valores da função $I_m(x; D^2)$ contra os quantis da distribuição gama, em que $D^2 = (\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2)$ é a estimativa da distância de Mahalanobis Δ^2 , com os parâmetros desconhecidos estimados por máxima verossimilhança. Tal distribuição produziu também uma boa aproximação para $I_m(x; D^2)$.

Depois dessa verificação, o autor passou para a análise dos dados de um exemplo de Campbell e Mahon (1974) citado em Campbell (1978), no qual foram examinadas as diferenças morfológicas entre duas espécies de caranguejos. Foram medidas 5 características em 100 animais de cada espécie (denominadas azul e laranja). Para esses dados, as matrizes de covariâncias para as duas espécies eram muito similares e $D^2 = 27,8$.

O objetivo dessa análise foi o de mostrar como a função de influência e as estatísticas relacionadas podem ser utilizadas para determinar observações discrepantes.

Determinou-se inicialmente o efeito da retirada da i -ésima observação na medida D^2 . Tal efeito foi quantificado através de $D^2 - D_{-i}^2$, sendo que D_{-i}^2 é a distância de Mahalanobis amostral calculada sem o i -ésimo ponto. Esta quantidade foi analisada em função do escore discriminante padronizado $c_p'(\mathbf{x}_i - \bar{\mathbf{x}}_j)$, sendo que $c_p = \frac{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' S^{-1}}{D}$ é o vetor padronizado de coeficientes discriminantes amostrais, \mathbf{x}_i é a i -ésima observação (a observação que vai ser eliminada), $\bar{\mathbf{x}}_j$ é o vetor de médias do j -ésimo grupo ($j = 1, 2$), os pesos w_j são dados por $\frac{n_j}{n_1 + n_2}$ e Δ^2 é substituído por D^2 .

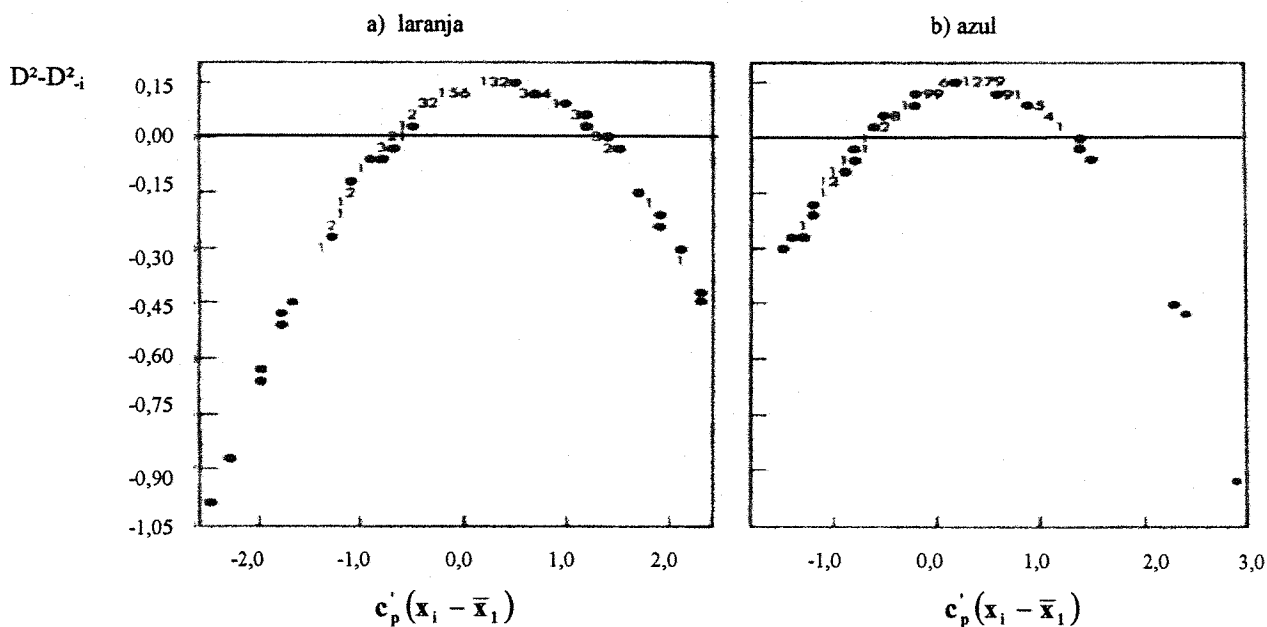


Figura 2.1: Gráfico de $D^2 - D_{-i}^2$ contra $c_p'(x_i - \bar{x}_1)$. O símbolo • representa um indivíduo e os números indicam a quantidade de sobreposições, sendo que o número 9 representa pelo menos 9 sobreposições.

Fonte: Figura 1 - Campbell (1978).

A Figura 2.1 mostra o gráfico de $D^2 - D_{-i}^2$ contra $c_p'(x_i - \bar{x}_1)$ exibindo uma tendência quadrática. Observa-se que, para ambos os grupos, o ponto de máximo ocorre em torno de $c_p'(x_i - \bar{x}_1) = 0,04$.

Os valores de $D^2 - D_{-i}^2$ são baixos para as observações com escores discriminantes entre $-0,5$ e $1,5$, enquanto que essa distância aumenta quando se exclui uma observação com escore discriminante fora desse intervalo. O maior aumento dessa medida (em valor absoluto) ocorre com os escores mais distantes dos obtidos para a média da outra espécie ($-2,4$ para a espécie laranja e $3,0$ para a azul).

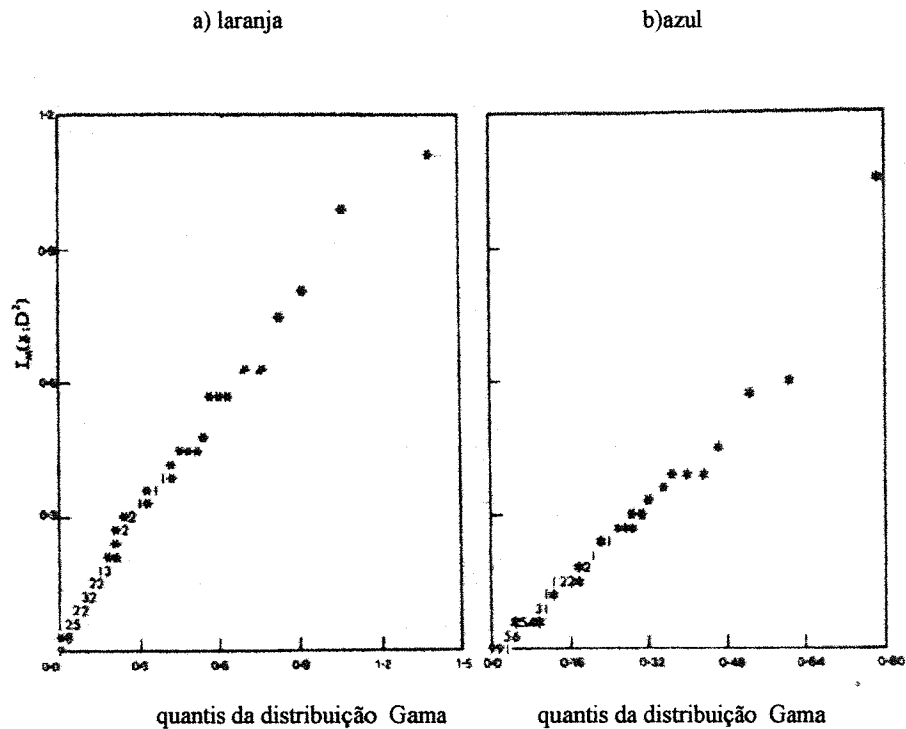


Figura 2.2: Gráfico QQ-Gama para $I_m(\mathbf{x}; D^2)$.

Fonte: Figura 2 - Campbell (1978)

Na Figura 2.2, os gráficos do tipo “QQ-Gama” de $I_m(\mathbf{x}; D^2)$ com parâmetros desconhecidos estimados por máxima verossimilhança apresentam aspecto linear. A inclinação dos pontos para o gráfico da espécie laranja está próxima da reta $y = x$, bem como a dos pontos da espécie azul, se eliminado o *outlier*.

Nenhum dos dois gráficos indicam observações atípicas para a espécie laranja. Já para a espécie azul nota-se que três observações devem ser examinadas, sendo que o gráfico da Figura 2.2.b mostra que uma delas é claramente atípica. Os autores observaram que essas três observações tinham em comum largura de carapaça maior do que o esperado, em relação ao lábio frontal do animal, comparando-se com as demais observações. Notaram também que observações para a espécie laranja, com escore discriminante semelhante a dois dos três *outliers* da espécie azul, têm influência mínima em D^2 .

O conjunto de dados utilizado já havia sido analisado através de técnicas para detecção de *outliers* multivariados e nenhum *outlier* sério tinha sido encontrado. Do mesmo modo, eliminando-se o mais óbvio dos possíveis *ouliers*, a estatística D^2 aumenta somente em 0,9 (de 27,8 para 28,7), o que é mínimo para este caso.

Por isso, deve-se examinar com cuidado a influência absoluta das observações suspeitas, pois mesmo que alguns pontos pareçam ter grande influência, eles podem não afetar significativamente as estatísticas de interesse.

O mais importante fato exibido nos gráficos é o modo assimétrico no qual as observações influenciam D^2 , e que, surpreendentemente, a inclusão de uma observação com medida mais distante da média do outro grupo, pode diminuir a estatística D^2 ao invés de aumentá-la. Este último fato pode ser justificado pelo aumento nas variâncias e mudança no valor das correlações ocasionados pela inclusão dessa observação.

É importante lembrar que os *outliers* podem “atrair” a média amostral em sua direção, bem como aumentar as variâncias. Sendo assim, os métodos clássicos de identificação, que são baseados nessas medidas, podem não identificar os *outliers*, ocorrendo o chamado “*masking effect*”. Rousseeuw e van Zomeren (1990) sugerem medidas baseadas em estimadores robustos de locação e covariância para detectar esses pontos discrepantes em dados multivariados.

Do estudo feito, Campbell (1978) conclui que o exame apenas do escore discriminante pode não ser adequado, uma vez que as suposições de normalidade em que são baseados podem não valer. Além disso, *outliers* óbvios podem indicar medidas incorretas ou alocação indevida no grupo.

2.3 Principais estatísticas no diagnóstico em análise discriminante

Nesta seção, são apresentadas duas estatísticas propostas por Fung (1992), fundamentais no diagnóstico em análise discriminante, sendo que várias medidas de influência dependem delas.

Na análise discriminante, o maior interesse é no grau de separação entre os grupos. Da Seção 1.1 do Apêndice A, sabe-se que a maior separação entre duas populações (usando combinações lineares das observações) é determinada para a combinação linear $y = \mathbf{l}'\mathbf{x}$ e o máximo valor do quociente $\frac{(\mu_{1y} - \mu_{2y})^2}{\sigma_y^2}$ coincide com a distância de Mahalanobis (Δ^2).

A partir da regra discriminante de Fisher (A-1.8), uma observação \mathbf{x} é alocada na população π_1 se

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \mathbf{x} \geq \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \bar{\mathbf{x}}_1 + \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \bar{\mathbf{x}}_2 \Leftrightarrow$$

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \mathbf{x} \geq (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \bar{\mathbf{x}}_1 - \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \bar{\mathbf{x}}_1 + \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \bar{\mathbf{x}}_2 \Leftrightarrow$$

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \mathbf{x} \geq (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \bar{\mathbf{x}}_1 - \frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \Leftrightarrow$$

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) \geq -\frac{1}{2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \Leftrightarrow$$

$$\hat{\mathbf{l}}' (\mathbf{x} - \bar{\mathbf{x}}_1) \geq -\frac{1}{2} D^2 \tag{2.14}$$

em que $\hat{\mathbf{l}} = \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ é o vetor de coeficientes discriminantes amostral.

Deste modo, usando (2.6), a regra fica:

alocar \mathbf{x}_0 para π_1 se $\hat{\phi} \geq -\frac{1}{2}D^2$ e alocar \mathbf{x}_0 para π_2 caso contrário.

Tomando a regra discriminante de Fisher (A-1.9) e assumindo os custos de classificação incorreta iguais, aloca-se \mathbf{x}_0 para π_1 se :

$$\hat{I}'\left[\mathbf{x}_0 - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2}\right] > \ln\left(\frac{1-q}{q}\right).$$

A quantidade $\hat{I}'\left[\mathbf{x}_0 - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2}\right] - \ln\left(\frac{1-q}{q}\right)$ é denominada logaritmo

da razão das probabilidades estimadas (*log-odds*). Em particular, na maioria dos casos, admite-se $1 - q = q$, então essa quantidade reduz-se à função discriminante para a observação \mathbf{x}_0 que, no caso da regra (2.14) descrita anteriormente, fica

$$\hat{\phi} + \frac{1}{2}D^2. \quad (2.15)$$

De (2.8) pode-se obter a função de influência amostral trocando-se Δ^2 e ϕ pelos seus estimadores D^2 e $\hat{\phi} = \hat{I}'(\mathbf{x}_0 - \bar{\mathbf{x}}_1)$ respectivamente:

$$\hat{I}(\mathbf{x}; \Delta^2) = w_1 D^2 + 2\hat{\phi} - w_1 \hat{\phi}^2. \quad (2.16)$$

A medida (2.13) sugerida por Campbell (1978) também pode ser escrita substituindo-se ϕ_p por $\frac{\phi}{\Delta}$, como

$$\begin{aligned} I_m(\mathbf{x}; \Delta^2) &= w_1 \Delta^2 + w_1^{-1} - \left(w_1 \Delta^2 + 2\Delta \frac{\phi}{\Delta} - w_1 \Delta^2 \frac{\phi^2}{\Delta^2} \right) \\ &= w_1^{-1} - 2\phi + w_1 \phi^2 = w_1 (\phi^2 - 2\phi w_1^{-1} + w_1^{-2}), \end{aligned}$$

de modo que

$$\hat{l}_m(\mathbf{x}; \Delta^2) = w_1 (\hat{\phi} - w_1^{-1})^2. \quad (2.17)$$

Observando-se (2.14) a (2.17), nota-se que $\hat{\phi}$ é importante não só na discriminação como também na análise de influência.

Healy (1968) utiliza a estatística

$$d_{ji}^2 = (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)' \mathbf{S}^{-1} (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j) \quad i = 1, \dots, n_j, \quad j = 1, 2$$

em testes de normalidade e também com o objetivo de detectar *outliers* sob normalidade.

Assim como o resíduo e o ponto de alavanca em regressão, Fung (1992) aponta d_{ji}^2 e $\hat{\phi}_{ji} = \hat{l}'(\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)$ como estatísticas básicas para detectar *outliers* e observações influentes em análise discriminante, uma vez que várias medidas de diagnóstico são escritas em função dessas duas estatísticas.

Para facilitar a notação serão usados os termos d_i^2 e ϕ_i nas próximas seções.

2.3.1 Distribuições assintóticas

Na Seção 2.2.1. verificou-se que $\frac{\phi_i}{\Delta} = \frac{\boldsymbol{\delta}' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_j)}{\Delta}$ tem distribuição normal padrão.

Fung (1995-a) usou o teorema a seguir para mostrar que, assintoticamente, d_i^2 é distribuída como χ_p^2 ($p =$ dimensão do vetor \mathbf{X}) e $\frac{\hat{\phi}_i}{D}$ tem distribuição $N(0, 1)$.

Assim, a partir desse teorema, obtém-se as distribuições assintóticas das medidas propostas e também de muitas outras medidas que são funções de d_i^2 e $\hat{\phi}_i$.

Na demonstração, serão utilizados os Resultados 2.1 e 2.2.

Resultado 2.1 (Graybill (1976)):

Sejam o vetor aleatório \mathbf{X} com distribuição normal de média $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, em que $\boldsymbol{\Sigma}$ tem posto p ($r(\boldsymbol{\Sigma}) = p$) e \mathbf{A} uma matriz quadrada de posto completo p .

Nessas condições $\mathbf{X}'\mathbf{A}\mathbf{X}$ tem distribuição quiquadrado com $r(\mathbf{A})$ graus de liberdade e parâmetro de não centralidade $\frac{1}{2}\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$ se e somente se $\mathbf{A}\boldsymbol{\Sigma}$ é idempotente.

Além disso, $tr(\mathbf{A}) = r(\mathbf{A})$ quando \mathbf{A} é idempotente.

Resultado 2.2 Independência entre forma quadrática e linear (Graybill (1976)):

Sejam o vetor aleatório \mathbf{X} com distribuição normal de média $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, em que $\boldsymbol{\Sigma}$ tem posto p , e as matrizes \mathbf{A} , quadrada de posto completo p e \mathbf{B} uma matriz $q \times p$.

Nessas condições, a forma quadrática $\mathbf{X}'\mathbf{A}\mathbf{X}$ é independente da forma linear $\mathbf{B}\mathbf{X}$, se e somente se $\mathbf{B}\Sigma\mathbf{A} = \mathbf{0}$.

Teorema: As estatísticas $DIF = d_i^2 - \left(\frac{\hat{\phi}_i}{D}\right)^2$ e $\frac{\hat{\phi}_i}{D} = \frac{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1}(\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)}{D}$ são assintoticamente independentes e têm distribuição quiquadrado com $p-1$ graus de liberdade e normal padrão.

Prova: Sabe-se que as estatísticas $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{S}$ e D^2 convergem quase certamente para $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \Sigma$ e Δ^2 respectivamente.

Por essa razão,

$$\frac{\hat{\phi}_i}{D} \stackrel{d}{=} \frac{\phi_i}{\Delta} \quad (\text{a distribuição assintótica de } \frac{\hat{\phi}_i}{D} \text{ é a mesma de } \frac{\phi_i}{\Delta}) \quad e$$

$$d_i^2 \stackrel{d}{=} (\mathbf{x}_{ji} - \boldsymbol{\mu}_j)' \Sigma^{-1} (\mathbf{x}_{ji} - \boldsymbol{\mu}_j) .$$

Como Σ é de posto completo, então pode ser decomposta como $\Sigma = \Gamma\Gamma'$ onde Γ é a matriz triangular inferior não singular. Sejam

$$\lambda = \frac{\Gamma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\Delta}, \quad \varphi_i = \Gamma^{-1}(\mathbf{x}_{ji} - \boldsymbol{\mu}_j) \quad e \quad t_i^2 = (\mathbf{x}_{ji} - \boldsymbol{\mu}_j)' \Sigma^{-1} (\mathbf{x}_{ji} - \boldsymbol{\mu}_j) .$$

Como consequência, $\varphi_i \sim N(\mathbf{0}, \mathbf{I})$.

Desse modo,

$$t_i^2 = \phi_i' \phi_i, \text{ pois } t_i^2 = (\mathbf{x}_{ji} - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{ji} - \boldsymbol{\mu}_j) = (\mathbf{x}_{ji} - \boldsymbol{\mu}_j)' (\boldsymbol{\Gamma}')^{-1} \boldsymbol{\Gamma}^{-1} (\mathbf{x}_{ji} - \boldsymbol{\mu}_j) = \phi_i' \phi_i$$

$$\text{e } \frac{\phi_i}{\Delta} = \frac{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_{ji} - \boldsymbol{\mu}_j)}{\Delta} = \boldsymbol{\lambda}' \phi_i.$$

O termo $t_i^2 - \frac{\phi_i^2}{\Delta^2}$ é escrito como $\phi_i' \mathbf{A} \phi_i$ com $\mathbf{A} = \mathbf{I} - \boldsymbol{\lambda} \boldsymbol{\lambda}'$.

A matriz \mathbf{A} é idempotente e seu traço é $p-1$ já que

$$\text{a) } (\mathbf{I} - \boldsymbol{\lambda} \boldsymbol{\lambda}') \cdot (\mathbf{I} - \boldsymbol{\lambda} \boldsymbol{\lambda}') = \mathbf{I} - \mathbf{I} \boldsymbol{\lambda} \boldsymbol{\lambda}' - \boldsymbol{\lambda} \boldsymbol{\lambda}' \mathbf{I} + (\boldsymbol{\lambda} \boldsymbol{\lambda}') \cdot (\boldsymbol{\lambda} \boldsymbol{\lambda}'),$$

$$\text{mas } \boldsymbol{\lambda}' \boldsymbol{\lambda} = \frac{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' (\boldsymbol{\Gamma}^{-1})' \boldsymbol{\Gamma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\Delta^2} = \frac{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}{\Delta^2} = \frac{\Delta^2}{\Delta^2} = 1$$

$$\text{então, } (\mathbf{I} - \boldsymbol{\lambda} \boldsymbol{\lambda}') \cdot (\mathbf{I} - \boldsymbol{\lambda} \boldsymbol{\lambda}') = \mathbf{I} - \boldsymbol{\lambda} \boldsymbol{\lambda}' - \boldsymbol{\lambda} \boldsymbol{\lambda}' + \boldsymbol{\lambda} \boldsymbol{\lambda}' = \mathbf{I} - \boldsymbol{\lambda} \boldsymbol{\lambda}', \text{ e}$$

$$\text{b) } \text{tr}(\mathbf{I} - \boldsymbol{\lambda} \boldsymbol{\lambda}') = \text{tr}(\mathbf{I}) - \text{tr}(\boldsymbol{\lambda} \boldsymbol{\lambda}').$$

$$\text{Como } \text{tr}(\boldsymbol{\lambda} \boldsymbol{\lambda}') = \text{tr}(\boldsymbol{\lambda}' \boldsymbol{\lambda}), \text{ então } \text{tr}(\mathbf{I}) - \text{tr}(\boldsymbol{\lambda}' \boldsymbol{\lambda}) = p - 1.$$

Desta forma, lembrando que $t_i^2 - \frac{\phi_i^2}{\Delta^2} = \phi_i' \mathbf{A} \phi_i$, $\phi_i \sim N(\mathbf{0}, \mathbf{I})$ e visto

que $\mathbf{A} \mathbf{I} = \mathbf{A}$, que é idempotente, então do Resultado 2.1, tem-se que

$$t_i^2 - \frac{\phi_i^2}{\Delta^2} \sim \chi_{p-1}^2.$$

Do Resultado 2.2 , para provar a independência entre $DIF = d_i^2 - \left(\frac{\hat{\phi}_i}{D}\right)^2$ e

$$\frac{\hat{\phi}_i}{D} = \frac{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_j)}{D}, \text{ é preciso mostrar que } \boldsymbol{\lambda}' \mathbf{I} \mathbf{A} = \mathbf{0}.$$

Realmente, $\boldsymbol{\lambda}' \mathbf{I} \mathbf{A} = \boldsymbol{\lambda}' \mathbf{I}(\mathbf{I} - \boldsymbol{\lambda} \boldsymbol{\lambda}') = \boldsymbol{\lambda}' - \boldsymbol{\lambda}' \boldsymbol{\lambda} \boldsymbol{\lambda}' = \boldsymbol{\lambda}' - \boldsymbol{\lambda}' = \mathbf{0}$, então,

$\frac{\hat{\phi}_i}{\Delta}$ e $t_i^2 - \frac{\phi_i^2}{\Delta^2}$ são variáveis aleatórias independentes com distribuição normal

padrão e quiquadrado com $p-1$ graus de liberdade, respectivamente.

Dessa forma, o teorema está provado porque

$$\frac{\hat{\phi}_i}{D} \stackrel{d}{=} \frac{\phi_i}{\Delta}, \quad d_i^2 \stackrel{d}{=} t_i^2 \quad \text{e} \quad d_i^2 - \frac{\hat{\phi}_i^2}{D^2} \stackrel{d}{=} t_i^2 - \frac{\phi_i^2}{\Delta^2}. \quad \square$$

Note-se também que d_i^2 é assintoticamente distribuído como χ_p^2 ,

pois usando o teorema sabe-se que

$$d_i^2 \stackrel{d}{=} u + v^2 \quad \text{e} \quad \frac{\hat{\phi}_i}{D} \stackrel{d}{=} v$$

sendo que $u = DIF$ e v são independentemente distribuídas como χ_{p-1}^2 e $N(\mathbf{0}, 1)$ respectivamente.

Esses resultados assintóticos são muito importantes uma vez que calculando-

se $DIF = d_i^2 - \left(\frac{\hat{\phi}_i}{D}\right)^2$ e $\frac{\hat{\phi}_i}{D} = \frac{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}_j)}{D}$ para os elementos amostrais,

valores maiores que os respectivos valores críticos da distribuição quiquadrado e da distribuição normal são indicadores de pontos influentes.

2.4 Medidas relacionadas com a probabilidade de classificação incorreta

2.4.1 Medidas baseadas na curva de influência

No Apêndice A, foi feita uma breve revisão sobre as probabilidades de classificação incorreta. De (A-1.11), a probabilidade mínima de classificação incorreta, denominada agora simplesmente *PCI*, é dada por $\Phi\left(-\frac{\Delta}{2}\right)$.

Verifica-se que a função de influência também pode ser escrita como

$\lim_{\varepsilon \rightarrow 0} \frac{T(F_\varepsilon) - T(F)}{\varepsilon} = \frac{\delta}{d\varepsilon} T(F_\varepsilon) \Big|_{\varepsilon=0}$, ou seja, a curva de influência é a derivada do funcional $T(F_\varepsilon)$ com respeito a ε calculada para $\varepsilon = 0$.

Lembrando que $\Delta^{2*} \cong (1 + \varepsilon w_1)\Delta^2 + 2\varepsilon\phi - \varepsilon w_1\phi^2$,

$$PCI^* = \Phi\left(-\frac{\Delta^*}{2}\right) = \Phi\left(-\frac{1}{2}\sqrt{(\Delta^2 + \varepsilon w_1\Delta^2 + 2\varepsilon\phi - \varepsilon w_1\phi^2)}\right),$$

e é possível calcular a função de influência para *PCI* através dessa forma alternativa.

Derivando em relação a ε , segue que

$$I(\mathbf{x}; PCI) = f\left(-\frac{1}{2}\sqrt{\Delta^2 + \varepsilon w_1\Delta^2 + 2\varepsilon\phi - \varepsilon w_1\phi^2}\right) \cdot -\frac{1}{2} \cdot \frac{1}{2} (\Delta^2 + \varepsilon w_1\Delta^2 + 2\varepsilon\phi - \varepsilon w_1\phi^2)^{-\frac{1}{2}} \cdot I(\mathbf{x}; \Delta^2)$$

em que $f(\cdot)$ é a função densidade da distribuição normal padrão.

Assim, para $\varepsilon = 0$ segue que

$$I(\mathbf{x}; PCI) = -I(\mathbf{x}; \Delta^2) f\left(-\frac{\Delta}{2}\right) / 4\Delta,$$

que é proporcional a $I(\mathbf{x}; \Delta^2)$.

Dessa forma $I(\mathbf{x}; PCI)$ e $I(\mathbf{x}; \Delta^2)$ (proposta por Campbell (1978)) fornecem basicamente a mesma informação do ponto de vista de influência .

Tomando-se a função de influência amostral de Cook e Weisberg (1982 p.110), para $\varepsilon = -\frac{1}{n_1 - 1}$, obtém-se

$$FIA_{(i)} = -(n_1 - 1) \left(T(\hat{F}_{(i)}) - T(\hat{F}) \right),$$

em que $\hat{F}_{(i)}$ é a função de distribuição amostral após perturbação, ou seja, eliminando-se a observação i do grupo 1 .

Trocando-se $\hat{F}_{(i)}$ e \hat{F} pelas funções $\Phi\left(-\frac{D_{(i)}}{2}\right)$ e $\Phi\left(-\frac{D}{2}\right)$

respectivamente, tem-se:

$$FI(\mathbf{x}; PCI) = -(n_1 - 1) \left[\Phi\left(-\frac{D_{(i)}}{2}\right) - \Phi\left(-\frac{D}{2}\right) \right], \quad (2.18)$$

sendo que $D_{(i)}$ é a estimativa de máxima verossimilhança para Δ , sem a i -ésima observação: $D_{(i)} = \left[(\bar{\mathbf{x}}_{1(i)} - \bar{\mathbf{x}}_2) \mathbf{S}_0^{-1} (\bar{\mathbf{x}}_{1(i)} - \bar{\mathbf{x}}_2) \right]^{\frac{1}{2}}$.

As probabilidades de classificação incorreta estimadas da expressão (2.18) são calculadas sob duas regras de classificação:

$\Phi\left(-\frac{D}{2}\right)$ é obtida sob a regra discriminante de Fisher com base na amostra toda e

$\Phi\left(-\frac{D_{(i)}}{2}\right)$ sob a regra discriminante de Fisher com base na amostra reduzida (sem

a i -ésima observação), ou seja, alocar \mathbf{x}_0 em π_1 se

$$\hat{\mathbf{I}}_{(i)} \left[\mathbf{x}_0 - \frac{1}{2}(\bar{\mathbf{x}}_{1(i)} + \bar{\mathbf{x}}_2) \right] > 0, \quad (2.19)$$

caso contrário, alocar \mathbf{x}_0 em π_2 ,

sendo que

$\hat{\mathbf{I}}_{(i)}$ é o vetor de coeficientes discriminantes calculado eliminando-se a observação i ,

ou seja, $\hat{\mathbf{I}}_{(i)} = \mathbf{S}_{(i)}^{-1}(\bar{\mathbf{x}}_{1(i)} - \bar{\mathbf{x}}_2)$ em que

$\bar{\mathbf{x}}_{1(i)}$ é o vetor de médias da amostra da população π_1 calculado sem a i -ésima observação e

$\mathbf{S}_{(i)}^{-1}$ é a matriz de covariância ponderada calculada sem a observação i .

Avaliando a probabilidade de classificação incorreta sob essa segunda regra, usando a distribuição $N(\bar{\mathbf{x}}_j, \mathbf{S})$ $j=1, 2$, de (A-1.10) tem-se

$$PCI = \frac{1}{2} \left(P(1/2)_{(i)} + P(2/1)_{(i)} \right) \quad \text{em que}$$

$$P(2/1)_{(i)} = P\left(Y_{(i)} < \hat{\mathbf{I}}_{(i)}' \frac{1}{2}(\bar{\mathbf{x}}_{1(i)} + \bar{\mathbf{x}}_2) \right), \quad P(1/2)_{(i)} = P\left(Y_{(i)} > \hat{\mathbf{I}}_{(i)}' \frac{1}{2}(\bar{\mathbf{x}}_{1(i)} + \bar{\mathbf{x}}_2) \right)$$

sendo que $P(2/1)_{(i)}$ é a probabilidade de classificar incorretamente um elemento em π_2 quando ele é de π_1 , $P(1/2)_{(i)}$ é a probabilidade de classificar incorretamente um elemento em π_1 quando ele é de π_2 e $Y_{(i)} = \hat{\mathbf{I}}_{(i)}' \mathbf{x}$. Ambas as probabilidades são calculadas eliminando-se o i -ésimo elemento da amostra da população 1.

Usando (A-1.1)

$$\begin{aligned}
P(2/1)_{(i)} &= P\left(\frac{Y_{(i)} - \hat{\mathbf{I}}_{(0)} \bar{\mathbf{x}}_1}{\sqrt{\hat{\mathbf{I}}_{(0)} \mathbf{S} \hat{\mathbf{I}}_{(0)}}} < \frac{\frac{1}{2} \hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_{1(0)} + \bar{\mathbf{x}}_2) - \hat{\mathbf{I}}_{(0)} \bar{\mathbf{x}}_1}{\sqrt{\hat{\mathbf{I}}_{(0)} \mathbf{S} \hat{\mathbf{I}}_{(0)}}}\right) \\
&= P\left(Z < \frac{\hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_{1(0)} + \bar{\mathbf{x}}_2) - 2\hat{\mathbf{I}}_{(0)} \bar{\mathbf{x}}_1}{2G}\right) = P\left(Z < \frac{-\hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})}{2G}\right) \\
&= \Phi\left[\frac{-\hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})}{2G}\right], \text{ sendo que } G^2 = \hat{\mathbf{I}}_{(0)} \mathbf{S} \hat{\mathbf{I}}_{(0)}. \quad (2.20)
\end{aligned}$$

Do mesmo modo,

$$\begin{aligned}
P(1/2)_{(i)} &= P\left(Z > \frac{\hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_{1(0)} + \bar{\mathbf{x}}_2) - 2\hat{\mathbf{I}}_{(0)} \bar{\mathbf{x}}_2}{2G}\right) \\
&= 1 - \Phi\left(\frac{\hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_{1(0)} + \bar{\mathbf{x}}_2) - 2\hat{\mathbf{I}}_{(0)} \bar{\mathbf{x}}_2}{2G}\right) = \Phi\left(\frac{-\hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_{1(0)} + \bar{\mathbf{x}}_2) + 2\hat{\mathbf{I}}_{(0)} \bar{\mathbf{x}}_2}{2G}\right) = \Phi\left(\frac{-\hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) + \hat{\mathbf{I}}_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})}{2G}\right). \quad (2.21)
\end{aligned}$$

Fung (1992) propõe uma medida de influência com base na diferença das estimativas das probabilidades de classificação incorreta, semelhante a (2.18), definida como

$$DPCI_i = \left[\frac{1}{2} (P(1/2)_{(i)} + P(2/1)_{(i)}) \right] - \left[\Phi\left(-\frac{D}{2}\right) \right]. \quad (2.22)$$

Comparando-se os diferentes termos presentes nas medidas (2.18) e (2.22), ou seja, $\Phi\left(-\frac{D_{(i)}}{2}\right)$ e $\frac{1}{2}(P(1/2)_{(i)} + P(2/1)_{(i)})$, nota-se que são formas distintas de cálculo das estimativas da probabilidade de classificação incorreta sob a mesma regra linear discriminante (2.19), após a retirada da i -ésima observação: $\Phi\left(-\frac{D_{(i)}}{2}\right)$ é calculada sob as distribuições $N(\bar{\mathbf{x}}_{1(i)}, \mathbf{S}_{(i)})$ e $N(\bar{\mathbf{x}}_2, \mathbf{S}_{(i)})$ que dependem de i , $\frac{1}{2}(P(1/2)_{(i)} + P(2/1)_{(i)})$ é calculada baseada na distribuição original $N(\bar{\mathbf{x}}_j, \mathbf{S})$ $j = 1, 2$.

Na expressão (2.22), Fung (1992) usou a distribuição $N(\bar{\mathbf{x}}_j, \mathbf{S})$ $j = 1, 2$, que é baseada na amostra inteira, para o cálculo das probabilidades de classificação incorreta. No entanto, observou em suas experiências práticas que, usando as distribuições $N(\bar{\mathbf{x}}_{1(i)}, \mathbf{S}_{(i)})$ e $N(\bar{\mathbf{x}}_2, \mathbf{S}_{(i)})$, obtém-se informações muito próximas a $DPCI_i$ calculado com base na amostra inteira.

Essas duas medidas podem ser usadas na pesquisa de observações influentes em análise discriminante. Altos valores de $DPCI_i$ indicam grandes diferenças entre as estimativas da probabilidade de classificação incorreta na presença e ausência do ponto, sugerindo uma alta influência do mesmo.

Após cálculos (Cálculo B-5 do Apêndice B), verifica-se que as expressões (2.20) e (2.21) podem ser reescritas como:

$$P(2/1)_{(i)} = \Phi \left[\left\{ -\frac{n-3}{n-2} \left[D^2 - \frac{b\hat{\phi}_i^2}{1+bd_i^2} - \frac{\hat{\phi}_i^2}{(1+bd_i^2)(n_1-1)} \right] - \frac{(n-3)[\hat{\phi}_i - d_i^2/(n_1-1)]}{(n-2)(n_1-1)(1+bd_i^2)} \right\} / 2G \right]$$

e

$$P(1/2)_{(i)} = \Phi \left[\left\{ -\frac{n-3}{n-2} \left[D^2 - \frac{b\hat{\phi}_i^2}{1+bd_i^2} - \frac{\hat{\phi}_i^2}{(1+bd_i^2)(n_1-1)} \right] + \frac{(n-3)[\hat{\phi}_i - d_i^2/(n_1-1)]}{(n-2)(n_1-1)(1+bd_i^2)} \right\} / 2G \right]$$

em que

$$G^2 = \left(\frac{n-3}{n-2} \right)^2 \left[D^2 - 2\hat{\phi}_i \left\{ \frac{b\hat{\phi}_i}{1+bd_i^2} + \frac{1-bd_i^2}{(n_1-1)(1+bd_i^2)} \right\} + \left\{ \frac{b\hat{\phi}_i}{1+bd_i^2} + \frac{1}{(n_1-1)(1+bd_i^2)} \right\}^2 d_i^2 \right],$$

$$b = \frac{-n_1}{(n_1-1)(n-2)}, \quad \hat{\phi}_i = \hat{\mathbf{1}}'(\mathbf{x}_{11} - \bar{\mathbf{x}}_1) \quad e \quad d_i^2 = (\mathbf{x}_{11} - \bar{\mathbf{x}}_1)' \mathbf{S}^{-1}(\mathbf{x}_{11} - \bar{\mathbf{x}}_1)$$

A partir dessas expansões, observa-se que as estatísticas básicas já apontadas, $\hat{\phi}_i$ e d_i^2 , têm papel importante no cálculo de $DPCI_i$.

Fung (1992) expandiu as medidas $P(1/2)_{(i)}$ e $P(2/1)_{(i)}$ em série de Taylor ao redor de $(-1/2 D)$ até segunda ordem, obtendo

$$DPCI_i \cong \frac{\Phi\left(-\frac{1}{2}D\right)}{4D(n_1-1)^2} \left[\left(1 - w_1 \hat{\phi}_i\right)^2 \left(d_i^2 - \hat{\phi}_i^2 / D^2\right) + \frac{1}{4} \hat{\phi}_i^2 \right]. \quad (2.23)$$

Tal aproximação será utilizada no exemplo a seguir.

O autor usou o conhecido conjunto de dados da *iris versicolour* (observações de 1 a 50) e *virginica* (observações 51 a 100) utilizado por Fisher, analisando apenas a largura da pétala e da sépala. Na Figura 2.3 estão as medidas dessas duas variáveis juntamente com o gráfico da reta discriminante:

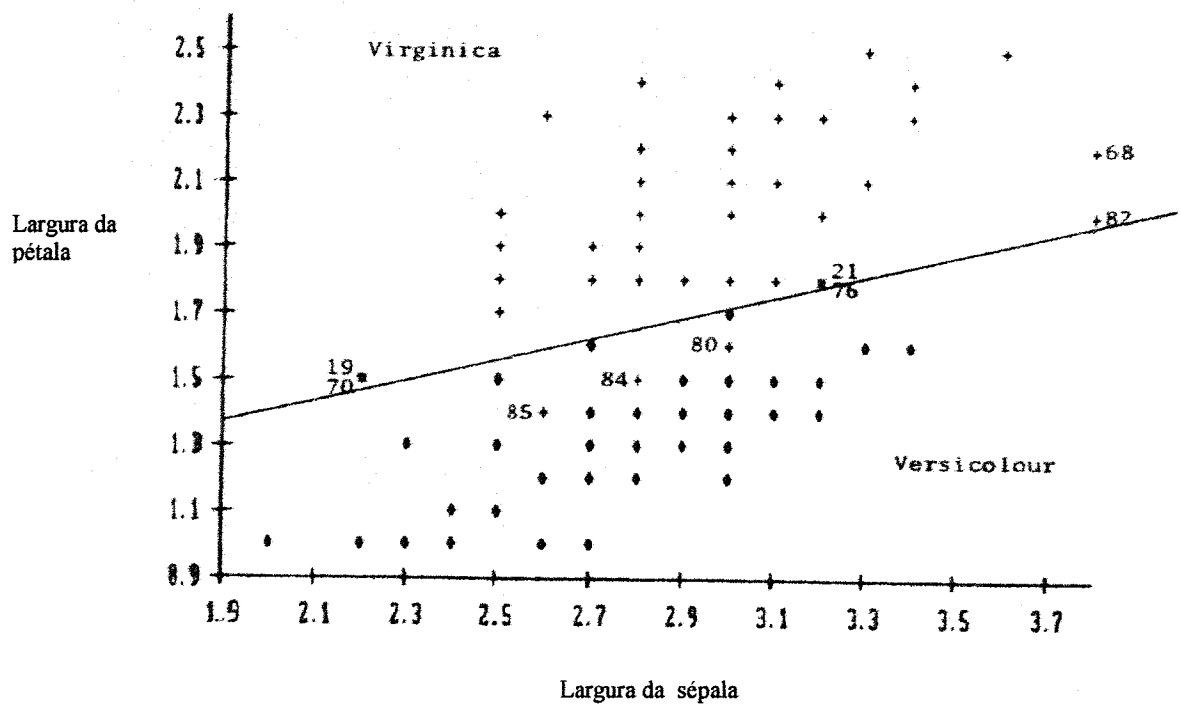


Figura 2.3 : Largura de sépala e pétala das plantas da espécie *Íris Versicolour* e *Íris Virginica*.

Fonte : Figura 1 - Fung (1992).

A Figura 2.4 mostra o gráfico de $\hat{I}_m(\mathbf{x}, \Delta^2)$ para todas as observações ; não indica qualquer observação que tenha significativa influência na estimação da distância entre os grupos, mas mostra que as observações 19, 21, 80, 84 e 85 têm os maiores valores de $\hat{I}_m(\mathbf{x}, \Delta^2)$. É importante observar que são essas observações classificadas incorretamente na análise discriminante executada, como mostra a Figura 2.3.

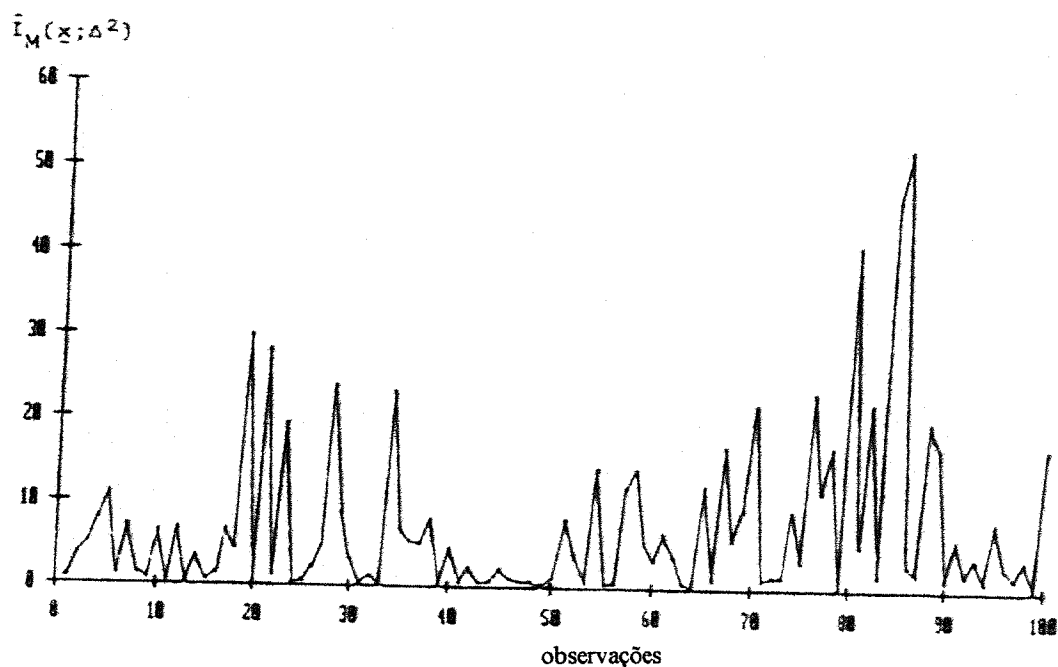


Figura 2.4: Gráfico de $\hat{I}_m(\mathbf{x}, \Delta^2)$.

Fonte: Figura 2 - Fung (1992).

Observa-se adicionalmente, de (2.17), que $\hat{I}_m(\mathbf{x}; \Delta^2) = w_1(\hat{\phi} - w_1^{-1})^2$ é nulo para $\hat{\phi} = \hat{I}'(\mathbf{x} - \bar{\mathbf{x}}_1) = w_1^{-1}$. Assim, observações distantes do plano $\hat{I}'(\mathbf{x} - \bar{\mathbf{x}}_1) = w_1^{-1}$ terão maior efeito na estimação de Δ^2 , enquanto que observações como as de número 68 e 82, que estão próximas do plano, têm menor influência na estimação da distância entre as populações.

Na Figura 2.5, o gráfico de $DPCI_i$ para cada observação mostra que as de número 19, 70 e 82 são as mais influentes. Observando a Figura 2.3, pode-se notar que essas observações estão nos limites inferior e superior da largura da sépala e bem próximas à reta discriminante.

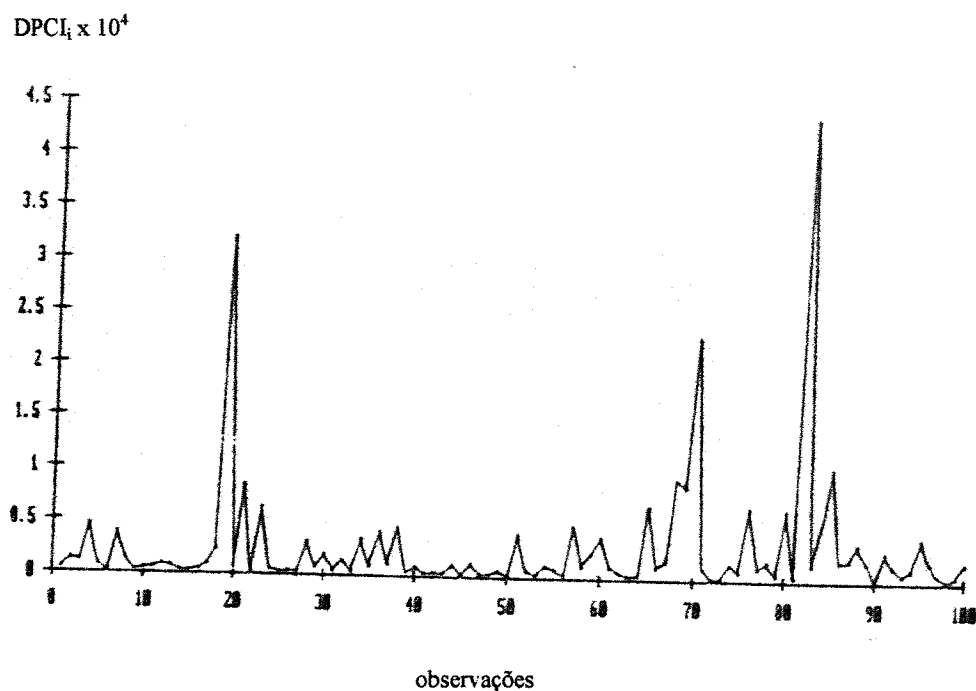


Figura 2.5: Gráfico de $DPCI_i \times 10^4$.

