

INFLUÊNCIA LOCAL E ANÁLISE
DE RESÍDUOS EM MODELOS DE
REGRESSÃO VON MISES

Francisco Antonio Morais de Souza

TESE APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA A OBTENÇÃO DO GRAU

DE

DOUTOR EM ESTATÍSTICA

Área de Concentração: **Estatística**

Orientador: **Prof. Dr. Gilberto Alvarenga Paula**

São Paulo, 5 de fevereiro de 1999

INFLUÊNCIA LOCAL E ANÁLISE DE RESÍDUOS
EM MODELOS DE REGRESSÃO VON MISES

Este exemplar corresponde à redação final
da tese devidamente corrigida, apresentada por
Francisco Antonio Morais de Souza e
aprovada pela Comissão Julgadora.

São Paulo, 5 de fevereiro de 1999

Banca Examinadora

Prof. Dr. Gilberto Alvarenga Paula (IME-USP)

Prof. Dr. Rinaldo Artes (IME-USP)

Prof. Dr. Josemar Rodrigues (ICMC-USP)

Prof. Dr. Gauss Moutinho Cordeiro (UFPE)

Prof. Dr. Manuel Galea Rojas (UNIV. VALPARAISO)

Agradecimentos

- Em primeiro lugar, quero agradecer aos colegas do Departamento de Matemática e Estatística da Universidade Federal da Paraíba - Campus de Campina Grande, pelo incentivo e pelo apoio que me foram dados;
- Registro ainda meu agradecimento ao Prof. Gilberto Alvarenga Paula pela confiança, como também pela forma solidária, consistente e segura de conduzir a orientação dessa tese;
- A Antonieta, Cardoso e Eliana, que ingressaram comigo nesse programa de doutorado, agradeço por ter desfrutado de suas amizades como também pelas boas e proveitosas trocas de idéias ao longo desses quatro anos;
- Ao curso de pós-graduação do IME - USP pela oportunidade de poder concretizar uma importante etapa de minha vida;
- À CAPES, agradeço pelo apoio financeiro, através do PICD (Programa Institucional de Capacitação Docente);
- De forma especial, agradeço aos meus pais Antonio e Geralda, que nunca mediram esforços na tarefa de educar-me;
- Agradeço ainda à minha mulher Rosana e ao meu filho Ricardo que acompanharam, incentivaram e tanto torceram pelo sucesso desta tese;
- Finalmente, e acima de tudo, agradeço a Deus por ter permitido a concretização de um sonho.

Resumo

Uma etapa importante após a formulação e ajuste de um modelo de regressão é a análise de diagnóstico. Neste trabalho são tratados alguns aspectos de diagnóstico em modelos de regressão von Mises. É empregada a abordagem de influência local sobre a *likelihood displacement* para o modelo misto, onde os dois parâmetros da distribuição von Mises são modelados através de funções de ligação apropriadas. Considerando o modelo de médias von Mises, são propostos dois resíduos padronizados obtidos a partir dos componentes da função desvio. Nós mostramos que as distribuições de probabilidade desses resíduos podem ser aproximadas pela distribuição normal padrão e apresentamos o resultado de um estudo de simulação desenvolvido com o objetivo de avaliar o comportamento de algumas características dessas distribuições. Os resultados são aplicados a dois conjuntos de dados reais.

Abstract

In this work we consider some diagnostic aspects in von Mises regression models. By using the likelihood displacement we employ the local influence approach in the case where both of the von Mises distribution parameters are modeled by suitable link functions. On considering the von Mises mean model, we propose two standardised residuals which were obtained from the residual deviance. We proved that the probability distributions of these residuals are close to the standard normal distribution. We present a simulation study made to compare some distributional measures of these residuals. The results were applied to two real data sets.

Conteúdo

1	Introdução	1
1.1	Influência local	1
1.2	Dados direcionais	2
1.3	Apresentação dos capítulos	4
2	Análise de Influência	6
2.1	Motivação	6
2.2	Diagnóstico em modelos de regressão	7
2.3	Afastamento pela verossimilhança	14
2.4	Influência local	18
2.4.1	Outras expressões para C_d	24

3	Introdução a Dados Circulares	27
3.1	Preliminares	27
3.2	Dados circulares	28
3.3	Medidas descritivas no círculo	30
3.4	Momentos trigonométricos	33
3.5	Modelos probabilísticos para dados circulares	35
3.5.1	A Distribuição uniforme circular - \mathcal{U}_c	36
3.5.2	A Distribuição normal arqueada - $\mathcal{WN}(\mu, \sigma^2)$	36
3.5.3	A Distribuição von Mises - $\mathcal{VM}(\mu, \lambda)$	37
3.5.4	Estimação dos parâmetros da distribuição von Mises	40
3.5.5	Outras distribuições circulares	42
4	Influência Local em Dados Circulares	43
4.1	Modelos de regressão para dados circulares	43
4.2	Modelos von Mises de regressão	45
4.2.1	O Modelo von Mises de médias	46
4.2.2	O Modelo von Mises de dispersão	48