Fonte: Figura 3 - Fung (1992).

A Figura 2.6 apresenta o gráfico da estatística $\hat{\phi}_i$ em função de d_i^2 . As curvas presentes nesse gráfico representam os contornos de $DPCI_i \times \frac{4D(n_i-1)^2}{\Phi\left(-\frac{1}{2}D\right)}$ obtidos por meio da aproximação (2.23). Observa-se que não existem valores extremos para $\hat{\phi}_i$ e d_i^2 . Além disso, pares com altos valores de $\hat{\phi}_i$ e d_i^2 não fornecem necessariamente grandes valores de $DPCI_i$.

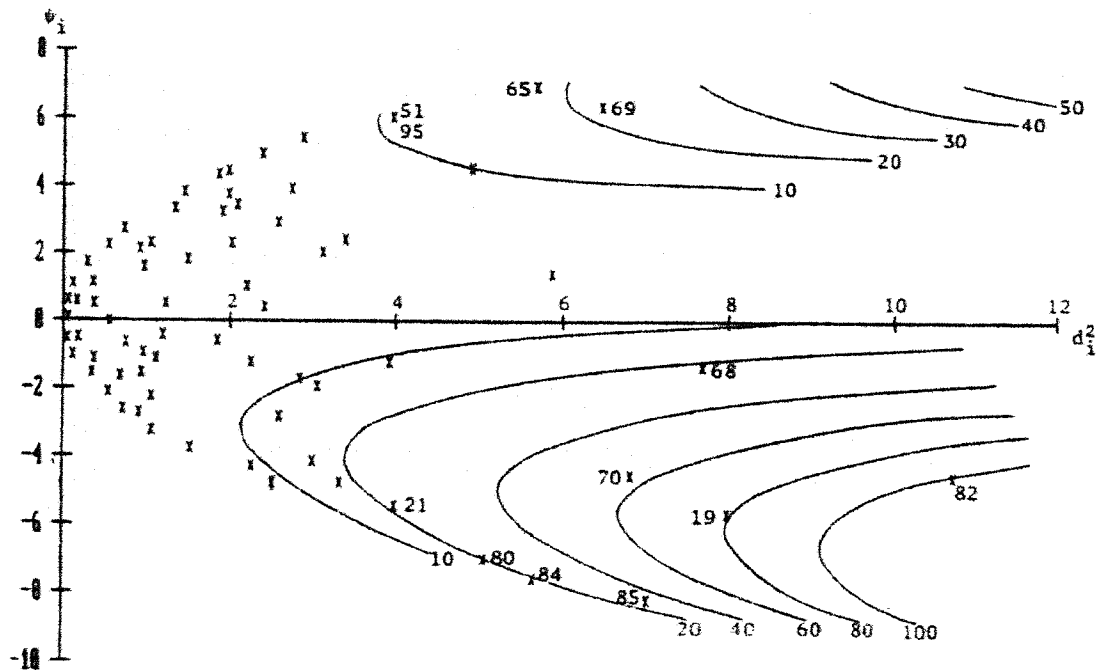


Figura 2.6: Gráfico de $\hat{\phi}_i$ e d_i^2 , com contornos aproximados da constante

$$DPCI_i \times 4D(n_i-1)^2 / \Phi\left(-\frac{1}{2}D\right)$$

Fonte: Figura 4 - Fung (1992)

Assim, um grande d_i^2 pode implicar que a observação i é *outlier*, mas não necessariamente influente do ponto de vista da análise discriminante. No entanto observações com grande d_i^2 e $\hat{\phi}_i$ negativo podem ter um grande $DPCI_i$ (ver Figura 2.6). Uma observação que está longe de sua média e mais perto da média

do outro grupo, tendo portanto uma maior chance de ser classificada incorretamente, apresentaria uma grande influência. É o que ocorre com as observações 19, 70 e 82, que uma vez excluídas, tiveram grande influência na regra de classificação e portanto, são altamente influentes na probabilidade de classificação incorreta.

2.4.2 Medidas alternativas

Critchley e Vitiello (1991), independentemente de Fung (1992), determinaram duas estatísticas principais, que compõem algumas medidas de influência das observações. São elas a “atipicalidade” e a diferença entre o escore discriminante linear da observação e o escore discriminante de sua média amostral.

Verifica-se que o escore discriminante de Fisher associado a uma nova observação \mathbf{x} de origem desconhecida, $FD(\mathbf{x}) = \mathbf{1}' \left(\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \right) =$

$= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \left(\mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \right)$, pode ser escrito como

$$\theta(\mathbf{x}) = \frac{1}{2} \{\alpha_2(\mathbf{x})\}^2 - \frac{1}{2} \{\alpha_1(\mathbf{x})\}^2 \quad \text{com} \quad \{\alpha_j(\mathbf{x})\}^2 = (\mathbf{x} - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}_j), \quad j = 1, 2.$$

A medida $\{\alpha_j(\mathbf{x})\}^2$ é interpretada como a “atipicalidade” da observação \mathbf{x} com relação à população j . Nesse sentido, seria análoga à quantidade d_i^2 , obtida por Fung (1992) e descrita na Seção 2.3.

Para Critchley e Vitiello (1991), o objetivo principal era considerar o efeito em $\hat{\Delta}^2$ produzido pela eliminação de uma única observação i da amostra proveniente de π_1 , ou seja, calcular $\hat{\Delta}_{(i)}^2 = (\bar{\mathbf{x}}_{1(i)} - \bar{\mathbf{x}}_2)' \mathbf{S}_{(i)}^{-1} (\bar{\mathbf{x}}_{1(i)} - \bar{\mathbf{x}}_2)$. Com a restrição de que

$n_1 + n_2 - 3 \geq p$ e por intermédio do cálculo B-7 do Apêndice B, chega-se a:

$$\hat{\Delta}_{(i)}^2 = \frac{n-1}{n} \left\{ \hat{\Delta}^2 - \frac{2e(\mathbf{x}_{11})}{n_1-1} + \frac{\hat{\alpha}_1^2(\mathbf{x}_{11})}{(n_1-1)^2} + \frac{1}{w_1^{-1}(n_1-1) - \hat{\alpha}_1^2(\mathbf{x}_{11})} \times \right. \\ \left. \left[(e(\mathbf{x}_{11}))^2 - \frac{2e(\mathbf{x}_{11})\hat{\alpha}_1^2(\mathbf{x}_{11})}{n_1-1} + \frac{(\hat{\alpha}_1^2(\mathbf{x}_{11}))^2}{(n_1-1)^2} \right] \right\} \quad (2.24)$$

sendo que

$$n = n_1 + n_2, \quad w_1 = \frac{n_1}{n}, \quad \hat{\alpha}_1^2(\mathbf{x}_{11}) = (\mathbf{x}_{11} - \bar{\mathbf{x}}_1)' \mathbf{S}^{-1} (\mathbf{x}_{11} - \bar{\mathbf{x}}_1), \quad e(\mathbf{x}_{11}) = \hat{\mathbf{I}}' (\mathbf{x}_{11} - \bar{\mathbf{x}}_1).$$

Observa-se que a eliminação de \mathbf{x}_{11} implica na mudança dos valores dos graus de liberdade, de $\bar{\mathbf{x}}_1$ e de \mathbf{S}_1 . A expressão (2.24) pode ser escrita alternativamente como :

$$\hat{\Delta}_{(i)}^2 = T_{gl} (\hat{\Delta}^2 + T_{\bar{\mathbf{x}}_1} + T_{\mathbf{S}_1} + T_{\bar{\mathbf{x}}_1, \mathbf{S}_1})$$

$$\text{em que } T_{gl} = \frac{n_1 + n_2 - 1}{n_1 + n_2}, \quad T_{\bar{\mathbf{x}}_1} = \frac{-2e(\mathbf{x}_{11})}{n_1 - 1} + \frac{\hat{\alpha}_1^2(\mathbf{x}_{11})}{(n_1 - 1)^2}, \quad T_{\mathbf{S}_1} = \frac{e^2(\mathbf{x}_{11})}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(\mathbf{x}_{11})}$$

e

$$T_{\bar{\mathbf{x}}_1, \mathbf{S}_1} = \frac{\hat{\alpha}_1^4(\mathbf{x}_{11}) - 2(n_1 - 1)\hat{\alpha}_1^2(\mathbf{x}_{11})e(\mathbf{x}_{11})}{(n_1 - 1)^2 \{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(\mathbf{x}_{11})\}}$$

Nota-se-se que T_{gl} reflete a mudança nos graus de liberdade com a retirada de \mathbf{x}_{11} , os termos $T_{\bar{\mathbf{x}}_1}$ e $T_{\mathbf{S}_1}$ são devidos à mudança em $\bar{\mathbf{x}}_1$ e \mathbf{S}_1 respectivamente e $T_{\bar{\mathbf{x}}_1, \mathbf{S}_1}$ é o termo de interação, que representa a alteração conjunta em $\bar{\mathbf{x}}_1$ e \mathbf{S}_1 .

Mantendo somente os termos até $O(n^{-1})$ na expressão $\hat{\Delta}_{(i)}^2$ escrita na forma alternativa, segue que

$$\hat{\Delta}_{1(i)}^2 = \left(\frac{n_1 + n_2 - 1}{n_1 + n_2} \right) \left[\hat{\Delta}^2 - \frac{2e(\mathbf{x}_{1i})}{(n_1 - 1)} + \frac{e^2(\mathbf{x}_{1i})}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(\mathbf{x}_{1i})} \right]$$

e somando e subtraindo $\frac{(w_1^{-1})^2}{w_1^{-1}(n_1 - 1)}$ ao segundo termo, segue que

$$\hat{\Delta}_{1(i)}^2 = \left(\frac{n_1 + n_2 - 1}{n_1 + n_2} \right) \left[\hat{\Delta}^2 - \frac{1}{w_1(n_1 - 1)} + \frac{\{e(\mathbf{x}_{1i}) - w_1^{-1}\}^2}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(\mathbf{x}_{1i})} \right]. \quad (2.25)$$

Do mesmo modo, eliminando uma observação da amostra proveniente de π_2

$$e(\mathbf{x}_{2i}) = F\hat{D}(\mathbf{x}_{2i}) - F\hat{D}(\bar{\mathbf{x}}_2) \quad \text{com} \quad e(\mathbf{x}_{2i}) = \hat{l}'(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2) \quad \text{tem-se:}$$

$$\hat{\Delta}_{2(i)}^2 = \left(\frac{n_1 + n_2 - 1}{n_1 + n_2} \right) \left[\hat{\Delta}^2 - \frac{1}{w_2(n_2 - 1)} + \frac{\{e(\mathbf{x}_{2i}) - w_2^{-1}\}^2}{w_2^{-1}(n_2 - 1) - \hat{\alpha}_2^2(\mathbf{x}_{2i})} \right].$$

Dessa forma, o efeito de eliminar uma observação na estimativa da probabilidade de classificação incorreta $\left(\Phi\left(-\frac{1}{2}\hat{\Delta}\right) \right)$, que é uma função decrescente de $\hat{\Delta}$, é determinado somente em função das duas quantidades:

- a) $\hat{\alpha}_j^2(\mathbf{x}_{ji})$, $j=1, 2$, que é a estimativa da *atipicalidade* da observação com relação à própria população a qual ela pertence, e
- b) $e(\mathbf{x}_{ji})$, $j=1, 2$, que é a diferença entre o escore discriminante da observação em questão e o escore discriminante da média amostral do grupo ao qual a observação pertence. Esta quantidade coincide com a estatística descrita por Fung (1992) como $\hat{\phi}_i = \delta'z$ (Seção 2.2).

As medidas descritas em a) e b) são similares respectivamente aos pontos de alavanca e resíduos em modelos de regressão linear, mas existem importantes diferenças.

Em modelos de regressão as observações com h_{ii} alto (Seção A-2 do Apêndice A) são os chamados *pontos de alavanca*, pois, de modo geral, trazem o plano de regressão em sua direção, o que tende a produzir pequenos resíduos (o resíduo mede a diferença entre o valor observado e o valor ajustado). Já na análise discriminante, o termo resíduo $e(\mathbf{x}_{1i})$ pode variar independentemente do termo alavanca $\hat{\alpha}_1^2(\mathbf{x}_{1i})$.

Observando a equação (2.25) nota-se que a melhoria na estimativa de probabilidade de classificação incorreta se dará quando:

- aumenta o quadrado da diferença entre o “resíduo” $e(\mathbf{x}_{1i})$ e o inverso do número de observações na amostra (w_1^{-1}) e
- aumenta a estimativa da *atipicalidade* da observação (maior distância dessa observação em relação à média de seu grupo). Este fato é mais intuitivo do que o resultado descrito em a).

A análise apresentada confirma os resultados surpreendentes de Campbell (1978), apresentados na Seção 2.2.3. Tomando a expressão (2.7) da Seção 2.2.1,

$\Delta^{2*} = (1 + \varepsilon w_1)\Delta^2 + 2\varepsilon\phi - \varepsilon w_1\phi^2$ e substituindo ε por $-\frac{1}{n_1 - 1}$ e ϕ por $e(\mathbf{x}_{1i})$, obtém-se

$$\Delta^{2*} = \left(1 - \frac{w_1}{n_1 - 1}\right)\hat{\Delta}^2 - \frac{2e(\mathbf{x}_{1i})}{n_1 - 1} + \frac{w_1(e(\mathbf{x}_{1i}))^2}{n_1 - 1}.$$

Multiplicando e dividindo o segundo e terceiro termos por w_1^{-1} e somando e subtraindo $\frac{(w_1^{-1})^2}{w_1^{-1}(n_1 - 1)}$ à equação, seque que

$$\Delta^{2*} \cong \left(1 - \frac{w_1}{n_1 - 1}\right) \hat{\Delta}^2 - \frac{1}{w_1(n_1 - 1)} + \frac{\{e(\mathbf{x}_{1i}) - w_1^{-1}\}^2}{w_1^{-1}(n_1 - 1)}, \quad (2.26)$$

expressão essa que não envolve $\hat{\alpha}_1^2(\mathbf{x}_{1i})$.

Analisando-se a expressão $\{e(x_{1i}) - w_1^{-1}\}^2$, nota-se que as observações afetam $\hat{\Delta}^2$ assimetricamente, ou seja, o efeito é menor quando é retirada uma observação com resíduo $e(\mathbf{x}_i)$ positivo, do que quando o resíduo é negativo e de mesmo valor absoluto.

Do mesmo modo que descrito no cálculo de (2.25), a equação (2.26) pode ser escrita como $\hat{\Delta}^{2*} = T_g \hat{\Delta}^2 + T_{\bar{x}_1} + T_{s_1}$, com uma decomposição aditiva em termos de

$$T_{\bar{x}_1} \sim \frac{-2e(\mathbf{x}_{1i})}{n_1 - 1} \quad \text{e} \quad T_{s_1} \sim \frac{e^2(\mathbf{x}_{1i})}{w_1^{-1}(n_1 - 1)}.$$

Para explicar o motivo da assimetria, verifica-se que, embora a magnitude de $T_{\bar{x}_1}$ e T_{s_1} sejam as mesmas quando $e(\mathbf{x}_{1i})$ muda de sinal, o valor da soma

$$T_{\bar{x}_1} + T_{s_1} \sim \frac{\{e(\mathbf{x}_{1i}) - w_1^{-1}\}^2}{w_1^{-1}(n_1 - 1)}$$

não é o mesmo quando da mudança de sinal. Isso porque ao contrário de T_{s_1} , $T_{\bar{x}_1}$ muda de sinal quando $e(\mathbf{x}_{1i})$ muda o sinal.

Observa-se ainda que $T_{\bar{x}_1}$ e T_{s_1} são de mesmo sinal quando $e(\mathbf{x}_{1i}) < 0$, mas têm sinal contrário quando $e(\mathbf{x}_{1i}) > 0$. Como T_{s_1} é quadrático em $e(\mathbf{x}_{1i})$ enquanto $T_{\bar{x}_1}$ é linear em $e(\mathbf{x}_{1i})$, T_{s_1} domina quando $e(\mathbf{x}_{1i})$ aumenta. Assim, excluindo uma observação \mathbf{x}_{1i} com $e(\mathbf{x}_{1i})$ positivo, $\hat{\Delta}^2$ sofre uma redução devido à mudança na média, mas esse efeito é compensado pelo aumento em $\hat{\Delta}^2$ causado por T_{s_1} .

2.5 Outras medidas de diagnóstico

2.5.1 Medidas baseadas na média quadrática

Tomando novamente a regra discriminante de Fisher (A-1.8), Fung (1995-a) define várias medidas de diagnóstico. Para isso, denotou o escore discriminante (que também é o logaritmo da razão das probabilidades estimadas para $q = 1 - q$)

$\hat{\mathbf{l}}' \mathbf{x} - \frac{1}{2} \hat{\mathbf{l}}' (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)$ como

$$\hat{\mathbf{y}}' \mathbf{y} \text{ em que } \hat{\mathbf{y}}' = \left(-\hat{\mathbf{l}}' \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2}, \hat{\mathbf{l}}' \right) \text{ e } \mathbf{y}' = (1, \mathbf{x}') . \quad (2.27)$$

Com o objetivo de analisar o efeito da eliminação da observação i na estimativa $\hat{\mathbf{y}}'$, o autor propõe o uso da medida

$$E(\hat{\mathbf{y}}' \mathbf{y} - \hat{\mathbf{y}}'_{(i)} \mathbf{y})^2 , \quad (2.28)$$

ou seja, a média do quadrado da diferença entre os escores discriminantes da amostra toda e da amostra sem a observação i , sendo que

$$\hat{\mathbf{y}}' \mathbf{y} - \hat{\mathbf{y}}'_{(i)} \mathbf{y} = \left(-\hat{\mathbf{l}}' \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} + \hat{\mathbf{l}}' \mathbf{x} \right) - \left(-\hat{\mathbf{l}}'_{(i)} \frac{(\bar{\mathbf{x}}_{1(i)} + \bar{\mathbf{x}}_2)}{2} + \hat{\mathbf{l}}'_{(i)} \mathbf{x} \right)$$

é a diferença entre os logaritmos da razão das probabilidades estimadas, denotada por DLO_i .

Para o cálculo de $E(DLO_i)^2$, admitiu-se que \mathbf{X} é distribuída como $w_1 N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) + w_2 N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ e usou-se as estimativas *plug-in* $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$ e \mathbf{S} para os parâmetros $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ e $\boldsymbol{\Sigma}$.

O princípio *plug-in* (Efron e Tibshirani (1993)) é um método simples de estimar parâmetros através de amostras. A estimativa *plug-in* de um parâmetro $\theta = t(F)$, é definida como $\hat{\theta} = t(\hat{F})$, em que se estima a função $\theta = t(F)$ da função de distribuição F pela mesma função da distribuição empírica \hat{F} , $\hat{\theta} = t(\hat{F})$.

O método *plug-in* é excelente quando a única informação que se tem sobre F vem da amostra.

Observa-se que, neste caso, \bar{x}_1 , \bar{x}_2 e S são considerados constantes para o cálculo de $E(DLO_i)^2$:

$$E(DLO_i)^2 = Var(DLO_i) + E^2(DLO_i).$$

No entanto,

$$\begin{aligned} Var(DLO_i) &= Var\left(\frac{-\hat{i}'(\bar{x}_1 + \bar{x}_2)}{2} + \frac{\hat{i}'_{(0)}(\bar{x}_{1(0)} + \bar{x}_2)}{2} + (\hat{i}' - \hat{i}'_{(0)})\mathbf{X}\right) = (\hat{i}' - \hat{i}'_{(0)})Var(\mathbf{X})(\hat{i}' - \hat{i}'_{(0)}) \\ &= (\hat{i}' - \hat{i}'_{(0)})S(\hat{i}' - \hat{i}'_{(0)}), \end{aligned} \quad (2.29)$$

e

$$\begin{aligned} E(DLO_i) &= E\left(\hat{i}'\mathbf{X} - \hat{i}'\frac{(\bar{x}_1 + \bar{x}_2)}{2} - \hat{i}'_{(0)}\mathbf{X} + \hat{i}'_{(0)}\frac{(\bar{x}_{1(0)} + \bar{x}_2)}{2}\right) \\ &= \hat{i}'(w_1\bar{x}_1 + (1-w_1)\bar{x}_2) - \frac{\hat{i}'(\bar{x}_1 + \bar{x}_2)}{2} - \hat{i}'_{(0)}(w_1\bar{x}_1 + (1-w_1)\bar{x}_2) + \frac{\hat{i}'_{(0)}(\bar{x}_{1(0)} + \bar{x}_2)}{2} \\ &= \left\{2\hat{i}'w_1\bar{x}_1 + 2\hat{i}'\bar{x}_2 - 2\hat{i}'w_1\bar{x}_2 - \hat{i}'\bar{x}_1 - \hat{i}'\bar{x}_2 - 2\hat{i}'_{(0)}w_1\bar{x}_1 - 2\hat{i}'_{(0)}\bar{x}_2 + 2\hat{i}'_{(0)}w_1\bar{x}_2 + \hat{i}'_{(0)}\bar{x}_{1(0)} + \hat{i}'_{(0)}\bar{x}_2\right\}/2. \end{aligned}$$

Acrescentando e subtraindo $\hat{i}'_{(0)}\bar{x}_1$ e posteriormente $w_1\hat{i}'_{(0)}(\bar{x}_1 - \bar{x}_{1(0)})/2$ à última expressão, obtém-se

$$\begin{aligned} E(DLO_i) &= \left\{2w_1(\hat{i}' - \hat{i}'_{(0)})(\bar{x}_1 - \bar{x}_2) + \hat{i}'\bar{x}_2 - \hat{i}'\bar{x}_1 - \hat{i}'_{(0)}\bar{x}_2 + \hat{i}'_{(0)}\bar{x}_{1(0)} - \hat{i}'_{(0)}\bar{x}_1 + \hat{i}'_{(0)}\bar{x}_1\right\}/2 \\ &= \left\{2w_1(\hat{i}' - \hat{i}'_{(0)})(\bar{x}_1 - \bar{x}_2) - (\hat{i}' - \hat{i}'_{(0)})(\bar{x}_1 - \bar{x}_2) - \hat{i}'_{(0)}(\bar{x}_1 - \bar{x}_{1(0)})\right\}/2 \end{aligned}$$

$$= w_1 \left(\frac{(\hat{\mathbf{i}}' - \hat{\mathbf{i}}'_{(0)}) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \hat{\mathbf{i}}'_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})}{2} \right) + (1 - w_1) \left(\frac{-(\hat{\mathbf{i}}' - \hat{\mathbf{i}}'_{(0)}) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \hat{\mathbf{i}}'_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})}{2} \right).$$

Assim, $E^2(DLO_i) = (w_1 B_1 + w_2 B_2)^2$ e

$$E(DLO_i)^2 = (w_1 B_1 + w_2 B_2)^2 + V, \quad (2.30)$$

sendo que $w_1 = \frac{n_1}{n}$, $w_2 = 1 - w_1$

$$B_1 = \frac{(\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(0)}) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{2} - \frac{\hat{\mathbf{i}}'_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})}{2},$$

$$B_2 = \frac{-(\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(0)}) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{2} - \frac{\hat{\mathbf{i}}'_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})}{2} \text{ e}$$

$$V = (\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(0)})' \mathbf{S} (\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(0)}). \quad (2.31)$$

O Cálculo B-6 do Apêndice B mostra que as quantidades B_1 , B_2 e V também podem ser escritas em função das estatísticas fundamentais d_i^2 e $\hat{\phi}_i$.

Usando a função de distribuição empírica não-paramétrica para o cálculo da esperança em (2.28) determina-se a medida

$$F2 = \sum_j \frac{[(\hat{y} - \hat{y}_{(0)}) \mathbf{x}_j]^2}{n} \quad \text{ou, de forma equivalente}$$

$$F2 = \frac{(\hat{y} - \hat{y}_{(0)})' \mathbf{X}' \mathbf{X} (\hat{y} - \hat{y}_{(0)})}{n}, \quad (2.32)$$

que é semelhante à estatística D de Cook (A-2.2) utilizada em diagnóstico em regressão.

Se usada a função de distribuição empírica sem a observação i , a equação (2.32) é escrita como

$$F2I = (\hat{y} - \hat{y}_{(i)}) \mathbf{X}'_{(i)} \mathbf{X}_{(i)} \frac{(\hat{y} - \hat{y}_{(i)})}{n-1} \quad \text{que é semelhante à estatística de Welsh (1982).}$$

Assim como $E^2(DLO)$, $F2$ e $F2I$ também podem ser expressas em função de d_i^2 e $\hat{\phi}_i$. Verifica-se ainda que essas estatísticas servem para detectar observações com grande influência no logaritmo da razão das probabilidades estimadas (score discriminante).

2.5.2 Medidas de influência para regressão logística – uma abordagem para análise discriminante

É conhecida a analogia entre análise discriminante e regressão logística (Efron (1975)). No modelo de regressão logística, várias medidas de influência já foram sugeridas com o objetivo de identificar observações influentes na estimação do vetor de coeficientes de regressão e no *deviance*. Johnson (1985) propõe novas medidas com o intuito de detectar a influência relativa na determinação de estimativas de probabilidades e classificação de futuras observações.

Dada uma amostra aleatória de N observações $(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \dots, (y_N, \mathbf{x}_N)$, com \mathbf{x}_j vetor $1 \times (p + 1)$ de covariáveis, sendo que a primeira coordenada é *um* para a inclusão do termo constante, o modelo de regressão logística é tal que

$$P(y_j = 1 / \mathbf{x}_j) = \frac{e^{\mathbf{x}_j \beta}}{1 + e^{\mathbf{x}_j \beta}}, \quad j = 1, 2, \dots, N$$

sendo que y_j tem distribuição de Bernoulli com parâmetro p_j e β é um vetor $(p + 1) \times 1$ de coeficientes de regressão desconhecido.

A população correspondente a sucesso ($y = 1$) é π_1 e a correspondente ao fracasso ($y = 0$) é π_2 . O logaritmo da função de verossimilhança de β é indicado por $l(\mathbf{X}\beta)$ e a estimativa de máxima verossimilhança de β é $\hat{\beta}$.

Seja $\mathbf{X}^f = (\mathbf{x}_1^f, \mathbf{x}_2^f, \dots, \mathbf{x}_M^f)$ (em que o termo f denota observação futura) o conjunto de covariáveis para M indivíduos que ainda não foram classificados em π_2 ou π_1 .

A estimativa de probabilidade de $y = 1$ para a k -ésima observação futura é :

$$\hat{P}(y = 1 / \mathbf{x}_k^f) = \frac{e^{\mathbf{x}_k^f \hat{\beta}}}{1 + e^{\mathbf{x}_k^f \hat{\beta}}} = \hat{p}_k^f, \quad k = 1, 2, \dots, M$$

Quando o objetivo é classificar observações, é comum utilizar-se o *logito*

$$\ln \frac{\hat{p}_k^f}{1 - \hat{p}_k^f} = \mathbf{x}_k^f \hat{\beta}$$

e nesse caso a regra de classificação fica:

alocar o k -ésimo elemento em π_1 se $\ln \frac{\hat{p}_k^f}{1 - \hat{p}_k^f} > 0 \Leftrightarrow \frac{\hat{p}_k^f}{1 - \hat{p}_k^f} > 1$ e

em π_2 se $\ln \frac{\hat{p}_k^f}{1 - \hat{p}_k^f} < 0 \Leftrightarrow \frac{\hat{p}_k^f}{1 - \hat{p}_k^f} < 1$.

Para medir o efeito de cada observação amostral na determinação do vetor de probabilidades \hat{p}_k^f , Johnson (1985) propôs o uso da medida de divergência simétrica de Kullback-Leibler (1951) :

$$J(\hat{p}_k^f, \hat{p}_{k(i)}^f) = I(\hat{p}_k^f, \hat{p}_{k(i)}^f) + I(\hat{p}_{k(i)}^f, \hat{p}_k^f), \quad \begin{matrix} k = 1, 2, \dots, M \\ i = 1, 2, \dots, N \end{matrix}$$

em que $I(\cdot, \cdot)$ é a divergência direcionada de Kullback-Leibler entre duas populações Bernoulli (Seção B-8 do Apêndice B) , e \hat{p}_k^f e $\hat{p}_{k(i)}^f$ são os valores de \hat{p}_k^f calculados respectivamente com e sem a i -ésima observação.

Se f_1 e f_2 são funções densidade e E_{f_l} indica a esperança calculada com respeito à densidade f_l , $l = 1, 2$, as divergências direcionadas de Kullback-Leibler são definidas como

$$I(f_1, f_2) = E_{f_1} \ln \left(\frac{f_1}{f_2} \right) \quad \text{e} \quad I(f_2, f_1) = E_{f_2} \ln \left(\frac{f_2}{f_1} \right).$$

Analisando-se essas expressões, verifica-se que medem distâncias entre as densidades f_1 e f_2 , já que são nulas para $f_2 = f_1$.

Para avaliar o efeito que a observação i tem na previsão dos M elementos futuros , o autor definiu:

$$\tilde{D}_i^f = \sum_{k=1}^M J(\hat{p}_k^f, \hat{p}_{k(i)}^f), \quad i = 1, 2, \dots, N.$$

Neste cálculo, está-se comparando as estimativas das probabilidades de observações futuras na presença e ausência da i -ésima observação.

É possível ainda a utilização de pesos, de modo que

$$\tilde{D}_i^f(c) = \sum_{k=1}^M c_k J(\hat{p}_k^f, \hat{p}_{k(i)}^f)$$

em que $c = (c_1, c_2, \dots, c_M)$ pode ser escolhido de modo a refletir a importância relativa associada à estimação de determinadas probabilidades. Por exemplo, uma

escolha poderia ser $c_k = n_k^f$, indicando que casos futuros com maior número de indivíduos, têm maior peso do que aqueles casos com um número menor. Por outro lado, a escolha $\mathbf{c} = (1, 1, \dots, 1)$ resultaria em \tilde{D}_i^f .

Após cálculos, presentes na Seção B-9 do Apêndice B, verifica-se que:

$$\begin{aligned} \tilde{D}_i^f(\mathbf{c}) &= \sum_{k=1}^M c_k (\hat{p}_k^f - \hat{p}_{k(i)}^f) \ln \left[\frac{\frac{\hat{p}_k^f}{1 - \hat{p}_k^f}}{\frac{\hat{p}_{k(i)}^f}{1 - \hat{p}_{k(i)}^f}} \right] \\ &= \sum_{k=1}^M c_k (\hat{p}_k^f - \hat{p}_{k(i)}^f) \mathbf{x}_k^T (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}), \end{aligned}$$

em que $\hat{\boldsymbol{\beta}}_{(i)}$ é o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ no modelo de regressão logística com base na amostra excluindo-se a i -ésima observação.

Uma diferença básica entre regressão (linear e logística) e análise discriminante é que em regressão, \mathbf{Y} é aleatório e \mathbf{X} é não aleatório, mas na análise discriminante ocorre o contrário, pois a observação \mathbf{x} é que deve ser identificada (modelada) dado o valor de y (Kshirsagar (1972, Sec. 6.4)). Assim sendo, medidas de diagnóstico em regressão, que são construídas sob as suposições de aleatoriedade em y , não são apropriadas para a análise discriminante.

Capítulo 3

Detecção de observações influentes no contexto Bayesiano

3.1 Densidades preditivas e divergência de Kullback-Leibler

Utilizando densidades preditivas, que são a base da análise discriminante Bayesiana, estuda-se, neste capítulo, a influência que as observações têm sobre aspectos inferenciais tais como alocação de futuras observações, separação entre populações e determinação da probabilidade de uma observação pertencer a certa população. Para esse estudo são analisadas as estatísticas de influência definidas por Johnson (1987).

Em adição à notação já utilizada, sejam

- o vetor de parâmetros (μ_1, μ_2, Σ) , que será denominado por θ ,
- X_{ji} com distribuição normal p -variada condicional aos parâmetros,
- d , definido como o conjunto de todos os dados observados X_{ji} , $j = 1, 2$,
 $i = 1, 2 \dots n_j$,
- $p(\theta / \pi_j)$, $j = 1, 2$, a densidade *a priori* para θ sob π_j ,

- $p(\boldsymbol{\theta} / \mathbf{d}, \pi_j)$, densidade *a posteriori* para $\boldsymbol{\theta}$ sob π_j e
- \mathbf{Y} vetor aleatório de observações futuras, independente do conjunto de dados, com distribuição condicional aos parâmetros $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$ sob π_j .

No enfoque Bayesiano, o problema de previsão de observações futuras é resolvido utilizando-se $f(\mathbf{y} / \mathbf{d})$, que é a densidade *a posteriori* da quantidade a ser prevista, \mathbf{Y} , dada a amostra observada \mathbf{d} .

Essa densidade, denominada densidade preditiva *a posteriori* de \mathbf{Y} dado \mathbf{d} , é definida como sendo a esperança *a posteriori* da densidade amostral para \mathbf{Y} :

$$f_j(\mathbf{y}) = f(\mathbf{y} / \mathbf{d}, \pi_j) = \int N(\mathbf{y} / \boldsymbol{\theta}, \pi_j) p(\boldsymbol{\theta} / \mathbf{d}, \pi_j) d\boldsymbol{\theta}, \quad j = 1, 2,$$

esperança essa tomada com relação à distribuição *a posteriori* de $\boldsymbol{\theta}$ dada a amostra \mathbf{d} .

Quando o vetor de parâmetros $\boldsymbol{\theta}$ é conhecido, a densidade preditiva para \mathbf{Y} (sob π_j) será justamente a densidade amostral $N(\cdot / \boldsymbol{\theta}, \pi_j)$.

Quando $\boldsymbol{\theta}$ é desconhecido, Johnson (1987) obtém densidades preditivas aproximadas, substituindo os parâmetros da densidade normal pelas estimativas *plug in*. Outra possibilidade seria o uso da *priori* de referência

$$p(\boldsymbol{\theta} / \pi_j) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(p+1)} \quad j = 1, 2, \quad (3.1)$$

sendo que, com base nessa distribuição, Geisser (1964, 1966) obteve a seguinte densidade preditiva:

$$\mathbf{Y} / \mathbf{x}, \pi_j \sim St(v, \bar{\mathbf{x}}_j, (1 + n_j^{-1}) \tilde{\mathbf{S}})$$

que é a distribuição *t-Student* p -variada com v graus de liberdade, vetor de locação $\bar{\mathbf{x}}_j$ e matriz de dispersão $(1 + n_j^{-1}) \tilde{\mathbf{S}}$, em que

$$v = n - p - 1$$

$$\bar{\mathbf{x}}_j = \sum_{i=1}^{n_j} \frac{\mathbf{x}_{ji}}{n_j}$$

$$\tilde{\mathbf{S}} = v^{-1} \sum_{j=1}^2 \sum_{i=1}^{n_j} (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j)(\mathbf{x}_{ji} - \bar{\mathbf{x}}_j) = v^{-1}(n-2)\mathbf{S}. \quad (3.2)$$

Com o objetivo de avaliar a discrepância entre duas densidades preditivas, Johnson (1987) usou a medida de divergência de Kullback-Leibler $I(f_i, f_j)$ (Seção B-8 do Apêndice B).

Para as funções densidade f_1 e f_2 , define-se as divergências direcionadas:

$$I(f_i, f_j) = E_{f_i} \ln(f_i f_j^{-1}) = \int_{\mathbb{R}^p} f_i(\mathbf{y}) \ln \left(\frac{f_i(\mathbf{y})}{f_j(\mathbf{y})} \right) d\mathbf{y}.$$

Quando f_i é a distribuição $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, verifica-se (Seção B-8 do Apêndice B) que:

$$2I(f_i, f_j) = (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}_j^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j) + \text{tr} \boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j^{-1} - \ln |\boldsymbol{\Sigma}_i \boldsymbol{\Sigma}_j^{-1}| - p \quad (3.3)$$

Quando f_i for a densidade *t-Student* com η_j graus de liberdade, vetor de locação $\boldsymbol{\alpha}_j$ e matriz de dispersão $\boldsymbol{\lambda}_j$, não é possível calcular $I(f_i, f_j)$ explicitamente; face à complexidade na solução das integrais envolvidas no cálculo, aproxima-se tal densidade por uma densidade normal multivariada apropriada.

Johnson (1987) propõe aproximar a densidade *t-student* por uma densidade normal com parâmetros $\boldsymbol{\mu}_j$ e $\boldsymbol{\Sigma}_j$ de modo a minimizar $I(\text{St}(\eta_j, \boldsymbol{\alpha}_j, \boldsymbol{\lambda}_j), N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j))$. Verifica-se que o ínfimo dessa expressão é atingido para $\boldsymbol{\mu}_j = \boldsymbol{\alpha}_j$ e

$$\boldsymbol{\Sigma}_j = \frac{\eta_j}{\eta_j - 2} \boldsymbol{\lambda}_j.$$

Dessa forma a estimativa normal ótima para a densidade *t-Student* $St(\eta_j, \alpha_j, \lambda_j)$ é

$$\tilde{f}_j = N\left(\alpha_j, \frac{\eta_j}{(\eta_j - 2)} \lambda_j\right)$$

e uma aproximação para $I(f_i, f_j)$ é dada por $\tilde{I}(f_i, f_j) = I(\tilde{f}_i, \tilde{f}_j)$.

3.2 Definição dos objetivos inferenciais

Tomado \mathbf{y} , vetor p -variado observado de origem desconhecida, consideram-se três objetivos inferenciais:

- alocar \mathbf{y} em π_1 ou π_2 ,
- estimar a probabilidade *a posteriori* de que a origem de \mathbf{y} seja π_1 , denotada por $\Pr(\mathbf{y} \in \pi_1 / \mathbf{d}) = p(\mathbf{y})$,
- medir o grau de separação entre as populações.

A correta alocação de \mathbf{y} , bem como a determinação de $p(\mathbf{y})$, depende do grau de separação entre as populações, assim, os três objetivos são relacionados entre si.

Quando o objetivo é alocar \mathbf{y} em π_1 ou π_2 , a regra ótima consiste em alocar \mathbf{y} em π_1 (Seção A-1.2 do Apêndice A) se :

$$o(\mathbf{y}) = \frac{\Pr(\mathbf{y} \in \pi_1 / \mathbf{d})}{\Pr(\mathbf{y} \in \pi_2 / \mathbf{d})} > \frac{c(1/2)}{c(2/1)} \cdot \frac{1-q}{q}$$

em que q é a probabilidade *a priori* de π_1 e $1 - q$ é a probabilidade *a priori* de π_2 .

Verifica-se que,

$$o(\mathbf{y}) = \frac{q}{(1-q)} \cdot \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})}.$$

Quando os parâmetros são conhecidos, utiliza-se

$$\begin{aligned} U(\mathbf{y}) &= \ln(o(\mathbf{y})) = \ln \frac{q}{(1-q)} + \left\{ -\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{y} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}_2) \right\} \\ &= \ln \frac{q}{1-q} + \left\{ -\frac{1}{2} [\mathbf{y}' \boldsymbol{\Sigma}^{-1} \mathbf{y} - \mathbf{y}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_1' \boldsymbol{\Sigma}^{-1} \mathbf{y} + \boldsymbol{\mu}_1' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 + \mathbf{y}' \boldsymbol{\Sigma}^{-1} \mathbf{y} - \mathbf{y}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2 - \boldsymbol{\mu}_2' \boldsymbol{\Sigma}^{-1} \mathbf{y} + \boldsymbol{\mu}_2' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2] \right\} \\ &= \ln \frac{q}{(1-q)} + \left\{ \left(\mathbf{y} - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \right)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \right\} \\ &= \ln \frac{q}{(1-q)} + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2), \end{aligned}$$

que é a função discriminante de Fisher (expressão (A-1.8) do Apêndice A) quando as probabilidades q e $1 - q$ forem iguais.

Quando $\boldsymbol{\theta}$ é desconhecido, os parâmetros presentes em $U(\mathbf{y})$ são substituídos por suas estimativas *plug-in*, de modo que

$$V(\mathbf{y}) = \ln \frac{q}{(1-q)} + \left\{ \left[\mathbf{y} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \right]' \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \right\} \text{ é a estimativa de } U(\mathbf{y}).$$

Ainda quando $\boldsymbol{\theta}$ é desconhecido, usando a *priori* de referência (3.1), calcula-se o logaritmo da razão das probabilidades estimadas (*log odds*) *a posteriori* como:

$$W(\mathbf{y}) = \ln \frac{q}{1-q} + \ln \left(\frac{St(\mathbf{y}/v, \bar{\mathbf{x}}_1, \kappa_1 \tilde{\mathbf{S}})}{St(\mathbf{y}/v, \bar{\mathbf{x}}_2, \kappa_2 \tilde{\mathbf{S}})} \right) \quad (3.4)$$

com $\kappa_1 = 1 + n_1^{-1}$ e $\kappa_2 = 1 + n_2^{-1}$.

Quando o objetivo é prever a origem de uma observação futura, o autor reformula o problema de alocação considerando o vetor aleatório (T, \mathbf{Y}) em que $T \sim \text{Bernoulli}(q)$ e $(\mathbf{Y} / T = t, \boldsymbol{\theta}) \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma})$. Dessa forma, T seria a variável aleatória que indica a origem de \mathbf{Y} .

Uma vez que o objetivo é prever T , encontra-se $\hat{T}(\mathbf{y})$ tal que $E_{(T, \mathbf{Y})} \left\{ (T - \hat{T}(\mathbf{Y}))^2 / \mathbf{d} \right\}$ é mínima, sendo que a esperança é tomada com base na densidade preditiva de (T, \mathbf{Y}) .

Essa quantidade é minimizada para

$$\hat{T} = E[T(\mathbf{y}) / \mathbf{d}] = 0.P(T = 0 / \mathbf{d}) + 1.P(T = 1 / \mathbf{d}) = P(T = 1 / \mathbf{d}) = p(\mathbf{y} \in \pi_1 / \mathbf{d}) = p(\mathbf{y})$$

Assim, o preditor ótimo da origem de uma observação \mathbf{y} é a probabilidade *a posteriori* $p(\mathbf{y})$.

O grau de separação entre populações afeta as regras de alocação e predição. No contexto Bayesiano, uma forma de medir esse grau de separação é medir a discrepância entre as densidades preditivas f_1 e f_2 usando as medidas de divergência de Kullback-Leibler. Johnson (1987) obteve a aproximação:

$$2\tilde{I}(f_1, f_2) = c_1 \hat{\Delta}^2 + c_2$$

que depende somente de $\hat{\Delta}^2$, a menos de constantes.

3.3 Medida de influência global, local e individual

Johnson (1987) definiu os conceitos de medida de influência global, local e individual.

Uma medida de influência é dita global quando avalia o quanto uma observação afeta dois ou mais aspectos da análise. É uma medida de influência definida para as duas densidades preditivas (f_1 e f_2) pois se uma observação tem um grande impacto nessas densidades, presumidamente, irá afetar vários aspectos da análise relacionados aos objetivos inferenciais tais como a alocação e separação.

A medida de influência é local quando mede o efeito de uma observação em um objetivo inferencial específico, tal como por exemplo, alguma estatística de separação.

Já a medida de influência individual é aquela que avalia o efeito da eliminação de determinada observação em quantidades como probabilidades individuais ou logaritmo da razão das probabilidades estimadas (*log-odds*). Este estudo é de muita utilidade porque pode-se observar mudanças muito grandes na estimativa da probabilidade de um elemento pertencer a π_1 depois da eliminação da observação.

Medidas globais

Uma vez que a densidade preditiva $f_j(\mathbf{y} / \mathbf{d}, \pi_j) = \int N(\mathbf{y} / \boldsymbol{\theta}, \pi_j) p(\boldsymbol{\theta} / \mathbf{d}, \pi_j) d\boldsymbol{\theta}$ envolve vários elementos – funções densidade, probabilidades, dados, parâmetros – analisando o efeito que as observações têm sobre a densidade preditiva, obtém-se medidas de influência globais.

O autor mediu o efeito de eliminar a observação \mathbf{x}_i , $i = 1, 2, \dots, n_j$, na j -ésima densidade preditiva, empregando a divergência simétrica de Kullback-Leibler

$$J(f_j, f_{j(i)}) = I(f_j, f_{j(i)}) + I(f_{j(i)}, f_j)$$

em que $f_{j(i)}(\cdot) = f_j(\cdot / \mathbf{x}_{(i)})$ e $f_j(\cdot / \mathbf{x}_{(i)})$ é a densidade *a posteriori* de y dada a amostra sem a observação \mathbf{x}_i . A partir daí, definiu a quantidade $J_i = J(f_1, f_{1(i)}) + J(f_2, f_{2(i)})$, que avalia o efeito que a observação \mathbf{x}_i tem em ambas as densidades preditivas.

Na impossibilidade de calcular $J(f_j, f_{j(i)})$ explicitamente, é necessário usar aproximações. Assim, define-se $\tilde{J}(f_j, f_{j(i)}) = J(\tilde{f}_j, \tilde{f}_{j(i)})$ e $\tilde{J}_{(i)} = \sum_{j=1}^2 \tilde{J}(f_j, f_{j(i)})$.

Como visto anteriormente, se utilizada a *priori* de referência (3.1), a densidade preditiva obtida é $St(\eta_j, \alpha_j, \lambda_j)$, em que

$$\eta_j = n - p - 1,$$

$$\lambda_j = \left(1 + n_j^{-1}\right) \tilde{\mathbf{S}} \text{ e}$$

$$\alpha_j = \bar{\mathbf{x}}_j.$$

Então a aproximação normal ótima para a esta densidade é

$$N\left(\alpha_j, \frac{\eta_j}{(\eta_j - 2)} \lambda_j\right), \text{ sendo que}$$

$$\alpha_j = \bar{\mathbf{x}}_j,$$

e a matriz de covariância é dada por

$$\frac{n-p-1}{n-p-1-2} \cdot \left(1 + \frac{1}{n_j}\right) \tilde{\mathbf{S}} = \frac{n-p-1}{n-p-3} \cdot \left(\frac{n_j+1}{n_j}\right) \frac{n-2}{n-p-1} \mathbf{S} \cong \frac{n-2}{n-p-3} \mathbf{S}.$$

Portanto a densidade *t-student* multivariada será aproximada por \tilde{f}_j , distribuída como $N\left(\bar{\mathbf{x}}_j, \frac{(n-2)}{(n-p-3)}\mathbf{S}\right)$. Esta aproximação é utilizada no cálculo de $J(f_j, f_{j(i)})$ e devido a sua complexidade algébrica, não será aqui exibida. Verifica-se no entanto que $J(\tilde{f}_j, \tilde{f}_{j(i)})$ depende dos dados apenas através da estatística D .

Medidas locais

Johnson (1987) definiu medidas locais relativas a estimação do logaritmo da razão das probabilidades estimadas (*log-odds*), às medidas de separação ($\hat{\Delta}^2$) e aos coeficientes discriminantes de Fisher ($\hat{\mathbf{I}} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1}$ e

$$\hat{m} = \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)' \mathbf{S}^{-1}(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)).$$

Para medir o efeito de uma observação no logaritmo da razão das probabilidades estimadas, o autor considerou a esperança do quociente dessas medidas na presença e ausência da observação. Tal medida, denominada esperança do logaritmo da razão das probabilidades estimadas é definida como:

$$ELO_i = E \ln \left\{ \frac{o(\mathbf{y})}{o_{(i)}(\mathbf{y})} \right\} \quad (3.5)$$

sendo que a esperança é calculada em relação a densidade preditiva com todos os dados,

$$f(\mathbf{y}) = qf_1(\mathbf{y}) + (1-q)f_2(\mathbf{y}) \quad e$$

$$o(\mathbf{y}) = \frac{q}{(1-q)} \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})}.$$

Se a estatística ELO_i for positiva, indica que a propensão para alocar uma futura observação \mathbf{y} em π_1 é menor depois de eliminada a observação i . Por outro

lado, se a estatística ELO_i for negativa, a propensão de alocar y em π_1 é maior depois da exclusão da observação i .

Se usadas aproximações para as densidades preditivas $f(y)$, obtém-se a esperança do logaritmo da razão das probabilidades estimadas “aproximada”:

$$E\tilde{L}O_i = E_q \ln \left\{ \frac{\tilde{o}(\mathbf{Y})}{\tilde{o}_{(i)}(\mathbf{Y})} \right\} . \quad (3.6)$$

Para medir o efeito da observação i no logaritmo da razão das probabilidades estimadas amostral, o autor considera o logaritmo da razão das probabilidades estimadas relativo amostral, definido como :

$$LORAM_i = n^{-1} \sum_{j=1}^2 \sum_{k=1}^{n_j} \ln \left\{ \frac{o(\mathbf{x}_{jk})}{o_{(i)}(\mathbf{x}_{jk})} \right\}, \quad i = 1, \dots, n_1 . \quad (3.7)$$

Verifica-se no entanto que essa medida pode ficar muito próxima do zero, em virtude dos cancelamentos dos efeitos individuais. Sendo assim é mais sensato considerar a medida absoluta do logaritmo da razão das probabilidades estimadas amostral

$$LORAB_i = n^{-1} \sum_{j=1}^2 \sum_{k=1}^{n_j} \left| \ln \left\{ \frac{o(\mathbf{x}_{jk})}{o_{(i)}(\mathbf{x}_{jk})} \right\} \right|, \quad i = 1, \dots, n_1 .$$

Com o objetivo de avaliar o efeito que a observação i tem na alocação das observações amostrais é possível ainda utilizar a função linear discriminante de Fisher ao invés das densidades preditivas. Assim, define-se:

$$\begin{aligned} LORAM_i^F &= n^{-1} \sum_{j=1}^2 \sum_{k=1}^{n_j} FD(\mathbf{x}_{jk}) - FD_{(i)}(\mathbf{x}_{jk}) \quad e \\ LORAB_i^F &= n^{-1} \sum_{j=1}^2 \sum_{k=1}^{n_j} |FD(\mathbf{x}_{jk}) - FD_{(i)}(\mathbf{x}_{jk})| \end{aligned} \quad (3.8)$$

sendo que $FD(\mathbf{x}_{jk}) = \hat{\mathbf{I}}' \mathbf{x}_{jk} + \hat{m}$ e

$$FD_{(i)}(\mathbf{x}_{jk}) = \hat{\mathbf{I}}'_{(i)} \mathbf{x}_{jk} + \hat{m}_{(i)},$$

com $\hat{\mathbf{I}}_{(0)} = (\bar{\mathbf{x}}_{1(0)} - \bar{\mathbf{x}}_2) \mathbf{S}_{(0)}^{-1}$ e $\hat{m}_{(i)} = \frac{1}{2} (\bar{\mathbf{x}}_{1(0)} + \bar{\mathbf{x}}_2) \mathbf{S}_{(0)}^{-1} (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_{1(0)})$.

Efeitos individuais

Depois de identificada a observação que tem influência global ou local, deve-se analisar o seu efeito individual. Se o objetivo de uma análise é a alocação de uma observação y de origem desconhecida ou a determinação da probabilidade de y pertencer a determinada população, pode-se perguntar qual o efeito que essa observação da amostra, por exemplo \mathbf{x}_i , tem na determinação de $p(y)$, $o(y)$ e também na função linear discriminante de Fisher. Calcula-se então esses três valores para o conjunto de dados sem a observação i , e compara-se com os valores obtidos quando calculados para o conjunto inteiro. Se as diferenças forem grandes pode-se afirmar que \mathbf{x}_i é influente em relação àquele objetivo inferencial.

3.4 Exemplo

Johnson (1987) considerou o conjunto de dados de Johnson e Wichern (1992) pg 564, em que foram tomadas quatro variáveis relativas a 21 empresas dois anos antes de sua falência (π_0), e para 25 empresas sólidas financeiramente (π_1).

As variáveis são:

X_1 = fluxo de caixa / total dos débitos

X_2 = lucro bruto / ativo total

X_3 = ativo circulante / passivo circulante

X_4 = ativo circulante / vendas brutas

mas a variável X_4 foi excluída em virtude do seu pequeno efeito na inferência.

O autor empregou $q_0 = 1 - q_0 = \frac{1}{2}$, que produz resultados mais comparáveis com os clássicos. No entanto, intuitivamente sabe-se que q_0 é muito menor que $\frac{1}{2}$, assim posteriormente, foi utilizado $q_0 = 0,05$ no cálculo da medida ELO_i .

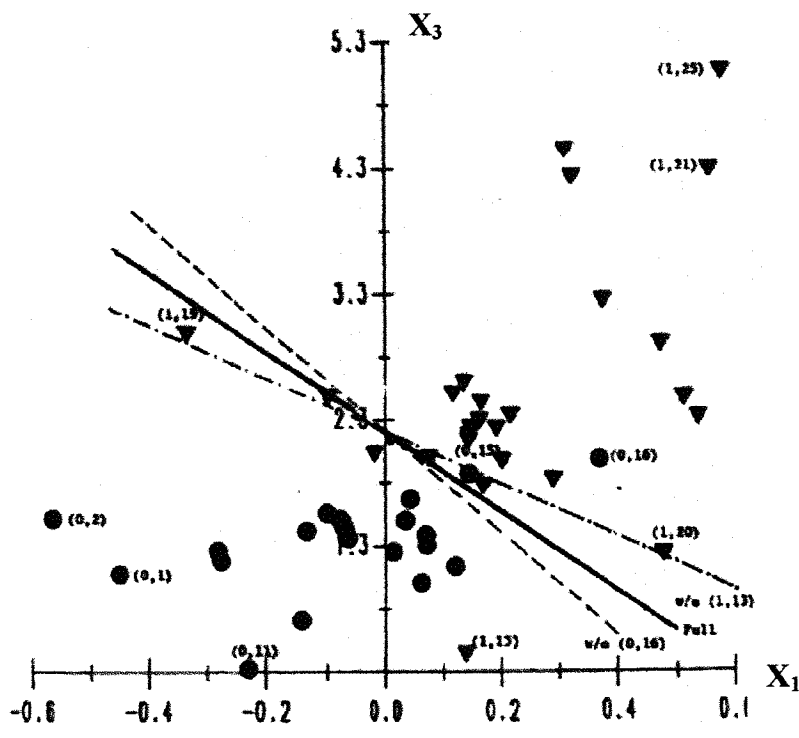


Figura 3.1 – Gráfico para as variáveis X_3 e X_1 ; círculo denota observação pertencente a π_0 e triângulo a π_1 . A linha cheia representa a reta discriminante com base no conjunto inteiro dos dados, enquanto que a linha tracejada é a reta discriminante sem o caso (0,16) e a linha com ponto e traços é a reta discriminante sem a observação (1,13).

Fonte: Figura 1- Johnson (1987).

A Figura 3.1 mostra o gráfico com as variáveis X_1 e X_3 (pode-se usar também X_1 e X_2 , ou X_2 e X_3). Usando a notação (l, k) para indicar que a observação se refere ao k -ésimo dado da população l , nota-se que a observação $(0, 16)$ pertence a π_0 , mas está alocada em π_1 e a observação $(1, 13)$ pertence a π_1 mas está alocada em π_0 . Já os dados $(0, 1)$ e $(1, 25)$, embora alocados corretamente, estão muito distantes da média de seu grupo.

Na Figura 3.2 (a-c) são exibidos os gráficos das observações *versus* as medidas \tilde{J}_i , $\tilde{E}L\tilde{O}_i$, e $LORAB_i^F$ para as empresas que faliram, enquanto a Figura 3.3 (a-c) mostra os mesmos gráficos para as empresas sólidas.

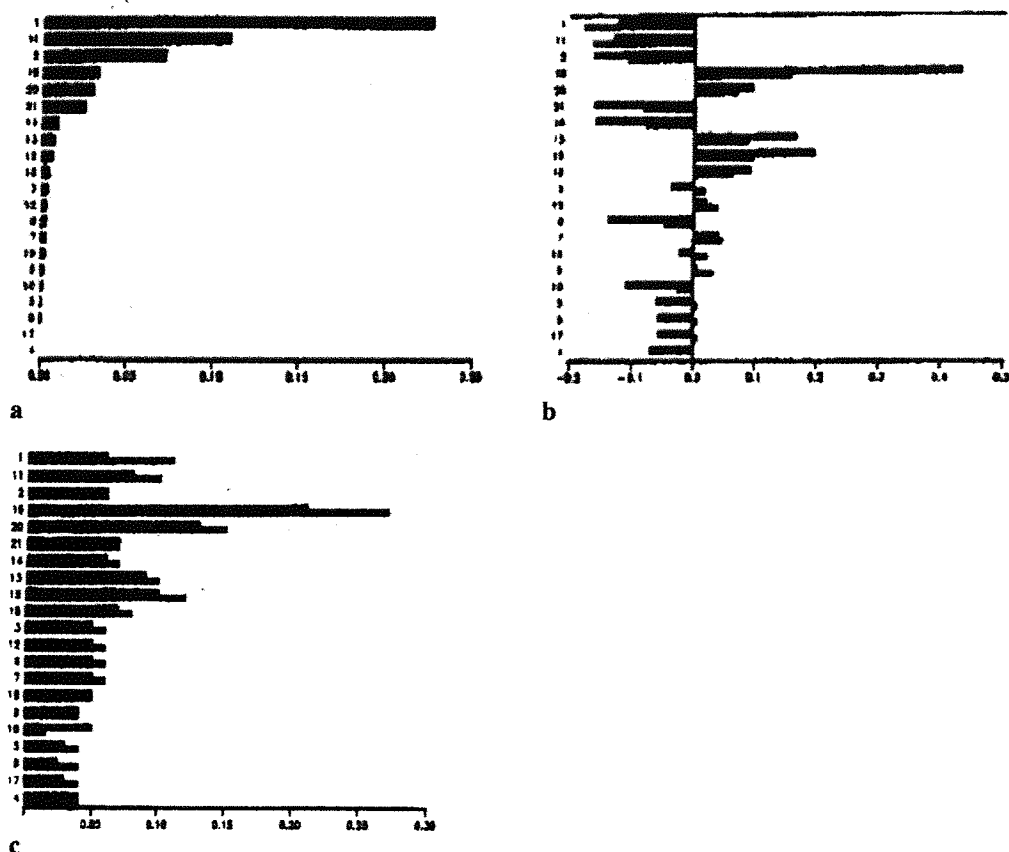


Figura 3.2 - Gráfico das medidas para as empresas falidas a: \tilde{J}_i , b: $\tilde{E}L\tilde{O}_i$ com $q_0 = 0,05$ e c: $LORAB_i^F$
 Fonte: Figura 2 (a,b e d) - Johnson (1987)

Analisando-se a medida \tilde{J}_i , verifica-se que as observações mais influentes são (0, 1), (0, 11), (1, 25) e (1, 19).

Quanto à medida $E\tilde{L}O_i$ (com $q_0 = 0,05$), observa-se que os dados (0, 16), (0, 1), (0, 11), (1, 25), (1, 13) e (1, 21) são os mais influentes. Notou-se que as observações que são influentes para $q_0 = 1/2$ também o são para $q_0 = 0,05$, embora os efeitos sejam maiores para este último. Na observação (1, 13), quando da mudança no valor de q_0 , muda também o sinal da medida de influência.

Nas Figuras 3.2-c e 3.3-c, que mostram os valores de $LORAB_i^F$, nota-se que as observações mais influentes são (0, 16), (0, 20), (1, 13), (1, 19) e (1, 25).

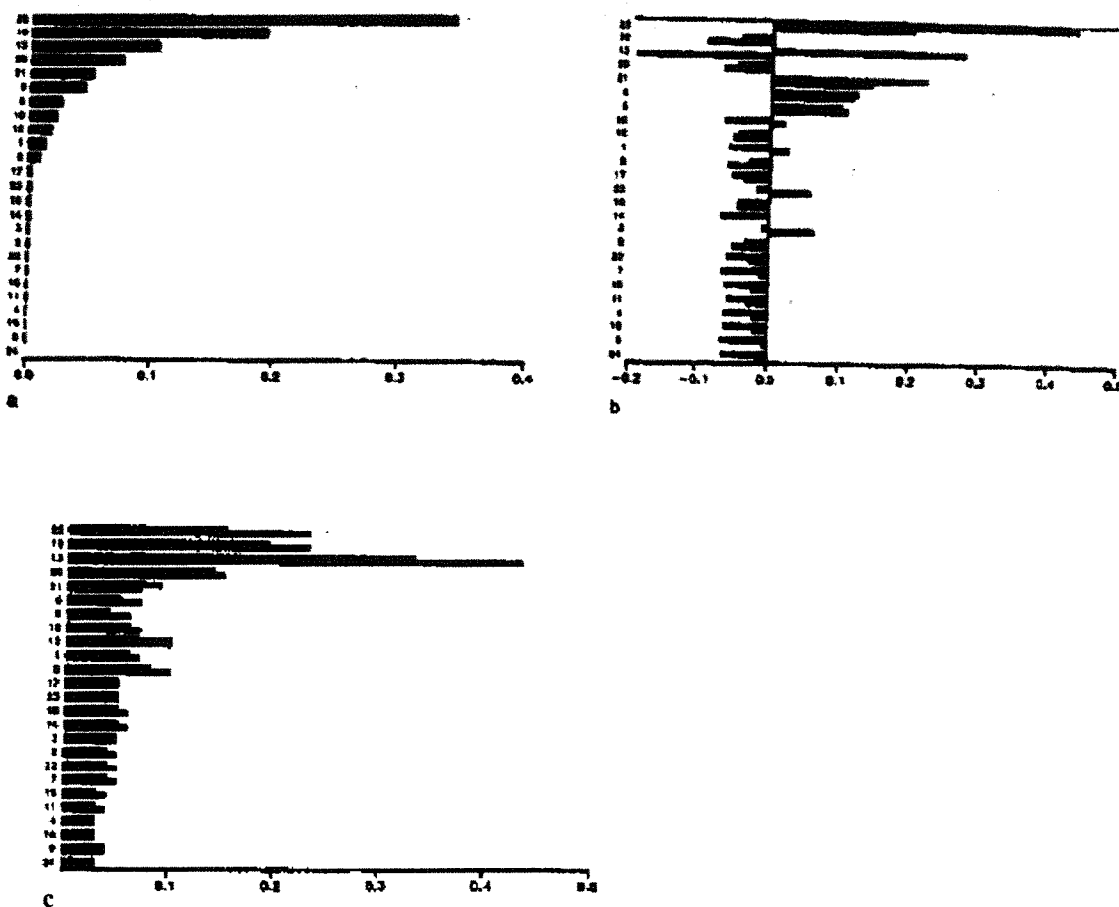


Figura 3.3 - Gráfico das medidas para as empresas financeiramente sólidas, a: J_i , b: $E\tilde{L}O_i$ com $q_0 = 0,05$ e c: $LORAB_i^F$.

Fonte: Figura 3 (a,b e d) - Johnson (1987)

O autor analisou as observações (0 , 1), (0 , 16) , (1 , 13) e (1 , 25) , uma vez que elas aparecem como influentes em mais de uma medida.

As observações (0 , 1) e (1 , 25) são as mais distantes de suas médias. Já as observações (0 , 16) e (1 , 13) são as mais influentes quanto a efeitos locais (Figura 3.2 b e c, e 3.3 b e c) mas nenhuma delas está distante da média de seu grupo embora estejam alocadas incorretamente.

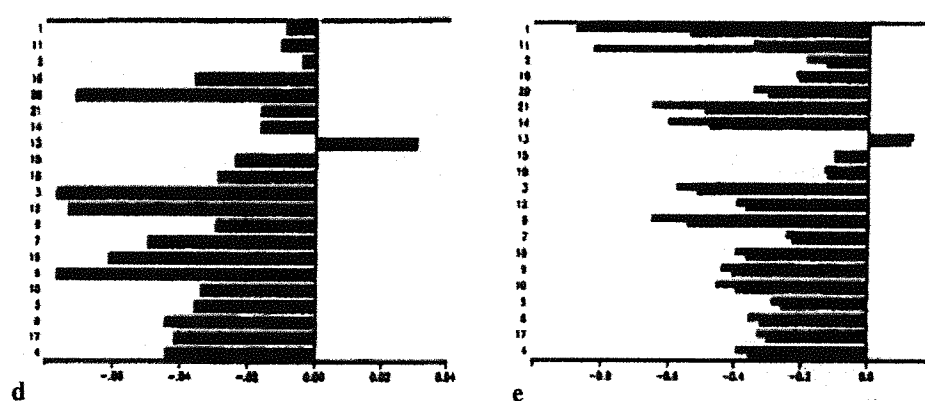


Figura 3.4 – Gráfico para as empresas falidas

d: diferenças de probabilidade com e sem a observação (1,13) ($p - p_{(1,13)}$)

e: diferenças dos escores discriminantes com e sem a observação (1,13) ($FD - FD_{(1,13)}$)

Fonte: Figura 2 (e-f)– Johnson (1987)

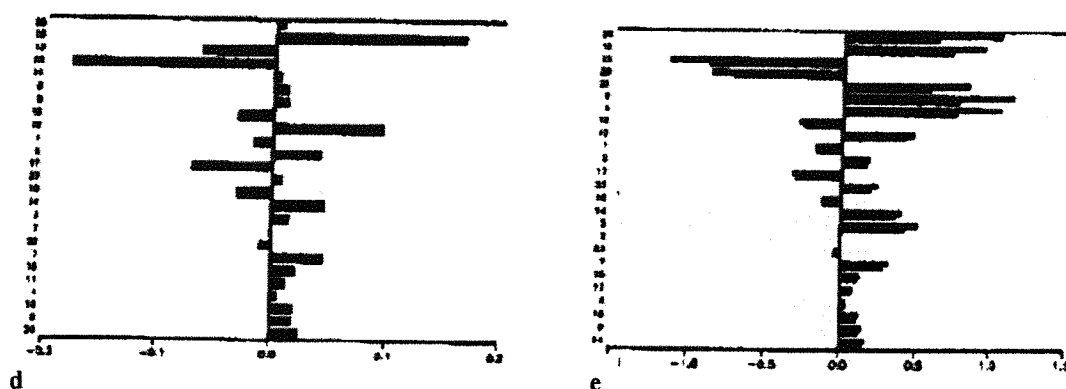


Figura 3.5 – Gráfico para as empresas financeiramente sólidas

d: diferenças de probabilidade com e sem a observação (1,13) ($p - p_{(1,13)}$)

e: diferenças dos escores discriminantes com e sem a observação (1,13) ($FD - FD_{(1,13)}$)

Fonte: Figura 3(e-f) – Johnson (1987)

Obtendo $E\tilde{L}O_{(0,16)} > 0$ para $q_0 = \frac{1}{2}$, a eliminação de (0, 16) implicou em uma maior propensão para alocar uma observação futura em π_1 . Já com $E\tilde{L}O_{(1,13)} < 0$, a retirada de (1, 13) resulta em maior chance de alocar a observação futura em π_0 .

Com a remoção da observação (1, 13), a probabilidade de alocação no grupo das empresas falidas aumenta para todas as firmas falidas (π_0), exceto para a observação (0, 13) (Figura 3.4). Aumenta também a probabilidade de alocação em falidas para sete das empresas sólidas (Figura 3.5), enquanto que para as demais observações essa probabilidade diminuiu. Assim, o efeito de (1, 13) nas empresas sólidas é maior do que nas empresas falidas.

Nota-se que as observações que são mais influentes com relação à medida \tilde{J}_i não são as mais influentes de acordo com as medidas locais.

O autor finaliza concluindo que medidas baseadas apenas na distância de Mahalanobis não são suficientes para uma análise de influência mais aprofundada.

Capítulo 4

Comparações entre algumas medidas de diagnóstico

Johnson (1987) construiu medidas de diagnóstico calculando a esperança (com relação à densidade preditiva de x) da diferença dos logaritmos da razão das probabilidades estimadas (*log odds*) de x baseados na amostra inteira e na amostra sem a observação i (Seção 3.3, expressão (3.5)).

Utilizando uma abordagem similar, Fung (1996-b) considerou medidas de diagnóstico condicionais ao fato que a observação x está próxima do plano discriminante. Tal estudo tem interesse porque em análise discriminante esse tipo de observação é difícil de classificar, despertando maior atenção.

Este capítulo tem por objetivo comparar medidas obtidas através de abordagens diferentes pelos dois autores, mas que produzem resultados semelhantes. São comparadas também estatísticas obtidas por um mesmo autor, Fung (1992 e 1998), que, calculadas de formas distintas, fornecem resultados similares.

Como visto nos capítulos anteriores, admite-se que a observação retirada pertence à primeira população, e tomando o logaritmo da razão das probabilidades estimadas definido na Seção 2.3

$$\hat{I} \{ \mathbf{x} - (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) / 2 \} - \log \left\{ \frac{1-q}{q} \right\}$$

o efeito da observação i nesta quantidade pode ser estudado usando a medida

$$DLO_i(\mathbf{x}) = \hat{I} \{ \mathbf{x} - (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) / 2 \} - \hat{I}_{(0)} \{ \mathbf{x} - (\bar{\mathbf{x}}_{1(0)} + \bar{\mathbf{x}}_2) / 2 \} \quad (\text{Seção 2.5}). \quad (4.1)$$

Observa-se que essa diferença do logaritmo da razão das probabilidades estimadas independe da probabilidade a priori q .

As medidas propostas por Johnson (1987), ELO_i e $LORAB_i$ ((3.5) e (3.8)) são basicamente $E(DLO_i)$ e $E(|DLO_i|)$ nessa ordem, sendo que as esperanças são tomadas em relação à densidade preditiva para \mathbf{x} e definidas incondicionalmente.

Complementando esse estudo, Fung (1996-b) propôs a medida condicional à situação em que a observação \mathbf{x} está próxima do plano discriminante, definida como

$$E2c_i = E [DLO_i^2 / \lambda(\mathbf{x}) = \hat{I} \{ \mathbf{X} - (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) / 2 \} = k] \quad (4.2)$$

e representa a média do quadrado da diferença do logaritmo da razão das probabilidades estimadas quando \mathbf{x} permanece no plano discriminante, ou seja

$$\lambda(\mathbf{x}) = k = \log \left\{ \frac{1-q}{q} \right\}.$$

Variando k em $\lambda(\mathbf{x})$ tem-se planos paralelos ao plano discriminante. Pode-se considerar também $k = \lambda(\bar{\mathbf{x}}_j)$, $j = 1, 2$. Além disso, se $\alpha(q) = \log \frac{1-q}{q}$, tomando $k = \alpha(q)$ com diferentes valores de q , $k = \alpha(0,5) = 0$ é de grande interesse, pois

$\lambda(x) = 0$ corresponde à função discriminante de Fisher com probabilidades *a priori* iguais. Portanto, as medidas condicionais para observações próximas do plano discriminante podem ser calculadas através de $E2c_i$ com escolhas adequadas para k .

Para $\mathbf{x} \in \pi_j, j = 1, 2$, utilizando-se a distribuição $N(\bar{\mathbf{x}}_j, \mathbf{S})$, calcula-se a média quadrática condicional (4.2) do seguinte modo:

$$E2c_i = E \left[DLO_i^2 / \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} = k \right] = Var \left[DLO_i / \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} = k \right] + E^2 \left[DLO_i / \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} = k \right].$$

O vetor aleatório $(DLO_i, \hat{\mathbf{I}}' \{ \mathbf{X} - (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)/2 \})$ tem distribuição normal bivariada, pois é combinação linear de \mathbf{x} , que tem distribuição normal multivariada. Como consequência, usando novamente as estimativas *plug-in* $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2$ e \mathbf{S} para os parâmetros $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ e $\boldsymbol{\Sigma}$, segue que

$$Var \left[DLO_i / \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} = k \right] = \hat{\Sigma}_{11} - \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21}$$

com $\hat{\Sigma}_{11} = Var(DLO_i)$,

$$\hat{\Sigma}_{12} = Cov \left(DLO_i, \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} \right)$$

$$\hat{\Sigma}_{22} = Var \left(\hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} \right) \text{ e}$$

$$\hat{\Sigma}_{21} = \hat{\Sigma}_{12}'.$$

De (2.29) sabe-se que $Var(DLO_i) = (\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)})' \mathbf{S} (\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)}) = \hat{\Sigma}_{11}$.

Adicionalmente,

$$\begin{aligned}
\hat{\Sigma}_{12} &= Cov\left(DLO_i, \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\}\right) \\
&= Cov\left(\hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} - \hat{\mathbf{I}}_{(0)}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_{1(0)} + \bar{\mathbf{x}}_2)}{2} \right\}, \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\}\right) \\
&= Cov\left(\hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\}, \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\}\right) - Cov\left(\hat{\mathbf{I}}_{(0)}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_{1(0)} + \bar{\mathbf{x}}_2)}{2} \right\}, \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\}\right) \\
&= \hat{\mathbf{I}}' \hat{\mathbf{S}} \hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)}' \mathbf{S} \hat{\mathbf{I}} = (\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)})' \mathbf{S} \hat{\mathbf{I}} \quad e \\
\hat{\Sigma}_{22} &= Var\left(\hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\}\right) = \hat{\mathbf{I}}' \hat{\mathbf{S}} \hat{\mathbf{I}} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \mathbf{S} \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = D^2.
\end{aligned}$$

Portanto,

$$Var\left[DLO_i / \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} = k\right] = (\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)})' \mathbf{S} (\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)}) - \frac{\left((\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)})' \mathbf{S} \hat{\mathbf{I}}\right)^2}{D^2}.$$

Do mesmo modo que na Seção 2.5, admitiu-se que \mathbf{X} é distribuída como $w_1 N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) + w_2 N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$, utilizando-se as estimativas *plug-in* para $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ e $\boldsymbol{\Sigma}$. Observa-se que $\bar{\mathbf{x}}_1$, $\bar{\mathbf{x}}_2$ e \mathbf{S} são considerados constantes neste cálculo e neste caso, admitiu-se $w_1 = w_2 = 1/2$. Como conseqüência, $E(\mathbf{X}) = \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2}$, e a partir daí,

$$E\left(DLO_i^2 / \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} = k\right) = E 2c_i.$$

Tem-se que

$$E\left(DLO_i / \hat{\mathbf{I}}' \left\{ \mathbf{X} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right\} = k\right) = \mu_1 + \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} (k - \mu_2)$$

$$\begin{aligned} \text{com } \mu_1 &= E(DLO_i) = E\left(\hat{\mathbf{i}}_i \mathbf{x} - \hat{\mathbf{i}}_i \frac{(\bar{x}_1 + \bar{x}_2)}{2} - \hat{\mathbf{i}}_{(0)} \mathbf{x} + \hat{\mathbf{i}}_{(0)} \frac{(\bar{x}_{1(0)} + \bar{x}_2)}{2}\right) \\ &= \hat{\mathbf{i}}_i \frac{(\bar{x}_1 + \bar{x}_2)}{2} - \hat{\mathbf{i}}_i \frac{(\bar{x}_1 + \bar{x}_2)}{2} - \hat{\mathbf{i}}_{(0)} \frac{(\bar{x}_1 + \bar{x}_2)}{2} + \hat{\mathbf{i}}_{(0)} \frac{(\bar{x}_{1(0)} + \bar{x}_2)}{2} = \frac{\hat{\mathbf{i}}_{(0)} (\bar{x}_{1(0)} - \bar{x}_1)}{2} \end{aligned}$$

$$\text{e } \mu_2 = E\left[\hat{\mathbf{i}}_i \mathbf{x} - \hat{\mathbf{i}}_i \frac{(\bar{x}_1 + \bar{x}_2)}{2}\right].$$

Se $E(\mathbf{X}) = \frac{(\bar{x}_1 + \bar{x}_2)}{2}$, então $\mu_2 = 0$ e com isto,

$$E\left(DLO_i / \hat{\mathbf{i}}_i \left\{ \mathbf{x} - \frac{(\bar{x}_1 + \bar{x}_2)}{2} \right\} = k\right) = \frac{\hat{\mathbf{i}}_{(0)} (\bar{x}_{1(0)} - \bar{x}_1)}{2} + \frac{(\hat{\mathbf{i}}_i - \hat{\mathbf{i}}_{(0)}) \mathbf{s} \hat{\mathbf{i}}_i}{D^2}$$

e $E2c_i = B_i^2 + V_i$ em que

$$B_i = \frac{\hat{\mathbf{i}}_{(0)} (\bar{x}_{1(0)} - \bar{x}_1)}{2} + \frac{k(\hat{\mathbf{i}}_i - \hat{\mathbf{i}}_{(0)}) \mathbf{s} \hat{\mathbf{i}}_i}{D^2}$$

$$V_i = (\hat{\mathbf{i}}_i - \hat{\mathbf{i}}_{(0)}) \mathbf{s} (\hat{\mathbf{i}}_i - \hat{\mathbf{i}}_{(0)}) - \frac{\{(\hat{\mathbf{i}}_i - \hat{\mathbf{i}}_{(0)}) \mathbf{s} \hat{\mathbf{i}}_i\}^2}{D^2}.$$

Grandes valores da medida $E2c_i$ indicam que a observação i é influente.

Essa medida também pode ser escrita em função das duas estatísticas fundamentais $\hat{\phi}_i$ e d_i^2 , conforme demonstrado nas Seções B-10 e B-11 do Apêndice B.

As medidas de diagnóstico propostas na Seção 3.3 são calculadas no contexto Bayesiano com base em distribuições preditivas aproximadas pela distribuição $N\left(\bar{\mathbf{x}}_j, \frac{(n-2)}{(n-p-3)}\mathbf{S}\right)$. Já na análise clássica, a distribuição usada é $N(\bar{\mathbf{x}}_j, \mathbf{S})$ $j = 1, 2$. Entretanto, essa pequena alteração nas matrizes de covariância tem pouco efeito nas medidas propostas, em virtude dos cancelamentos quando se calcula a diferença dos logaritmos da razão das probabilidades estimadas. Portanto, na prática, as duas abordagens fornecem resultados similares.

Equivalência assintótica entre $DPCI_i$ e $E2c_i^0$

Partindo do fato que a maioria das medidas de diagnóstico em análise discriminante podem ser escritas em função das estatísticas fundamentais $\hat{\phi}_i$ e d_i^2 , nesta seção, mostra-se que as medidas $DPCI_i$ (Seção 2.4, expressão (2.22)) e $E2c_i$ (expressão (4.2)) são assintoticamente equivalentes.

Como já foi visto, a probabilidade de classificação incorreta de uma regra de alocação é de fundamental importância na análise discriminante. Fung (1992) estudou o efeito da observação i , depois de sua eliminação, na estimativa dessa probabilidade, obtendo a medida $DPCI_i$:

$$DPCI_i = \left\{ \left[\Phi\left[\frac{-A_i - B_i}{2G_i}\right] + \Phi\left[\frac{-A_i + B_i}{2G_i}\right] \right] / 2 \right\} - \Phi(-D/2) \quad (4.3)$$

em que $A_i = \hat{\mathbf{I}}_{(0)}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$, $B_i = \hat{\mathbf{I}}_{(0)}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})$ e $G_i^2 = \hat{\mathbf{I}}_{(0)}' \mathbf{S} \hat{\mathbf{I}}_{(0)}$.

Para as observações próximas do plano discriminante, que de modo geral, são de difícil classificação, Fung (1996-b) propôs a medida de diagnóstico $E2c_i^0$:

$$\begin{aligned}
E2c_i^0 &= E\left\{DLO_i^2 / \hat{\mathbf{I}} \left[\mathbf{x} - \frac{(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)}{2} \right] = 0 \right\} \\
&= \left[\hat{\mathbf{I}}_{(0)} \frac{(\bar{\mathbf{x}}_{1(0)} - \bar{\mathbf{x}}_1)}{2} \right]^2 + (\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)})' \mathbf{S} (\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)}) - \frac{[(\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)})' \mathbf{S} \hat{\mathbf{I}}]^2}{D^2}. \tag{4.4}
\end{aligned}$$

Nota-se das expressões (4.3) e (4.4) que as medidas dependem da observação i somente através de $\bar{\mathbf{x}}_{1(0)}$ e $\hat{\mathbf{I}}_{(0)} = \mathbf{S}_{(0)}^{-1}(\bar{\mathbf{x}}_{1(0)} - \bar{\mathbf{x}}_2)$.

Após os cálculos B-10 e B-11 do Apêndice B mostra-se que:

$$E2c_i^0 = \left[chf_i(\hat{\phi}_i - hd_i^2)/(2b) \right]^2 + c^2 f_i^2(\hat{\phi}_i + h/b)^2 (d_i^2 - \hat{\phi}_i^2 / D^2).$$

Substituindo pelas expressões do Cálculo B-5 do Apêndice B :

$$\begin{aligned}
E2c_i^0 &= \left[\frac{n-3}{n-2} \cdot \frac{1}{n_1-1} \cdot \frac{\frac{-n_1}{(n_1-1)(n-2)}}{1 - \frac{n_1}{(n_1-1)(n-2)} d_i^2} \cdot \left(\hat{\phi}_i - \frac{1}{n_1-1} d_i^2 \right) / \frac{-2n_1}{(n_1-1)(n-2)} \right]^2 + \\
&+ \left(\frac{n-3}{n-2} \right)^2 \cdot \left(\frac{\frac{-n_1}{(n_1-1)(n-2)}}{1 - \frac{n_1}{(n_1-1)(n-2)} d_i^2} \right)^2 \cdot \left(\hat{\phi}_i + \left(\frac{1}{n_1-1} / \frac{-n_1}{(n_1-1)(n-2)} \right) \right)^2 \cdot (d_i^2 - \hat{\phi}_i^2 / D^2).
\end{aligned}$$

Assintoticamente, tomando n_1 e n tendendo para infinito na mesma "velocidade", ou seja, de modo que $\frac{n_1}{n}$ tende para w_1 , segue que

$$E2c_i^0 \cong \left[\frac{1}{n_1} \cdot \frac{-w_1}{n_1} \left(\hat{\phi}_i \cdot \frac{-n_1}{2w_1} \right) \right]^2 + \left(\frac{-w_1}{n_1} \right)^2 \cdot \left(\hat{\phi}_i - \frac{1}{w_1} \right)^2 \cdot (d_i^2 - \hat{\phi}_i^2 / D^2)$$

$$\begin{aligned}
&= \left(\frac{\hat{\phi}_i}{2n_1} \right)^2 + \frac{w_1^2}{n_1^2} \cdot \left(\hat{\phi}_i - \frac{1}{w_1} \right)^2 \left(d_i^2 - \hat{\phi}_i^2 / D^2 \right) \\
&= \left(\frac{\hat{\phi}_i^2}{4n_1^2} \right) + \frac{w_1^2}{n_1^2} \cdot \left(\frac{\hat{\phi}_i^2 w_1^2 - 2\hat{\phi}_i w_1 + 1}{w_1^2} \right) \left(d_i^2 - \hat{\phi}_i^2 / D^2 \right) \\
&= \left[\frac{\hat{\phi}_i^2}{4} + (1 - w_1 \hat{\phi}_i)^2 \left(d_i^2 - \hat{\phi}_i^2 / D^2 \right) \right] / n_1^2.
\end{aligned}$$

Como visto na Seção 2.4, Fung (1992) usou expansão em série de Taylor até segunda ordem em torno de $(-1/2D)$, para a função de distribuição normal padrão $\Phi(\cdot)$ obtendo, assintoticamente, a medida $DPCI_i$

$$DPCI_i = \frac{\Phi(-D/2)}{4D(n_1 - 1)^2} \left[(1 - w_1 \hat{\phi}_i)^2 \left(d_i^2 - \hat{\phi}_i^2 / D^2 \right) + \hat{\phi}_i^2 / 4 \right].$$

Está claro que $DPCI_i$ e $E2c_i^0$ são assintoticamente proporcionais e suas distribuições assintóticas podem ser obtidas com base nos resultados dados na Seção 2.3.1 sobre as distribuições de $\hat{\phi}_i$ e $(d_i^2 - \hat{\phi}_i^2 / D^2)$.

A relação entre as medidas $DPCI_i$ e $E2c_i^0$ foi analisada por Fung (1998) em dois conjuntos de dados. No primeiro, os tamanhos de amostras eram $n_1 = 19$ e $n_2 = 20$ e o coeficiente de correlação entre as medidas resultou em 0,981. No segundo conjunto, $n_1 = n_2 = 50$ e o coeficiente de correlação entre $DPCI_i$ e $E2c_i^0$ é igual a 1. Estes resultados mostram a forte relação linear entre as medidas, tanto para amostras pequenas, quanto para amostras de tamanhos moderados.

Com essa comparação entre as medidas que fornecem resultados similares, encerra-se o estudo de diagnóstico em análise discriminante para duas populações com matrizes de covariância iguais.

Capítulo 5

Observações influentes em análise discriminante quadrática

Neste capítulo, são propostas várias medidas de diagnóstico em análise discriminante quadrática, caso em que as matrizes de covariância são diferentes (Seção A-1.3-b do Apêndice A). Tais medidas envolvem estimativas das funções densidade de probabilidades, da probabilidade de uma observação pertencer a determinada população e do logaritmo da razão das probabilidades estimadas (*log odds*).

Partindo de duas populações normais p -variadas com vetor de médias μ_j e matriz de covariância Σ_j , e considerando amostras aleatórias dessas duas populações, sejam \bar{x}_j , $\hat{\Sigma}_j$, $j = 1, 2$ os estimadores não viciados dos vetores de médias e matrizes de covariância. A estimativa da densidade normal correspondente à população j é

$$\hat{f}_j(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_j|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_j)' \hat{\Sigma}_j^{-1} (\mathbf{x} - \bar{\mathbf{x}}_j)\right]. \quad (5.1)$$

Assumindo, como nas seções anteriores, probabilidades *a priori* e custos de classificação incorreta iguais, a regra discriminante quadrática aloca uma observação \mathbf{x}_0 (de origem desconhecida) para o grupo com maior estimativa da densidade, ou seja aloca \mathbf{x}_0 em π_1 se

$$(\mathbf{x}_0 - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_1) + \log |\hat{\Sigma}_1| < (\mathbf{x}_0 - \bar{\mathbf{x}}_2)' \hat{\Sigma}_2^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_2) + \log |\hat{\Sigma}_2| \quad (5.2)$$

e caso contrário, aloca \mathbf{x}_0 em π_2 .

Na análise discriminante quadrática, considerando duas populações normais multivariadas, uma medida de “atipicalidade” é a distância de Mahalanobis de uma observação com relação à média do seu grupo:

$$d_i^2 = (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1) \quad i = 1, \dots, n_1,$$

tomada para a população 1. De forma similar, seria definida a medida para elementos da população 2.

Assim, uma observação com um grande valor de d_i^2 é considerada *outlier* e pode ser influente em análise discriminante quadrática.

Uma vez que a regra discriminante quadrática (5.2) depende da função densidade (5.1), analisa-se o efeito das observações nas densidades estimadas.

Considerando-se a eliminação da observação i do grupo 1, têm-se as estimativas $\bar{\mathbf{x}}_{1(i)}$ e $\hat{\Sigma}_{1(i)}$, sem essa observação, produzindo a densidade $\hat{f}_{1(i)}$, normal multivariada com média $\bar{\mathbf{x}}_{1(i)}$ e matriz de covariância $\hat{\Sigma}_{1(i)}$. Nesta análise, as estimativas de parâmetros da população 2 não são afetadas.

Segundo Johnson (1987), Fung (1996-a) propõe o estudo da influência da observação i através da medida de divergência de Kullback-Leibler :

$$\begin{aligned} J(\hat{f}_1, \hat{f}_{1(i)}) &= E_{\hat{f}_1} \ln \left(\frac{\hat{f}_1(\mathbf{x})}{\hat{f}_{1(i)}(\mathbf{x})} \right) + E_{\hat{f}_{1(i)}} \ln \left(\frac{\hat{f}_{1(i)}(\mathbf{x})}{\hat{f}_1(\mathbf{x})} \right) \\ &= \int_{\mathfrak{R}} \hat{f}_1(\mathbf{x}) \ln \left(\frac{\hat{f}_1(\mathbf{x})}{\hat{f}_{1(i)}(\mathbf{x})} \right) d\mathbf{x} + \int_{\mathfrak{R}} \hat{f}_{1(i)}(\mathbf{x}) \ln \left(\frac{\hat{f}_{1(i)}(\mathbf{x})}{\hat{f}_1(\mathbf{x})} \right) d\mathbf{x} . \end{aligned}$$

Com base na Seção B-8 do Apêndice B,

$$\begin{aligned} J(\hat{f}_1, \hat{f}_{1(i)}) &= \{(\bar{\mathbf{x}}_{1(i)} - \bar{\mathbf{x}}_1) \hat{\Sigma}_{1(i)}^{-1} (\bar{\mathbf{x}}_{1(i)} - \bar{\mathbf{x}}_1) + tr \hat{\Sigma}_1 \hat{\Sigma}_{1(i)}^{-1} - \ln |\hat{\Sigma}_1 \hat{\Sigma}_{1(i)}^{-1}| - p\} / 2 + \\ &+ \{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(i)}) \hat{\Sigma}_1^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(i)}) + tr \hat{\Sigma}_{1(i)} \hat{\Sigma}_1^{-1} - \ln |\hat{\Sigma}_{1(i)} \hat{\Sigma}_1^{-1}| - p\} / 2 \\ &= \{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(i)}) (\hat{\Sigma}_1^{-1} + \hat{\Sigma}_{1(i)}^{-1}) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(i)}) + tr \hat{\Sigma}_1 \hat{\Sigma}_{1(i)}^{-1} + tr \hat{\Sigma}_{1(i)} \hat{\Sigma}_1^{-1} - 2p - \ln |\hat{\Sigma}_1 \hat{\Sigma}_{1(i)}^{-1}| - \ln |\hat{\Sigma}_{1(i)} \hat{\Sigma}_1^{-1}| \} / 2 . \end{aligned}$$

$$\begin{aligned} \text{Mas } & -\ln |\Sigma_1 \Sigma_{1(i)}^{-1}| - \ln |\Sigma_{1(i)} \Sigma_1^{-1}| = -\ln (\Sigma_1 | \Sigma_{1(i)}^{-1}) - \ln (\Sigma_{1(i)} | \Sigma_1^{-1}) \\ &= -(\ln |\Sigma_1| + \ln |\Sigma_{1(i)}^{-1}|) - (\ln |\Sigma_{1(i)}| + \ln |\Sigma_1^{-1}|) = -(\ln |\Sigma_1| + \ln |\Sigma_1^{-1}|) - (\ln |\Sigma_{1(i)}^{-1}| + \ln |\Sigma_{1(i)}|) \\ &= -\ln (\Sigma_1 | \Sigma_1^{-1}) - \ln (\Sigma_{1(i)}^{-1} | \Sigma_{1(i)}) = -\ln |\Sigma_1 \Sigma_1^{-1}| - \ln |\Sigma_{1(i)}^{-1} \Sigma_{1(i)}| = \ln |\mathbf{I}| - \ln |\mathbf{I}| = -\ln 1 - \ln 1 = 0 \end{aligned}$$

$$\text{então } J(\hat{f}_1, \hat{f}_{1(i)}) = \{(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(i)}) (\hat{\Sigma}_1^{-1} + \hat{\Sigma}_{1(i)}^{-1}) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(i)}) + tr \hat{\Sigma}_1 \hat{\Sigma}_{1(i)}^{-1} + tr \hat{\Sigma}_{1(i)} \hat{\Sigma}_1^{-1} - 2p\} / 2 .$$

No entanto, essa medida não leva em conta a estrutura da análise discriminante. O autor propõe o uso de medidas que levem em conta essa estrutura, definidas nos cálculos a seguir.