4.2.3	O Modelo von Mises misto	51
4.3	Influência local no modelo von Mises misto	53
4.3.1	Ponderação de casos	56
4.3.2	Perturbando a variável resposta y	57
4.3.3	Perturbando uma covariável individualmente	59
4.4	Influência local sobre predições	62
5	Estudo de Resíduos em Dados Circulares	65
5.1	Introdução	65
5.2	Componente da função desvio	66
5.2.1	Resíduo d_i^*	67
5.2.2	Resíduo r_i	68
5.3	Simulação	72
6	Aplicações	76
6.1	Aplicação 1	76
6.2	Aplicação 2	90

7	Considerações Finais	98
7.1	Conclusões	98
7.2	Futuros Estudos	100
A	Intervalos de Confiança para os parâmetros da distribuição von Mises	101
A.1	Intervalos de confiança para μ	101
A.2	Intervalos de confiança para λ	102
B	Alguns resultados sobre as funções de Bessel	103
B.1	Derivadas	103
B.2	Avaliação de I_0 e I_1	104
B.3	Fórmulas de recorrência	105
C	Geração de Envelopes	106

Lista de Figuras

2.1	Gráficos de dispersão e retas ajustadas.	10
2.2	Gráfico de influência tridimensional.	19
3.1	Representação gráfica de uma amostra de 3 observações circulares.	31
3.2	Densidade da distribuição von Mises para $\mu = 0^\circ$ e $\lambda = 1/2, 1, 2$ e 4	38
3.3	Densidade da distribuição von Mises para $\lambda = 1$ e $\mu = 0^\circ, 90^\circ$ e 180°	39
6.1	Localização dos pombos no desaparecimento.	78
6.2	Localização dos pombos: Local de soltura.	79
6.3	Localização dos pombos: Período da soltura.	80
6.4	Distância entre as direções dos pombos (MANHÃ e TARDE).	81
6.5	Gráfico dos envelopes para os resíduos r e d^*	82

6.6	Gráfico de dispersão da diferença entre as direções dos pombos.	83
6.7	Diferença entre as direções dos pombos.	84
6.8	Gráfico dos envelopes para os resíduos r e d^*	84
6.9	Gráfico dos envelopes para o resíduo r	86
6.10	Gráfico dos envelopes para o resíduo d^*	87
6.11	Influência das observações sobre as estimativas dos parâmetros.	88
6.12	Influência das observações sobre as estimativas de γ	89
6.13	Gráfico dos envelopes para o resíduo d^* (modelo de médias).	91
6.14	Gráfico dos envelopes para o resíduo r (modelo de médias).	92
6.15	Gráfico dos envelopes para o resíduo d^* (modelo misto).	93
6.16	Gráfico dos envelopes para o resíduo r (modelo misto).	93
6.17	Influência dos casos sobre as estimativas dos parâmetros.	94
6.18	Influência das observações sobre as estimativas de γ	95
6.19	Influência dos casos sobre as previsões.	96
6.20	Influência das respostas sobre as previsões.	96

Lista de Tabelas

2.1	Dados hipotéticos.	9
5.1	Medidas de locação e escala para os resíduos padronizados r_i e d_i^*	75
6.1	Estimativas dos parâmetros do modelo von Mises misto.	82
6.2	Estimativas dos parâmetros do modelo von Mises de médias.	85
6.3	Estimativas dos parâmetros do modelo von Mises de dispersão.	85
6.4	Medidas de distância percorrida e direção escolhida por 31 caramujos (<i>Nodilittorina unifasciata</i>) após o deslocamento do seu <i>habitat</i> natural.	90

Capítulo 1

Introdução

1.1 Influência local

Quando se está ajustando um modelo a um conjunto de dados, é imprescindível que as estimativas obtidas a partir do modelo proposto sejam resistentes a “pequenas” perturbações nas observações. Se o modelo ajustado não apresentar uma boa descrição dos dados que foram observados, o mesmo pode conduzir a inferências errôneas .

Assim sendo, é importante que se faça um estudo sobre a robustez dos resultados obtidos, em termos dos vários aspectos que envolvem a formulação do modelo, e as estimativas dos seus parâmetros, ou seja, fazer uma *análise de diagnóstico*, que consiste de métodos para avaliar o grau de sensibilidade das inferências a pequenas perturbações nos dados ou mesmo no modelo proposto.