A probabilidade estimada que uma observação \mathbf{x} pertença ao grupo j é

$$\hat{p}_j(\mathbf{x}) = \frac{\hat{f}_j(\mathbf{x})}{\hat{f}_1(\mathbf{x}) + \hat{f}_2(\mathbf{x})}, \quad \text{para } j = 1, 2..$$

Estuda-se o efeito da observação i nessa probabilidade através do cálculo de $E_{\mathbf{x}}[\hat{p}_j(\mathbf{x}) - \hat{p}_{j(i)}(\mathbf{x})]^2$ $j = 1, 2$, em que $\hat{p}_{j(i)}(\mathbf{x})$ é a probabilidade estimada que uma observação \mathbf{x} pertença ao grupo j depois de eliminada a observação i . Calculando-se essa esperança com base na distribuição empírica, define-se a medida de diagnóstico de probabilidade quadrática relativa :

$$PQR_i = \frac{1}{n} \sum_{j=1}^2 \sum_{k=1}^{n_j} [\hat{p}_1(\mathbf{x}_{jk}) - \hat{p}_{1(i)}(\mathbf{x}_{jk})]^2, \quad i = 1, 2, \dots, n_1.$$

Para medir a influência da i -ésima observação no logaritmo da razão das probabilidades estimadas, a medida proposta pelo autor é:

$$LOQR_i = \frac{1}{n} \sum_{j=1}^2 \sum_{k=1}^{n_j} \left[\log \left\{ \frac{\hat{p}_1(\mathbf{x}_{jk})}{\hat{p}_2(\mathbf{x}_{jk})} \right\} - \log \left\{ \frac{\hat{p}_{1(i)}(\mathbf{x}_{jk})}{\hat{p}_{2(i)}(\mathbf{x}_{jk})} \right\} \right]^2.$$

Calculando inicialmente $R(\mathbf{x}) = \log \left\{ \frac{\hat{p}_1(\mathbf{x})}{\hat{p}_2(\mathbf{x})} \right\} - \log \left\{ \frac{\hat{p}_{1(i)}(\mathbf{x})}{\hat{p}_{2(i)}(\mathbf{x})} \right\}$, obtém-se

$$\begin{aligned}
R(\mathbf{x}) &= \log \left\{ \left[\frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_1|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) \right] \div \left(\frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_1|^{1/2}} \right. \right. \right. \\
&\quad \left. \left. \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) \right] + \frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_2|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_2)' \hat{\Sigma}_2^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) \right] \right) \right] \div \\
&\quad \left[\frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_2|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_2)' \hat{\Sigma}_2^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) \right] \div \left(\frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_1|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) \right] + \right. \right. \\
&\quad \left. \left. \frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_2|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_2)' \hat{\Sigma}_2^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) \right] \right) \right] \right\} - \\
&\quad - \log \left\{ \left[\frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_{1(0)}|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_{1(0)})' \hat{\Sigma}_{1(0)}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_{1(0)}) \right] \div \left(\frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_{1(0)}|^{1/2}} \right. \right. \right. \\
&\quad \left. \left. \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_{1(0)})' \hat{\Sigma}_{1(0)}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_{1(0)}) \right] + \frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_2|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_2)' \hat{\Sigma}_2^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) \right] \right) \right] \div \\
&\quad \left[\frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_2|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_2)' \hat{\Sigma}_2^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) \right] \div \left(\frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_{1(0)}|^{1/2}} \right. \right. \\
&\quad \left. \left. \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_{1(0)})' \hat{\Sigma}_{1(0)}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_{1(0)}) \right] + \frac{1}{(2\pi)^{p/2} |\hat{\Sigma}_2|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_2)' \hat{\Sigma}_2^{-1} (\mathbf{x} - \bar{\mathbf{x}}_2) \right] \right) \right] \right\} \\
&= -\log |\hat{\Sigma}_1|^{1/2} - \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x} - \bar{\mathbf{x}}_1) + \log |\hat{\Sigma}_{1(0)}|^{1/2} + \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_{1(0)})' \hat{\Sigma}_{1(0)}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_{1(0)}).
\end{aligned}$$

Retornando à primeira expressão, tem-se que

$$LOQR_i = \frac{1}{n} \sum_{j=1}^2 \sum_{k=1}^{n_j} R_j^2(\mathbf{x}_{jk}) \quad \text{com}$$

$$2R_j(\mathbf{x}_{jk}) = -\log |\hat{\Sigma}_1| - (\mathbf{x}_{jk} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x}_{jk} - \bar{\mathbf{x}}_1) + \log |\hat{\Sigma}_{1(0)}| + (\mathbf{x}_{jk} - \bar{\mathbf{x}}_{1(0)})' \hat{\Sigma}_{1(0)}^{-1} (\mathbf{x}_{jk} - \bar{\mathbf{x}}_{1(0)})$$

O termo $R_j^2(\mathbf{x}_{jk})$ pode ser escrito como $[\log \hat{f}_1(\mathbf{x}) - \log \hat{f}_{1(i)}(\mathbf{x})]^2$, devido ao cancelamento da estimativa da densidade para o grupo 2, que não é afetado. Sendo assim, é mais razoável propor medidas de diagnóstico para o grupo da observação eliminada como

$$LOQRG_i = \frac{1}{n_1} \sum_{k=1}^{n_1} [\log \hat{f}_1(\mathbf{x}_{1k}) - \log \hat{f}_{1(i)}(\mathbf{x}_{1k})]^2 = \frac{1}{n_1} \sum_k R_1^2(\mathbf{x}_{1k}) \quad (5.3)$$

em que

$$2R_1(\mathbf{x}_{1k}) = -\log |\hat{\Sigma}_1| - (\mathbf{x}_{1k} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x}_{1k} - \bar{\mathbf{x}}_1) + \log |\hat{\Sigma}_{1(i)}| + (\mathbf{x}_{1k} - \bar{\mathbf{x}}_{1(i)})' \hat{\Sigma}_{1(i)}^{-1} (\mathbf{x}_{1k} - \bar{\mathbf{x}}_{1(i)}).$$

Pode-se definir também as correspondentes medidas de diferenças simples e diferenças absolutas, no entanto, numa análise com dados reais, Fung (1996-a) verificou que as medidas sugeridas de diferenças simples, absoluta e quadrada forneceram resultados similares.

Tais medidas foram calculadas sob uma abordagem frequentista, mas poderiam ser generalizadas para uma abordagem preditiva Bayesiana. No entanto, devido a uma simplificação nos cálculos, a abordagem Bayesiana forneceria praticamente os mesmos resultados.

Capítulo 6

Observações influentes em análise discriminante para g grupos

6.1 Observações influentes na probabilidade estimada de uma observação pertencer a determinado grupo

Neste capítulo são apresentadas as medidas que detectam observações influentes em análise discriminante múltipla, ou seja, quando existem mais que duas populações. O capítulo é composto por três seções. Inicialmente aborda-se a influência das observações na probabilidade estimada que uma observação pertença a determinado grupo. Posteriormente, são estudadas a probabilidade de classificação incorreta e a função de influência discutida no Capítulo 2.

Como já foi visto, a estimativa da probabilidade de uma observação de origem desconhecida pertencer a um certo grupo tem papel importante em análise discriminante.

Fung (1995-b) sugeriu quatro medidas de diagnóstico para avaliar o quanto estas estimativas podem ser afetadas pelas observações.

Seja \mathbf{x}_{jk} , $j = 1, 2, \dots, g$, $k = 1, \dots, n_j$ uma amostra de tamanho $n = n_1 + n_2 + \dots + n_g$ selecionada aleatoriamente de g populações, de modo que \mathbf{x}_{jk} tem distribuição normal p-variada com média μ_j e matriz de covariância Σ .

A média amostral do grupo j é denotada por $\bar{\mathbf{x}}_j$ e a estimativa da matriz de covariância é $\hat{\Sigma} = \sum_j \sum_k (\mathbf{x}_{jk} - \bar{\mathbf{x}}_j)(\mathbf{x}_{jk} - \bar{\mathbf{x}}_j)' / (n - g)$.

Com base nessas quantidades, a estimativa da densidade normal para uma observação da população j é:

$$\hat{f}_j(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\hat{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_j)' \hat{\Sigma}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_j)\right]. \quad (6.1)$$

Seja q_j a probabilidade *a priori* de que uma observação \mathbf{x}_0 de origem desconhecida venha da população j , $j = 1, 2, \dots, g$. A probabilidade estimada de que \mathbf{x}_0 pertença ao grupo j é:

$$\hat{p}_j(\mathbf{x}_0) = \frac{q_j \hat{f}_j(\mathbf{x}_0)}{\sum_j q_j \hat{f}_j(\mathbf{x}_0)} \quad j = 1, \dots, g$$

e substituindo (6.1) em \hat{f}_j , obtém-se

$$\hat{p}_j(\mathbf{x}_0) = \frac{q_j \exp[-(\mathbf{x}_0 - \bar{\mathbf{x}}_j)' \hat{\Sigma}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_j)/2]}{\sum_j q_j \exp[-(\mathbf{x}_0 - \bar{\mathbf{x}}_j)' \hat{\Sigma}^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}}_j)/2]} \quad j = 1, 2, \dots, g. \quad (6.2)$$

Essa quantidade é muito importante uma vez que, quando os custos de classificação incorreta são iguais, a regra de classificação é obtida alocando-se \mathbf{x}_0 no grupo j tal que $\hat{p}_j(\mathbf{x}_0)$ é máxima.

Quando os custos não forem iguais, aloca-se \mathbf{x}_0 na população π_j com $\sum_{\substack{i=1 \\ i \neq j}}^g q_i p_i(\mathbf{x}) c(j/i)$ mínima (A-1.4).

Para determinar elementos influentes no cálculo de $\hat{p}_j(\mathbf{x})$ será usado novamente o método da omissão de observações. Se \mathbf{x}_{ri} , i -ésima observação do grupo r , $r = 1, 2, \dots, g$, $i = 1, 2, \dots, n_r$, é a observação eliminada da amostra, a média do r -ésimo grupo fica:

$$\bar{\mathbf{x}}_{r(r)} = \sum_{\substack{j=1 \\ j \neq i}}^{n_r} \frac{\mathbf{x}_{rj}}{n_r - 1}$$

e as demais médias não são afetadas. A matriz de covariância resultante é denotada como $\hat{\Sigma}_{(ri)}$.

Após a retirada de \mathbf{x}_{ri} , a probabilidade estimada, $\hat{p}_j(\mathbf{x}_0)$, é calculada através da expressão (6.2) usando as estimativas para a amostra reduzida. Indicando esta estimativa por $\hat{p}_{j(r)}(\mathbf{x})$, a influência da observação retirada é calculada como

$$\hat{p}_j(\mathbf{x}_0) - \hat{p}_{j(r)}(\mathbf{x}_0)$$

e com base nela, Fung (1995-b) propõe as seguintes medidas de diagnóstico:

a) a diferença absoluta das probabilidades estimadas

$$DAP_{ri} = \sum_{j=1}^g \sum_{k=1}^{n_j} \frac{|\hat{p}_j(\mathbf{x}_{jk}) - \hat{p}_{j(r)}(\mathbf{x}_{jk})|}{n}, \quad r = 1, 2, \dots, g, \quad i = 1, 2, \dots, n_r$$

- b) a diferença quadrática das probabilidades estimadas

$$DQP_{ri} = \sum_{j=1}^g \sum_{k=1}^{n_j} \frac{[\hat{p}_j(\mathbf{x}_{jk}) - \hat{p}_{j(r)}(\mathbf{x}_{jk})]^2}{n}, \quad r = 1, 2, \dots, g, \quad i = 1, 2, \dots, n_r$$

- c) a diferença absoluta dos logaritmos das probabilidades estimadas

$$DALP_{ri} = \sum_{j=1}^g \sum_{k=1}^{n_j} \frac{|\log \hat{p}_j(\mathbf{x}_{jk}) - \log \hat{p}_{j(r)}(\mathbf{x}_{jk})|}{n}, \quad r = 1, 2, \dots, g, \quad i = 1, 2, \dots, n_r$$

- d) a diferença quadrática dos logaritmos das probabilidades estimadas

$$DQLP_{ri} = \sum_{j=1}^g \sum_{k=1}^{n_j} \frac{[\log \hat{p}_j(\mathbf{x}_{jk}) - \log \hat{p}_{j(r)}(\mathbf{x}_{jk})]^2}{n}, \quad r = 1, 2, \dots, g, \quad i = 1, 2, \dots, n_r$$

Na definição destas medidas, foi usado o estimador não viciado para Σ . No entanto, Fung (1995-b) verificou que, se usadas outras estimativas como a de máxima verossimilhança ou estimativas Bayesianas, na prática os resultados são similares.

Os diagnósticos de influência propostos medem a influência total, pois são definidos sobre todas as observações da amostra, mas pode-se também usar medidas para um grupo específico, por exemplo,

$$DQLPG_{j(r)} = \sum_{k=1}^{n_j} \frac{[\log \hat{p}_j(\mathbf{x}_{jk}) - \log \hat{p}_{j(r)}(\mathbf{x}_{jk})]^2}{n_j},$$

que mede o efeito da eliminação da observação x_{ri} no logaritmo da probabilidade estimada para os elementos do grupo j , $j = 1, \dots, g$. As outras medidas também podem ser definidas desse modo.

Depois de identificadas as potenciais observações influentes, pode ser interessante considerar a mudança nas probabilidades estimadas para casos individuais, depois de omitidas essas observações. Assim, o efeito da observação x_{ri} na estimativa da probabilidade individual de x_{jk} ser corretamente classificada no grupo j pode ser analisado como

$$\hat{p}_j(\mathbf{x}_{jk}) - \hat{p}_{j(n)}(\mathbf{x}_{jk}) \quad \text{ou}$$

$$\log \hat{p}_j(\mathbf{x}_{jk}) - \log \hat{p}_{j(n)}(\mathbf{x}_{jk}) \quad k = 1, \dots, n_j \quad j = 1, \dots, g.$$

Fung (1995-b) verificou que as medidas propostas foram satisfatórias para um conjunto de dados real e também sob um estudo de simulação. Usando medidas de influência para grupos específicos para todos os pares possíveis dos g grupos, obteve informações adicionais que poderiam ter sido “mascaradas” no cálculo das medidas para todos os grupos.

No estudo de simulação, foi observado que as medidas quadráticas DQP e $DQLP$ fornecem resultados similares entre si e melhores do que as medidas absolutas DAP e $DALP$, e as medidas $DALP$ e $DQLP$, que utilizam logaritmos, são superiores. Sendo assim, a medida $DQLP$ seria preferível dentre as quatro medidas propostas. Foram usados também diferentes conjuntos de probabilidades *a priori*, e os resultados foram próximos, sendo que a análise com as probabilidades *a priori* iguais se aproximou mais do conjunto de probabilidades *a priori* real, o que indica o uso de probabilidades *a priori* iguais quando não se tem informações sobre as probabilidades *a priori* verdadeiras.

6.2 Medidas baseadas na probabilidade de classificação incorreta

Prosseguindo sua pesquisa em análise discriminante múltipla, Fung (1996-c) calculou a probabilidade de classificação incorreta, sob a seguinte regra de classificação:

Um vetor aleatório \mathbf{x}_0 , de origem desconhecida, é alocado em π_j se $\mathbf{x}_0 \in R_j$, sendo que

$$R_j = \bigcap_{\substack{k=1 \\ k \neq j}}^g [2(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} \mathbf{x}_0 + \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j - \boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k < 0] \quad (6.3)$$

que é uma forma alternativa de definir a regra discriminante de Fisher.

A probabilidade de classificação incorreta para essa regra de classificação está apresentada na seguinte proposição.

Proposição 6.1 - A probabilidade de classificação incorreta para $\mathbf{x}_0 \in \pi_j$, $j = 1, 2, \dots, g$, sob a regra de alocação (6.3) é

$$PCI_j = 1 - G(b_{j1}, \dots, b_{j,j-1}, b_{j,j+1}, \dots, b_{jg}) \quad (6.4)$$

em que $b_{jk} = \frac{\Delta_{jk}}{2}$, $\Delta_{jk}^2 = (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)$ é a distância de Mahalanobis entre as populações j e k , para $k = 1, 2, \dots, g$ e $k \neq j$ (6.5)

$G(b_{j1}, \dots, b_{j,j-1}, b_{j,j+1}, \dots, b_{jg})$ é a função distribuição de uma normal multivariada, de

modo que $G(b_{j1}, \dots, b_{j,j-1}, b_{j,j+1}, \dots, b_{jg}) = P \left[\bigcap_{k \neq j} z_k < b_{jk} \right]$, (6.6)

$$z_k \sim N(0, 1) \quad \text{e} \quad \text{cov}(z_k, z_l) = \frac{(\Delta_{jk}^2 + \Delta_{jl}^2 - \Delta_{kl}^2)}{2\Delta_{jk}\Delta_{jl}}, \quad k \neq j, \quad l \neq j.$$

Prova: Sob a regra de classificação (6.3), a probabilidade de classificação incorreta PCI_j é $1 - P(R_j / \mathbf{x}_0 \in \pi_j)$. Quanto ao conjunto R_j , tem-se que

$$\begin{aligned}
 R_j &= \bigcap_{\substack{k=1 \\ k \neq j}}^g [2(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} \mathbf{x}_0 + \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j - \boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k < 0] \\
 &= \bigcap_{\substack{k=1 \\ k \neq j}}^g [2(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} \mathbf{x}_0 + (\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_k + \boldsymbol{\mu}_j) < 0] \\
 &= \bigcap_{\substack{k=1 \\ k \neq j}}^g [2(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} \mathbf{x}_0 - (\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_k + \boldsymbol{\mu}_j) + (\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_j) < \Delta_{jk}^2] \\
 &= \bigcap_{\substack{k=1 \\ k \neq j}}^g [2(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} \mathbf{x}_0 + (\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_j - \boldsymbol{\mu}_k - \boldsymbol{\mu}_j) < \Delta_{jk}^2] \\
 &= \bigcap_{\substack{k=1 \\ k \neq j}}^g \frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}_j)}{\Delta_{jk}} < \frac{\Delta_{jk}}{2}.
 \end{aligned}$$

Portanto, denotando $\frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}_j)}{\Delta_{jk}}$ por z_k e $\frac{\Delta_{jk}}{2}$ por b_{jk} ,

segue que

$$P(R_j / \mathbf{x}_0 \in \pi_j) = P\left(\bigcap_{k \neq j} z_k < b_{jk}\right).$$

Se $\mathbf{x}_0 \in \pi_j$,

$$E(z_j) = E\left(\frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_0 - \boldsymbol{\mu}_j)}{\Delta_{jk}}\right) = \frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_j)}{\Delta_{jk}} = 0,$$

$$Var(z_j) = Var\left(\frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_0 - \boldsymbol{\mu}_j)}{\Delta_{jk}}\right) = \frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} Var(\mathbf{X}_0) \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)}{\Delta_{jk}^2} = 1$$

Desta forma, a normalidade de z_k é consequência do fato que \mathbf{X}_0 tem distribuição normal multivariada. Portanto, a variável z_k tem distribuição normal padrão, com

$$Cov(z_k, z_l) = Cov\left(\frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_0 - \boldsymbol{\mu}_j)}{\Delta_{jk}}, \frac{(\boldsymbol{\mu}_l - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\mathbf{X}_0 - \boldsymbol{\mu}_j)}{\Delta_{jl}}\right)$$

$$= \frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} Cov(\mathbf{X}_0, \mathbf{X}_0) \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_l - \boldsymbol{\mu}_j)}{\Delta_{jk} \Delta_{jl}} = \frac{(\boldsymbol{\mu}_k - \boldsymbol{\mu}_j)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_l - \boldsymbol{\mu}_j)}{\Delta_{jk} \Delta_{jl}}$$

$$= \frac{\boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l - \boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j - \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l + \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j}{\Delta_{jk} \Delta_{jl}}$$

$$= \frac{2\boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l - 2\boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j - 2\boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l + 2\boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j}{2\Delta_{jk} \Delta_{jl}}$$

Somando e subtraindo os termos $\boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k$ e $\boldsymbol{\mu}_l' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l$ no numerador e agrupando convenientemente,

$$Cov(z_k, z_l) =$$

$$= \frac{2\boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l - 2\boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j - 2\boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l + 2\boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k - \boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \boldsymbol{\mu}_l' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l - \boldsymbol{\mu}_l' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l}{2\Delta_{jk} \Delta_{jl}}$$

$$= \left\{ \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j - \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k - \boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j - \boldsymbol{\mu}_j' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l - \boldsymbol{\mu}_l' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \boldsymbol{\mu}_l' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l - \left[\boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k - \boldsymbol{\mu}_k' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l - \boldsymbol{\mu}_l' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_k + \boldsymbol{\mu}_l' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_l \right] \right\} / 2\Delta_{jk} \Delta_{jl}$$

$$= \frac{\Delta_{jk}^2 + \Delta_{jl}^2 - \Delta_{kl}^2}{2\Delta_{jk}\Delta_{jl}}.$$

Nota-se então que a probabilidade de classificação incorreta da Proposição 6.1 é função somente das distâncias entre as médias populacionais.

Influência de uma observação

Tomando a amostra \mathbf{x}_{jk} , $j = 1, 2, \dots, g$, $k = 1, 2, \dots, n_j$, descrita no início do capítulo, mas estimando a matriz de covariância por máxima verossimilhança,

$$\hat{\Sigma} = \sum_{j=1}^g \sum_{k=1}^{n_j} \frac{(\mathbf{x}_{jk} - \bar{\mathbf{x}}_j)(\mathbf{x}_{jk} - \bar{\mathbf{x}}_j)'}{n}, \quad (6.7)$$

obtem-se a regra de classificação amostral substituindo os parâmetros por suas estimativas em (6.3). Sob essa regra amostral, a probabilidade de classificação incorreta para o grupo j , $PCIE_j$ (6.4), pode ser calculada de maneira análoga, trocando os parâmetros por suas estimativas na Proposição 6.1. Desta forma, a estimativa da probabilidade de classificação incorreta é

$$PCIE = \sum_{j=1}^g \frac{PCIE_j}{g}.$$

Uma vez que o critério mais comum para estudar a influência de uma observação é o método da omissão, seja \mathbf{x}_H a observação i do grupo r eliminada da amostra. Sob essa amostra reduzida obtém-se nova regra de classificação bem como diferentes estimativas para os parâmetros e para a probabilidade de classificação incorreta que fica

$$PCIE_{(ri)} = \sum_{j=1}^g \frac{PCIE_{j(ri)}}{g}$$

em que $PCIE_{j(ri)}$ é a probabilidade de classificação incorreta estimada para um elemento do grupo j eliminando-se a observação i , do grupo r . O efeito dessa observação na estimativa da probabilidade de classificação incorreta é avaliado pela diferença $PCIE - PCIE_{j(ri)}$.

6.3 Uso de funções de influência para as medidas de probabilidade de classificação incorreta

Nesta seção, será obtida a função de influência para a probabilidade de classificação incorreta.

Conforme notação introduzida na Seção 2.2, os vetores de média e matriz de covariância para múltiplos grupos são

$$\mu_j = \int x dF_j$$

(6.8)

$$\Sigma = w_1 \Sigma_{F_1} + \dots + w_g \Sigma_{F_g}$$

(6.9)

com $\Sigma_{F_1} = \Sigma_{F_2} = \dots = \Sigma_{F_g}$ e $\Sigma w_j = 1$, $w_j > 0$, sendo que

(6.10)

$$\Sigma_{F_j} = \int (x - \mu_j)(x - \mu_j)' dF_j.$$

Além disso, seja $\theta = T(F_1, F_2, \dots, F_k, \dots, F_g)$ um parâmetro geral, obtido a partir das funções de distribuição da normal multivariada, F_k , $k = 1, \dots, g$.

Com base nesses elementos, Fung (1996-c) propôs o seguinte lema:

Lema 6.1 - Se a r -ésima distribuição é perturbada, $r = 1, 2, \dots, g$, então a função de influência para Δ_{jk}^2 é dada por

$$I(\mathbf{x}; \Delta_{jk}^2) = \begin{cases} w_r \Delta_{jk}^2 - w_r \phi_{jk}^2 & j \neq r, k \neq r \\ w_r \Delta_{jk}^2 + 2\phi_{jk} - w_r \phi_{jk}^2 & j = r \\ w_r \Delta_{jk}^2 - 2\phi_{jk} - w_r \phi_{jk}^2 & k = r \end{cases} \quad (6.11)$$

em que $\phi_{jk} = (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)' \boldsymbol{\Sigma}^{-1} \mathbf{z}$ e $\mathbf{z} = \mathbf{x} - \boldsymbol{\mu}_r$.

Prova: Do mesmo modo que na Seção 2.2.1, fórmulas (2.1), (2.3), (2.4) e (2.5), calculando (6.8), (6.9) e (6.10) para a distribuição perturbada $\tilde{F}_r = (1 - \varepsilon)F_r + \varepsilon \delta_{\mathbf{x}}$ obtém-se:

$$\boldsymbol{\mu}_r^* = \boldsymbol{\mu}_r + \varepsilon \mathbf{z}$$

$$\boldsymbol{\Sigma}_{F_r}^* = (1 - \varepsilon)\boldsymbol{\Sigma}_{F_r} + \varepsilon \mathbf{z}\mathbf{z}'$$

$$\boldsymbol{\Sigma}^* = (1 - \varepsilon w_r)\boldsymbol{\Sigma} + \varepsilon w_r \mathbf{z}\mathbf{z}'$$

$$\boldsymbol{\Sigma}^{-1*} = (1 + \varepsilon w_r)\boldsymbol{\Sigma}^{-1} - \varepsilon w_r \boldsymbol{\Sigma}^{-1} \mathbf{z}\mathbf{z}' \boldsymbol{\Sigma}^{-1}$$

i) Para $j \neq r$ e $k \neq r$

$$\begin{aligned} \Delta_{jk}^{2*} &= (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)' [(1 + \varepsilon w_r)\boldsymbol{\Sigma}^{-1} - \varepsilon w_r \boldsymbol{\Sigma}^{-1} \mathbf{z}\mathbf{z}' \boldsymbol{\Sigma}^{-1}] (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k) \\ &= (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)' (1 + \varepsilon w_r)\boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k) - (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)' \varepsilon w_r \boldsymbol{\Sigma}^{-1} \mathbf{z}\mathbf{z}' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_k) \\ &= (1 + \varepsilon w_r)\Delta_{jk}^2 - \varepsilon w_r \phi_{jk}^2 \end{aligned}$$

e calculando a função de influência, tem-se

$$I(\mathbf{x}; \Delta_{jk}^2) = \lim_{\varepsilon \rightarrow 0} \frac{\Delta_{jk}^{2*} - \Delta_{jk}^2}{\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{(1 + \varepsilon w_r) \Delta_{jk}^2 - \varepsilon w_r \phi_{jk}^2 - \Delta_{jk}^2}{\varepsilon}$$

$$= w_r \Delta_{jk}^2 - w_r \phi_{jk}^2 \quad \text{para } j \neq r \text{ e } k \neq r .$$

ii) Para $j = r$

$$\begin{aligned} \Delta_{jk}^{2*} &= (\boldsymbol{\mu}_r^* - \boldsymbol{\mu}_k) [(1 + \varepsilon w_r) \boldsymbol{\Sigma}^{-1} - \varepsilon w_r \boldsymbol{\Sigma}^{-1} \mathbf{z} \mathbf{z}' \boldsymbol{\Sigma}^{-1}] (\boldsymbol{\mu}_r^* - \boldsymbol{\mu}_k) \\ &= (\boldsymbol{\mu}_r + \varepsilon \mathbf{z} - \boldsymbol{\mu}_k) [(1 + \varepsilon w_r) \boldsymbol{\Sigma}^{-1} - \varepsilon w_r \boldsymbol{\Sigma}^{-1} \mathbf{z} \mathbf{z}' \boldsymbol{\Sigma}^{-1}] (\boldsymbol{\mu}_r + \varepsilon \mathbf{z} - \boldsymbol{\mu}_k) \end{aligned}$$

e usando o cálculo do tópico a) da Seção 2.2.1

$$\Delta_{jk}^{2*} = (1 + \varepsilon w_r) \Delta_{jk}^2 + 2\varepsilon \phi_{jk} - \varepsilon w_r \phi_{jk}^2 .$$

Calculando a função de influência para Δ_{jk}^2 ,

$$I(\mathbf{x}; \Delta_{jk}^2) = \lim_{\varepsilon \rightarrow 0} \frac{\Delta_{jk}^{2*} - \Delta_{jk}^2}{\varepsilon} = w_r \Delta_{jk}^2 + 2\phi_{jk} - w_r \phi_{jk}^2, \quad \text{para } j = r .$$

iii) para $k = r$

$$\boldsymbol{\delta}^* = (\boldsymbol{\mu}_j - \boldsymbol{\mu}_r^*) = (\boldsymbol{\mu}_j - \boldsymbol{\mu}_r - \varepsilon \mathbf{z}) = \boldsymbol{\delta} - \varepsilon \mathbf{z},$$

de modo que

$$\Delta_{jk}^{2*} = (\boldsymbol{\delta} - \varepsilon \mathbf{z}) [(1 + \varepsilon w_r) \boldsymbol{\Sigma}^{-1} - \varepsilon w_r \boldsymbol{\Sigma}^{-1} \mathbf{z} \mathbf{z}' \boldsymbol{\Sigma}^{-1}] (\boldsymbol{\delta} - \varepsilon \mathbf{z}) .$$

Efetuando as multiplicações obtém-se:

$$\begin{aligned} \Delta_{jk}^{2*} &= \delta' \Sigma^{-1} \delta + \delta' \varepsilon w_r \Sigma^{-1} \delta - \varepsilon z' \Sigma^{-1} \delta - \varepsilon z' \varepsilon w_r \Sigma^{-1} \delta - \delta' \Sigma^{-1} \varepsilon z - \delta' \varepsilon w_r \Sigma^{-1} \varepsilon z + \\ &+ \varepsilon z' \Sigma^{-1} \varepsilon z + \varepsilon z' \varepsilon w_r \Sigma^{-1} \varepsilon z - \delta' \varepsilon w_r \Sigma^{-1} z z' \Sigma^{-1} \delta - \varepsilon z' \varepsilon w_r \Sigma^{-1} z z' \Sigma^{-1} \delta + \\ &+ \delta' \varepsilon w_r \Sigma^{-1} z z' \Sigma^{-1} \varepsilon z - \varepsilon z' \varepsilon w_r \Sigma^{-1} z z' \Sigma^{-1} \varepsilon z \end{aligned}$$

e lembrando que $\phi_{jk} = \delta' \Sigma^{-1} z$,

$$\begin{aligned} \Delta_{jk}^{2*} &= \delta' \Sigma^{-1} \delta (1 + \varepsilon w_r) - \varepsilon \phi_{jk}' + \varepsilon^2 w_r \phi_{jk}' - \varepsilon \phi_{jk} - \varepsilon^2 w_r \phi_{jk} + \varepsilon^2 z' \Sigma^{-1} z + \varepsilon^3 w_r z' \Sigma^{-1} z - \\ &- \varepsilon w_r \phi_{jk} \phi_{jk}' - \varepsilon^2 w_r z' \Sigma^{-1} z \phi_{jk}' + \varepsilon^2 w_r \phi_{jk} z' \Sigma^{-1} z - \varepsilon^3 w_r (z' \Sigma^{-1} z)^2. \end{aligned}$$

Eliminando os termos de ordem maior ou igual a ε^2

$$\Delta_{jk}^{2*} = (1 + \varepsilon w_r) \Delta_{jk}^2 - 2\varepsilon \phi_{jk} - \varepsilon w_r \phi_{jk}^2$$

e calculando a função de influência, obtém-se

$$I(\mathbf{x}; \Delta_{jk}^2) = w_r \Delta_{jk}^2 - 2\phi_{jk} - w_r \phi_{jk}^2, \text{ para } k = r.$$

O autor apresenta a função de influência para a probabilidade de classificação incorreta para um elemento da população j , com a r -ésima distribuição perturbada, $I(\mathbf{x}; PCI_j)$, escrevendo-a a partir de $I(\mathbf{x}; \Delta_{jk}^2)$ obtida anteriormente,

$$I(\mathbf{x}; PCI_j) = - \sum_{\substack{k=1 \\ k \neq j}}^g \frac{\delta G(b_{j1}, \dots, b_{j,j-1}, b_{j,j+1}, \dots, b_{jg})}{\delta b_{jk}} \cdot \frac{I(\mathbf{x}; \Delta_{jk}^2)}{4\Delta_{jk}} \quad (6.12)$$

em que $I(\mathbf{x}; \Delta_{jk}^2)$ é dado em (6.11) e $\frac{\delta G(\cdot)}{\delta b_{jk}}$ é a derivada parcial de

$$G(\cdot) = P \left(\bigcap_{\substack{l \neq j \\ l \neq k}} z_l < b_{jl} \right), \text{ em que } z_l \sim N(0,1) \text{ e } \text{cov}(z_l, z_k) = \frac{(\Delta_{jl}^2 + \Delta_{jk}^2 - \Delta_{lk}^2)}{2\Delta_{jl}\Delta_{jk}}, \quad l \neq j, k \neq j.$$

A função de influência para a probabilidade de classificação incorreta total é definida como

$$I(\mathbf{x}; PCI) = \sum_{j=1}^g \frac{I(\mathbf{x}; PCI_j)}{g}.$$

Uma vez que as funções de distribuição F_j , $j = 1, 2, \dots, g$ são frequentemente desconhecidas, calcula-se a função de influência empírica, $IE(\mathbf{x}; PCI)$, obtida usando as estimativas amostrais $\bar{\mathbf{x}}_j$ e $\hat{\Sigma}$ para os parâmetros em (6.12).

No cálculo de $I(\mathbf{x}; \Delta_{jk}^2)$ e portanto no de $I(\mathbf{x}; PCI)$ foi assumido $\varepsilon \rightarrow 0$. No entanto, tomando-se um elemento amostral $\mathbf{x} = \mathbf{x}_{ri}$ da distribuição perturbada e $\varepsilon = \frac{-1}{n_r - 1}$, obtém-se a chamada função de influência amostral, $IA(\mathbf{x}_{ri}; PCI)$, que coincide com a que seria obtida através do método da omissão de observações ou seja

$$IA(\mathbf{x}_{ri}; PCI) = -(n_r - 1)(PCIE_{(ri)} - PCIE). \quad (6.13)$$

Observa-se que a função de influência empírica geralmente dá uma boa aproximação para a função de influência amostral, especialmente quando n_r é grande. Além disso com o uso da função de influência empírica o trabalho computacional é reduzido.

A função de influência empírica e a função de influência amostral são calculadas para a PCI total. Mas, pode ser de interesse examinar a influência de uma observação na PCI de determinado grupo, $IE(\mathbf{x}_{ri}; PCI_j)$ e $IA(\mathbf{x}_{ri}; PCI_j)$, $j = 1, 2, \dots, g$, o que revelaria uma informação de influência mais detalhada, além de diminuir o trabalho computacional.

Aplicação ao conjunto de dados da Íris de Fisher

Fung (1996-c) utilizou o conjunto de dados da Íris de Fisher (Johnson e Wichern (1992, p. 566)) para ilustrar algumas das medidas definidas. Os dados consistem de três grupos com 50 observações cada um, numeradas de 1 a 50 para o grupo 1 da espécie Íris setosa, de 51 a 100 para o grupo 2 da espécie Íris versicolor e de 101 a 150 para o grupo 3 da espécie Íris virginica. São registrados para cada observação os valores de quatro variáveis, a saber, comprimento da sépala, largura da sépala, comprimento da pétala, largura da pétala. Como o tamanho dos grupos é constante, os pesos w_1 , w_2 e w_3 para estimar a matriz de covariância são iguais a $1/3$.

O autor observou que escolhendo diferentes denominadores para estimar a matriz de covariância, n (estimativa de máxima verossimilhança), $(n-g)$ (estimativa não viciada) e $(n-g-p-1)$ (estimativa Bayesiana), os resultados obtidos são muito similares.

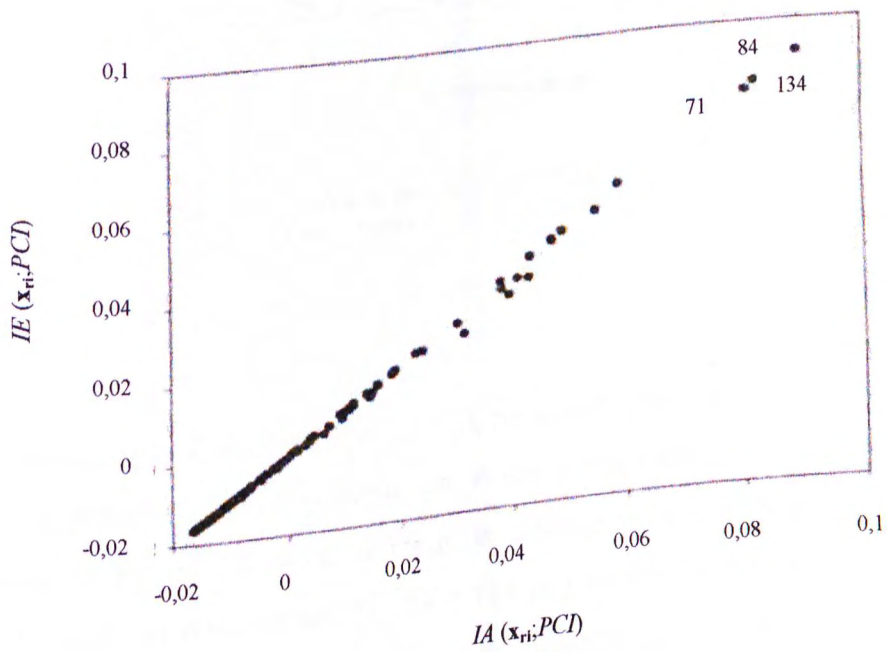


Figura 6.1: Gráfico de $IA(x_{ri}; PCI)$ versus $IE(x_{ri}; PCI)$
Fonte: Figura 1 - Fung (1996-c)

Na Figura 6.1 encontra-se o gráfico de $IA(\mathbf{x}_{ri}; PCI)$ e $IE(\mathbf{x}_{ri}; PCI)$. Sua análise sugere que as informações fornecidas por essas duas medidas são praticamente idênticas.

A Figura 6.2 exibe o gráfico dos valores de $|IE(\mathbf{x}_{ri}; PCI)|$ para cada observação. Verifica-se que, de acordo com a medida PCI , as observações 71 e 84 (do grupo 2) e 134 (do grupo 3) têm a maior influência.

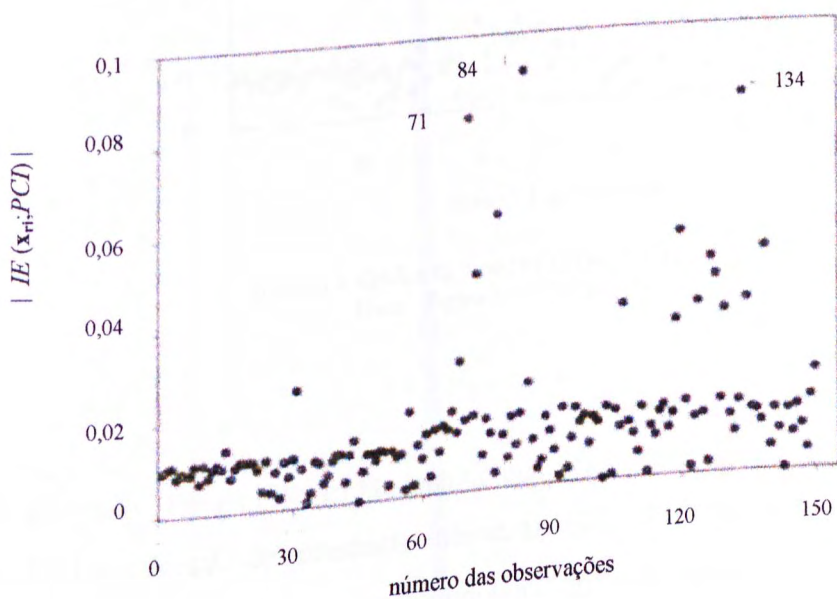


Figura 6.2: Gráfico da medida $|IE(\mathbf{x}_{ri}; PCI)|$
 Fonte: Figura 2 - Fung (1996-c)

Observando a medida $IE(\mathbf{x}_{ri}; PCI_j)$, para cada grupo, o autor verificou que os valores de influência para o primeiro grupo são muito menores que dos outros dois. O gráfico da Figura 6.3 exibe as medidas calculadas para a PCI com relação ao grupo 2, onde as observações 71, 84 e 134 são detectadas como mais influentes. Nota-se também que estas medidas são muito similares às da Figura 6.2, baseadas na função de influência empírica para a PCI total, com pequena diferença na escala do eixo das ordenadas.

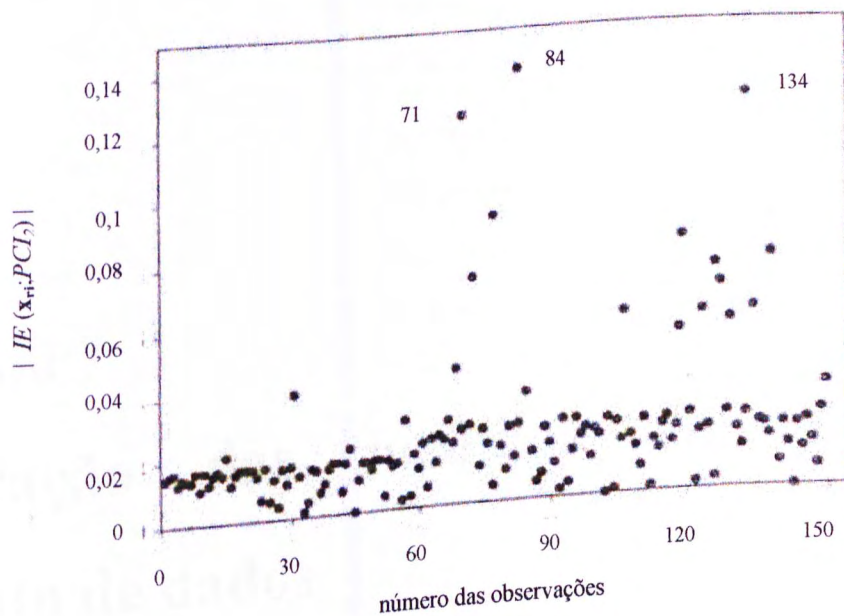


Figura 6.3: Gráfico da medida $|IE(x_{ri}; PCI_2)|$
 Fonte: Figura 3 - Fung (1996-c)

Finalizando, Fung (1996-c) faz uma comparação entre a função de influência empírica (IE) e a função de influência amostral (IA). Na função empírica, o cálculo de $G(\cdot)$ dado em (6.12) exige integrais múltiplas de dimensão $(g-2)$. Um total de $g(g-1)$ integrais devem ser calculadas para a obtenção de $IE(x; PCI)$ já que estas integrais não dependem dos valores de x_{ri} . Frequentemente essa quantidade, $g(g-1)$, é menor do que $g(n+1)$, que é o número de integrais de dimensão $(g-1)$ necessárias para o cálculo de $IA(x_{ri}; PCI)$, pois, de modo geral, o número de elementos da amostra (n) é muito maior que o número de grupos (g). Fica claro então a importância da utilização da função de influência empírica (IE) na redução da quantidade de cálculos necessários.

Capítulo 7

Verificação das medidas em um conjunto de dados

O objetivo deste capítulo é o da aplicação das medidas apresentadas a um conjunto de dados reais. Para esse fim foi utilizado um subconjunto do conjunto de dados obtido no Centro de Estatística Aplicada (CEA) do Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP).

Trata-se dos dados referentes ao Relatório de análise estatística (RAE-CEA-9415) sobre o projeto: Biologia de uma comunidade de marsupiais e roedores, em floresta atlântica de montanha, no Parque Estadual da Serra do Tabuleiro, Santa Catarina, Brasil, (Leite, Singer e Lourenço (1994)).

O estudo teve como fonte o Parque Estadual da Serra do Tabuleiro, no Estado de Santa Catarina, com o objetivo de analisar espécies de marsupiais e roedores (pequenos mamíferos) nativos da região. Os dados foram obtidos de amostras resultantes de capturas-recapturas desses animais, com a utilização de armadilhas, dispersas pelo parque, através da estratificação de uma região do parque em três níveis: nível do solo, a dois metros e a nove metros e meio de altura, em média, durante oito meses (abril/91 a novembro/91).

A partir desses dados, foi estimado o número de animais da espécie *Akodon montensis* e obtida uma regra discriminante para facilitar a classificação de futuras capturas em marsupiais e roedores. Verificou-se também a importância da amostragem estratificada pois se ela não tivesse sido utilizada as espécies *Thomasomys petipes* e *Echimys aff. dasythrix* (estratos arborícolas) não ocorreriam e a espécie *Marmosa cinerea* (estratos terrestre + arborícolas) teria sido subestimada. Constatou-se também que o maior número de capturas-recapturas e de mortes, nas armadilhas, ocorreram nos meses de inverno na região (junho a agosto).

Em cada animal capturado, foram medidas 26 características. Na análise discriminante feita no projeto foram testados 8 conjuntos de funções discriminantes e respectivas variáveis, sendo que o modelo utilizado neste exemplo é o que apresentou melhor relação custo/benefício pois utiliza apenas 4 medidas, com 1% de animais classificados incorretamente, e se baseia na informação de 88 animais, 13 do grupo 1, de marsupiais, e 75 do grupo 2, de roedores. As quatro variáveis em questão são comprimento total (mm), comprimento da cauda (mm), comprimento do pé (mm) e comprimento do pelo do dorso (mm).

Assumindo que os dados têm distribuição normal multivariada, a análise de diagnóstico, objetivo do presente capítulo, inicia-se com o teste de homogeneidade de matrizes de covariâncias (Morrison (1976, pg 252)) para verificar qual a medida mais adequada ao conjunto de dados. Toda a programação foi feita no software S-plus e se encontra no Apêndice C.

Uma vez que o teste apontou para matrizes de covariância diferentes, utilizou-se primeiramente a medida $LOQRG_i$ de (5.3), para análise discriminante quadrática:

$$LOQRG_i = \frac{1}{n_1} \sum_{k=1}^n \left[\log \hat{f}_1(\mathbf{x}_{1k}) - \log \hat{f}_{1(i)}(\mathbf{x}_{1k}) \right]^2 = \frac{1}{n_1} \sum_k R_1^2(\mathbf{x}_{1k})$$

em que

$$2R_1(\mathbf{x}_{1k}) = -\log \left| \hat{\Sigma}_1 \right| - (\mathbf{x}_{1k} - \bar{\mathbf{x}}_1)' \hat{\Sigma}_1^{-1} (\mathbf{x}_{1k} - \bar{\mathbf{x}}_1) + \log \left| \hat{\Sigma}_{1(i)} \right| + (\mathbf{x}_{1k} - \bar{\mathbf{x}}_{1(i)})' \hat{\Sigma}_{1(i)}^{-1} (\mathbf{x}_{1k} - \bar{\mathbf{x}}_{1(i)}).$$

Tal medida apontou as observações de números 3, 8 e 13 do grupo 1 e 6, 9, 10, 20, 21, 29, 41, 67 e 73 do grupo 2 como as mais influentes no logaritmo da razão das probabilidades estimadas, conforme Figura 7.1.

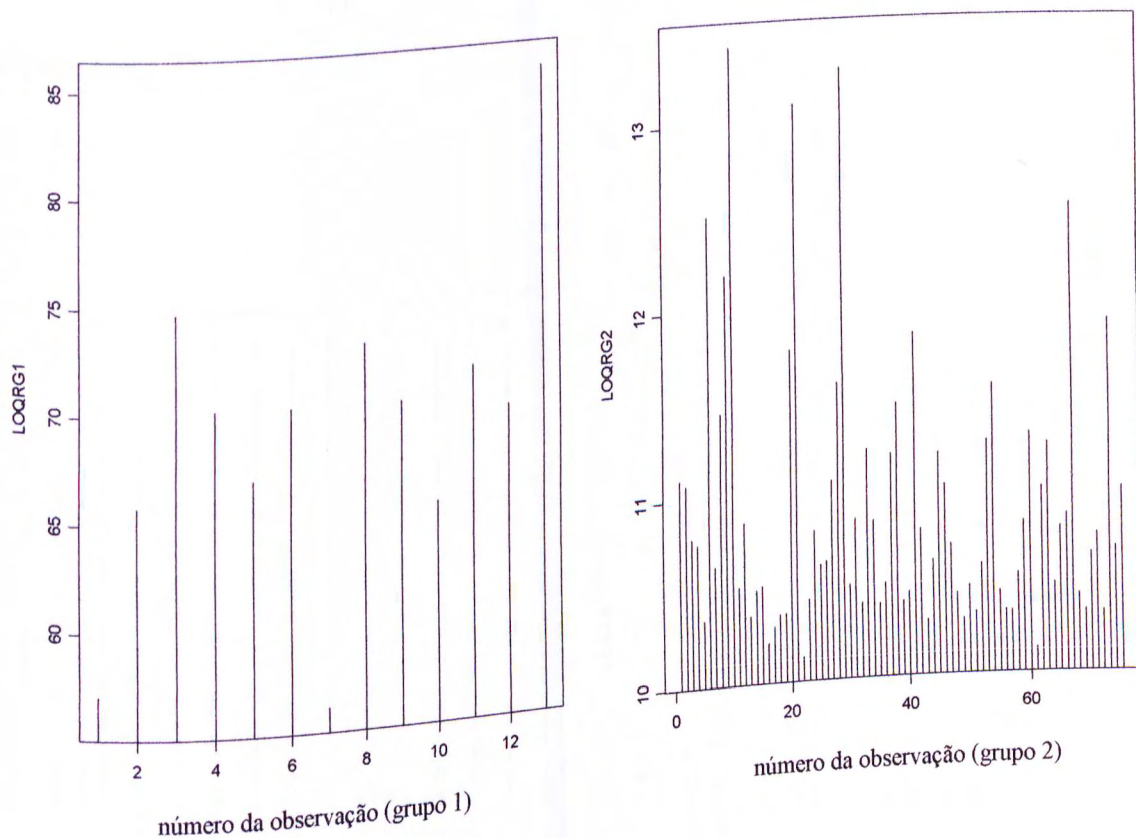


Figura 7.1 Gráficos da estatística *LOQRG*.

Apesar do teste de homogeneidade de matrizes de covariâncias indicar covariâncias diferentes, para efeito de ilustração, foi verificado também o desempenho das demais medidas.

As Figuras 7.2 a 7.10 são os gráficos dos valores destas medidas de diagnóstico para ambos os grupos.

- $$LORAM_i = n^{-1} \sum_{j=1}^2 \sum_{k=1}^{n_j} \ln \left\{ \frac{o(\mathbf{x}_{jk})}{o_{(i)}(\mathbf{x}_{jk})} \right\}$$

em que $\ln(o(\mathbf{y})) = \ln \frac{q}{(1-q)} + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$,

expressão (3.7)

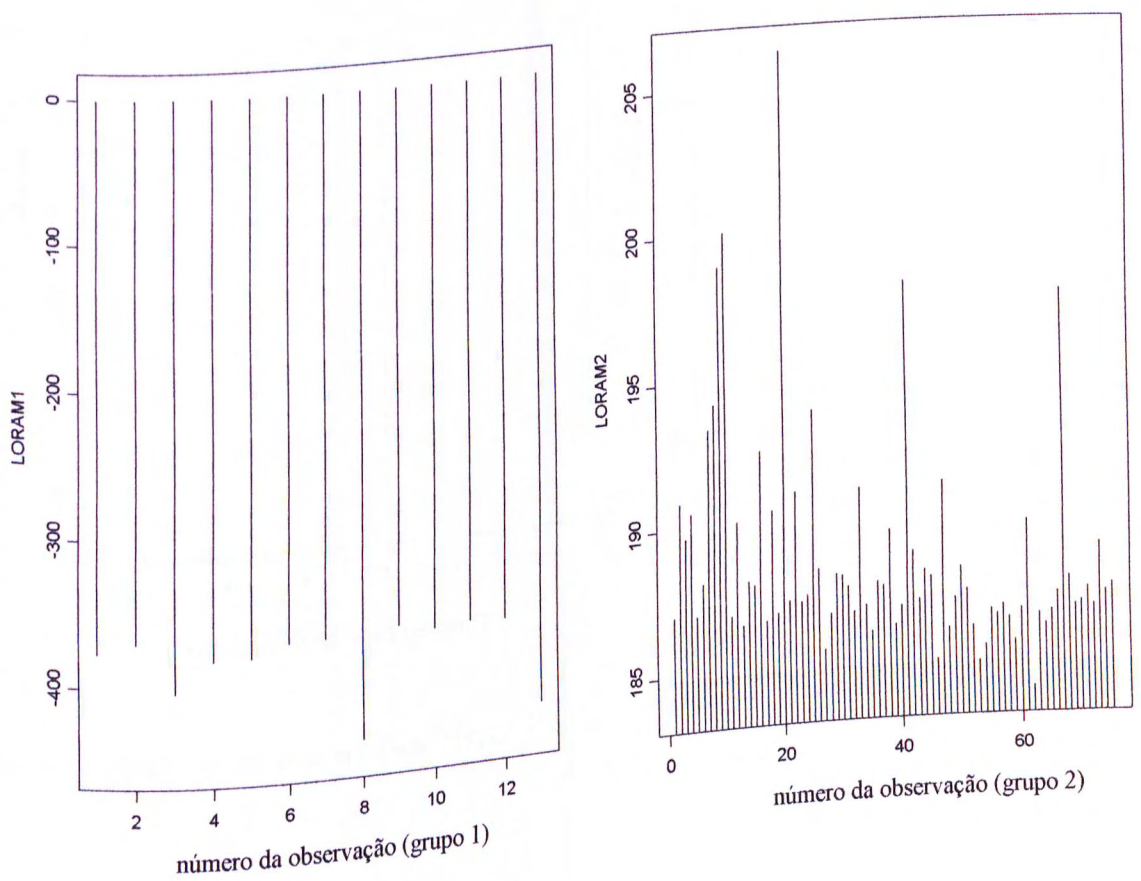


Figura 7.2 Gráficos da estatística *LORAM* para o grupo 1 e grupo 2.

- $I(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + 2\phi - w_1 \phi^2$, expressão (2.8)

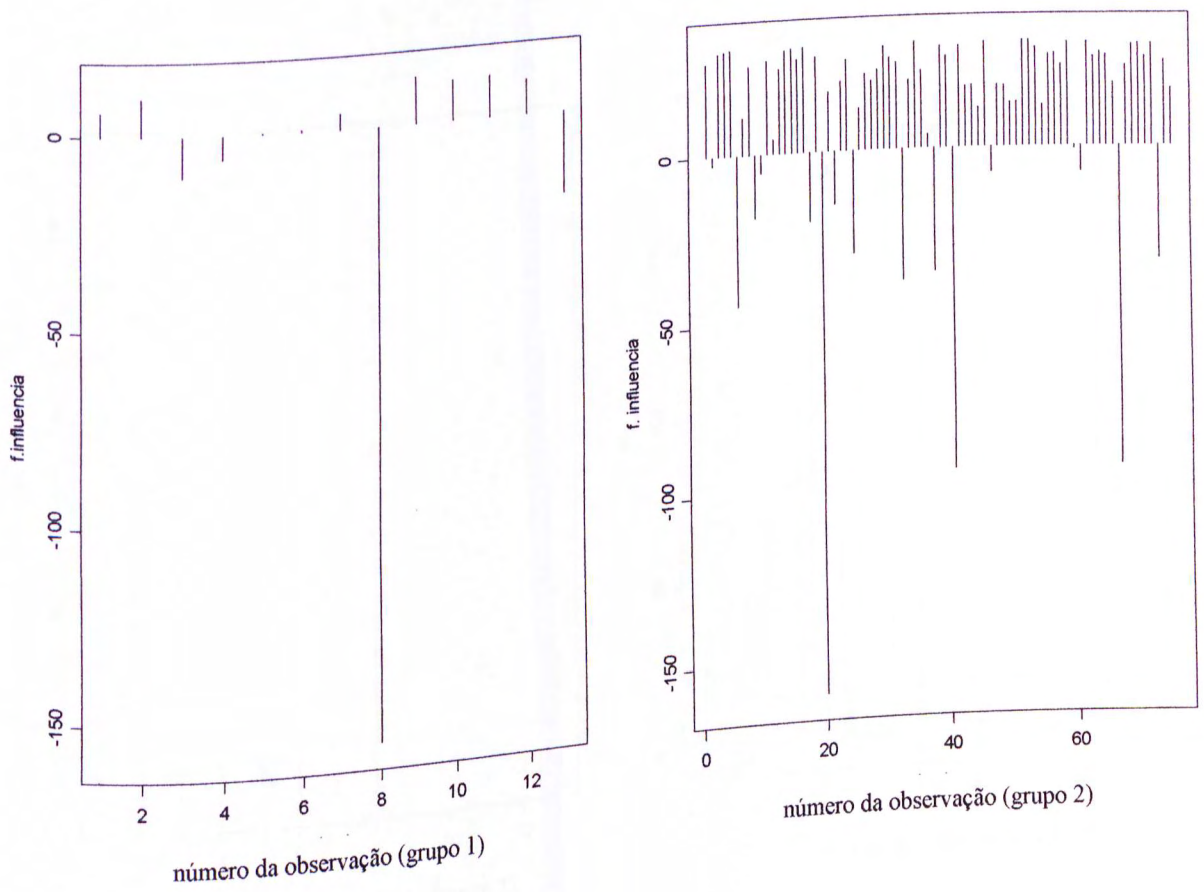


Figura 7.3 – Gráfico da função de influência $\hat{I}(\mathbf{x}; \Delta^2)$.

- $\hat{I}_m(\mathbf{x}; \Delta^2) = w_1 (\hat{\phi} - w_1^{-1})^2$, expressão (2.17)

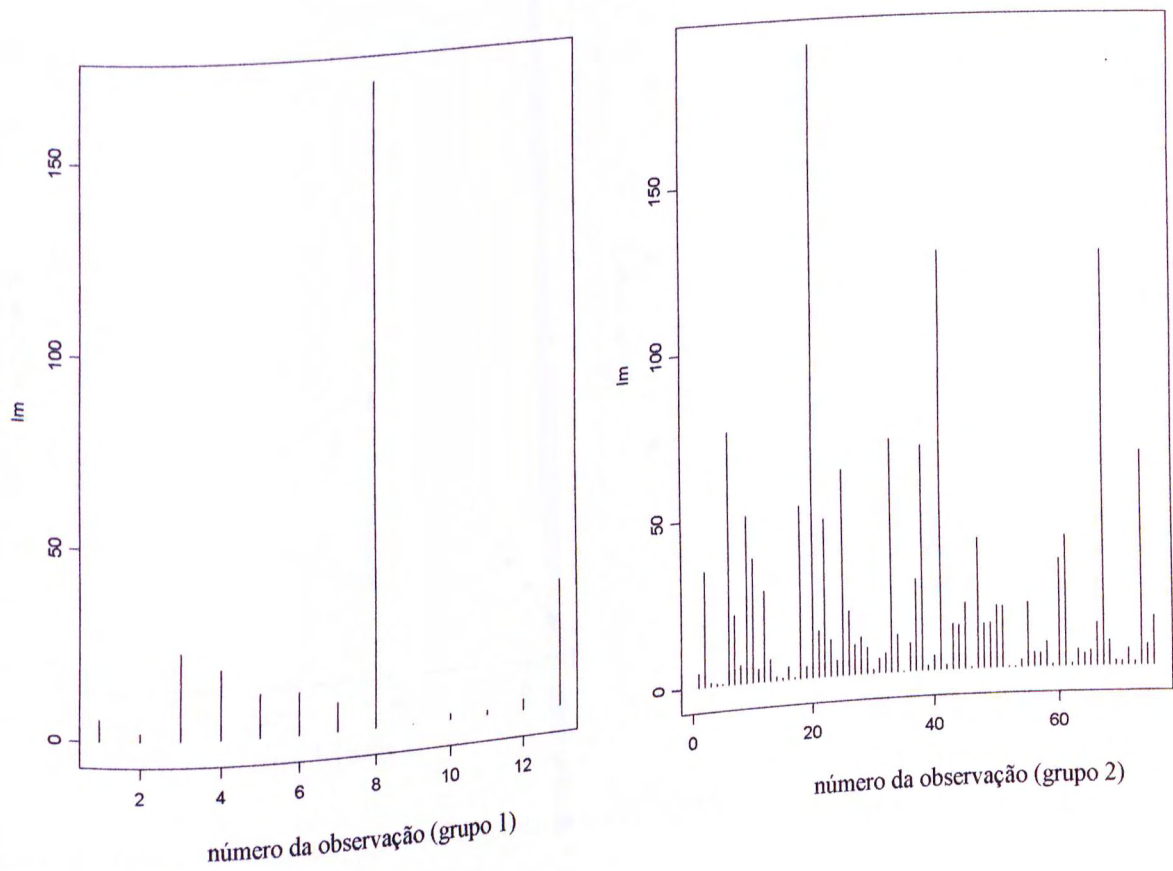


Figura 7.4 – Gráfico da estatística $\hat{I}_m(\mathbf{x}; \Delta^2)$ para o grupo 1 e grupo 2.

O gráfico a seguir apresenta a medida de influência $I_m(\mathbf{x}; \Delta^2)$ sugerida por Campbell em função da medida $I(\mathbf{x}; \Delta^2)$:

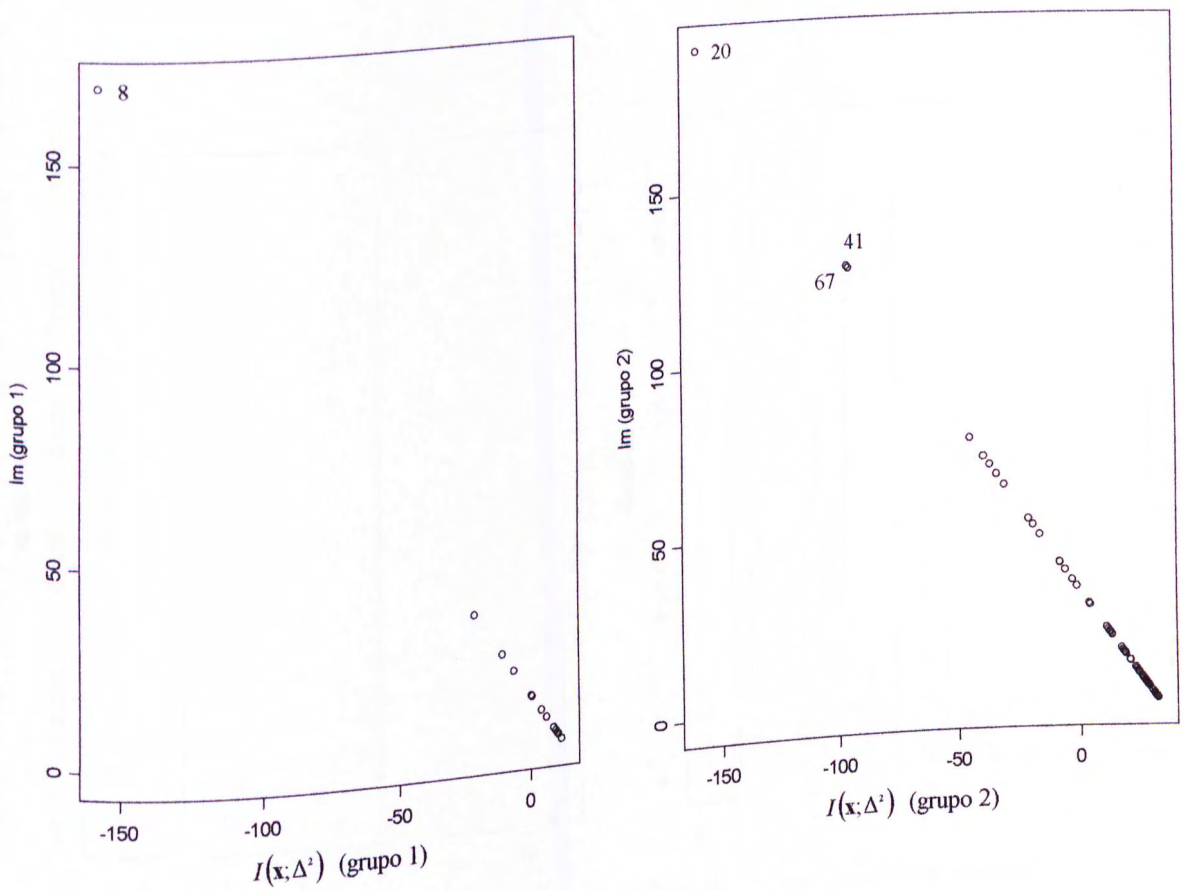


Figura 7.5 – Gráfico da estatística $I_m(\mathbf{x}; \Delta^2)$ em função da estatística $I(\mathbf{x}; \Delta^2)$.

•
$$DPCI_i \cong \frac{\Phi\left(-\frac{1}{2}D\right)}{4D(n_1-1)^2} \left[(1-w_1\hat{\phi}_i)^2 (d_i^2 - \hat{\phi}_i^2 / D^2) + \frac{1}{4}\hat{\phi}_i^2 \right], \text{ expressão (2.23)}$$

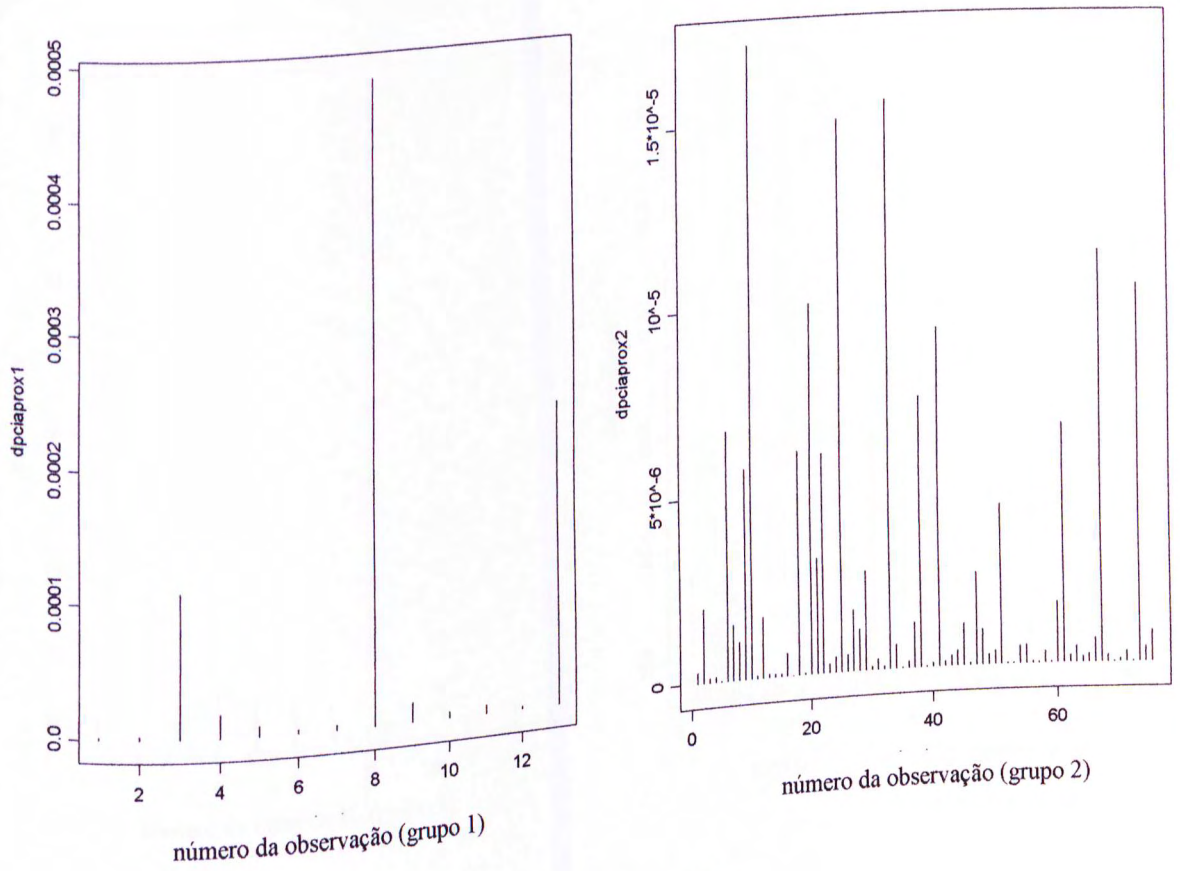


Figura 7.6 – Gráfico da estatística *DPCI* aproximada para o grupo 1 e grupo 2.

•
$$\hat{\Delta}_{1(i)}^2 = \left(\frac{n_1 + n_2 - 1}{n_1 + n_2} \right) \left[\hat{\Delta}^2 - \frac{1}{w_1(n_1 - 1)} + \frac{\{e(\mathbf{x}_{1i}) - w_1^{-1}\}^2}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(\mathbf{x}_{1i})} \right], \text{expressão (2.25)}$$

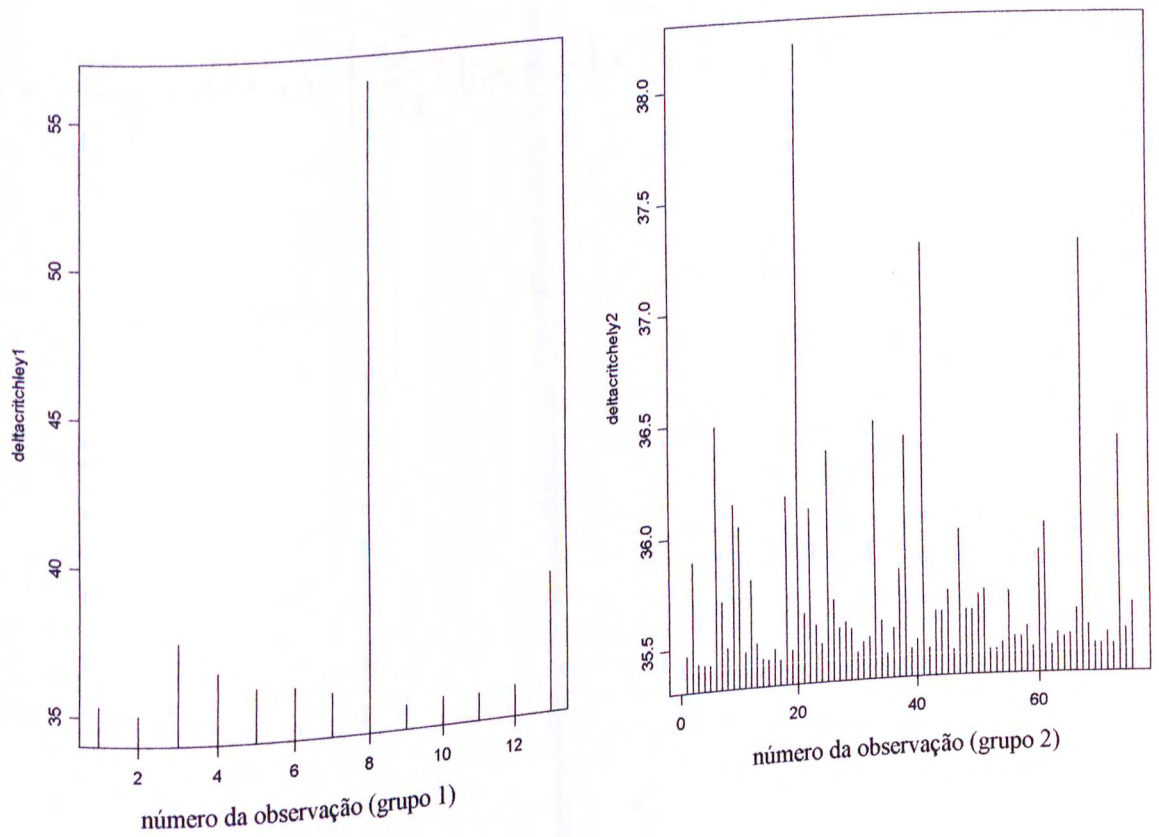


Figura 7.7 – Gráfico da medida $\Delta_{1(i)}^2$.

• $E(DLO_i)^2 = (w_1 B_1 + w_2 B_2)^2 + V$, expressão (2.30)

em que $B_1 = \frac{A_1}{2} - \frac{A_2}{2}$ e $B_2 = \frac{-A_1}{2} - \frac{A_2}{2}$

$$A_1 = \frac{D^2}{n-2} + c f_i (\hat{\phi}_i)^2 + c h \hat{\phi}_i - c h f_i d_i^2 \hat{\phi}_i$$

$$A_2 = c \cdot h \cdot (\hat{\phi}_i - h \cdot d_i^2) g_i$$

$$V = \frac{D^2}{(n-2)^2} + c (f_i \hat{\phi}_i + h g_i) \left[\frac{2 \hat{\phi}_i}{n-2} + (f_i \hat{\phi}_i + h g_i) c d_i^2 \right]$$

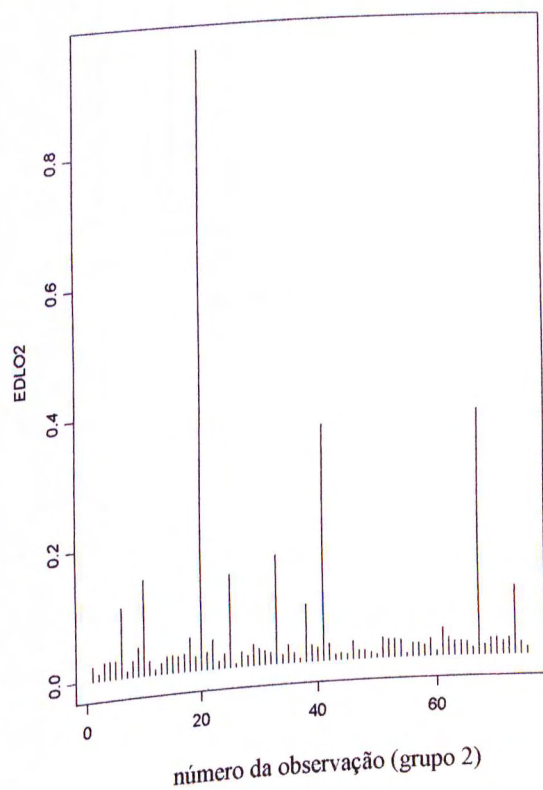
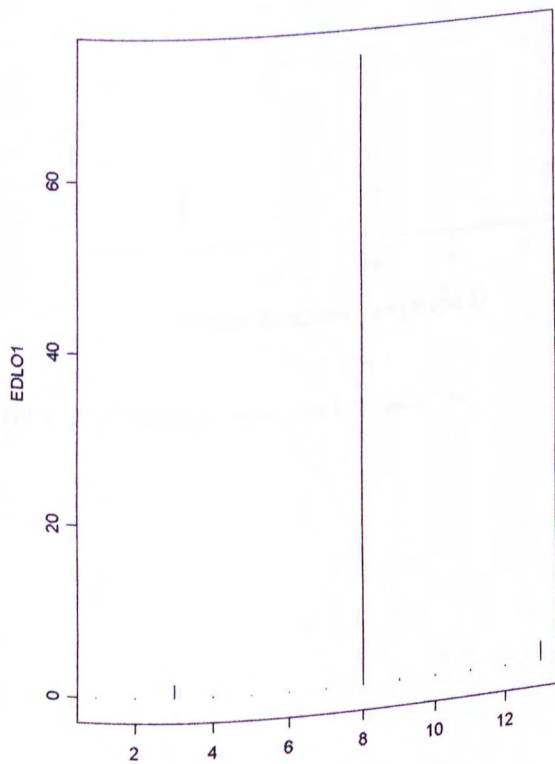


Figura 7.8 – Gráfico da estatística $E(DLO)^2$.

- $$E2c_i^0 = \left[chf_i(\hat{\phi}_i - hd_i^2)/(2b) \right]^2 + c^2 f_i^2 (\hat{\phi}_i + h/b)^2 (d_i^2 - \hat{\phi}_i^2 / D^2)$$

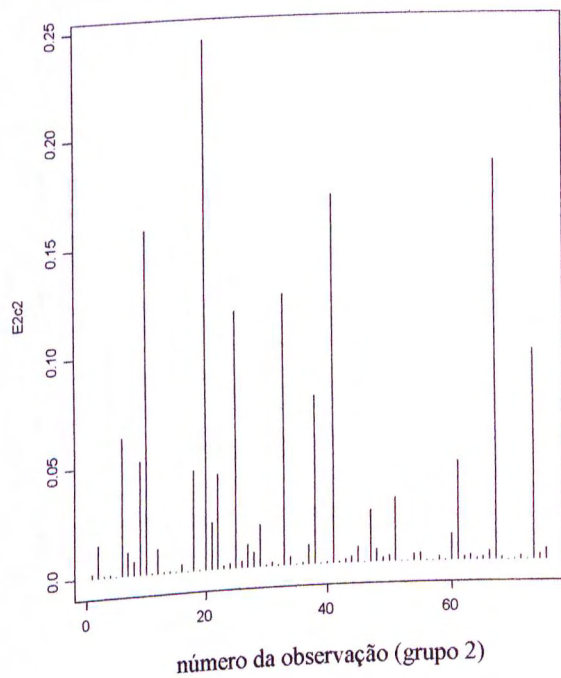
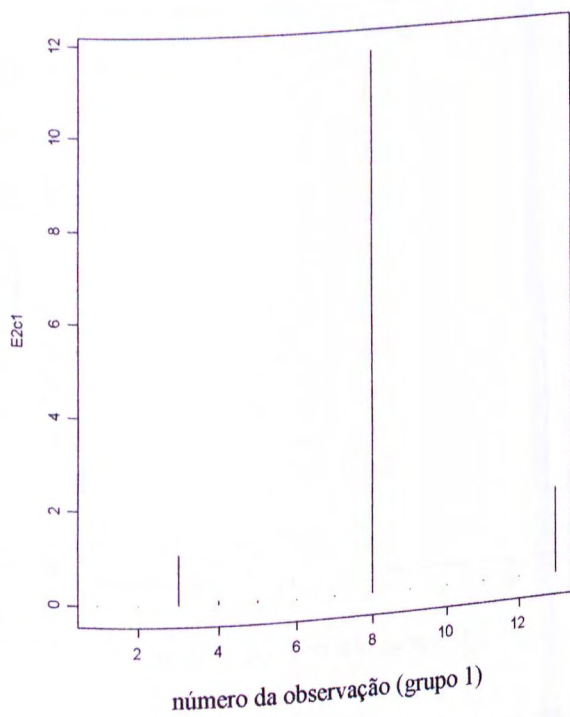


Figura 7.9 – Gráfico da estatística E2 condicional.

- $$E2c_i^0 = \left[\frac{\hat{\phi}_i^2}{4} + (1 - w_1 \hat{\phi}_i)^2 (d_i^2 - \hat{\phi}_i^2 / D^2) \right] / n_1^2$$

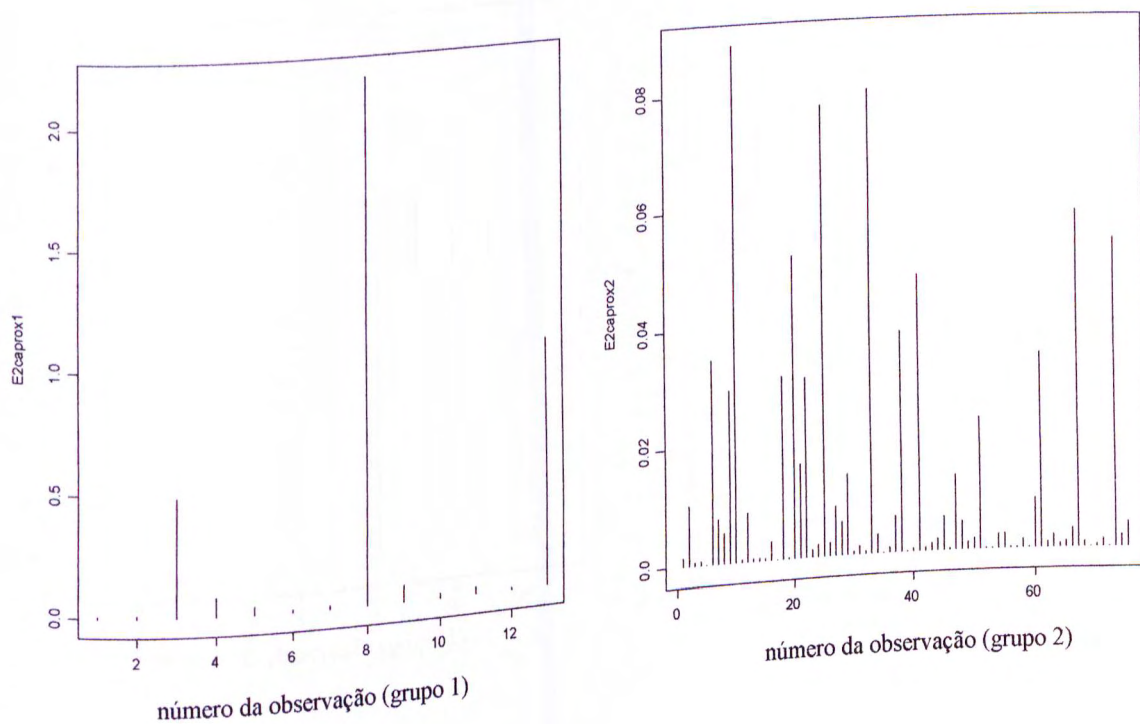


Figura 7.10 – Gráfico da estatística E2 condicional aproximada.

São apresentados nas Figuras 7.11 e 7.12 os gráficos das duas estatísticas principais $\hat{\phi}_i$ e d_i^2 , respectivamente.

- $\phi_i = (\mu_1 - \mu_2)' \Sigma^{-1} z$ em que $z = x_i - \mu_1$.

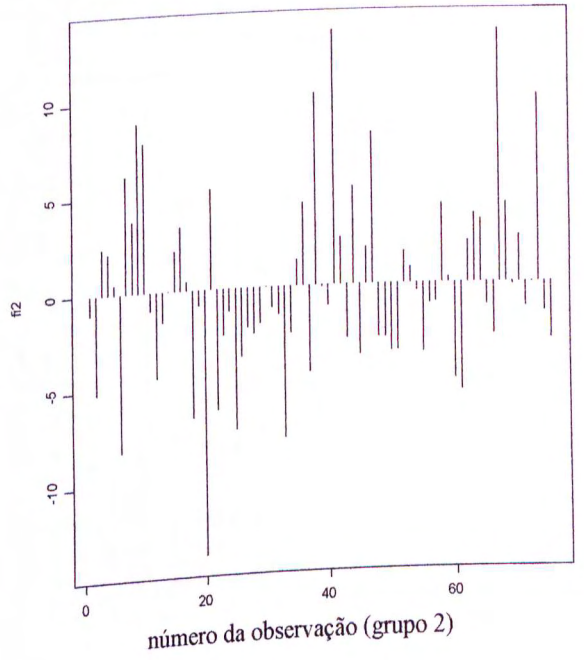
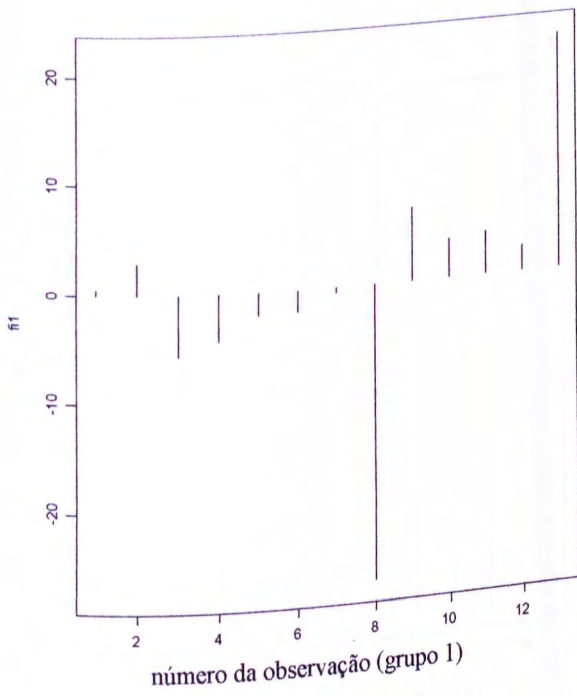


Figura 7.11 – Gráfico da Estatística $\hat{\phi}_i$.

- $$d_i^2 = (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j) \mathbf{S}^{-1} (\mathbf{x}_{ji} - \bar{\mathbf{x}}_j) \quad i=1, \dots, n_j, \quad j=1, 2 .$$

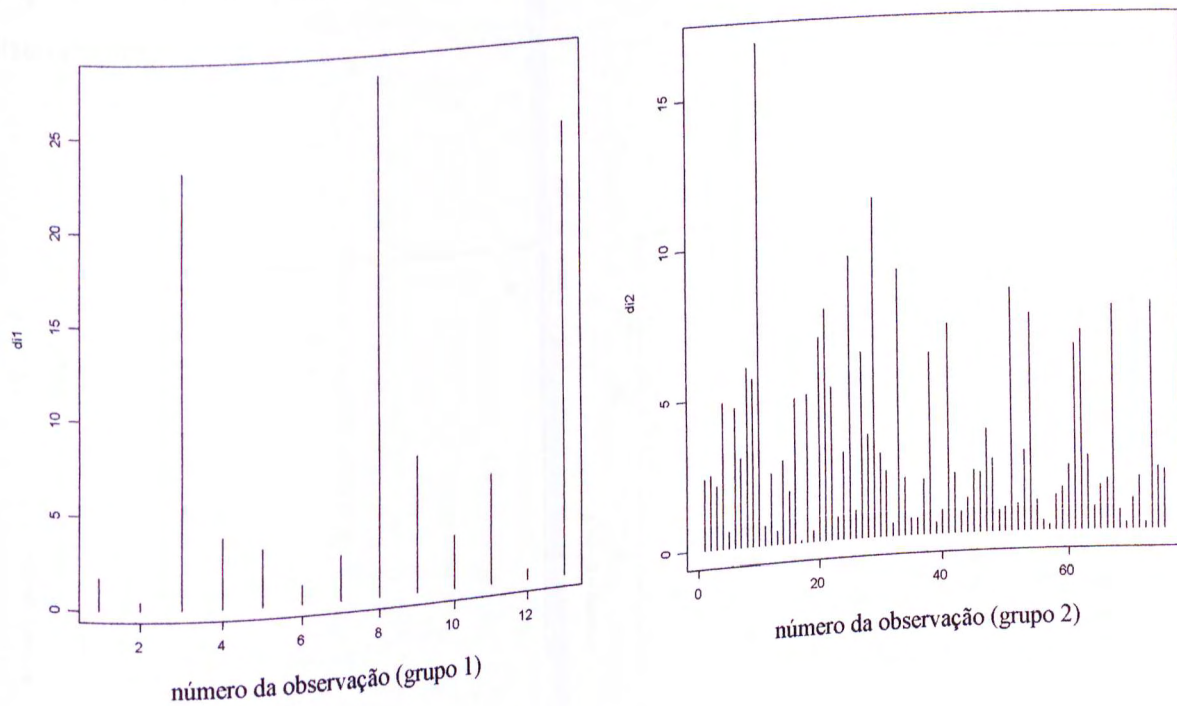


Figura 7.12 – Gráfico da estatística d_i^2 .

Complementando a análise gráfica, são apresentados a seguir alguns gráficos do tipo *qqplot*. A Figura 7.13 mostra os valores da função $I_m(\mathbf{x}; \Delta^2)$ para os dois grupos, contra os quantis da distribuição gama, destacando as observações discrepantes.

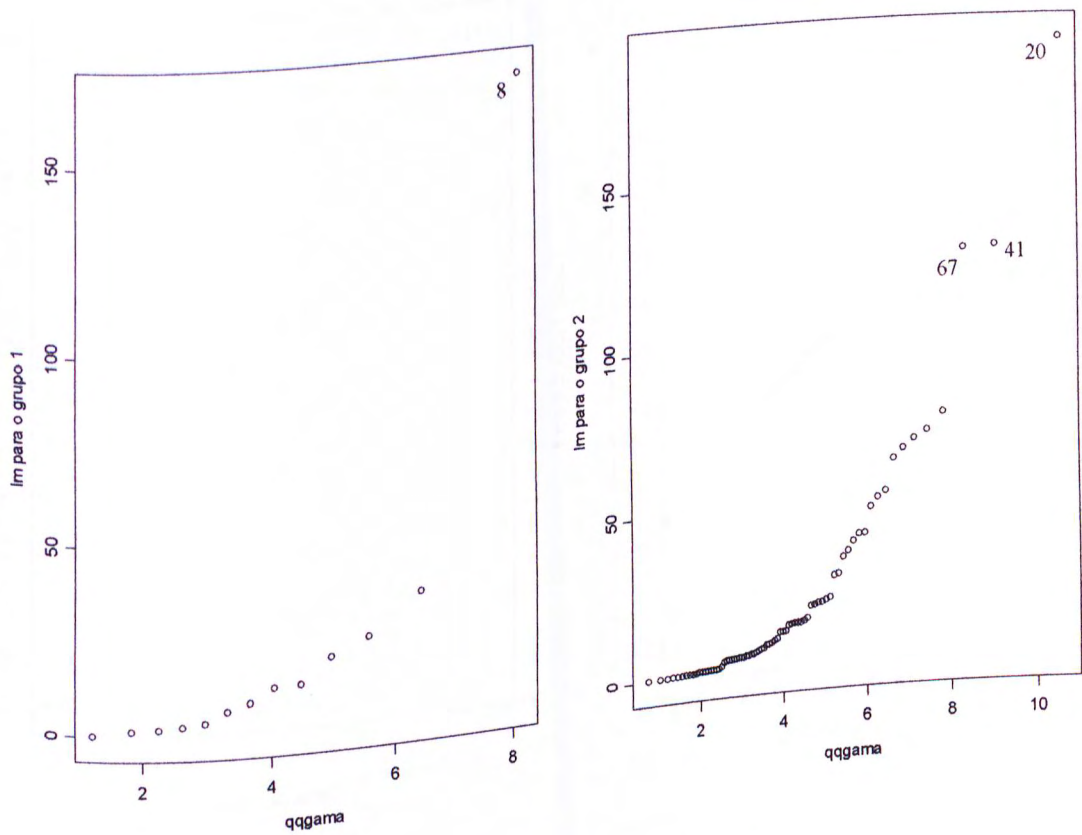


Figura 7.13 – Gráfico da estatística $I_m(\mathbf{x}; \Delta^2)$ contra quantis da distribuição gama.

Apresenta-se na Figura 7.14 os valores da estatística $\frac{\hat{\phi}_i}{D}$ contra os quantis

da distribuição normal padrão.

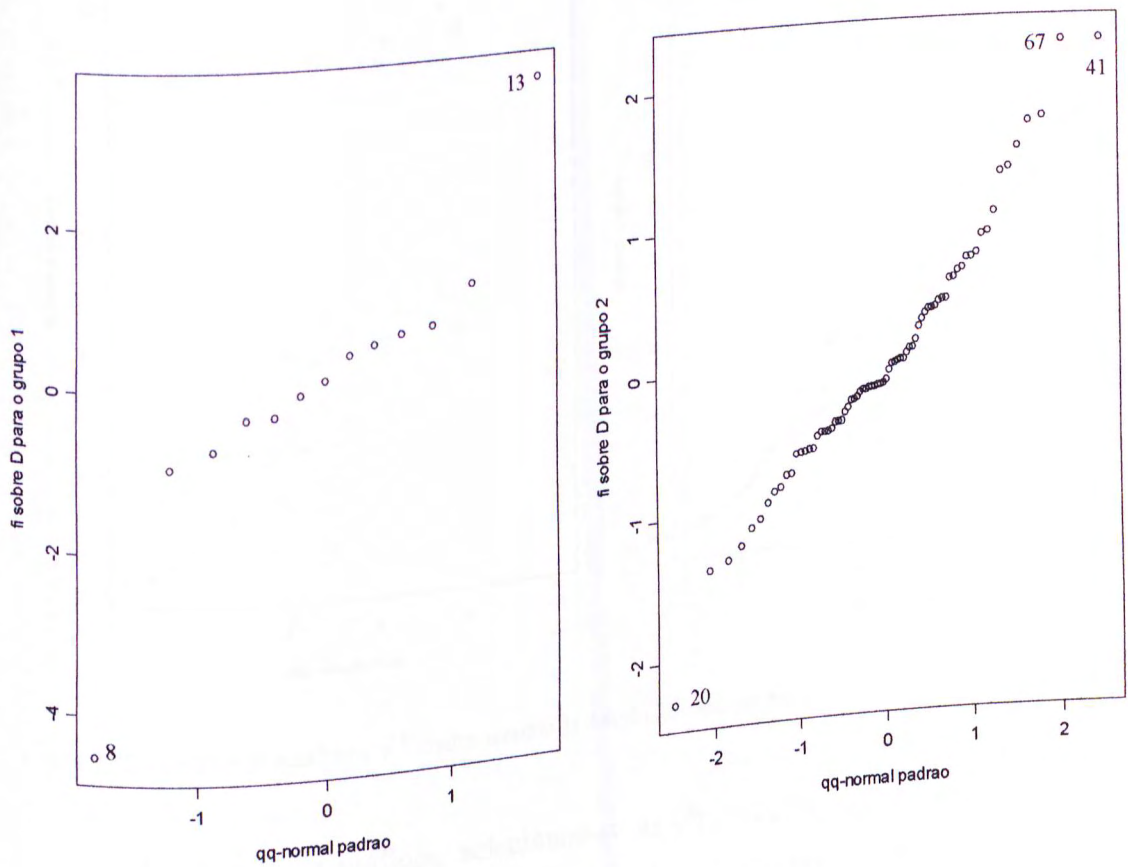


Figura 7.14 – Gráfico da estatística $\frac{\hat{\phi}_i}{D}$ contra os quantis da distribuição normal padrão.

No gráfico da Figura 7.15 observa-se os valores da estatística d_i^2 contra os quantis da distribuição quiquadrado.

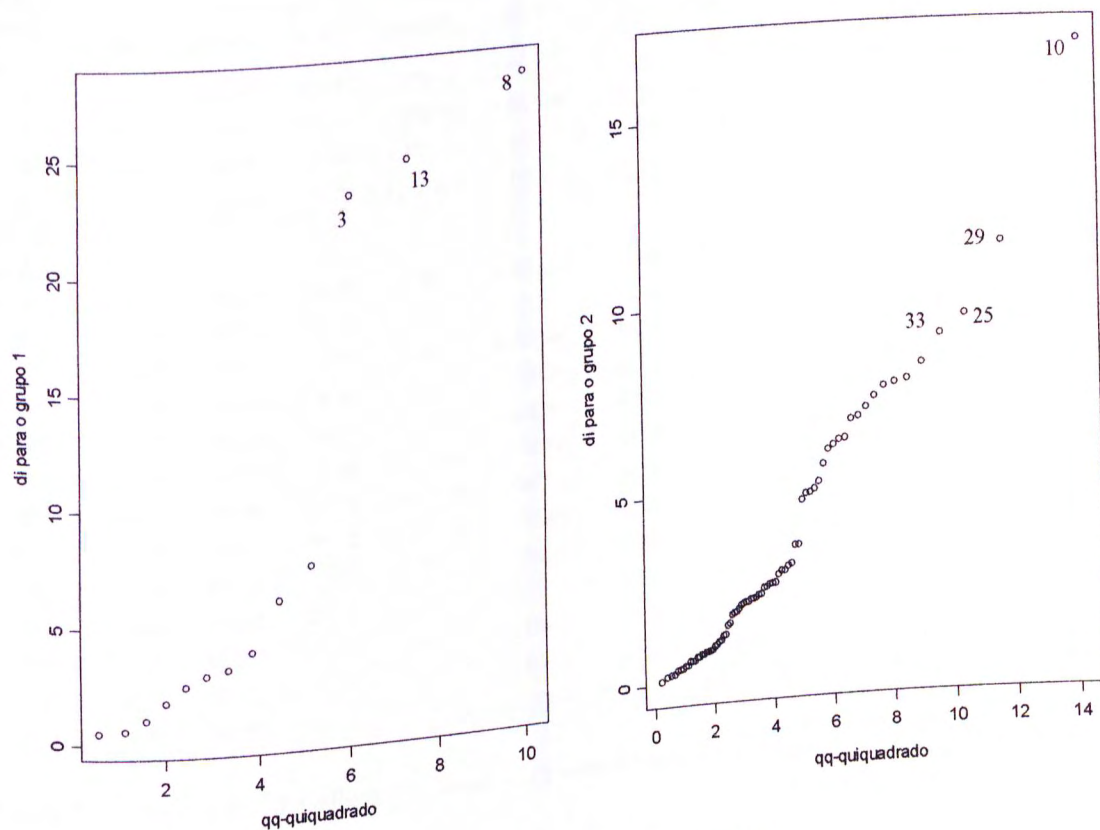


Figura 7.15 – Gráfico da estatística d_i^2 contra quantis da distribuição quiquadrado.

Observando os gráficos, selecionou-se as observações que apresentaram os maiores valores para as medidas em relação às demais, dentro de cada grupo.

Nota-se que, para o grupo 1, as observações de número 8 e 13 foram apontadas por todas as medidas, sendo que a observação de número 3 deixou de ser apontada apenas pela estatística $I(x; \Delta^2)$. Já para o grupo 2, as observações 20 e 67 foram indicadas em todas as medidas e a observação 73 deixou de ser indicada apenas pela medida $I(x; \Delta^2)$. No grupo 1, destacam-se as observações 3, 8 e 13, e no grupo 2, as observações 6, 20, 25, 33, 38, 41, 67 e 73.

A Tabela 7.1 apresenta as observações que se mostraram mais influentes e os correspondentes valores de cada medida de influência.

Casos	Estatísticas											
	LOQRG	LORAM	$\hat{I}(x, \Delta^2)$	$\hat{I}_m(x, \Delta^2)$	DPCI	$\hat{\Delta}_{1(t)}^2$	EDLO	E2c	E2ca	$\hat{\phi}_i$	d_i^2	
Grupo 1												
3	74,74	-403,40	-10,67	2,27E+01	1,09E-04	37,49	1,57	1,08E+00	0,49	-5,64	23,17	
8	73,06	-442,90	-157,23	1,69E+02	4,86E-04	56,13	73,70	1,17E+01	2,19	-27,08	27,93	
9	70,22	-367,70	12,06	2,83E-04	1,48E-04	34,87	0,18	1,67E-05	0,07	6,73	7,33	
13	85,32	-432,30	-21,29	3,34E+01	2,29E-04	38,81	2,30	1,84E+00	1,03	21,8	24,53	
Grupo 2												
6	12,50	188,09	-44,35	76,06	6,75E-06	36,49	0,11	6,28E-02	0,03	-8,27	4,66	
8	11,45	194,18	26,20	5,52	1,02E-06	35,49	0,03	6,19E-03	0,01	3,72	5,97	
9	12,19	198,89	-18,75	50,46	5,70E-06	36,14	0,05	5,17E-02	0,03	8,87	5,58	
10	13,41	200,08	-5,65	37,37	1,72E-06	36,03	0,15	1,57E-01	0,09	7,8	16,85	
20	11,77	206,22	-159,73	191,45	1,00E-05	38,19	0,96	2,45E-01	0,05	-13,82	6,82	
21	13,09	187,30	17,68	14,03	3,12E-06	35,62	0,00	2,13E-02	0,02	5,23	7,76	
25	10,61	193,78	-30,54	62,25	1,50E-05	36,34	0,14	1,18E-01	0,08	-7,37	9,46	
29	13,28	188,11	23,72	7,99	2,68E-06	35,53	0,03	1,90E-02	0,01	-1,89	11,39	
33	11,22	191,02	-38,91	70,62	1,55E-05	36,46	0,17	1,25E-01	0,08	-7,93	8,95	
38	11,46	189,52	-36,51	68,22	7,35E-06	36,39	0,09	7,71E-02	0,04	10,12	6,08	
41	11,84	198,06	-95,32	127,03	9,21E-06	37,26	0,37	1,70E-01	0,05	13,38	7,05	
67	12,52	197,65	-94,66	126,37	1,13E-05	37,26	0,38	1,85E-01	0,06	13,35	7,59	
73	11,89	188,87	-33,64	65,35	1,03E-05	36,37	0,11	9,71E-02	0,06	9,93	7,68	

Tabela 7.1 - Observações mais influentes e respectivos valores das medidas de diagnóstico.

Complementando o estudo, foram feitas análises discriminantes linear e quadrática para o conjunto de dados, através do software Minitab.

Na análise discriminante quadrática para o conjunto completo dos dados, a proporção de classificação correta foi de 0,989, sendo apontada apenas a observação 8 do grupo 1, classificada incorretamente, com probabilidade 0,33 de ser do grupo 1 e 0,67 de pertencer ao grupo 2.

A distância de Mahalanobis é 35,8264, e na análise discriminante linear a proporção correta de classificação foi de 0,989, apontando também a observação do grupo 1 como classificada incorretamente.

A função linear discriminante para o grupo 1 é:

$$y_1 = -25,056 + 0,261 \times \text{comprimentototal} - 0,144 \times \text{cauda} - 0,723 \times \text{pé} + 0,956 \times$$

pelo

e para o grupo 2:

$$y_2 = -21,947 + 0,151 \times \text{comprimentototal} - 0,379 \times \text{cauda} + 1,273 \times \text{pé} + 1,559 \times$$

pelo.

Procedeu-se à análise discriminante na ausência dos pontos 8 do grupo 1 e 20, 41 e 67 do grupo 2, por serem os mais discrepantes.

Eliminando-se somente a observação 8 do grupo 1 tanto na análise discriminante quadrática quanto na análise discriminante linear, obteve-se uma proporção de 100% de classificação correta. A distância de Manalanobis aumentou para 56,6779, o que, na análise discriminante linear implica numa menor probabilidade de classificação incorreta. As funções discriminantes lineares obtidas foram:

grupo 1:

$$y_1 = -34,981 + 0,278 \times \text{comprimentototal} + 0,010 \times \text{cauda} - 1,552 \times \text{pé} + 1,304 \times$$

pelo

grupo 2:

$$y_2 = -21,696 + 0,149 \times \text{comprimentototal} - 0,377 \times \text{cauda} + 1,278 \times \text{pé} + 1,534 \times$$

pelo.

Com a retirada da observação 20 do grupo 2, a proporção de classificação correta é 0,989 tanto na análise linear quanto na quadrática, (observação 8 do grupo 1 classificada incorretamente) . A distância de Mahalanobis ficou 37,42 , valor este que difere pouco da distância D^2 do conjunto de dados completo, que é de 35,8264.

Excluindo-se a observação 41 do grupo 2, verificam-se as mesmas proporções de classificação correta obtidas quando da retirada da observação 20 do grupo 2, o mesmo ocorrendo na ausência da observação 67 do grupo 2.

Excluída a observação 41 do grupo 2, a distância de Mahalanobis é de 38,0738 e excluída a observação 67 do grupo 2, essa distância é de 38,0813.

Uma vez que a observação 10 do grupo 2 foi apontada pela medida *LOQRG* (quadrática) e apontada também por algumas medidas (Bayesiana, *DPCI*, *E2* condicional e outras), foi feita a análise discriminante quadrática e linear sem esse ponto. Observa-se que a distância de Mahalanobis nessa situação é de 36,5536 que é bem próxima da distância de Mahalanobis para o conjunto de dados completo, e a única observação apontada como classificada incorretamente é a de número 8 do grupo 1.

Observou-se também os elementos 21 e 29 do grupo 2, apontados pela medida *LOQRG* como mais influentes. Na análise discriminante linear realizada na ausência desses pontos, obteve-se a distância de Mahalanobis de 35,9158 e 35,4021, respectivamente, valores estes bem próximos de 35,8264 que é a distância de Mahalanobis para o conjunto completo dos dados. As funções lineares discriminantes também foram próximas àquelas para o conjunto completo, e tanto na análise discriminante quadrática, quanto na análise discriminante linear é apontada apenas a observação 8 como classificada incorretamente.

Nota-se que os três pontos apontados pela estatística *LOQRG* como os mais influentes, casos 10, 21 e 29 do grupo 2, são aqueles que apresentam altos valores para d_i^2 , no entanto procedendo-se à análise discriminante na ausência desses pontos, as alterações não são muito significativas.

Examinando a observação 8 do grupo 1, verifica-se que apresenta comprimento total 155 mm, comprimento da cauda 57mm, comprimento do pé 17mm e comprimento do pelo 4 mm. Se excluído esse caso do grupo 1, os valores

mínimos para essas variáveis seriam 250mm, 148mm, 20mm e 8 mm respectivamente. Portanto, todos os valores das variáveis da observação 8 do grupo 1 estão abaixo desses valores mínimos, identificando-a claramente como um *outlier*.

No grupo 2, os valores mínimos são de 130mm, 46mm, 17mm e 5 mm, sendo que a observação 8, a princípio, estaria melhor classificada nesse grupo. Sugere-se então que se verifique as outras variáveis medidas nesse elemento a fim de explicar o porquê de sua inclusão no grupo 1, e questionar se essa classificação é realmente correta, uma vez que sua permanência no grupo 1 causa grande impacto não só na distância de Mahalanobis como também na probabilidade de classificação incorreta.

Deste exemplo, pode-se concluir que as medidas de diagnóstico estudadas têm o potencial de apontar observações atípicas, que podem influir na análise. Desta forma, o pesquisador deve examinar com mais cautela essas observações, verificando a necessidade ou não da exclusão desses casos. Por outro lado, a escolha das medidas de discrepância depende do objetivo inferencial e também das características do problema apresentado.

Considerando o estudo de todas as medidas apresentadas no presente trabalho, sugere-se o uso da medida $\hat{\Delta}_{1(i)}^2$ de Critchley e Vitiello (1991):

$$\hat{\Delta}_{1(i)}^2 = \left(\frac{n_1 + n_2 - 1}{n_1 + n_2} \right) \left[\hat{\Delta}^2 - \frac{1}{w_1(n_1 - 1)} + \frac{\{e(\mathbf{x}_{1i}) - w_1^{-1}\}^2}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(\mathbf{x}_{1i})} \right],$$

uma vez que ela fornece diretamente a distância de Mahalanobis quando da eliminação da observação, e, consequentemente, pode-se avaliar também a alteração da probabilidade de classificação incorreta, visto que $PCI = \Phi\left(-\frac{\Delta}{2}\right)$.

Capítulo 8

Considerações finais

Neste trabalho, procurou-se descrever as técnicas de diagnóstico em análise discriminante. Essas técnicas são exaustivamente estudadas em regressão mas pouco exploradas em análise discriminante. O objetivo foi agrupar em um só texto as medidas de diagnóstico que foram propostas pelos principais pesquisadores do assunto nos últimos 26 anos.

Foram apresentadas várias medidas usadas quando as matrizes de covariâncias das populações são iguais, a comparação entre algumas dessas medidas, e também a abordagem Bayesiana para o assunto. Tratou-se ainda das medidas para a análise discriminante quadrática e múltipla. Finalizando, foi utilizado um conjunto de dados reais para testar a eficiência das medidas propostas. Face a inexistência de programas que calculem essas medidas diretamente, indicase o uso do Apêndice C, em que algumas das principais medidas de diagnóstico para análise discriminante foram calculadas através do software S-plus

Concluindo, sugere-se como primeiro passo da análise examinar aquelas observações identificadas como mais influentes, pois algumas delas podem se tratar de erros, devendo ser corrigidas ou descartadas. Para as outras observações influentes

que permanecerem no conjunto de dados, é importante verificar como as inferências de interesse seriam alteradas se esses casos não tivessem sido observados. Sendo assim, o pesquisador poderá decidir sobre a permanência ou não desses *outliers* em seu estudo, e talvez até partir para uma análise mais aprofundada dessas observações.

No presente trabalho procedeu-se à análise de influência eliminando-se uma observação de cada vez. No entanto os dados da análise discriminante podem conter observações de influência múltipla. Examinar a influência conjunta de várias observações é sempre mais difícil do que a análise de uma observação individualmente, isso porque além do problema computacional, existe o problema dos possíveis efeitos de "mascaramento" (*masking effect*).

Propõem-se então, para um trabalho futuro, o estudo de técnicas de diagnóstico quando da eliminação em bloco das observações, para analisar uma possível influência múltipla.

Apêndice A

A-1 Conceitos básicos em análise discriminante

Esta seção do Apêndice foi elaborada tomando por base o Capítulo 11 de Johnson e Wichern (1992).

A-1.1 Função discriminante de Fisher para duas populações

Análise discriminante é uma importante técnica multivariada. Um exemplo prático de sua utilização é aquele em que um médico, com base nos sintomas de seu paciente e resultados de exames preliminares, deve decidir se o paciente é portador ou não de certa doença.

Neste caso, $\mathbf{X}' = [x_1, x_2, \dots, x_p]$ é o vetor aleatório contendo as p variáveis para cada paciente, π_1 é a população de pessoas portadoras da doença e π_2 a população de não-portadores da doença.

Sejam π_1 e π_2 as duas populações e suas funções densidade $f_1(\mathbf{x})$ (com média $\boldsymbol{\mu}_1$) e $f_2(\mathbf{x})$ (com média $\boldsymbol{\mu}_2$) respectivamente, e matriz de covariância positiva definida Σ , comum às duas populações.

Seja $\mathbf{x}' = [x_1, x_2, \dots, x_p]$ os valores observados de \mathbf{X} na amostra.

A função discriminante de Fisher tem valor histórico como a primeira solução específica para a análise discriminante. A idéia de Fisher foi transformar as observações multivariadas \mathbf{x} em observações univariadas y (derivadas de π_1 e π_2) que, quando calculadas para elementos de populações distintas, sejam o mais separadas possível. Para essa transformação, Fisher usou combinações lineares de \mathbf{x} pela facilidade de manuseio. Os coeficientes da combinação linear eram obtidos visando a maximização do quociente entre a diferença dos valores de y (calculados para as médias populacionais μ_{1y} e μ_{2y}), e o desvio padrão de y estimado dentro das amostras, o que significa maximizar a distância entre as duas populações.

O método de Fisher não assume que as populações têm distribuição normal, mas supõe que as matrizes de covariância são iguais. Define-se então:

$y = \mathbf{l}'\mathbf{x}$ combinação linear das variáveis originais em que \mathbf{l} é um vetor $p \times 1$ de constantes e \mathbf{x} possui dimensão $p \times 1$, de modo que y é unidimensional,

(A-1.1)

$$\mu_{1y} = E(\mathbf{l}'\mathbf{x} / \pi_1) = \mathbf{l}'\boldsymbol{\mu}_1$$

$$\mu_{2y} = E(\mathbf{l}'\mathbf{x} / \pi_2) = \mathbf{l}'\boldsymbol{\mu}_2$$

$$\sigma_y^2 = \text{Var}(\mathbf{l}'\mathbf{x}) = \mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}$$

Se $\boldsymbol{\delta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, o objetivo é determinar o vetor \mathbf{l} que maximiza

$$\frac{(\mu_{1y} - \mu_{2y})^2}{\sigma_y^2} = \frac{(\mathbf{l}'\boldsymbol{\mu}_1 - \mathbf{l}'\boldsymbol{\mu}_2)^2}{\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}} = \frac{(\mathbf{l}'\boldsymbol{\delta})^2}{\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}}$$

tomando $\mathbf{l} = c\boldsymbol{\Sigma}^{-1}\boldsymbol{\delta}$, para qualquer $c \neq 0$.

Escolhendo $c = 1$, obtém-se

(A-1.2)

$$y = \mathbf{l}'\mathbf{x} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}\mathbf{x}$$

que é a função discriminante de Fisher.

Verifica-se que, para $\mathbf{l} = \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$, o valor máximo de $\frac{(\mathbf{l}'\boldsymbol{\delta})^2}{\mathbf{l}'\Sigma\mathbf{l}}$ é dado

$$(A-1.3)$$

$$\text{por } \Delta^2 = \boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)'\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2),$$

que é a distância de Mahalanobis.

Desta forma, as quantidades de interesse envolvidas na análise são

$$\begin{aligned} \text{vetor de coeficientes:} & \quad \mathbf{l} = \Sigma^{-1}\boldsymbol{\delta} \\ \text{média discriminante:} & \quad \mathbf{l}'\boldsymbol{\mu}_j, \quad j=1, 2 \\ \text{distância de Mahalanobis:} & \quad \Delta^2 = \boldsymbol{\delta}'\Sigma^{-1}\boldsymbol{\delta} \end{aligned}$$

No entanto, em situações práticas $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ e Σ não são conhecidos e são substituídos por seus estimadores não viciados, descritos a seguir.

Sejam n_1 observações de π_1 , n_2 observações de π_2 , com $n_1 + n_2 - 2 \geq p$.

As matrizes de dados são

$\mathbf{X}_{1(p \times n_1)} = (\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{1n_1})$ e $\mathbf{X}_{2(p \times n_2)} = (\mathbf{x}_{21}, \mathbf{x}_{22}, \dots, \mathbf{x}_{2n_2})$, em que

$\mathbf{x}_{1i} = (x_{1i1}, x_{1i2}, \dots, x_{1ip})$ $i = 1, \dots, n_1$ e $\mathbf{x}_{2i} = (x_{2i1}, x_{2i2}, \dots, x_{2ip})$ $i = 1, \dots, n_2$ ou seja,

cada coluna da matriz representa um elemento da amostra com suas p variáveis;

$$\hat{\boldsymbol{\mu}}_1 = \bar{\mathbf{x}}_{1(p \times 1)} = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{x}_{1i}, \quad \hat{\boldsymbol{\mu}}_2 = \bar{\mathbf{x}}_{2(p \times 1)} = \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{x}_{2i} \quad \text{e}$$

$$(A-1.4)$$

$$\hat{\Sigma} = \mathbf{S}_{(\text{ponderada})} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{(n_1 + n_2 - 2)},$$

que será denotada apenas como \mathbf{S} ,

$$\text{sendo que } \mathbf{S}_{1(p \times p)} = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)'$$

$$\mathbf{S}_{2(p \times p)} = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)'$$

A distância de Mahalanobis é estimada por

$$D^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2). \quad (\text{A-1.5})$$

Uma vez encontrada a função discriminante, é necessário estabelecer uma regra de classificação para uma nova observação \mathbf{x}_0 .

Seja $y_0 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{x}_0$ o valor da função discriminante para uma nova observação \mathbf{x}_0 e $m = \frac{1}{2}(\mu_{1y} + \mu_{2y})$.

$$\quad (\text{A-1.6})$$

Escrevendo m em função de $\boldsymbol{\mu}_j$,

$$m = \frac{1}{2}(\mathbf{1}' \boldsymbol{\mu}_1 + \mathbf{1}' \boldsymbol{\mu}_2) = \frac{1}{2}[\mathbf{1}'(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)] = \frac{1}{2}((\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)). \quad (\text{A-1.7})$$

Calculando a distância da média de y_0 em relação a m para \mathbf{x}_0 pertencente a π_1 , obtém-se

$$\begin{aligned} E(y_0 / \pi_1) - m &= \left[(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1 - \frac{1}{2} [(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)] \right] \\ &= \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \geq 0, \text{ pois } \boldsymbol{\Sigma} \text{ é positiva definida.} \end{aligned}$$

Assim, se $E(y_0 / \pi_1) - m \geq 0$, aloca-se \mathbf{x}_0 em π_1 e analogamente, se $E(y_0 / \pi_1) - m < 0$, aloca-se \mathbf{x}_0 em π_2 .

Da mesma forma que descrito anteriormente, na prática usam-se os estimadores não viciados dos parâmetros desconhecidos. Desse modo a regra de alocação baseada na função discriminante de Fisher fica:

alocar \mathbf{x}_0 em π_1 se $FD(\mathbf{x}) = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} \mathbf{x}_0 - \hat{m} \geq 0$

$$\quad (\text{A-1.8})$$

e alocar \mathbf{x}_0 em π_2 se $FD(\mathbf{x}) < 0$,

sendo que $\hat{m} = \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)$.

A-1.2 Probabilidade de classificação incorreta

As regras de classificação não estão livres de erros, pois as características que distinguem as duas populações podem não ser tão evidentes, surgindo assim a possibilidade de classificar incorretamente uma observação em π_2 quando na verdade ela pertence a π_1 , ou classificar como π_1 , quando a observação pertence a π_2 . A idéia então é criar uma regra que minimize a chance de cometer esses erros.

Outro ponto a ser observado é o custo, ou seja, em muitos casos é mais sério cometer um tipo de erro do que o outro. Por exemplo, diagnosticar erradamente que o paciente não tem uma doença fatal é muito mais grave do que concluir que o paciente tem a doença quando na realidade não tem. Além disso, uma população pode ter maior probabilidade de ocorrência do que outra (denominada de probabilidade *a priori*).

Em resumo, um bom procedimento de classificação além de minimizar as probabilidades de erro, deve levar em conta as probabilidades *a priori* e os custos associados às classificações incorretas.

Designa-se então por R_1 ao conjunto de valores de \mathbf{X} para os quais as observações são classificadas como π_1 e por $R_2 = \Omega - R_1$, ao conjunto de valores restantes de \mathbf{X} , para os quais as observações são classificadas como π_2 , em que

$$R_1 \cap R_2 = \emptyset$$

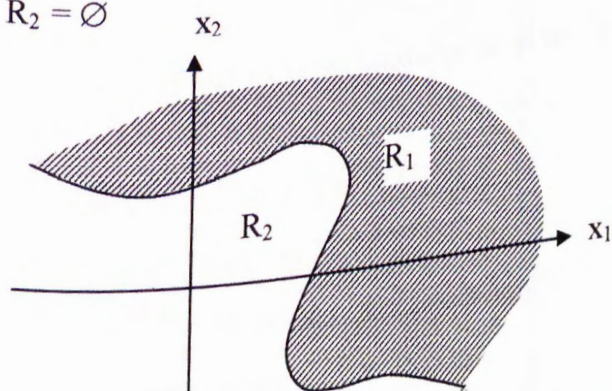


Figura A-1.1 - Divisão do espaço amostral (Ω) em regiões de classificação

São assumidas as seguintes notações:

q = probabilidade *a priori* de π_1

$1 - q$ = probabilidade *a priori* de π_2

$P(i/j)$ = probabilidade de classificar uma observação em π_i quando ela é de π_j

$C(i/j)$ = custo de classificar uma observação em π_i quando de fato ela pertence a π_j ,

$i, j = 1, 2, i \neq j$.

Assim, a probabilidade condicional de classificar uma observação como π_1 , quando na realidade é de π_2 é:

$$P(1/2) = P(\mathbf{X} \in R_1 / \pi_2) = \int_{R_1} f_2(\mathbf{x}) d\mathbf{x}$$

e, do mesmo modo, a probabilidade condicional de classificar uma observação como π_2 , quando de fato ela é de π_1 é:

$$P(2/1) = P(\mathbf{X} \in R_2 / \pi_1) = \int_{R_2} f_1(\mathbf{x}) d\mathbf{x}$$

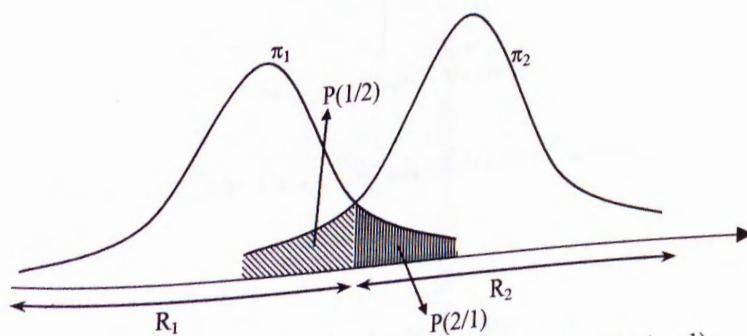


Figura A-1.2 - Probabilidade de classificação incorreta ($p = 1$)

Pode-se calcular as probabilidades de classificação certa e errada do seguinte modo:

$$P(\text{classificação correta em } \pi_1) = P(1/1) q,$$

$$P(\text{classificação correta em } \pi_2) = P(2/2) (1 - q),$$

$$P(\text{classificação incorreta em } \pi_1) = P(1/2) (1 - q),$$

$$P(\text{classificação incorreta em } \pi_2) = P(2/1) q.$$

As regiões R_1 e R_2 devem ser determinadas de forma a minimizar a probabilidade total de classificação incorreta denotada por PCI ,

$$\begin{aligned}
 PCI &= P(\text{classificar incorretamente em } \pi_1) + P(\text{classificar incorretamente em } \pi_2) \\
 &= P(1/2)(1-q) + P(2/1)q \\
 &= (1-q) \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} + q \int_{R_2} f_1(\mathbf{x}) d\mathbf{x}.
 \end{aligned}$$

Mas, como já foi mencionado, deve-se levar em conta o custo do erro na classificação. Conforme tabela a seguir, tem-se custo zero para classificações corretas ($C(i/i) = 0$), e um custo para classificações incorretas ($C(i/j) > 0$ para $i \neq j$).

População	Classificação	
	π_1	π_2
π_1	0	$C(2/1)$
π_2	$C(1/2)$	0

Tabela A-1.1 - Custos de Classificação

Dessa forma, o custo médio de classificação incorreta, $CMCI$, pode ser calculado como:

$$CMCI = C(1/2)P(1/2)(1-q) + C(2/1)P(2/1)q.$$

Uma boa regra de classificação deve ter um $CMCI$ tão pequeno quanto possível.

Um resultado importante é que as regiões R_1 e R_2 que minimizam $CMCI$

são assim definidas:

$$R_1: \text{ região dos valores de } \mathbf{X} \text{ em que } \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{C(1/2)}{C(2/1)} \cdot \frac{1-q}{q}$$

$$R_2: \text{ região dos valores de } \mathbf{X} \text{ em que } \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{C(1/2)}{C(2/1)} \cdot \frac{1-q}{q}$$

Uma separação significativa entre as populações não implica em uma boa regra de alocação. A eficácia de uma regra de classificação pode ser avaliada pelo cálculo de sua *taxa de erro* (ou probabilidade de classificação incorreta).

A-1.3 Duas populações normais multivariadas

a) Matrizes de covariância iguais ($\Sigma_1 = \Sigma_2$)

Seja $f_1(\mathbf{x})$ uma densidade normal multivariada com vetor de médias μ_1 e matriz de covariância Σ_1 e $f_2(\mathbf{x})$ densidade normal multivariada com vetor de médias μ_2 e matriz de covariância Σ_2 , sendo que $\Sigma_1 = \Sigma_2 = \Sigma$.

A função densidade de probabilidade de \mathbf{X} na população π_j é dada por:

$$f_j(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu_j)' \Sigma^{-1}(\mathbf{x} - \mu_j)\right] \quad \text{para } j=1, 2,$$

em que $|\Sigma|$ é o determinante de Σ .

Verifica-se que as regiões que minimizam CMCI são:

$$R_1: \exp\left[-\frac{1}{2}(\mathbf{x} - \mu_1)' \Sigma^{-1}(\mathbf{x} - \mu_1) + \frac{1}{2}(\mathbf{x} - \mu_2)' \Sigma^{-1}(\mathbf{x} - \mu_2)\right] \geq \frac{C(1/2)}{C(2/1)} \cdot \frac{1-q}{q}$$

$$R_2: \exp\left[-\frac{1}{2}(\mathbf{x} - \mu_1)' \Sigma^{-1}(\mathbf{x} - \mu_1) + \frac{1}{2}(\mathbf{x} - \mu_2)' \Sigma^{-1}(\mathbf{x} - \mu_2)\right] < \frac{C(1/2)}{C(2/1)} \cdot \frac{1-q}{q}$$

Desenvolvendo o primeiro membro das desigualdades acima, usando logaritmo e reagrupando convenientemente tem-se:

$$-\frac{1}{2}(\mathbf{x}'\Sigma^{-1} - \boldsymbol{\mu}'_1\Sigma^{-1})(\mathbf{x} - \boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x}'\Sigma^{-1} - \boldsymbol{\mu}'_2\Sigma^{-1})(\mathbf{x} - \boldsymbol{\mu}_2) =$$

$$-\frac{1}{2}(\mathbf{x}'\Sigma^{-1}\mathbf{x} - \mathbf{x}'\Sigma^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}'_1\Sigma^{-1}\mathbf{x} + \boldsymbol{\mu}'_1\Sigma^{-1}\boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x}'\Sigma^{-1}\mathbf{x} - \mathbf{x}'\Sigma^{-1}\boldsymbol{\mu}_2 - \boldsymbol{\mu}'_2\Sigma^{-1}\mathbf{x} + \boldsymbol{\mu}'_2\Sigma^{-1}\boldsymbol{\mu}_2)$$

então a regra de alocação que minimiza *CMCI* é:

$$\text{alocar } \mathbf{x}_0 \text{ em } \pi_1 \text{ se} \quad (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} \mathbf{x}_0 - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq \ln \left[\frac{C(1/2)}{C(2/1)} \cdot \frac{1-q}{q} \right] \quad (\text{A-1.9})$$

$$\text{e alocar } \mathbf{x}_0 \text{ em } \pi_2 \text{ se} \quad (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} \mathbf{x}_0 - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) < \ln \left[\frac{C(1/2)}{C(2/1)} \cdot \frac{1-q}{q} \right].$$

Quando R_1 e R_2 são determinadas adequadamente a fim de se obter o valor mínimo para a *PCI*, tem-se então a taxa de erro ótima (*TEO*).

Supondo $q = (1-q) = \frac{1}{2}$ e $C(1/2) = C(2/1)$, as regras para o custo médio esperado mínimo, *CMCI*, e para a mínima *PCI* coincidem.

Assim, com $\ln \left[\frac{C(1/2)}{C(2/1)} \cdot \frac{1-q}{q} \right] = 0$, tem-se:

$$R_1: (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} \mathbf{x}_0 - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq 0$$

$$R_2: (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} \mathbf{x}_0 - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) < 0$$

que podem ser expressos em termos de $y = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1} \mathbf{x} = \mathbf{l}' \mathbf{x}$ como

$$R_{1(y)}: y \geq \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \quad \text{e}$$

$$R_{2(y)}: y < \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \Sigma^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2).$$

Mas Y é uma combinação linear de variáveis aleatórias, então as densidades de probabilidade de Y , $f_1(y)$ e $f_2(y)$, são normais univariadas (Resultado B-3 do Apêndice B) com médias e variâncias dadas por:

$$\mu_{1y} = \mathbf{l}' \boldsymbol{\mu}_1 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1$$

$$\mu_{2y} = \mathbf{l}' \boldsymbol{\mu}_2 = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_2$$

$$\sigma_y^2 = \mathbf{l}' \boldsymbol{\Sigma} \mathbf{l} = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \Delta^2$$

$$\text{e } PCI = \frac{1}{2} P(2/1) + \frac{1}{2} P(1/2). \quad (\text{A-1.10})$$

Como $P(2/1) = P\left(Y < \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right)$,

padronizando-se, tem-se

$$P(2/1) = P\left(\frac{Y - \mu_{1y}}{\sigma_y} < \frac{\frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_1}{\Delta}\right)$$

$$= P\left(Z < \frac{-\frac{1}{2} \Delta^2}{\Delta}\right) = \Phi\left(\frac{-\Delta}{2}\right)$$

sendo que $\Phi(\cdot)$ é a função de distribuição acumulada da distribuição normal padrão.

Do mesmo modo,

$$P(1/2) = P\left(Y \geq \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right)$$

$$= P\left(Z \geq \frac{\Delta}{2}\right) = 1 - \Phi\left(\frac{\Delta}{2}\right) = \Phi\left(\frac{-\Delta}{2}\right).$$

(A-1.11)

Dessa forma, $PCI = \frac{1}{2} \Phi\left(\frac{-\Delta}{2}\right) + \frac{1}{2} \Phi\left(\frac{-\Delta}{2}\right) = \Phi\left(\frac{-\Delta}{2}\right).$

b) **Matrizes de covariância diferentes ($\Sigma_1 \neq \Sigma_2$)**

A regra de classificação fica mais complicada quando as matrizes de covariância não são iguais, pois o termo $-\frac{1}{2} \mathbf{x}' (\Sigma_1^{-1} - \Sigma_2^{-1}) \mathbf{x}$ permanece quando do cálculo de $\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})}$, diferentemente do caso onde $\Sigma_1 = \Sigma_2$.

A regra de alocação que minimiza o CMCI é dada por: alocar \mathbf{x}_0 em π_1 , se

$$-\frac{1}{2} \mathbf{x}_0' (\Sigma_1^{-1} - \Sigma_2^{-1}) \mathbf{x}_0 + (\mu_1' \Sigma_1^{-1} - \mu_2' \Sigma_2^{-1}) \mathbf{x}_0 - k \geq \ln \left[\frac{C(1/2) \cdot \frac{1-q}{q}}{C(2/1)} \right],$$

em que $k = \frac{1}{2} \ln \left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) + \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2)$, e alocar \mathbf{x}_0 em π_2 , caso contrário.

Como no caso em que $\Sigma_1 = \Sigma_2$ usa-se, na prática, os estimadores não viciados $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{S}_1$ e \mathbf{S}_2 para μ_1, μ_2, Σ_1 e Σ_2 , respectivamente.

Observa-se portanto que as regiões de classificação são agora definidas por funções quadráticas.

Regras de classificação com funções quadráticas para mais que duas dimensões podem gerar resultados incoerentes, além de não ser um processo robusto na ausência de normalidade. Já a regra de classificação linear é a mais utilizada nos softwares disponíveis. Sua estrutura é mais simples e baseada em suposições de normalidade e igualdade das matrizes de covariância.

Antes de se adotar uma regra, deve-se testar primeiro as suposições de normalidade. Se os dados não seguem uma distribuição normal multivariada, pode-se transformá-los aproximando-os para uma distribuição normal. Depois disso testa-se a igualdade das matrizes de covariância para a escolha do método mais adequado:

linear (para matrizes de covariâncias iguais) ou quadrático (quando as matrizes de covariâncias são diferentes).

O ideal é sempre verificar como se comporta o método de classificação escolhido. Um modo de se fazer essa verificação é dividir a amostra usando uma parte para estabelecer a regra de alocação (conjunto teste) e a outra para avaliar o desempenho da regra escolhida (conjunto de validação).

A-1.4 Discriminação entre g populações

Análise discriminante múltipla é o método de classificação de uma observação \mathbf{x}_0 , de origem desconhecida, em uma de g populações. Tais populações são independentes, com vetores de médias μ_j e matriz de covariância Σ_j , $j = 1, 2, \dots, g$. A probabilidade *a priori* da população π_j é q_j , $j = 1, 2, \dots, g$, e o custo de alocar uma observação em π_k quando na realidade ela pertence a π_j é $C(k/j)$, para $k, j = 1, 2, \dots, g$, $k \neq j$.

A análise discriminante múltipla pode ser realizada de acordo com os seguintes métodos principais.

a) Método do custo médio de classificação incorreta

Com g populações, o custo médio de classificar uma observação que pertence a π_1 , em uma das $g-1$ populações é dado por:

$$CMCI_1 = P(2/1)C(2/1) + P(3/1)C(3/1) + \dots + P(g/1)C(g/1) = \sum_{j=2}^g P(j/1)C(j/1).$$

Esse custo vai ocorrer com probabilidade q_1 que é a probabilidade *a priori* de π_1 . Então o cálculo do *CMCI* total será:

$$CMCI = q_1 CMCI_1 + q_2 CMCI_2 + \dots + q_g CMCI_g = \sum_{i=1}^g q_i \left(\sum_{\substack{j=1 \\ j \neq i}}^g P(j/i) C(j/i) \right).$$

O problema é definir as regiões de classificação R_1, R_2, \dots, R_g que tornem esse custo mínimo. Verifica-se que essas regiões de classificação são obtidas alocando-se a observação \mathbf{x}_0 na população π_j para a qual $\sum_{\substack{i=1 \\ i \neq j}}^g q_i f_i(\mathbf{x}_0) C(j/i)$ é a

menor.

Se todos os custos de classificação incorreta são iguais, a regra que minimiza o custo médio de classificação incorreta é a mesma que a regra da menor probabilidade de classificação incorreta, ou seja para a menor $\sum_{\substack{i=1 \\ i \neq j}}^g q_i f_i(\mathbf{x}_0)$.

Assim, essa soma será mínima quando a parcela omitida $q_j f_j(\mathbf{x}_0)$ é máxima. Dessa forma, a regra do mínimo custo médio de classificação incorreta com custos iguais fica:

alocar \mathbf{x}_0 para π_j se $q_j f_j(\mathbf{x}_0) > q_i f_i(\mathbf{x}_0)$, para todo $i \neq j$.

Quando π_j tem distribuição normal multivariada, considerando todos os custos de classificação incorreta iguais, tem-se:

$$\ln q_j f_j(\mathbf{x}) = \ln q_j - \left(\frac{p}{2} \right) \ln(2\pi) - \frac{1}{2} \ln |\Sigma_j| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_j)' \Sigma_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j).$$

Eliminando a constante $\left(\frac{p}{2} \right) \ln(2\pi)$, que é a mesma para as g populações,

o escore de classificação quadrática é definido por

$$d_j^Q(\mathbf{x}) = -\frac{1}{2} \ln |\Sigma_j| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_j)' \Sigma_j^{-1} (\mathbf{x} - \boldsymbol{\mu}_j) + \ln q_j, \quad j = 1, 2, \dots, g.$$

A regra de alocação definida anteriormente é equivalente a alocar \mathbf{x}_0 na população para a qual $d_j^Q(\mathbf{x}_0)$ é o maior.

Uma vez que nas aplicações usa-se $\bar{\mathbf{x}}_j$ e \mathbf{S}_j , a regra estimada para a mínima probabilidade de classificação incorreta para várias populações normais com variâncias diferentes fica:

alocar \mathbf{x}_0 para π_j se o escore quadrático $\hat{d}_j^Q(\mathbf{x})$ é o maior dos $d_1^Q(\mathbf{x}), d_2^Q(\mathbf{x}), \dots, d_g^Q(\mathbf{x})$,

sendo que

$$\hat{d}_j^Q(\mathbf{x}) = -\frac{1}{2} \ln |\mathbf{S}_j| - \frac{1}{2} (\mathbf{x} - \bar{\mathbf{x}}_j)' \mathbf{S}_j^{-1} (\mathbf{x} - \bar{\mathbf{x}}_j) + \ln q_j, \quad j = 1, \dots, g.$$

Se as matrizes de covariâncias forem iguais ($\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$) o escore de classificação quadrática se reduz a

$$d_j^Q(\mathbf{x}) = -\frac{1}{2} \ln |\Sigma| - \frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{x} + \boldsymbol{\mu}_j' \Sigma^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}_j' \Sigma^{-1} \boldsymbol{\mu}_j + \ln q_j, \quad j = 1, \dots, g.$$

Da mesma forma que mencionado acima, ignorando os termos $-\frac{1}{2} \ln |\Sigma| - \frac{1}{2} \mathbf{x}' \Sigma^{-1} \mathbf{x}$ que são os mesmos para todos os valores de j , fica então definido o escore de classificação linear:

$$d_j(\mathbf{x}) = \boldsymbol{\mu}_j' \Sigma^{-1} \mathbf{x} - \frac{1}{2} \boldsymbol{\mu}_j' \Sigma^{-1} \boldsymbol{\mu}_j + \ln q_j, \quad j = 1, \dots, g.$$

Na prática usa-se $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_g$ para estimar $\boldsymbol{\mu}_j, j = 1, 2, \dots, g$ e

$$\mathbf{S} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + \dots + (n_g - 1)\mathbf{S}_g}{n_1 + n_2 + \dots + n_g - g}$$
 como estimador de Σ .

A regra estimada para a mínima probabilidade de classificação incorreta com matrizes de covariâncias iguais é dada por:

alocar \mathbf{x}_0 em π_j se o escore de classificação linear $\hat{d}_j(\mathbf{x})$ é o maior dentre os escores

$$\hat{d}_1(\mathbf{x}), \hat{d}_2(\mathbf{x}), \dots, \hat{d}_g(\mathbf{x}), \text{ em que } \hat{d}_j(\mathbf{x}) = \bar{\mathbf{x}}_j' \mathbf{S}^{-1} \mathbf{x} - \frac{1}{2} \bar{\mathbf{x}}_j' \mathbf{S}^{-1} \bar{\mathbf{x}}_j + \ln q_j \quad j = 1, \dots, g.$$

b) Método de Fisher

No método de Fisher, não é necessária a suposição de que as g populações tenham distribuição normal multivariada, mas é assumido que as matrizes de covariância sejam iguais e de posto completo.

$$\text{Seja } \bar{\boldsymbol{\mu}} = \frac{1}{g} \sum_{j=1}^g \boldsymbol{\mu}_j$$

$$\text{e } \mathbf{B}_0 = \sum_{j=1}^g (\boldsymbol{\mu}_j - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_j - \bar{\boldsymbol{\mu}})' \quad \text{a matriz da soma de quadrados entre os$$

grupos.

Considerando a combinação linear $Y = \mathbf{l}'\mathbf{x}$ com

$$E(Y) = E(\mathbf{l}'\mathbf{x}) = \mathbf{l}'\boldsymbol{\mu}_j = \mu_{jy} \quad \text{para cada população } \pi_j \text{ e}$$

$$\text{Var}(Y) = \mathbf{l}'\boldsymbol{\Sigma}\mathbf{l} = \sigma_y^2 \quad \text{para todas as populações,}$$

o objetivo é maximizar $\sum_{j=1}^g \frac{(\mu_{jy} - \bar{\mu}_y)^2}{\sigma_y^2}$ (extensão do caso com duas populações),

com $\bar{\mu}_y = \mathbf{l}'\bar{\boldsymbol{\mu}}$.

$$\text{Como } \sum_{j=1}^g \frac{(\mu_{jy} - \bar{\mu}_y)^2}{\sigma_y^2} = \sum_{j=1}^g \frac{(\mathbf{l}'\boldsymbol{\mu}_j - \mathbf{l}'\bar{\boldsymbol{\mu}})^2}{\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}} = \frac{\mathbf{l}'\mathbf{B}_0\mathbf{l}}{\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}}, \text{ verifica-se que o valor}$$

de \mathbf{l} que maximiza $\frac{\mathbf{l}'\mathbf{B}_0\mathbf{l}}{\mathbf{l}'\boldsymbol{\Sigma}\mathbf{l}}$ (variabilidade entre os grupos / variabilidade dentro dos grupos) é dado por $\mathbf{l}_j = \mathbf{e}_j$ onde \mathbf{e}_j são os autovalores padronizados ($\mathbf{e}_j'\boldsymbol{\Sigma}\mathbf{e}_j = 1$) correspondentes aos s autovetores de $\boldsymbol{\Sigma}^{-1}\mathbf{B}_0$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s > 0$ onde $s \leq [\min(g-1), p]$.

Ficam então definidas

$l_1' \mathbf{x}$ = primeira função discriminante

$l_2' \mathbf{x}$ = segunda função discriminante

\vdots

$l_s' \mathbf{x}$ = s-ésima função discriminante.

A regra de classificação baseada nas $r \leq s$ primeiras funções discriminantes é:

alocar \mathbf{x}_0 em π_j se $\sum_{j=1}^r (l_j'(\mathbf{x}_0 - \bar{\mathbf{x}}_j))^2 \leq \sum_{j=1}^r (l_j'(\mathbf{x}_0 - \bar{\mathbf{x}}_k))^2$ para todo $k \neq j$.

A-2 Conceitos básicos sobre diagnóstico em regressão

Análise de diagnóstico é uma etapa da análise de regressão em que se verifica os possíveis afastamentos das suposições iniciais do modelo, bem como as observações extremas (*outliers*) que afetam desproporcionalmente os resultados do ajuste.

Os métodos de diagnóstico tiveram início no começo dos anos 1970 com a análise de resíduos para detectar a presença de pontos extremos e avaliar a conformidade entre os dados e o modelo assumido, pois se o modelo ajustado não produz um conjunto de resíduos que parece ser razoável então algum aspecto do modelo não está bem adequado.

Em 1975 a análise de resíduos já era parte integrante da análise de regressão. Foi então que se desenvolveram métodos para calcular a influência de observações individuais, permitindo uma maior compreensão das análises de dados. Na análise de influência, uma pequena perturbação que produz um grande efeito deve ser levada em conta no procedimento estatístico.

Inferências baseadas nos estimadores de mínimos quadrados podem ser fortemente influenciadas por poucas observações extremas do conjunto de dados, e conseqüentemente, o modelo ajustado pode refletir características incomuns desses dados, quando o correto seria refletir o relacionamento global entre todas as variáveis e todos os dados. Se essas observações forem ignoradas, o resultado da análise pode mudar totalmente. São os chamados *pontos influentes*.

Portanto, a análise de diagnóstico tem dois objetivos principais:

- permitir o reconhecimento de fenômenos importantes pela detecção de *outliers*, que podem ter maior relevância do que o estudo do conjunto dos dados, e
- sugerir ações reparadoras para o modelo ajustado inicialmente.

São duas as principais estatísticas do diagnóstico na regressão linear: o ponto de alavanca (h_i) e o resíduo (r_i), sendo que várias medidas de diagnóstico podem ser expressas em função dessas duas estatísticas.

Pontos de alavanca e resíduos

Considera-se o modelo de regressão linear:

$$y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + e_i \quad i = 1, \dots, n$$

onde os erros e_i 's são variáveis aleatórias independentes, normalmente distribuídas com média zero e variância constante σ^2 .

Simplificando a notação para $y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ (onde y , $\boldsymbol{\beta}$ e $\boldsymbol{\varepsilon}$ são vetores e \mathbf{X} é matriz) o estimador de máxima verossimilhança para $\boldsymbol{\beta}$ é: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$,

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Na previsão: $\hat{y} = \mathbf{X}\hat{\boldsymbol{\beta}}$

$$\hat{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y = \mathbf{H}y$$

então $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ é denominada "matriz *hat*" porque $\mathbf{H}y = \hat{y}$ ou seja, "coloca chapéu" no y . Essa matriz é simétrica e idempotente ($\mathbf{H}^2 = \mathbf{H}$).

O resíduo, que mede a diferença entre o valor observado e o valor ajustado para a i -ésima observação é dado por $r_i = y_i - \hat{y}_i$, $i = 1, 2, \dots, n$.

Matricialmente,

$$\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = [\mathbf{I} - \mathbf{H}]\mathbf{y} \quad e$$

$$\text{Var}(\mathbf{r}) = (\mathbf{I} - \mathbf{H})\text{Var}(\mathbf{y})(\mathbf{I} - \mathbf{H})'$$

Como \mathbf{H} é idempotente,

$$\text{Var}(\mathbf{r}) = \sigma^2(\mathbf{I} - \mathbf{H})$$

$$\text{Var}(r_i) = \sigma^2(1 - h_{ii}) \quad (h_{ii} < 1)$$

$$\text{Cov}(r_i, r_j) = -\sigma^2 h_{ij}, \quad i \neq j$$

$$\text{Como } \hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \text{ então } \hat{y}_i = h_{ii}y_i + \sum_{j=1, j \neq i}^n h_{ij}y_j, \quad (\text{A-2.1})$$

de modo que h_{ij} mede a influência exercida por y_j em \hat{y}_i e h_{ii} mede a influência exercida por y_i em \hat{y}_i .

Devido ao fato que $\mathbf{H}\mathbf{H} = \mathbf{H}$, prova-se que se $h_{ii} = 0$ então $\hat{y}_i = 0, \forall y_i$ (\hat{y}_i não é afetado por y_i) e se $h_{ii} = 1$, $\hat{y}_i = y_i$.

Para valores altos de h_{ii} na expressão A-2.1 predomina a influência de y_i sobre o valor ajustado \hat{y}_i , por isso utiliza-se h_{ii} como uma medida de influência da i -ésima observação sobre o próprio valor ajustado. O efeito do i -ésimo caso na regressão é provavelmente maior se h_{ii} é alto, mas sua importância é incerta, pois depende dos y_j .

Portanto o elemento $h_{ii} = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$ desempenha um papel importante na construção de técnicas de diagnóstico.

Pontos com h_{ii} "alto" são chamados *pontos de alavancagem (high leverage)* por ter um peso desproporcional no próprio valor ajustado; geralmente destoam dos demais em alguma coordenada, trazendo o plano de regressão em sua direção, o que pode ou não distorcer algumas estimativas.

Existem critérios para considerar que o ponto é de alavancagem. Como \mathbf{H} é

idempotente, isso implica que $\text{posto}(\mathbf{H}) = \text{traço}(\mathbf{H}) = \sum_{i=1}^n h_{ii} = p$ e o valor médio de

h_{ii} é $\bar{h} = \frac{\sum_{i=1}^n h_{ii}}{n} = \frac{p}{n}$. Considera-se então como de alavanca aqueles pontos onde h_{ii}

é maior que duas vezes a sua média, ou seja $h_{ii} \geq 2\bar{h}$ ou $h_{ii} \geq \frac{2p}{n}$.

Em resumo, a análise de \mathbf{H} (que não depende de y), pode revelar pontos sensíveis, nos quais um valor discrepante em y pode ter um grande impacto no ajuste.

Medida de influência

Os métodos mais comuns para se detectar influência eliminam uma observação do conjunto de dados de cada vez.

Sejam $\hat{\beta}_{(i)}$ o estimador de β calculado sem a observação i

$\mathbf{X}_{(i)}$ a matriz $(n-1) \times p$ obtida excluindo-se a i -ésima linha de \mathbf{X}

$$\hat{\beta}_{(i)} = (\mathbf{X}_{(i)}' \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}' \mathbf{Y}_{(i)}$$

O problema é quantificar a distância entre $\hat{\beta}$ e $\hat{\beta}_{(i)}$. Uma possibilidade seria

utilizar

$$D_i = \frac{(\mathbf{x}\hat{\beta} - \mathbf{x}\hat{\beta}_{(i)})(\mathbf{x}\hat{\beta} - \mathbf{x}\hat{\beta}_{(i)})}{p \cdot s^2},$$

sendo que s^2 é o quadrado médio do resíduo após o ajuste do modelo.

Assim surge a medida D-Cook (distância de Cook):

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' (\mathbf{X}'\mathbf{X}) (\hat{\beta} - \hat{\beta}_{(i)})}{ps^2}, \quad i = 1, 2, \dots, n \quad (\text{A-2.2})$$

que também pode ser escrita em função das estatísticas h_{ii} e r_i como

$$D_i = \left(\frac{r_i}{s\sqrt{(1-h_{ii})}} \right)^2 \frac{h_{ii}}{1-h_{ii}} \frac{1}{p}.$$

Observa-se que D_i é crescente em r_i e h_{ii} , r_i reflete falta de ajuste no ponto i , h_{ii} reflete a posição de x_i com relação a \bar{x} e D_i alto é devido a $|r_i|$ alto ou h_{ii} alto ou ambos.

No entanto, a distância D_i pode não ser adequada quando r_i for grande e h_{ii} pequeno. Por esse motivo, Belsley, Kuh e Welsch (1980) propuseram a medida $DFFITs_i$, mais apropriada para medir influência nas estimativas dos coeficientes de um ponto aberrante com h_{ii} pequeno. Sua expressão é

$$DFFITs_i = \frac{|r_i|}{s_{(i)}\sqrt{(1-h_{ii})}} \left(\frac{h_{ii}}{1-h_{ii}} \right)^{\frac{1}{2}} \quad i = 1, 2, \dots, n \quad (\text{A-2.3})$$

em que $s_{(i)}^2$ é o quadrado médio do resíduo sem a observação i .

Apêndice B

Definições, resultados e cálculos

B-1 Definição de Ordem

Se $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ são seqüências de números reais, então $a_n = O(b_n)$ (a_n é de ordem b_n) se $\left| \frac{a_n}{b_n} \right|$ é limitada para $n \rightarrow \infty$, ou seja, se existir um número $m > 0$ e um inteiro $n_0 = n_0(m)$, tal que $\left| \frac{a_n}{b_n} \right| \leq m, \forall n \geq n_0$.

B-2 Distribuição quiquadrado não central

Se y tem distribuição normal p -variada com média μ e matriz de covariância Σ , $y \sim N_p(\mu, \Sigma)$, então $y' \Sigma^{-1} y$ tem distribuição quiquadrado não central com p graus de liberdade, e parâmetro de não centralidade $\mu' \Sigma^{-1} \mu$.

B-3 - Resultado

Se \mathbf{X} tem distribuição Normal p -variada de média $\boldsymbol{\mu}$ e matriz de covariância $\boldsymbol{\Sigma}$, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, então qualquer combinação linear das variáveis $\mathbf{a}'\mathbf{X} = a_1X_1 + a_2X_2 + \dots + a_pX_p$ é distribuída como normal univariada de média $\mathbf{a}'\boldsymbol{\mu}$ e variância $\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$.

Também, se $\mathbf{a}'\mathbf{X}$ é distribuída como $N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$ para todo \mathbf{a} , então \mathbf{X} tem distribuição $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

B-4 - Resultado

Se $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\mathbf{Z} = \mathbf{X} - \boldsymbol{\mu} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$, devido ao Resultado B-3, $\phi = \mathbf{1}'\mathbf{z} \sim N(0, \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1})$.

Dessa forma, $\frac{\mathbf{1}'\mathbf{z}}{\sqrt{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}} \sim N(0,1)$ e $\frac{\phi^2}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}} \sim \chi_1^2$.

Assim, $E\left(\frac{\phi^2}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right) = 1$, Mediana $\left(\frac{\phi^2}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right) = 0,455$,

$P\left(\frac{\phi^2}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}} \leq 0,455\right) = 0,5$ e $P(\phi^2 \leq 0,455 \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}) = 0,5$.

Então, $E(\phi^2) = \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}$, Mediana $(\phi^2) = 0,455 \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}$, ou seja, a média de ϕ^2 é maior que a mediana e portanto, ϕ^2 tem assimetria positiva.

Como consequência, $-w_1\Delta^2\phi_p^2$ tem assimetria negativa e como $w_1\Delta^2$ é constante e $2\Delta\phi_p$ é simétrica, então $I(x; \Delta^2)$ terá assimetria negativa.

B-5 Cálculos de $\hat{I}_{(0)}(\bar{x}_1 - \bar{x}_{1(0)})$, $\hat{I}_{(0)}(\bar{x}_1 - \bar{x}_2)$ e G^2

As fórmulas a seguir, apresentadas no Apêndice A de Fung (1995-a), foram utilizadas para os cálculos de $\hat{I}_{(0)}(\bar{x}_1 - \bar{x}_{1(0)})$, $\hat{I}_{(0)}(\bar{x}_1 - \bar{x}_2)$ e G^2 .

$$\hat{I}_{(0)} = c(\mathbf{I} - f_i \mathbf{S}^{-1} \mathbf{z}_i \mathbf{z}_i') (\mathbf{I} - h \mathbf{S}^{-1} \mathbf{z}_i) \text{ sendo que } c = \frac{n-3}{n-2}, \quad f_i = \frac{b}{1 + b d_i^2},$$

$$b = \frac{-n_1}{(n_1 - 1)(n - 2)}, \quad h = \frac{1}{n_1 - 1}, \quad \mathbf{z}_i = \mathbf{x}_{1i} - \bar{\mathbf{x}}_1$$

$$\bar{x}_1 - \bar{x}_{1(0)} = \frac{z_i}{n_1 - 1} \quad \text{e} \quad g_i = \frac{1}{1 + b d_i^2}.$$

B-5.1 Cálculo de $\hat{I}_{(0)}(\bar{x}_1 - \bar{x}_{1(0)})$:

$$\hat{I}_{(0)} = \frac{n-3}{n-2} \left(\mathbf{I} - \frac{b}{1 + b d_i^2} \mathbf{S}^{-1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1) (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)' \right) \cdot \left(\mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \frac{1}{(n_1 - 1)} \mathbf{S}^{-1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1) \right)$$

$$\hat{I}_{(0)} = \frac{n-3}{n-2} \left(\mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \frac{1}{n_1 - 1} \mathbf{S}^{-1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1) \right) \cdot \left(\mathbf{I} - \frac{b}{1 + b d_i^2} \mathbf{S}^{-1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1) (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)' \right)$$

$$\hat{I}_{(0)}(\bar{x}_1 - \bar{x}_{1(0)}) = \frac{n-3}{n-2} \left[\left((\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} - \frac{1}{n_1 - 1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)' \mathbf{S}^{-1} \right) \cdot \left(\mathbf{I} - \frac{b}{1 + b d_i^2} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1) (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)' \mathbf{S}^{-1} \right) \right] (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})$$

$$= \frac{n-3}{n-2} \left[(\bar{x}_1 - \bar{x}_2)' S^{-1} - \frac{(\bar{x}_1 - \bar{x}_2)' S^{-1} b (\mathbf{x}_{1i} - \bar{x}_1) (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1}}{1 + b d_i^2} - \frac{1}{n_1 - 1} (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1} + \frac{1}{n_1 - 1} (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1} \frac{b}{1 + b d_i^2} (\mathbf{x}_{1i} - \bar{x}_1) (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1} \right] (\bar{x}_1 - \bar{x}_{1(0)})$$

$$= \frac{n-3}{n-2} \left[(\bar{x}_1 - \bar{x}_2)' S^{-1} \frac{(\mathbf{x}_{1i} - \bar{x}_1)}{n_1 - 1} - (\bar{x}_1 - \bar{x}_2)' S^{-1} \frac{b}{1 + b d_i^2} (\mathbf{x}_{1i} - \bar{x}_1) (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1} \frac{(\mathbf{x}_{1i} - \bar{x}_1)}{n_1 - 1} - \frac{1}{n_1 - 1} (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1} \frac{(\mathbf{x}_{1i} - \bar{x}_1)}{n_1 - 1} + \frac{1}{n_1 - 1} (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1} \frac{b}{1 + b d_i^2} (\mathbf{x}_{1i} - \bar{x}_1) (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1} \frac{(\mathbf{x}_{1i} - \bar{x}_1)}{n_1 - 1} \right]$$

$$= \frac{n-3}{n-2} \left[(\bar{x}_1 - \bar{x}_2)' S^{-1} \frac{(\mathbf{x}_{1i} - \bar{x}_1)}{(n_1 - 1)} - (\bar{x}_1 - \bar{x}_2)' S^{-1} \frac{b}{1 + b d_i^2} (\mathbf{x}_{1i} - \bar{x}_1) \frac{d_i^2}{n_1 - 1} - \frac{1}{n_1 - 1} (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1} \frac{(\mathbf{x}_{1i} - \bar{x}_1)}{n_1 - 1} + \frac{1}{n_1 - 1} (\mathbf{x}_{1i} - \bar{x}_1)' S^{-1} \frac{b}{1 + b d_i^2} (\mathbf{x}_{1i} - \bar{x}_1) \frac{d_i^2}{n_1 - 1} \right]$$

$$= \frac{n-3}{n-2} \left[\frac{\hat{\phi}_i}{n_1 - 1} - \frac{\hat{\phi}_i}{n_1 - 1} \frac{b d_i^2}{1 + b d_i^2} - \frac{d_i^2}{(n_1 - 1)} + \frac{b \frac{(d_i^2)^2}{n_1 - 1}}{(n_1 - 1)(1 + b d_i^2)} \right]$$

$$= \frac{n-3}{n-2} \left[\frac{\hat{\phi}_i (1 + b d_i^2) - \hat{\phi}_i b d_i^2 - (1 + b d_i^2) \frac{d_i^2}{n_1 - 1} + \frac{b (d_i^2)^2}{n_1 - 1}}{(1 + b d_i^2)(n_1 - 1)} \right]$$

$$= \frac{n-3}{n-2} \left[\frac{\hat{\phi}_i - \frac{d_i^2}{n_1 - 1}}{(1 + b d_i^2)(n_1 - 1)} \right]$$

(B-5.1)

B-5.2 Cálculo de $\hat{I}_{(0)}(\bar{x}_1 - \bar{x}_2)$

$$\hat{I}_{(0)}(\bar{x}_1 - \bar{x}_2) = \frac{n-3}{n-2} \left[\left((\bar{x}_1 - \bar{x}_2)' S^{-1} - \frac{1}{n_1 - 1} (\mathbf{x}_{11} - \bar{x}_1)' S^{-1} \right) \left(\mathbf{I} - \frac{b}{1 + bd_i^2} (\mathbf{x}_{11} - \bar{x}_1)(\mathbf{x}_{11} - \bar{x}_1)' S^{-1} \right) \right] (\bar{x}_1 - \bar{x}_2)$$

$$= \frac{n-3}{n-2} \left[(\bar{x}_1 - \bar{x}_2)' S^{-1} - \frac{(\bar{x}_1 - \bar{x}_2)' S^{-1} b (\mathbf{x}_{11} - \bar{x}_1)(\mathbf{x}_{11} - \bar{x}_1)' S^{-1}}{1 + bd_i^2} - \frac{1}{n_1 - 1} (\mathbf{x}_{11} - \bar{x}_1)' S^{-1} + \frac{1}{n_1 - 1} (\mathbf{x}_{11} - \bar{x}_1)' S^{-1} \frac{b}{1 + bd_i^2} (\mathbf{x}_{11} - \bar{x}_1)(\mathbf{x}_{11} - \bar{x}_1)' S^{-1} \right] (\bar{x}_1 - \bar{x}_2)$$

$$= \frac{n-3}{n-2} \left[(\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)' S^{-1} b (\mathbf{x}_{11} - \bar{x}_1) \cdot \frac{(\mathbf{x}_{11} - \bar{x}_1)' S^{-1} (\bar{x}_1 - \bar{x}_2)}{1 + bd_i^2} + \frac{1}{n_1 - 1} (\mathbf{x}_{11} - \bar{x}_1)' S^{-1} (\bar{x}_1 - \bar{x}_2) + \frac{1}{n_1 - 1} (\mathbf{x}_{11} - \bar{x}_1)' S^{-1} \frac{b}{1 + bd_i^2} (\mathbf{x}_{11} - \bar{x}_1)(\mathbf{x}_{11} - \bar{x}_1)' S^{-1} (\bar{x}_1 - \bar{x}_2) \right]$$

$$= \frac{n-3}{n-2} \left(D^2 - \frac{b(\hat{\phi}_i)^2}{1 + bd_i^2} - \frac{\hat{\phi}_i}{n_1 - 1} + \frac{bd_i^2 \hat{\phi}_i}{(n_1 - 1)(1 + bd_i^2)} \right)$$

(B-5.2)

$$= \frac{n-3}{n-2} \left(D^2 - \frac{b(\hat{\phi}_i)^2}{1 + bd_i^2} - \frac{\hat{\phi}_i}{(1 + bd_i^2)(n_1 - 1)} \right)$$

B-5.3 Cálculo de G^2 :

$$G^2 = \left(\frac{n-3}{n-2}\right)^2 \left[(\bar{x}_1 - \bar{x}_2)' S^{-1} - (\bar{x}_1 - \bar{x}_2)' S^{-1} \frac{b}{1+bd_i^2} (x_{1i} - \bar{x}_1)(x_{1i} - \bar{x}_1)' S^{-1} - \right. \\ \left. - \frac{1}{n_1-1} (x_{1i} - \bar{x}_1)' S^{-1} + \frac{1}{n_1-1} (x_{1i} - \bar{x}_1)' S^{-1} \frac{b}{1+bd_i^2} (x_{1i} - \bar{x}_1)(x_{1i} - \bar{x}_1)' S^{-1} \right] \cdot S \\ \left[S^{-1}(\bar{x}_1 - \bar{x}_2) - \frac{b}{1+bd_i^2} S^{-1}(x_{1i} - \bar{x}_1)(x_{1i} - \bar{x}_1)' S^{-1}(\bar{x}_1 - \bar{x}_2) - \frac{1}{n_1-1} S^{-1}(x_{1i} - \bar{x}_1) + \right. \\ \left. + \frac{b}{1+bd_i^2} S^{-1}(x_{1i} - \bar{x}_1)(x_{1i} - \bar{x}_1)' \frac{1}{n_1-1} S^{-1}(x_{1i} - \bar{x}_1) \right]$$

$$G^2 = \left(\frac{n-3}{n-2}\right)^2 \left[D^2 - \frac{b(\hat{\phi}_i)^2}{1+bd_i^2} - \frac{\hat{\phi}_i}{n_1-1} + \frac{bd_i^2 \hat{\phi}_i}{(n_1-1)(1+bd_i^2)} - \frac{b(\hat{\phi}_i)^2}{1+bd_i^2} + \frac{b^2(\hat{\phi}_i)^2 d_i^2}{(1+bd_i^2)^2} + \right.$$

$$+ \frac{bd_i^2 \hat{\phi}_i}{(n_1-1)(1+bd_i^2)} - \frac{b^2(d_i^2)^2 \hat{\phi}_i}{(n_1-1)(1+bd_i^2)^2} - \frac{\hat{\phi}_i}{n_1-1} + \frac{bd_i^2 \hat{\phi}_i}{(n_1-1)(1+bd_i^2)} + \frac{d_i^2}{(n_1-1)^2} -$$

$$\frac{b(d_i^2)^2}{(n_1-1)^2(1+bd_i^2)} + \frac{b\hat{\phi}_i d_i^2}{(n_1-1)(1+bd_i^2)} - \frac{b^2 \hat{\phi}_i (d_i^2)^2}{(1+bd_i^2)(n_1-1)} - \frac{b(d_i^2)^2}{(n_1-1)^2(1+bd_i^2)} +$$

$$\left. + \frac{b^2(d_i^2)^3}{(n_1-1)^2(1+bd_i^2)^2} \right]$$

$$G^2 = \left(\frac{n-3}{n-2}\right)^2 \left\{ D^2 - 2\hat{\phi}_i \left[\frac{b\hat{\phi}_i}{1+bd_i^2} + \frac{(1-bd_i^2)}{(n_1-1)(1+bd_i^2)} \right] + \left[\frac{b\hat{\phi}_i}{1+bd_i^2} + \frac{1}{(n_1-1)(1+bd_i^2)} \right]^2 d_i^2 \right\} \quad (B-5.3)$$

B-6

Nas expressões presentes em (2.30), verificou-se que

$$B_1 = \frac{A_1}{2} - \frac{A_2}{2} \quad e \quad B_2 = \frac{-A_1}{2} - \frac{A_2}{2},$$

com $A_1 = (\hat{\mathbf{i}}' - \hat{\mathbf{i}}'_{(0)}) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$, $A_2 = \hat{\mathbf{i}}'_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})$ e $V = (\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(0)})' \mathbf{S} (\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(0)})$.

Será demonstrado a seguir que B_1 , B_2 e V são funções das estatísticas básicas d_i^2 e $\hat{\phi}_i$.

B-6.1 Cálculo de $A_1 = (\hat{\mathbf{i}}' - \hat{\mathbf{i}}'_{(0)}) (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$

$$A_1 = \hat{\mathbf{i}}' (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \hat{\mathbf{i}}'_{(0)} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

e do Cálculo B-5.2 do Apêndice B, tem-se que

$$A_1 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) - \frac{n-3}{n-2} \left[D^2 - \frac{b(\hat{\phi}_i)^2}{1+bd_i^2} - \frac{\hat{\phi}_i}{n_1-1} + \frac{bd_i^2 \hat{\phi}_i}{(n_1-1)(1+bd_i^2)} \right]$$

$$= D^2 - D^2 \frac{n-3}{n-2} + cf_i (\hat{\phi}_i)^2 + ch \hat{\phi}_i - chf_i d_i^2 \hat{\phi}_i$$

$$= \frac{D^2}{n-2} + cf_i (\hat{\phi}_i)^2 + ch \hat{\phi}_i - chf_i d_i^2 \hat{\phi}_i$$

em que c , f_i , b e h estão definidos na Seção B-5 do Apêndice B.

B-6.2 Cálculo de $A_2 = \hat{\mathbf{I}}_{(0)}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_{1(0)})$

Devido à expressão (B-5.1),

$$\begin{aligned}
 A_2 &= \frac{n-3}{n-2} \left[\frac{\hat{\phi}_i - \frac{d_i^2}{n_1-1}}{(n_1-1)(1+bd_i^2)} \right] \\
 &= \frac{n-3}{n-2} \frac{1}{n_1-1} \left[\hat{\phi}_i - \frac{1}{n_1-1} d_i^2 \right] \frac{1}{1+bd_i^2} \\
 &= c \cdot h \cdot (\hat{\phi}_i - h d_i^2) g_i.
 \end{aligned}$$

B-6.3 Cálculo de $V = (\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(0)})' \mathbf{S} (\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(0)})$

Substituindo $\hat{\mathbf{i}}_{(0)} = c(\mathbf{I} - f_i \mathbf{S}^{-1} \mathbf{z}_i \mathbf{z}_i') (\hat{\mathbf{i}} - h \mathbf{S}^{-1} \mathbf{z}_i)$ e $\hat{\mathbf{i}}_{(0)} = c(\hat{\mathbf{i}}' - h \mathbf{z}_i' \mathbf{S}^{-1}) (\mathbf{I} - f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1})$

na expressão de V , segue que

$$\begin{aligned}
 V &= (\hat{\mathbf{i}}' - [c(\hat{\mathbf{i}}' - h \mathbf{z}_i' \mathbf{S}^{-1}) (\mathbf{I} - f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1})]) \mathbf{S} (\hat{\mathbf{i}} - [c(\mathbf{I} - f_i \mathbf{S}^{-1} \mathbf{z}_i \mathbf{z}_i') (\hat{\mathbf{i}} - h \mathbf{S}^{-1} \mathbf{z}_i)]) \\
 &= (\hat{\mathbf{i}}' - [c(\hat{\mathbf{i}}' - \hat{\mathbf{i}}' f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1} - h \mathbf{z}_i' \mathbf{S}^{-1} + h \mathbf{z}_i' \mathbf{S}^{-1} f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1})]) \mathbf{S} \\
 &\quad (\hat{\mathbf{i}} - [c(\hat{\mathbf{i}} - h \mathbf{S}^{-1} \mathbf{z}_i - f_i \mathbf{S}^{-1} \mathbf{z}_i \mathbf{z}_i' \hat{\mathbf{i}} + f_i \mathbf{S}^{-1} \mathbf{z}_i \mathbf{z}_i' h \mathbf{S}^{-1} \mathbf{z}_i)])
 \end{aligned}$$

$$= (\hat{\mathbf{i}}' - c\hat{\mathbf{i}}' + c\hat{\mathbf{i}}' f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1} + c h \mathbf{z}_i' \mathbf{S}^{-1} - c h \mathbf{z}_i' \mathbf{S}^{-1} f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1}) \\ (\hat{\mathbf{S}}\hat{\mathbf{i}} - \mathbf{S}c\hat{\mathbf{i}} + c h \mathbf{z}_i + c f_i \mathbf{z}_i \mathbf{z}_i' \hat{\mathbf{i}} - c f_i \mathbf{z}_i \mathbf{z}_i' h \mathbf{S}^{-1} \mathbf{z}_i).$$

Substituindo $\hat{\mathbf{i}}$ por $\mathbf{S}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ e \mathbf{z}_i por $\mathbf{x}_{1i} - \bar{\mathbf{x}}_1$ e após vários cálculos obtém-se:

$$V = D^2 - cD^2 + ch\hat{\phi}_i + cf_i(\hat{\phi}_i)^2 - cf_i h \hat{\phi}_i d_i^2 - cD^2 + c^2 D^2 - c^2 h \hat{\phi}_i - c^2 f_i (\hat{\phi}_i)^2 + c^2 f_i h \hat{\phi}_i d_i^2 \\ + cf_i (\hat{\phi}_i)^2 - c^2 f_i (\hat{\phi}_i)^2 + c^2 f_i h \hat{\phi}_i d_i^2 + c^2 (f_i)^2 (\hat{\phi}_i)^2 d_i^2 - c^2 (f_i)^2 h \hat{\phi}_i (d_i^2)^2 + ch \hat{\phi}_i - c^2 h \hat{\phi}_i + \\ + c^2 h^2 d_i^2 + c^2 h f_i d_i^2 \hat{\phi}_i - c^2 h^2 f_i (d_i^2)^2 - ch f_i d_i^2 \hat{\phi}_i + c^2 h f_i d_i^2 \hat{\phi}_i - c^2 h^2 f_i (d_i^2)^2 - \\ - c^2 h (f_i)^2 (d_i^2)^2 \hat{\phi}_i + c^2 h^2 (f_i)^2 (d_i^2)^3$$

$$= D^2 - 2cD^2 + c^2 D^2 + 2ch\hat{\phi}_i + 2cf_i(\hat{\phi}_i)^2 - 2cf_i h \hat{\phi}_i d_i^2 - 2c^2 h \hat{\phi}_i - 2c^2 f_i (\hat{\phi}_i)^2 + \\ + 4c^2 f_i h \hat{\phi}_i d_i^2 + c^2 (f_i)^2 (\hat{\phi}_i)^2 d_i^2 - 2c^2 (f_i)^2 h \hat{\phi}_i (d_i^2)^2 + c^2 h^2 d_i^2 - 2c^2 h^2 f_i (d_i^2)^2 + \\ + c^2 h^2 (f_i)^2 (d_i^2)^3$$

$$= \frac{D^2}{(n-2)^2} + 2ch\hat{\phi}_i(1-c) + 2cf_i(\hat{\phi}_i)^2(1-c) - 2cf_i h \hat{\phi}_i d_i^2 + 4c^2 f_i h \hat{\phi}_i d_i^2 + c^2 (f_i)^2 (\hat{\phi}_i)^2 d_i^2 - \\ - 2c^2 (f_i)^2 h \hat{\phi}_i (d_i^2)^2 + c^2 h^2 d_i^2 (1 - 2f_i d_i^2 + (f_i)^2 (d_i^2)^2)$$

$$= \frac{D^2}{(n-2)^2} + \frac{2ch\hat{\phi}_i}{n-2} + \frac{2cf_i(\hat{\phi}_i)^2}{n-2} - 2cf_i h \hat{\phi}_i d_i^2 (1-c) + 2c^2 f_i h \hat{\phi}_i d_i^2 (1 - f_i d_i^2) + \\ + c^2 (f_i)^2 (\hat{\phi}_i)^2 d_i^2 + c^2 h^2 d_i^2 (1 - f_i d_i^2)^2$$

$$= \frac{D^2}{(n-2)^2} + \frac{2cf_i(\hat{\phi}_i)^2}{n-2} + \frac{2ch\hat{\phi}_i}{n-2} \left(1 - \frac{bd_i^2}{1+bd_i^2} \right) + c^2 f_i^2 (\hat{\phi}_i)^2 d_i^2 + 2c^2 f_i d_i^2 h \hat{\phi}_i \left(1 - \frac{bd_i^2}{1+bd_i^2} \right) + \\ + c^2 h^2 d_i^2 (1 - f_i d_i^2)^2$$

$$= \frac{D^2}{(n-2)^2} + \frac{2cf_i(\hat{\phi}_i)^2}{n-2} + c^2 f_i^2 (\hat{\phi}_i)^2 d_i^2 + 2c^2 f_i d_i^2 h \hat{\phi}_i g_i + \frac{2ch\hat{\phi}_i g_i}{n-2} + c^2 h^2 d_i^2 g_i^2$$

$$= \frac{D^2}{(n-2)^2} + (cf_i\hat{\phi}_i + chg_i) \left[\frac{2\hat{\phi}_i}{n-2} + f_i\hat{\phi}_i c d_i^2 + hg_i c d_i^2 \right]$$

$$V = \frac{D^2}{(n-2)^2} + c(f_i\hat{\phi}_i + hg_i) \left[\frac{2\hat{\phi}_i}{n-2} + (f_i\hat{\phi}_i + hg_i) c d_i^2 \right]. \quad (\text{B-6.3})$$

B-7

Usando os fatos presentes em Fung (1996-b)

$$\bar{\mathbf{x}}_{1(i)} = \bar{\mathbf{x}}_1 - \frac{\mathbf{z}_i}{(n_1 - 1)} \quad \text{e} \quad \mathbf{S}_{(i)}^{-1} = a(\mathbf{S}^{-1} - f_i \mathbf{S}^{-1} \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1})$$

$$\text{em que} \quad a = \frac{n-1}{n}, \quad f_i = \frac{b}{(1 + b d_i^2)}, \quad b = \frac{-n_1}{[(n_1 - 1)n]}, \quad \mathbf{z}_i = \mathbf{x}_{1i} - \bar{\mathbf{x}}_1$$

$$d_i^2 = \mathbf{z}_i' \mathbf{S}^{-1} \mathbf{z}_i = \hat{\alpha}_1^2(\mathbf{x}_{1i}) \quad \text{e} \quad e(\mathbf{x}_{1i}) = F\hat{D}(\mathbf{x}_{1i}) - F\hat{D}(\bar{\mathbf{x}}_1) = \hat{\Gamma}'(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)$$

calcula-se $\hat{\Delta}_{(i)}^2$:

$$\hat{\Delta}_{(i)}^2 = (\bar{\mathbf{x}}_{1(i)} - \bar{\mathbf{x}}_2)' \mathbf{S}_{(i)}^{-1} (\bar{\mathbf{x}}_{1(i)} - \bar{\mathbf{x}}_2)$$

$$= \left(\bar{x}_1 - \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1} - \bar{x}_2 \right)' a (S^{-1} - f_i S^{-1} z_i z_i' S^{-1}) \left(\bar{x}_1 - \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1} - \bar{x}_2 \right)$$

$$= a \left[\left((\bar{x}_1 - \bar{x}_2) - \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1} \right)' (S^{-1} - f_i S^{-1} z_i z_i' S^{-1}) \left((\bar{x}_1 - \bar{x}_2) - \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1} \right) \right]$$

$$= a \left[(\bar{x}_1 - \bar{x}_2)' S^{-1} - (\bar{x}_1 - \bar{x}_2)' f_i S^{-1} z_i z_i' S^{-1} - \frac{(x_{1i} - \bar{x}_1)'}{n_1 - 1} S^{-1} + \frac{(x_{1i} - \bar{x}_1)'}{n_1 - 1} f_i S^{-1} z_i z_i' S^{-1} \right]$$

$$\cdot \left((\bar{x}_1 - \bar{x}_2) - \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1} \right)$$

$$= a \left[(\bar{x}_1 - \bar{x}_2)' S^{-1} (\bar{x}_1 - \bar{x}_2) - (\bar{x}_1 - \bar{x}_2)' S^{-1} \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1} - (\bar{x}_1 - \bar{x}_2)' f_i S^{-1} z_i z_i' S^{-1} (\bar{x}_1 - \bar{x}_2) + \right.$$

$$+ (\bar{x}_1 - \bar{x}_2)' f_i S^{-1} z_i z_i' S^{-1} \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1} - \frac{(x_{1i} - \bar{x}_1)'}{n_1 - 1} S^{-1} (\bar{x}_1 - \bar{x}_2) + \frac{(x_{1i} - \bar{x}_1)'}{n_1 - 1} S^{-1} \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1} + \left. \right]$$

$$+ \frac{(x_{1i} - \bar{x}_1)'}{n_1 - 1} f_i S^{-1} z_i z_i' S^{-1} (\bar{x}_1 - \bar{x}_2) - \frac{(x_{1i} - \bar{x}_1)'}{n_1 - 1} f_i S^{-1} z_i z_i' S^{-1} \frac{(x_{1i} - \bar{x}_1)}{n_1 - 1} \left. \right]$$

$$= a \left[\hat{\Delta}^2 - \frac{\hat{\Gamma}'(x_{1i} - \bar{x}_1)}{n_1 - 1} - f_i \left(\hat{\Gamma}'(x_{1i} - \bar{x}_1) \right)^2 + f_i \frac{\hat{\Gamma}'(x_{1i} - \bar{x}_1) \hat{\alpha}_1^2}{n_1 - 1} - \frac{\hat{\Gamma}'(x_{1i} - \bar{x}_1)}{n_1 - 1} + \right.$$

$$\left. + \frac{\hat{\alpha}_1^2(x_{1i})}{(n_1 - 1)^2} + f_i \frac{\hat{\alpha}_1^2(x_{1i}) \hat{\Gamma}'(x_{1i} - \bar{x}_1)}{n_1 - 1} - f_i \frac{(\hat{\alpha}_1^2(x_{1i}))^2}{(n_1 - 1)^2} \right]$$

$$= a \left[\hat{\Delta}^2 - \frac{2\hat{\Gamma}'(x_{1i} - \bar{x}_1)}{n_1 - 1} + \frac{\hat{\alpha}_1^2(x_{1i})}{(n_1 - 1)^2} - f_i \left(\left(\hat{\Gamma}'(x_{1i} - \bar{x}_1) \right)^2 - \frac{2\hat{\Gamma}'(x_{1i} - \bar{x}_1) \hat{\alpha}_1^2(x_{1i})}{n_1 - 1} + \frac{(\hat{\alpha}_1^2(x_{1i}))^2}{(n_1 - 1)^2} \right) \right]$$

$$= a \left[\hat{\Delta}^2 - \frac{2e(x_{1i})}{n_1 - 1} + \frac{\hat{\alpha}_1^2(x_{1i})}{(n_1 - 1)^2} + \frac{1}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(x_{1i})} \left((e(x_{1i}))^2 - \frac{2e(x_{1i}) \hat{\alpha}_1^2(x_{1i})}{n_1 - 1} + \frac{(\hat{\alpha}_1^2(x_{1i}))^2}{(n_1 - 1)^2} \right) \right]$$

B-8 - Medida de divergencia de Kullback-Leibler

Na Seção 2.5.2, a medida de divergência de Kullback-Leibler (1951) foi utilizada para determinar pontos influentes no modelo de regressão logística. É uma medida de distância entre duas populações, definida do seguinte modo :

se f_1 e f_2 são funções densidade de probabilidade e E_{f_j} é a esperança com respeito às densidades f_j , $j = 1, 2$, têm-se :

- a) as divergências direcionadas

$$I(f_1, f_2) = E_{f_1} \ln \left(\frac{f_1}{f_2} \right) \quad e$$

$$I(f_2, f_1) = E_{f_2} \ln \left(\frac{f_2}{f_1} \right) .$$

- b) a divergência entre f_1 e f_2

$$J(f_1, f_2) = I(f_1, f_2) + I(f_2, f_1) .$$

Verifica-se que essas medidas são bem definidas e não negativas, sendo nulas apenas quando $f_1 = f_2$.

Além disso, se f_1 e f_2 são densidades de vetores aleatórios com distribuição $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ com $\boldsymbol{\Sigma}_j$ positiva definida, $j = 1, 2$, respectivamente,

$$2I(f_1, f_2) = (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \text{tr} \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1} - \ln |\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1}| - p$$

em que $|\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1}|$ representa o determinante da matriz $\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1}$.

Resultado 8.1 (Graybill (1976))

Se \mathbf{X} é um vetor aleatório com $E(\mathbf{X}) = \boldsymbol{\mu}$, $Var(\mathbf{X}) = \boldsymbol{\Sigma}$ e \mathbf{A} é uma matriz simétrica então $E(\mathbf{X}'\mathbf{A}\mathbf{X}) = tr(\mathbf{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$.

Resultado 8.2

Sejam f_1 e f_2 densidades de vetores aleatórios com distribuição normal p-variada com média $\boldsymbol{\mu}_j$ e matriz de covariância $\boldsymbol{\Sigma}_j$, positiva definida, $j = 1, 2$. Nessas condições, a divergência de Kullback-Leibler entre f_1 e f_2 , $I(f_1, f_2)$ é tal que

$$2I(f_1, f_2) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_2^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \left\{ r \boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1} - \ln |\boldsymbol{\Sigma}_1 \boldsymbol{\Sigma}_2^{-1}| - p \right\}.$$

Prova:

Sejam f_1 a densidade $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ e f_2 a densidade $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, de modo que

$$f_1(\mathbf{y}) = \left(\frac{1}{2\pi} \right)^{p/2} |\boldsymbol{\Sigma}_1|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1} (\mathbf{y} - \boldsymbol{\mu}_1) \right\},$$

$$f_2(\mathbf{y}) = \left(\frac{1}{2\pi} \right)^{p/2} |\boldsymbol{\Sigma}_2|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_2^{-1} (\mathbf{y} - \boldsymbol{\mu}_2) \right\} e$$

$$\ln \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})} = -\frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}_1| - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1} (\mathbf{y} - \boldsymbol{\mu}_1) + \frac{p}{2} \ln(2\pi) + \frac{1}{2} \ln |\boldsymbol{\Sigma}_2| +$$
$$+ \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_2^{-1} (\mathbf{y} - \boldsymbol{\mu}_2)$$

$$= \frac{1}{2} \ln \frac{|\boldsymbol{\Sigma}_2|}{|\boldsymbol{\Sigma}_1|} + \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_2^{-1} (\mathbf{y} - \boldsymbol{\mu}_2) - \frac{1}{2} (\mathbf{y} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}_1^{-1} (\mathbf{y} - \boldsymbol{\mu}_1).$$

De acordo com a definição, $I(f_1, f_2) = E_{f_1} \ln \left(\frac{f_1}{f_2} \right)$ e portanto, tomando esperança com respeito à f_1 ,

$$E_{f_1} \ln \left(\frac{f_1}{f_2} \right) = \int \ln \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})} f_1(\mathbf{y}) d\mathbf{y}$$

$$= \frac{1}{2} \ln \frac{|\Sigma_2|}{|\Sigma_1|} + \frac{1}{2} E_{f_1} \left[(\mathbf{Y} - \mu_2)' \Sigma_2^{-1} (\mathbf{Y} - \mu_2) - (\mathbf{Y} - \mu_1)' \Sigma_1^{-1} (\mathbf{Y} - \mu_1) \right].$$

Como, sob f_1 , $\mathbf{Y} \sim N_p(\mu_1, \Sigma_1)$, então $(\mathbf{Y} - \mu_1) \sim N_p(\mathbf{0}, \Sigma_1)$ e $(\mathbf{Y} - \mu_2) \sim N_p(\mu_1 - \mu_2, \Sigma_1)$ e assim, devido ao Resultado 8.1,

$$E_{f_1} [(\mathbf{Y} - \mu_2)' \Sigma_2^{-1} (\mathbf{Y} - \mu_2)] = \text{tr}(\Sigma_2^{-1} \Sigma_1) + (\mu_1 - \mu_2)' \Sigma_2^{-1} (\mu_1 - \mu_2) \quad \text{e}$$

$$E_{f_1} [(\mathbf{Y} - \mu_1)' \Sigma_1^{-1} (\mathbf{Y} - \mu_1)] = \text{tr}(\Sigma_1^{-1} \Sigma_1) = p.$$

Portanto,

$$E_{f_1} \left(\ln \frac{f_1}{f_2} \right) = \frac{1}{2} \ln |\Sigma_2 \Sigma_1^{-1}| + \frac{1}{2} \left[(\mu_1 - \mu_2)' \Sigma_2^{-1} (\mu_1 - \mu_2) + \text{tr}(\Sigma_2^{-1} \Sigma_1) - p \right].$$

Como Σ_1 e Σ_2 são simétricas de mesma ordem, $\text{tr}(\Sigma_2^{-1} \Sigma_1) = \text{tr}(\Sigma_1 \Sigma_2^{-1})$ então

$$E_{f_1} \left(\ln \frac{f_1}{f_2} \right) = \frac{1}{2} (\mu_1 - \mu_2)' \Sigma_2^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \left\{ \text{tr}(\Sigma_1 \Sigma_2^{-1}) - \ln |\Sigma_1 \Sigma_2^{-1}| - p \right\}. \quad \square$$

B-9

Sejam f_1 e f_2 duas distribuições de Bernoulli:

$$f_1(y) = p_1^y (1-p_1)^{1-y}$$

$$f_2(y) = p_2^y (1-p_2)^{1-y}$$

e

$$\begin{aligned} \ln\left(\frac{f_1}{f_2}\right) &= \ln\left(\frac{p_1^y (1-p_1)^{1-y}}{p_2^y (1-p_2)^{1-y}}\right) = \ln\left(\frac{p_1}{p_2}\right)^y + \ln\left(\frac{(1-p_1)}{(1-p_2)}\right)^{1-y} = \\ &= y \ln\left(\frac{p_1}{p_2}\right) + (1-y) \ln\left(\frac{1-p_1}{1-p_2}\right). \end{aligned}$$

No cálculo da medida de influência para regressão logística, as divergências direcionadas de Kullback-Leibler ficam

$$\begin{aligned} I(f_1, f_2) &= E_{f_1} \ln\left(\frac{f_1}{f_2}\right) = E\left[y \ln \frac{p_1}{p_2} + \ln \frac{1-p_1}{1-p_2} - y \ln\left(\frac{1-p_1}{1-p_2}\right)\right] \\ &= p_1 \ln \frac{p_1}{p_2} + \ln \frac{1-p_1}{1-p_2} - p_1 \ln\left(\frac{1-p_1}{1-p_2}\right) \end{aligned}$$

e

$$I(f_2, f_1) = E_{f_2} \ln\left(\frac{f_2}{f_1}\right) = p_2 \ln \frac{p_2}{p_1} + \ln \frac{1-p_2}{1-p_1} - p_2 \ln\left(\frac{1-p_2}{1-p_1}\right).$$

Substituindo p_1 por \hat{p}_k e p_2 por $\hat{p}_{k(i)}$, a medida sugerida por Johnson (1985) é

$$\tilde{D}_i^f(c) = \sum_{k=1}^M c_k J(\hat{p}_k^f, \hat{p}_{k(i)}^f), \quad i = 1, 2, \dots, n.$$

Calculando apenas $J(\hat{p}_k^f, \hat{p}_{k(i)}^f)$,

$$\begin{aligned}
J(\hat{p}_K^f, \hat{p}_{K(i)}^f) &= I(\hat{p}_K^f, \hat{p}_{K(i)}^f) + I(\hat{p}_{K(i)}^f, \hat{p}_K^f) = \\
&= \hat{p}_K^f \ln \hat{p}_K^f - \hat{p}_K^f \ln \hat{p}_{K(i)}^f + \ln(1 - \hat{p}_K^f) - \ln(1 - \hat{p}_{K(i)}^f) - \hat{p}_K^f \ln(1 - \hat{p}_K^f) + \hat{p}_K^f \ln(1 - \hat{p}_{K(i)}^f) + \hat{p}_{K(i)}^f \ln \hat{p}_{K(i)}^f - \\
&- \hat{p}_{K(i)}^f \ln \hat{p}_K^f + \ln(1 - \hat{p}_{K(i)}^f) - \ln(1 - \hat{p}_K^f) - \hat{p}_{K(i)}^f \ln(1 - \hat{p}_{K(i)}^f) + \hat{p}_{K(i)}^f \ln(1 - \hat{p}_K^f) \\
&= (\hat{p}_K^f - \hat{p}_{K(i)}^f) (\ln \hat{p}_K^f - \ln \hat{p}_{K(i)}^f - \ln(1 - \hat{p}_K^f) + \ln(1 - \hat{p}_{K(i)}^f)).
\end{aligned}$$

$$\text{Então } \tilde{D}_i^f(c) = \sum_{k=1}^M c_k J(\hat{p}_k^f, \hat{p}_{k(i)}^f) = \sum_{k=1}^M c_k (\hat{p}_k^f - \hat{p}_{k(i)}^f) \ln \left[\frac{\frac{\hat{p}_k^f}{1 - \hat{p}_k^f}}{\frac{\hat{p}_{k(i)}^f}{1 - \hat{p}_{k(i)}^f}} \right]$$

$$= \sum_{k=1}^M c_k (\hat{p}_k^f - \hat{p}_{k(i)}^f) (\mathbf{x}_k^f \hat{\boldsymbol{\beta}} - \mathbf{x}_k^f \hat{\boldsymbol{\beta}}_{(i)})$$

$$= \sum_{k=1}^M c_k (\hat{p}_k^f - \hat{p}_{k(i)}^f) \mathbf{x}_k^f (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)}).$$

B-10

Calculando o termo $\frac{k(\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(i)})' \mathbf{S} \hat{\mathbf{i}}}{D^2}$, têm-se do Cálculo B-6.3 do Apêndice B que

$$(\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(i)})' = \hat{\mathbf{i}}' - \hat{c}\mathbf{i}' + \hat{c}\mathbf{i}' f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1} + c h \mathbf{z}_i' \mathbf{S}^{-1} - c h \mathbf{z}_i' \mathbf{S}^{-1} f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1}, \quad \text{então}$$

$$(\hat{\mathbf{i}} - \hat{\mathbf{i}}_{(i)})' \mathbf{S} \hat{\mathbf{i}} = \hat{\mathbf{i}}' \mathbf{S} \hat{\mathbf{i}} - \hat{c}\mathbf{i}' \mathbf{S} \hat{\mathbf{i}} + \hat{c}\mathbf{i}' f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1} \mathbf{S} \hat{\mathbf{i}} + c h \mathbf{z}_i' \mathbf{S}^{-1} \mathbf{S} \hat{\mathbf{i}} - c h \mathbf{z}_i' \mathbf{S}^{-1} f_i \mathbf{z}_i \mathbf{z}_i' \mathbf{S}^{-1} \mathbf{S} \hat{\mathbf{i}}$$

$$\begin{aligned}
&= (\bar{x}_1 - \bar{x}_2)' S^{-1} S S^{-1} (\bar{x}_1 - \bar{x}_2) - c (\bar{x}_1 - \bar{x}_2)' S^{-1} S S^{-1} (\bar{x}_1 - \bar{x}_2) + \\
&+ c (\bar{x}_1 - \bar{x}_2)' S^{-1} f_i (x_{1i} - \bar{x}_1) (x_{1i} - \bar{x}_1)' S^{-1} S S^{-1} (\bar{x}_1 - \bar{x}_2) + ch (x_{1i} - \bar{x}_1)' S^{-1} S S^{-1} (\bar{x}_1 - \bar{x}_2) - \\
&- ch (x_{1i} - \bar{x}_1)' S^{-1} f_i (x_{1i} - \bar{x}_1) (x_{1i} - \bar{x}_1)' S^{-1} S S^{-1} (\bar{x}_1 - \bar{x}_2) \\
&= D^2 - cD^2 + c\hat{\phi}_i^2 f_i + ch\hat{\phi}_i - chd_i^2 \hat{\phi}_i f_i . \tag{B-10.1}
\end{aligned}$$

Multiplicando por k e dividindo por D^2 segue que

$$\begin{aligned}
\frac{k(\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)})' S \hat{\mathbf{I}}}{D^2} &= k - kc + \frac{kc\hat{\phi}_i^2 f_i + kch\hat{\phi}_i - kchd_i^2 \hat{\phi}_i f_i}{D^2} \\
&= k \left(1 - \frac{n-3}{n-2} \right) + \frac{kcf_i (\hat{\phi}_i^2 + h\hat{\phi}_i / f_i - hd_i^2 \hat{\phi}_i)}{D^2} \\
&= \frac{k}{n-2} + \frac{kcf_i \left(\hat{\phi}_i^2 + \frac{h\hat{\phi}_i (1 - f_i d_i^2)}{f_i} \right)}{D^2} = \frac{k}{n-2} + \frac{kcf_i \left(\hat{\phi}_i^2 + \frac{h\hat{\phi}_i (1 - bd_i^2 / 1 + bd_i^2)}{b/1 + bd_i^2} \right)}{D^2} \\
&= \frac{k}{n-2} + \frac{kcf_i \left(\hat{\phi}_i^2 + \frac{h\hat{\phi}_i}{b} \right)}{D^2} .
\end{aligned}$$

Do Cálculo B-6.2 do Apêndice B

$$\frac{\hat{\mathbf{I}}_{(0)}' (\bar{x}_{1(0)} - \bar{x}_1)}{2} = \frac{-ch(\hat{\phi}_i - hd_i^2)g_i}{2} = \frac{-chf_i(\hat{\phi}_i - hd_i^2)}{2b} , \text{ assim}$$

$$B_i = \frac{\hat{\mathbf{I}}_{(0)}' (\bar{x}_{1(0)} - \bar{x}_1)}{2} + \frac{k(\hat{\mathbf{I}} - \hat{\mathbf{I}}_{(0)})' S \hat{\mathbf{I}}}{D^2} = \frac{-chf_i(\hat{\phi}_i - hd_i^2)}{2b} + \left\{ \frac{k}{n-2} + \frac{kcf_i(\hat{\phi}_i^2 + h\hat{\phi}_i/b)}{D^2} \right\} .$$

Portanto B_i é função de $\hat{\phi}_i$ e d_i^2 .

B-11

Tem-se que

$$V = (\hat{i} - \hat{i}_0) S(\hat{i} - \hat{i}_0) - \frac{\{(\hat{i} - \hat{i}_0) S \hat{i}\}^2}{D^2}.$$

Devido ao cálculo da Seção B-6.3 do Apêndice B,

$$(\hat{i} - \hat{i}_0) S(\hat{i} - \hat{i}_0) = \frac{D^2}{(n-2)^2} + c(f_i \hat{\phi}_i + hg_i) \left[\frac{2\hat{\phi}_i}{n-2} + (f_i \hat{\phi}_i + hg_i) cd_i^2 \right],$$

e como $g_i = \frac{1}{1+bd_i^2}$,

$$\begin{aligned} (\hat{i} - \hat{i}_0) S(\hat{i} - \hat{i}_0) &= \frac{D^2}{(n-2)^2} + c \left(f_i \hat{\phi}_i + \frac{h}{1+bd_i^2} \cdot \frac{b}{b} \right) \left[\frac{2\hat{\phi}_i}{n-2} + \left(f_i \hat{\phi}_i + \frac{h}{1+bd_i^2} \cdot \frac{b}{b} \right) cd_i^2 \right] \\ &= \frac{D^2}{(n-2)^2} + c \left(f_i \hat{\phi}_i + \frac{hf_i}{b} \right) \left[\frac{2\hat{\phi}_i}{n-2} + \left(f_i \hat{\phi}_i + \frac{hf_i}{b} \right) cd_i^2 \right] \\ &= \frac{D^2}{(n-2)^2} + \left(cf_i \hat{\phi}_i + cf_i \frac{h}{b} \right) \left[\frac{2\hat{\phi}_i}{n-2} + cf_i \hat{\phi}_i d_i^2 + cf_i d_i^2 \frac{h}{b} \right] \\ &= \frac{D^2}{(n-2)^2} + \frac{2cf_i \hat{\phi}_i^2}{n-2} + c^2 f_i^2 \hat{\phi}_i^2 d_i^2 + c^2 f_i^2 \hat{\phi}_i d_i^2 \frac{h}{b} + \frac{2c\hat{\phi}_i f_i h}{(n-2)b} + c^2 f_i^2 \hat{\phi}_i \frac{h}{b} d_i^2 + c^2 f_i^2 d_i^2 \left(\frac{h}{b} \right)^2 \\ &= \frac{D^2}{(n-2)^2} + \frac{2cf_i \hat{\phi}_i^2}{n-2} + c^2 f_i^2 d_i^2 \left(\hat{\phi}_i^2 + 2\hat{\phi}_i \frac{h}{b} + \left(\frac{h}{b} \right)^2 \right) + \frac{2c\hat{\phi}_i f_i h}{(n-2)b} \end{aligned}$$

$$= \frac{D^2}{(n-2)^2} + \frac{2cf_i\hat{\phi}_i}{n-2} \left(\frac{h}{b} + \hat{\phi}_i \right) + c^2 f_i^2 d_i^2 \left(\hat{\phi}_i + \frac{h}{b} \right)^2.$$

Da expressão (B-10.1) do Apêndice B tem-se que

$$(\hat{l} - \hat{l}_{(0)}) S \hat{l} = D^2 - cD^2 + cf_i\hat{\phi}_i^2 + ch\hat{\phi}_i - chd_i^2\hat{\phi}_i f_i$$

$$= D^2 - \frac{n-3}{n-2} D^2 + cf_i\hat{\phi}_i \left(\hat{\phi}_i + \frac{h}{f_i} - hd_i^2 \right) = \frac{D^2(n-2) - (n-3)D^2}{n-2} + cf_i\hat{\phi}_i \left(\hat{\phi}_i + \frac{h(1-f_i d_i^2)}{f_i} \right)$$

$$= \frac{D^2}{n-2} + cf_i\hat{\phi}_i \left(\hat{\phi}_i + h \left(1 - \frac{bd_i^2}{1+bd_i^2} \right) \left(\frac{1+bd_i^2}{b} \right) \right) = \frac{D^2}{n-2} + cf_i\hat{\phi}_i \left(\hat{\phi}_i + \frac{h}{b} \right).$$

Portanto

$$\begin{aligned} \frac{[(\hat{l} - \hat{l}_{(0)}) S \hat{l}]^2}{D^2} &= \left\{ \frac{(D^2)^2}{(n-2)^2} + 2 \frac{D^2}{n-2} cf_i\hat{\phi}_i \left(\hat{\phi}_i + \frac{h}{b} \right) + c^2 f_i^2 \hat{\phi}_i^2 \left(\hat{\phi}_i + \frac{h}{b} \right)^2 \right\} / D^2 \\ &= \frac{D^2}{(n-2)^2} + \frac{2cf_i\hat{\phi}_i}{n-2} \left(\hat{\phi}_i + \frac{h}{b} \right) + \frac{c^2 f_i^2 \hat{\phi}_i^2 \left(\hat{\phi}_i + \frac{h}{b} \right)^2}{D^2}. \end{aligned}$$

Voltando a V, e subtraindo o segundo termo do primeiro :

$$\begin{aligned} V &= \frac{D^2}{(n-2)^2} + \frac{2cf_i\hat{\phi}_i}{n-2} \left(\frac{h}{b} + \hat{\phi}_i \right) + c^2 f_i^2 d_i^2 \left(\hat{\phi}_i + \frac{h}{b} \right)^2 - \frac{D^2}{(n-2)^2} - \frac{2cf_i\hat{\phi}_i}{n-2} \left(\hat{\phi}_i + \frac{h}{b} \right) - \\ &\quad - \frac{c^2 f_i^2 \hat{\phi}_i^2 \left(\hat{\phi}_i + \frac{h}{b} \right)^2}{D^2} \end{aligned}$$

$$V = c^2 f_i^2 \left(\hat{\phi}_i + \frac{h}{b} \right)^2 \left(d_i^2 - \frac{\hat{\phi}_i^2}{D^2} \right).$$

Apêndice C

Programas no S-plus

As medidas de diagnóstico estudadas foram calculadas computacionalmente utilizando o software S-plus (Venables e Ripley (1994) ; Krause e Olson (2002)).

Os dados utilizados (RAE-CEA-9415) encontram-se na Tabela C.1 e Tabela C.2 .

Tabela C.1 – Conjunto de dados referente ao grupo 1.

grupo	comprimento total (mm)	comprimento da cauda (mm)	comprimento do pé (mm)	comprimento do pelo do dorso (mm)
1	324	193	26	12
1	337	186	25	11
1	370	170	28	15
1	250	148	20	9
1	261	155	20	10
1	290	165	23	9
1	267	154	20	8
1	155	57	17	4
1	155	57	17	4
1	374	220	30	8
1	355	205	27	14
1	378	207	28	15
1	340	193	26	12
1	508	295	38	10

Tabela C.2 — Conjunto de dados referente ao grupo 2.

grupo	comprimento total (mm)	comprimento da cauda (mm)	comprimento do pé (mm)	comprimento do pelo do dorso (mm)
2	177	81	25	7
2	175	77	25	12
2	195	87	25	7
2	202	80	25	6
2	205	88	26	9
2	130	55	23	7
2	212	89	24	8
2	191	81	24	5
2	175	90	21	7
2	250	87	25	8
2	180	85	25	9
2	175	80	25	9
2	200	90	27	12
2	155	53	20	9
2	200	80	24	7
2	140	46	17	6
2	192	88	25	10
2	223	114	33	11
2	177	81	24	10
2	193	83	31	12
2	320	162	38	11
2	235	121	34	12
2	184	88	26	10
2	143	51	20	6
2	221	123	34	12
2	180	79	25	11
2	294	137	37	12
2	137	57	21	6
2	195	92	25	16
2	155	53	20	7
2	131	51	19	7
2	196	89	26	11
2	277	120	37	12
2	139	50	20	7
2	197	92	25	11
2	206	98	25	10
2	132	47	20	8
2	187	100	21	11
2	202	92	26	11
2	180	87	25	10
2	193	100	20	10
2	224	102	28	8
2	215	100	29	11
2	197	97	24	10
2	141	49	20	9
2	257	120	31	12
2	194	85	21	10
2	215	100	29	11
2	186	85	26	11

grupo	comprimento total (mm)	comprimento da cauda (mm)	comprimento do pé (mm)	comprimento do pelo do dorso (mm)
2	289	153	39	13
2	204	90	25	11
2	230	105	28	13
2	281	145	37	10
2	200	90	28	9
2	205	95	27	11
2	185	85	25	10
2	200	94	24	11
2	226	102	28	12
2	136	49	21	7
2	254	132	36	12
2	180	100	24	13
2	250	111	29	11
2	204	97	25	11
2	202	90	26	12
2	142	50	20	8
2	196	107	21	10
2	199	92	24	10
2	181	82	24	10
2	238	109	29	10
2	150	55	20	9
2	182	82	24	10
2	181	102	21	11
2	145	51	20	7
2	130	46	19	8

Inicialmente o programa lê o conjunto de dados e obtém as estatísticas resumo para cada grupo, bem como a matriz de covariância ponderada.

```
attach (DADOS) #lê o conjunto de dados #
dados<-DADOS[,-1] #dados recebe o conj de dados sem a coluna com o grupo #
n<-88 #numero total de observações#
n1<-13 #numero de observações de cada grupo#
n2<-75

dados1<-dados[1:13,] #separa o conj de dados em grupos#
dados2<-dados[14:88,]

medial<-array(0,4) #porque são 4 variáveis p = 4 #
for(i in 1:4) {medial[i]<-mean(dados1[,i])}
medial
#define o vetor de médias do grupo 1#
```

```

media2<-array(0,4)
for(i in 1:4) {media2[i]<-mean(dados2[,i])}
media2

difm<-medial-media2      #define o vetor da diferença das médias#
mcov<-((n1-1)*var(dados1)+(n2-1)*var(dados2))/(n-2)
#define a matriz de covariância ponderada #
mcovinv<-solve(mcov)     #inversa da matriz de covariância#

```

Para verificar a igualdade das matrizes de covariâncias de cada grupo procedeu-se ao teste de homogeneidade de variâncias (Morrison (1976, p.252)):

```

k<-2      #número de grupos#

library(Matrix)
ldmcovaux<- det.Matrix(mcov)
ldmcov<-ldmcovaux$modulus[1]*ldmcovaux$sign
ldmcov1aux<- det.Matrix(var(dados1))
ldmcov1<-ldmcov1aux$modulus[1]*ldmcov1aux$sign
ldmcov2aux<-det.Matrix(var(dados2))
ldmcov2<-ldmcov2aux$modulus[1]*ldmcov2aux$sign

M<- (n-k) * ldmcov - ((n1-1) *
ldmcov1 + (n2-1) * ldmcov2)

p<-4      #número de variáveis#

Claux1<- (2*p^2+3*p-1) / (6*(p+1)*(k-1))
Claux2<- (1/(n1-1)) + (1/(n2-1)) - (1/(n-k))
Claux<-Claux1*Claux2
C1<-1-Claux
M*C1
g1<-0.5*(k-1)*p*(p+1)

qchisq(0.95,g1)

```

Como $M*C1$ é muito maior que o valor crítico do quiquadrado rejeita-se a hipótese de que as variâncias são iguais.

- Cálculo da estatística ϕ_i :

```

medialrep<-matrix(media1,n1,4,byrow=T)#matriz c/ média do grupo 1 n1 vezes#
media2rep<-matrix(media2,n2,4,byrow=T)

dadosdif1<-dados1-medialrep #diferença entre as obs e a média de seu
  grupo#
dadosdif2<-dados2-media2rep

dadosdif1transp<-t(dadosdif1)
dadosdif2transp<-t(dadosdif2)

fi1<-difm**%mcovinv**%dadosdif1transp #cálculo de fi para o grupo 1#
fi2<-difm**%mcovinv**%dadosdif2transp

par(mfrow=c(1,2))

plot(x=1:n1, fi1,type="h", xlab="numero da observação (grupo 1)" ,
  ylab="fi1",xlim=c(1,n1))

plot(x=1:n2, fi2,type="h", xlab="numero da observação (grupo 2)" ,
  ylab="fi2",xlim=c(1,n2))

```

- Cálculo da estatística d_i^2 :

```

di1<-matrix(0,n1,1) #calculo da medida di para o grupo 1#
for (i in 1:n1)
di1[i,]<-dadosdif1[i,]**%mcovinv**%dadosdif1transp[,i]

di2<-matrix(0,n2,1) #calculo da medida di para o grupo 2#
for (i in 1:n2)
di2[i,]<-dadosdif2[i,]**%mcovinv**%dadosdif2transp[,i]

par(mfrow=c(1,2))

plot(x=1:n1, di1,type="h", xlab="numero da observação (grupo 1)" ,
  ylab="di1",xlim=c(1,n1))

plot(x=1:n2, di2,type="h", xlab="numero da observação (grupo 2)" ,
  ylab="di2",xlim=c(1,n2))

```

- Cálculo da medida *LOQRG* (Capítulo 5) utilizada quando as matrizes de covariâncias não são iguais :

$$LOQRG_i = \frac{1}{n_1} \sum_{k=1}^{n_1} \left[\log \hat{f}_1(\mathbf{x}_{1k}) - \log \hat{f}_{1(i)}(\mathbf{x}_{1k}) \right]^2 = \frac{1}{n_1} \sum_k R_1^2(\mathbf{x}_{1k}),$$

```

difmlinha<-matrix(difm,4,1) #inverte vetor difm#
delta2<-difm**%mcovinv**%difmlinha #calcula distância de Mahalanobis#
delta<-sqrt(delta2) #calcula raiz quadrada da dist de Mahalanobis#

w1<-n1/n #proporção de obs do grupo 1 em relação a todas obs#
w2<-n2/n

sigma1inv<-solve(var(dados1))
sigma2inv<-solve(var(dados2))

medialsemi<-matrix(0,n1,4) #calcula a media do grupo 1 sem a observação i
#
for ( i in 1:n1)
medialsemi[i,]<-medial-(dadosdif1[i,]/(n1-1))

media2semi<-matrix(0,n2,4) #calcula a media do grupo 2 sem a observação i
#
for ( i in 1:n2)
media2semi[i,]<-media2-(dadosdif2[i,]/(n2-1))

a<-(n-1)/n

mcovinvsemi1<-array(c(0), dim=c(4,4,n1)) #calcula a matriz de covariância
sem a obs i para o grupo 1#
for ( i in 1:n1)
mcovinvsemi1[,,i]<-a*(mcovinv-
(ffil[i]*mcovinv**%dadosdif1transp[,i]**%dadosdif1[i,]**%mcovinv))

mcovinvsemi2<-array(c(0), dim=c(4,4,n2)) #calcula a matriz de covariância
sem a obs i para o grupo 2#
for ( i in 1:n2)
mcovinvsemi2[,,i]<-a*(mcovinv-
(ffii2[i]*mcovinv**%dadosdif2transp[,i]**%dadosdif2[i,]**%mcovinv))

ldmcov1
ldmcov2

dados1transp<-t(dados1)
dados1semi<-array(c(0), dim=c(4,(n1-1),n1))
for(i in 1:n1)
dados1semi[,,i]<-dados1transp[, -i]

dados1semit<-array(c(0), dim=c((n1-1),4,n1))
for(i in 1:n1)
dados1semit[,,i]<-t(dados1semi[, ,i])

```



```

dados2transp<-t(dados2)
dados2semi<-array(c(0), dim=c(4, (n2-1), n2))
for(i in 1:n2)
dados2semi[, , i]<-dados2transp[, -i]

dados2semit<-array(c(0), dim=c((n2-1), 4, n2))
for(i in 1:n2)
dados2semit[, , i]<-t(dados2semi[, , i])

mcovsemi1<-array(c(0), dim=c(4, 4, n1))
for ( i in 1:n1)
mcovsemi1[, , i]<-var(dados1semit[, , i])

library(Matrix)
logdsigmalsemi<-array(0, n1)
logdsigmalsemi
for(i in 1:n1){
auxiliar1<- det.Matrix(mcovsemi1[, , i])
logdsigmalsemi[i]<-auxiliar1$modulus[1]*auxiliar1$sign
}
#monta um vetor com os valores dos log-determinantes sem a observ i do grupo
1#

mcovsemi2<-array(c(0), dim=c(4, 4, n2))
for ( i in 1:n2)
mcovsemi2[, , i]<-var(dados2semit[, , i])
mcovsemi2

logdsigma2semi<-array(0, n2)
logdsigma2semi
for(i in 1:n2){
auxiliar2<- det.Matrix(mcovsemi2[, , i])
logdsigma2semi[i]<-auxiliar2$modulus[1]*auxiliar2$sign
}

medialsemirep<-array(c(0), dim=c(n1, 4, n1))
for (i in 1:n1)
medialsemirep[, , i]<-matrix(medialsemi[i, ], n1, 4, byrow=T)#matriz c/ média do
grupo 1 sem o i n1 vezes#

media2semirep<-array(c(0), dim=c(n2, 4, n2))
for (i in 1:n2)
media2semirep[, , i]<-matrix(media2semi[i, ], n2, 4, byrow=T)#matriz c/ média do
grupo 2 sem o i n2 vezes#

dadosdif1semi<-array(c(0), dim=c(n1, 4, n1))
for (i in 1:n1)
dadosdif1semi[, , i]<-dados1-medialsemirep[, , i]

dadosdif1semit<-array(c(0), dim=c(4, n1, n1))
for (i in 1:n1)
dadosdif1semit[, , i]<-t(dadosdif1semi[, , i])

dadosdif2semi<-array(c(0), dim=c(n2, 4, n2))
for (i in 1:n2)
dadosdif2semi[, , i]<-dados2-media2semirep[, , i]

dadosdif2semit<-array(c(0), dim=c(4, n2, n2))

```

```

for (i in 1:n2)
dadosdif2semit[,i]<-t(dadosdif2semi[,i])
aux1grupo1<-matrix(0,n1,1)
for (i in 1:n1)
aux1grupo1[i,]<- -ldmcov1-dadosdif1[i,]**sigma1inv**dadosdif1transp[,i]

aux1grupo2<-matrix(0,n2,1)
for (i in 1:n2)
aux1grupo2[i,]<- -ldmcov2-dadosdif2[i,]**sigma2inv**dadosdif2transp[,i]
aux2grupo1<-array(c(0),dim=c(n1,1,n1))
for (i in 1:n1) { for (j in 1:n1) aux2grupo1[j,,i]<- logdsigma1semi[i]+
dadosdif1semi[j,,i]**mcovinvsemi1[,i]** dadosdif1semit[,j,i] }
aux2grupo2<-array(c(0),dim=c(n2,1,n2))
for (i in 1:n2) { for (j in 1:n2) aux2grupo2 [j,,i]<- logdsigma2semi[i]+
dadosdif2semi[j,,i]**mcovinvsemi2[,i]** dadosdif2semit[,j,i] }

loqrg1<-matrix(0,n1,1)
for (i in 1:n1) loqrg1[i,]<-
1/(4*n1)*((sum(aux1grupo1)+sum(aux2grupo1[,,i]))^2)

loqrg2<-matrix(0,n2,1)
for (i in 1:n2) loqrg2[i,]<-
1/(4*n2)*((sum(aux1grupo2)+sum(aux2grupo2[,,i]))^2)

par(mfrow=c(1,2))
plot(x=1:n1, loqrg1,type="h",xlab="numero da observação (grupo 1)" ,
ylab="LOQRG1",xlim=c(1,n1))
plot(x=1:n2, loqrg2,type="h", xlab="numero da observação (grupo 2)" ,
ylab="LOQRG2",xlim=c(1,n2))

```

- Cálculo da medida Bayesiana LORAM (Capítulo 3):

$$LORAM_i = n^{-1} \sum_{j=1}^2 \sum_{k=1}^{n_j} \ln \left\{ \frac{o(\mathbf{x}_{jk})}{o_{(i)}(\mathbf{x}_{jk})} \right\}$$

$$\text{em que } \ln(o(\mathbf{y})) = \ln \frac{q}{(1-q)} + (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{y} - \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)$$

```

dadostransp<-cbind(dados1transp,dados2transp)

oy<-matrix(0,n,1) #calcula primeira parte do o(y) #
for ( i in 1:n)
oy[i,]<-difm%*%mcovinv%*%dadostransp[,i]

mediale2<-matrix(media1+media2,4,1) #calcula segunda parte do o(y)#
mediale2
cte<-0.5*difm%*%mcovinv%*%mediale2

lnodey<-matrix(0,n,1)
for(i in 1:n)
lnodey[i,]<-oy[i,]-cte #calcula o log-odds das obs do grupo 1#

difmsemi1<-matrix(0,n1,4)
for (i in 1:n1)
difmsemi1[i,]<-media1semi[i,]-media2

difmsemi2<-matrix(0,n2,4)
for (i in 1:n2)
difmsemi2[i,]<-media2semi[i,]-media1

oysemi1<-array(c(0),dim=c(n,1,n1)) #calcula primeira parte do o(y) sem obs
  i grupo 1#
for ( i in 1:n1) {for (j in 1:n)
oysemi1[j,,i]<- difmsemi1[i,]*%*%mcovinvsemi1[,,i]*%*%dadostransp[,j]}

oysemi2<-array(c(0), dim=c(n,1,n2)) #calcula primeira parte do o(y) sem obs
  i grupo 2#
for ( i in 1:n2) {for (j in 1:n)
oysemi2[j,,i]<- difmsemi2[i,]*%*%mcovinvsemi2[,,i]*%*%dadostransp[,j]}

aux3<-matrix(0,n1,4)
for (i in 1:n1)
aux3[i,]<-media1semi[i]+media2
aux3t<-t(aux3)

aux4<-matrix(0,n2,4)
for ( i in 1:n2)
aux4[i,]<-media2semi[i]+media1
aux4t<-t(aux4)

ctesemi1<-matrix(0,n1,1)
for (i in 1:n1)
ctesemi1[i,]<-0.5*difmsemi1[i,]*%*%mcovinvsemi1[,,i]*%*%aux3t[,i]

ctesemi2<-matrix(0,n2,1)
for (i in 1:n2)
ctesemi2[i,]<-0.5*difmsemi2[i,]*%*%mcovinvsemi2[,,i]*%*%aux4t[,i]
lnodeysemi1<-array(c(0), dim=c(n,1,n1))

```

```

for(i in 1:n1) {for (j in 1:n)
lnodeysemi1[j,,i]<-oysemi1[j,,i]-ctesemi1[i]} #calcula o log-odds da obs
sem i grupo 1#

lnodeysemi2<- array(c(0), dim=c(n,1,n2))
for(i in 1:n2) { for (j in 1:n)
lnodeysemi2[j,,i]<-oysemi2[j,,i]-ctesemi2[i] } #calcula o log-odds da obs
sem i grupo 2#

lnodeysemi1g1<-matrix(0,n1,1)
for ( i in 1:n1)
lnodeysemi1g1[i,]<-sum(lnodeysemi1[,,i])

lnodeysemi2g2<-matrix(0,n2,1)
for ( i in 1:n2)
lnodeysemi2g2[i,]<-sum(lnodeysemi2[,,i])

LORAM1<-matrix(0,n1,1)
for (i in 1:n1)
LORAM1[i,]<-1/n*(sum(lnodey)-lnodeysemi1g1[i,])

LORAM2<-matrix(0,n2,1)
for (i in 1:n2)
LORAM2[i,]<-1/n*(sum(lnodey)-lnodeysemi2g2[i])

par(mfrow=c(1,2))
plot(x=1:n1, LORAM1,type="h", xlab="numero da observação (grupo 1)" ,
ylab="LORAM1",xlim=c(1,n1),ylim=c(-450,0))
plot(x=1:n2, LORAM2,type="h", xlab="numero da observação (grupo 2)" ,
ylab="LORAM2",xlim=c(1,n2))

```

- Cálculo de medida de influência para Δ^2 (Capítulo 2):

$$I(\mathbf{x}; \Delta^2) = w_1 \Delta^2 + 2\phi - w_1 \phi^2$$

```

infludelta1<-matrix(0,n1,1)
for (i in 1:n1)
infludelta1[i,]<-w1*delta2+2*(fi1[i])-w1*(fi1[i]^2)      #calcula função de#
#influencia para delta do grupo 1#

infludelta2<-matrix(0,n2,1)
for (i in 1:n2)
infludelta2[i,]<-w2*delta2+2*(fi2[i])-w2*(fi2[i]^2)      #calcula função de#
#influencia para delta do grupo 2#

par(mfrow=c(1,2))

plot(x=1:n1, infludelta1,type="h", xlab="numero da observação (grupo 1)" ,
     ylab=" f.influencia",xlim=c(1,n1))

plot(x=1:n2, infludelta2,type="h", xlab="numero da observação (grupo 2)" ,
     ylab="f. influencia",xlim=c(1,n2))

```

- Cálculo da medida $I_m(\mathbf{x}; \Delta^2)$ (Capítulo 2):

$$I_m(\mathbf{x}; \Delta^2) = w_1 (\phi - w_1^{-1})^2$$

```

Im1<-matrix(0,n1,1)
for(i in 1:n1)
Im1[i,]<-w1*((fi1[i]-1/w1)^2) #calcula o Im de Campbell para o grupo 1#

Im2<-matrix(0,n2,1)
for(i in 1:n2)
Im2[i,]<-w2*((fi2[i]-1/w2)^2) #calcula o Im de Campbell para o grupo 2#
par(mfrow=c(1,2))

plot(x=1:n1, Im1,type="h", xlab="numero da observação (grupo 1)" ,
     ylab="Im",xlim=c(1,n1))

plot(x=1:n2, Im2,type="h", xlab="numero da observação (grupo 2)" ,
     ylab="Im",xlim=c(1,n2))

```

- Cálculo de medida *DPCI*:

$$DPCI_i \cong \frac{\Phi\left(-\frac{1}{2}D\right)}{4D(n_1-1)^2} \left[(1-w_1\hat{\phi}_i)^2 (d_i^2 - \hat{\phi}_i^2 / D^2) + \frac{1}{4}\hat{\phi}_i^2 \right]$$

```
pci<- dnorm(-1/2*delta,0,1) #calcula a PCI#
pci
dpciaprox1<-matrix(0,n1,1)
for (i in 1:n1)
dpciaprox1[i,]<-( pci/(4*delta*((n1-1)^2))*((1-w1*fi1[i])^2 * (di1[i]-
((fi1[i]^2)/delta2))+ (1/4)*(fi1[i]^2))
dpciaprox2<-matrix(0,n2,1)
for (i in 1:n2)
dpciaprox2[i,]<-( pci/(4*delta*((n2-1)^2))*((1-w2*fi2[i])^2 * (di2[i]-
((fi2[i]^2)/delta2))+ (1/4)*(fi2[i]^2))
par(mfrow=c(1,2))
plot(x=1:n1, dpciaprox1,type="h", xlab="numero da observação (grupo 1)",
ylab="dpciaprox1 ",xlim=c(1,n1))
plot(x=1:n2, dpciaprox2,type="h", xlab="numero da observação (grupo 2)",
ylab="dpciaprox2 ",xlim=c(1,n2))
```

- Cálculo da medida de influência de Critchley:

$$\hat{\Delta}_{1(i)}^2 = \left(\frac{n_1 + n_2 - 1}{n_1 + n_2} \right) \left[\hat{\Delta}^2 - \frac{1}{w_1(n_1 - 1)} + \frac{\{e(x_{1i}) - w_1^{-1}\}^2}{w_1^{-1}(n_1 - 1) - \hat{\alpha}_1^2(x_{1i})} \right]$$

```
deltacritchley1<-matrix(0,n1,1)
for (i in 1:n1)
deltacritchley1[i,]<- ((n-1)/n) * ( delta2-(1/(w1*(n1-1)))+ ((fi1[i]-
1/w1)^2)/(((1/w1)*(n1-1))-di1[i]))
deltacritchley2<-matrix(0,n2,1)
for (i in 1:n2)
deltacritchley2[i,]<- ((n-1)/n) * ( delta2-(1/(w2*(n2-1)))+ (((fi2[i]-
1/w2)^2)/(((1/w2)*(n2-1))-di2[i])))
par(mfrow=c(1,2))
plot(x=1:n1, deltacritchley1,type="h", xlab="numero da observação (grupo 1)",
ylab="deltacritchley1 ",xlim=c(1,n1))
plot(x=1:n2, deltacritchley2,type="h", xlab="numero da observação (grupo 2)",
ylab="deltacritchely2 ",xlim=c(1,n2))
```

- Cálculo da medida $E(DLO_i)^2$:

$$E(DLO_i)^2 = (w_1 B_1 + w_2 B_2)^2 + V, \text{ expressão (2.30)}$$

$$\text{em que } B_1 = \frac{A_1}{2} - \frac{A_2}{2} \quad \text{e} \quad B_2 = \frac{-A_1}{2} - \frac{A_2}{2}$$

$$A_1 = \frac{D^2}{n-2} + c f_i (\hat{\phi}_i)^2 + c h \hat{\phi}_i - c h f_i d_i^2 \hat{\phi}_i$$

$$A_2 = c \cdot h \cdot (\hat{\phi}_i - h d_i^2) g_i$$

$$V = \frac{D^2}{(n-2)^2} + c \left(f_i \hat{\phi}_i + h g_i \left[\frac{2 \hat{\phi}_i}{n-2} + (f_i \hat{\phi}_i + h g_i) c d_i^2 \right] \right)$$

```
b1<-(-n1)/((n1-1)*(n-2))
b2<-(-n2)/((n2-1)*(n-2))
```

```
h1<-1/(n1-1)
h2<-1/(n2-1)
```

```
c<-(n-3)/(n-2)
```

```
gil<-matrix(0,n1,1)
gil<-1/(1+b1*dil)
gi2<-matrix(0,n2,1)
gi2<-1/(1+b2*di2)
```

```
ffil<-matrix(0,n1,1)
ffil<-b1*gil
```

```
ffi2<-matrix(0,n2,1)
ffi2<-b2*gi2
```

```
V1<-matrix(0,n1,1)
for (i in 1:n1)
V1[i,]<-(delta2/((n-2)^2)) + c*( ffil[i]*fil[i]+h1*gil[i])*
((2*fil[i])/(n-2)+ (ffil[i]*fil[i]+h1*gil[i])* c*dil[i] )
```

```
V2<-matrix(0,n2,1)
for (i in 1:n2)
V2[i,]<-(delta2/((n-2)^2)) + c*( ffi2[i]*fi2[i]+h2*gi2[i])*
((2*fi2[i])/(n-2)+ (ffi2[i]*fi2[i]+h2*gi2[i])* c*di2[i] )
```

```
Alg1<-matrix(0,n1,1)
for (i in 1:n1)
Alg1[i,]<-(delta2/(n-2))+(c*ffil[i]*(fil[i]^2))+(c*h1*fil[i])-(
c*h1*ffil[i]*dil[i]*fil[i])
```

```
Alg2<-matrix(0,n2,1)
for (i in 1:n2)
Alg2[i,]<-(delta2/(n-2))+(c*ffi2[i]*(fi2[i]^2))+(c*h2*fi2[i])-(
c*h2*ffi2[i]*di2[i]*fi2[i])
```

```

A2g1<-matrix(0,n1,1)
for (i in 1:n1)
A2g1[i,]<-c*h1*(fi1[i]-h1*di1[i])*gi1[i]

A2g2<-matrix(0,n2,1)
for (i in 1:n2)
A2g2[i,]<-c*h2*(fi2[i]-h2*di2[i])*gi2[i]

B1g1<-(A1g1/2)-(A2g1/2)
B1g2<-(A1g2/2)-(A2g2/2)

B2g1<-((-A1g1)/2)-(A2g1/2)
B2g2<-((-A1g2)/2)-(A2g2/2)

EDLO1<-matrix(0,n1,1)
EDLO1<-((w1*B1g1+w2*B2g1)^2)+V1

EDLO2<-matrix(0,n2,1)
EDLO2<-((w1*B1g2+w2*B2g2)^2)+V2
par(mfrow=c(1,2))

plot(x=1:n1, EDLO1,type="h", xlab="numero da observação (grupo 1)", ylab="EDLO1",xlim=c(1,n1))

plot(x=1:n2, EDLO2,type="h", xlab="numero da observação (grupo 2)", ylab="EDLO2",xlim=c(1,n2))

```

- Cálculo da medida E2 condicional :

$$E2c_i^0 = \left[chf_i(\hat{\phi}_i - hd_i^2)/(2b) \right]^2 + c^2 f_i^2 (\hat{\phi}_i + h/b)^2 (d_i^2 - \hat{\phi}_i^2 / D^2)$$

```

aux1g1<-matrix(0,n1,1)
for (i in 1:n1)
aux1g1[i,]<-(c*h1*ffi1[i]*(fi1[i]-h1*di1[i]))/2*b1

aux1g2<-matrix(0,n2,1)
for (i in 1:n2)
aux1g2[i,]<-(c*h2*ffi2[i]*(fi2[i]-h2*di2[i]))/2*b2

Vc1<-matrix(0,n1,1)
for(i in 1:n1)
Vc1[i,]<-c^2*(ffi1[i]^2)*((fi1[i]+h1/b1)^2)*(di1[i]-((ffi1[i]^2)/delta2))

Vc2<-matrix(0,n2,1)
for(i in 1:n2)
Vc2[i,]<-c^2*(ffi2[i]^2)*((fi2[i]+h2/b2)^2)*(di2[i]-((ffi2[i]^2)/delta2))

E2c1<-matrix(0,n1,1) #calcula a medida E2 condicional para o grupo 1#
for (i in 1:n1)
E2c1[i,]<-(aux1g1[i]^2)+Vc1[i]

E2c2<-matrix(0,n2,1) #calcula a medida E2 condicional para o grupo 2#
for (i in 1:n2)
E2c2[i,]<-(aux1g2[i]^2)+Vc2[i]

```



```

par(mfrow=c(1,2))
plot(x=1:n1, E2c1,type="h", xlab="numero da observação (grupo 1)" ,
ylab="E2c1 ",xlim=c(1,n1))
plot(x=1:n2, E2c2,type="h", xlab="numero da observação (grupo 2)" ,
ylab="E2c2",xlim=c(1,n2))

```

- Cálculo da medida E2 condicional aproximada:

$$E2c_i^0 = \left[\frac{\hat{\phi}_i^2}{4} + (1 - w_1 \hat{\phi}_i)^2 (d_i^2 - \hat{\phi}_i^2 / D^2) \right] / n_1^2$$

```

E2caprox1<-matrix(0,n1,1)
for (i in 1:n1)
E2caprox1[i,]<- ((1-w1*fi1[i])^2*(di1[i]-((fi1[i]^2)/delta2))+
(fi1[i]^2)/4)/(n1^2)

```

```

E2caprox2<-matrix(0,n2,1)
for (i in 1:n2)
E2caprox2[i,]<- ((1-w2*fi2[i])^2*(di2[i]-((fi2[i]^2)/delta2))+
(fi2[i]^2)/4)/(n2^2)

```

```

par(mfrow=c(1,2))
plot(x=1:n1, E2caprox1,type="h", xlab="numero da observação (grupo 1)" ,
ylab="E2caprox1 ",xlim=c(1,n1))
plot(x=1:n2, E2caprox2,type="h", xlab="numero da observação (grupo 2)" ,
ylab="E2caprox2 ",xlim=c(1,n2))

```

Apêndice D

Referências Bibliográficas

- Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. 2nd ed. New York : Wiley Series.
- Andrews, D. F. ; Bickel, P. J. ; Hampel, F.R. ; Huber, P. J. ; Rogers, W. H. ; Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press.
- Belsley, D. A.; Kuh, E.; Welsch, R. E. (1980). *Regression Diagnostics – Identifying Influential Data and Sources of Collinearity*. Wiley Series.
- Campbell, N. A. (1978). *The Influence Function as an Aid in Outlier Detection in Discriminant Analysis*. *Applied Statistics*, 27 nr 3: 251-258.
- Campbell, N.A.; Mahon, R.J. (1974). *A multivariate study of variation in two species of rock crab of the genus *Leptograpsus**. *Aust. J. Zool.*, 22, 417-425.

Cook, R. D.; Weisberg, S. (1982). *Residuals and Influence in Regression*. New York. Chapman & Hall.

Critchley, F. ; Vitiello, C. (1991). *The Influence of Observations Missclassification Probability Estimates in Linear Discriminant Analysis*. *Biometrika*, 78: 677-690.

Devlin, S. J.; Gnanadesikan, R. E.; Kettenring, J. R. (1975). *Robust Estimation and Outlier Detection with Correlation Coefficients*. *Biometrika*, 62: 531-545.

Efron, B. (1975). *The Efficiency of Logistic Regression Compared to Normal Discriminant Analysis*. *Journal of the American Statistical Association*, 70: 892-898.

Efron, B.; Tibshirani, R. (1993) . *An introduction to the bootstrap*. Chapman & Hall.

Fung, W. K. (1992). *Some Diagnostic Measures in Discriminant Analysis*. *Statistics and Probability Letters*, 13: 279-285.

Fung, W. K. (1993). *Unmasking Outliers and Leverage Points: a Confirmation*. *Journal of the American Statistical Association*, 88: 515-519.

Fung, W. K. (1995-a). *Diagnostics in Linear Discriminant Analysis*. *Journal of the American Statistical Association*, 90 nr 43: 952-956.

Fung, W. K. (1995-b). *Detecting Influential Observations for Estimated Probabilities in Multiple Discriminant Analysis*. *Computational Statistics & Data analysis*, 20: 557-568.

Fung, W. K. (1996-a). *Diagnosing Influential Analysis*. *Biometrics*, 52: 1235-1241.

- Fung, W. K. (1996-b). *The Influence of Observations for Local Log-odds in Linear Discriminant Analysis*. *Communications in Statistics – Theory and Methods*, 25: 257-268.
- Fung, W. K. (1996-c). *The Influence of Observations on Missclassification Probability in Multiple Discriminant Analysis*. *Communications in Statistics – Theory and Methods*, 25(8), 1917-1930.
- Fung, W. K. (1998). *On the Equivalence of Two Diagnostic Measures in Discriminant Analysis*. *Communications in Statistics – Theory and Methods*, 27(8): 1915-1922.
- Geisser, S. (1964). *Posterior Odds for Multivariate Normal Classifications*. *Journal of the Royal Statistical Society, Ser B*, 26: 69-76.
- Geisser, S. (1966). “*Predictive Discrimination*” in *Proceedings of the International Symposium on Multivariate Analysis*. New York: Academic Press, pg 149-163.
- Gnanadesikan, R. (1997). *Methods for Statistical Data Analysis of Multivariate Observations*. 2nd ed. Wiley Series, USA .
- Graybill, F. A. (1976). *Theory and Application of the Linear Model*. Duxbury Press, Belmont, California, Wadsworth Publishing Company Inc.
- Hampel, F. R. (1974). *The Influence Curve and its Role in Robust Estimation*. *Journal of the American Statistical Association*, 69: 383-393.
- Healy, M. J. R. (1968). *Multivariate Normal Plotting*. *Applied Statistics* 17 : 157-161.
- James, B. R. (1996). *Probabilidade: Um Curso em Nível Intermediário*. 2^a ed.. Rio de Janeiro – IMPA.

Johnson, R. A.; Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis*. 3rd ed. Prentice Hall, Upper Saddle River, New Jersey.

Johnson, W. (1985). *Influence Measures for Logistic Regression: Another Point of View*. *Biometrika*, 72: 59-65.

Johnson, W. (1987). *The Detection of Influential Observations for Allocation, Separation, and the Determination of Probabilities in a Bayesian Framework*. *Journal of Business & Economic Statistics* 5: 369-381.

Johnson, W.; Geisser, S. (1983). *A Predictive View of the Detection and Characterization of Influential Observations in Regression Analysis*. *Journal of the American Statistical Association*, 78: 137-144.

Krause, A.; Olson, M. (2002). *The Basics of S-plus*. Third edition. Springer-Verlag New York Inc.

Kshirsagar, A. M. (1972). *Multivariate Analysis*. Marcel Dekker.

Kullback, S.; Leibler, R. A. (1951). *On Information and Sufficiency*. *The Annals of Statistics*, 22: 79-86.

Leite, J.G.; Singer, J. M. e Lourenço, F.C. (1994). *Relatório de análise Estatística sobre o projeto: Biologia de uma comunidade de marsupiais e roedores, em Floresta Atlântica de montanha, no Parque Estadual da Serra do Tabuleiro, Santa Catarina, Brasil*. São Paulo, IME - USP, 33 p (RAE-CEA-94P11)

Mallows, C. L. (1973). *Influence Functions*. Unpublished Talk Presented at the Working Conference on Robust Regression at National Bureau of Economics Research in Cambridge, MA.

McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. New York ; John Wiley.

Morrison, D. F. (1976) . *Multivariate Statistical Methods* (2nd ed) New York McGraw Hill.

Rousseeuw, P. J.; van Zomeren, B. C. (1990). *Unmasking Multivariate Outliers and Leverage Points (with discussion)*. *Journal of the American Statistical Association*, 85: 633-651.

Searle, R. S. (1982) . *Matrix Algebra Useful for Statistics* . Wiley Series.

Venables, W. N.; Ripley, B. D. (1994). *Modern Applied Statistics with S-plus*. Springer-Verlag New York Inc.

Welsch, R. E. (1982). "Influence Functions and Regression Diagnostics" in *Modern Data Analysis* . Eds. R. L. Launer e A. F. Siegel, New York Academic Press.