Dentre as medidas geralmente utilizadas para detectar influência de observações sobre

o modelo ajustado, duas se destacam. Uma delas é a distância D_i , sugerida por Cook (1977) - um dos precursores do estudo de influência em modelos de regressão. A distância de Cook consiste em comparar o valor ajustado para a variável resposta, quando a i -ésima observação é excluída da análise, com o valor ajustado que se obtém, considerando o ajuste com todas as observações. Outra medida de influência, que também utiliza a estratégia da exclusão de casos, é conhecida por $DFFITs_i$, e foi proposta por Belsley, Kuh e Welsch (1980). Cook, Peña e Weisberg (1988) comparam as duas medidas e mostram que as mesmas não medem as mesmas coisas e portanto podem ser usadas conjuntamente.

Uma abordagem mais moderna é o estudo de influência local, um conceito introduzido por Cook (1986), que se baseia na análise das curvaturas das seções normais de uma determinada superfície, denominada *gráfico de influência*, numa vizinhança de um ponto especial.

Thomas e Cook (1990) aplicam o método de influência local para as predições nos modelos lineares generalizados e, posteriormente, Paula (1996) faz uma extensão desse estudo para os modelos próprios de dispersão, propostos por Jørgensen (1997).

1.2 Dados direcionais

As distribuições de probabilidade que são encontradas com maior frequência são distribuições geralmente apropriadas para descrever variáveis aleatórias que possuem como suporte um intervalo na reta. Entretanto, existem situações em que essa representação não ocorre de forma natural, por exemplo as *direções* medidas a partir de um determinado ponto de origem.

Em particular, a direção de deslocamento de um animal, ou a direção da rachadura de uma rocha, comuns em pesquisas nos campos da Zoologia e da Geologia, são situações em que não é apropriado considerar um intervalo da reta como suporte para a distribuição de probabilidade correspondente. Observações como essas, são naturalmente reportadas como *ângulos*, podendo variar entre 0 e 2π radianos, ou entre 0° e 360° , de forma que elas podem ser consideradas como pontos sobre a circunferência do círculo unitário. Em alguns casos podem ser consideradas sobre a superfície de uma esfera.

As distribuições de probabilidade univariadas que se aplicam a dados não direcionais são comumente chamadas de *lineares* pelo fato de seus gráficos serem plotados sobre a linha reta, como por exemplo a distribuição binomial, a gama, a normal, etc. Da mesma forma, as observações são também denominadas de *lineares*.

O interesse no desenvolvimento de técnicas para o tratamento de dados direcionais é bastante antigo. A teoria dos erros foi desenvolvida por Gauss (1823) com o objetivo inicial de analisar determinadas medidas direcionais em Astronomia. Tendo em vista que os erros envolvidos eram suficientemente pequenos, pode-se dizer que um *acidente histórico* o permitiu fazer uma aproximação linear. Com isto, Gauss desenvolveu toda uma teoria *linear* em vez de uma teoria *direcional*.

Uma questão que surge naturalmente está relacionada com a necessidade de se buscar técnicas especiais para analisarmos as observações direcionais. Em inúmeras situações nós nos deparamos com observações direcionais que não podem ser analisadas através de uma aproximação linear, pois poderemos obter conclusões contraditórias. Por que não analisarmos uma amostra de observações direcionais pelos métodos usuais utilizados para dados *lineares*? Afinal, podemos fazer um “corte” no círculo e considerar o intervalo $[0, 2\pi)$ como um intervalo na reta. Acontece que esse artifício pode causar problemas, principalmente se os valores observados estiverem dispersos em toda a circunferência, e

não simplesmente concentrados sobre um determinado arco do círculo.

Além disso, surge a questão de saber onde podemos “cortar” o círculo. Por exemplo, se o “corte” for feito no ponto correspondente a 0° e em seguida analisarmos os resultados como se fosse uma amostra de observações lineares, então quaisquer dois ângulos θ° e $360^\circ - \theta^\circ$ teriam uma média aritmética igual a 180° , mesmo que esses ângulos estejam mais próximos de 0° do que de 180° . Com isto, já podemos perceber que os métodos utilizados para variáveis lineares não podem ser estendidos para variáveis direcionais de forma imediata.

A álgebra de ângulos é de certa forma diferente da álgebra que rege os números reais. Como também, a escolha da direção *zero* (como o eixo X ou a direção norte) é feita arbitrariamente. Além disso, não devemos esquecer do fato de existir uma invariância modulo (2π) nas operações envolvendo ângulos.

1.3 Apresentação dos capítulos

No Capítulo 2 tratamos do problema de diagnóstico em modelos de regressão, com ênfase no método de influência local, através do afastamento pela verossimilhança (tradução livre de *likelihood displacement*, Cook e Weisberg (1982)). No Capítulo 3 apresentamos alguns conceitos básicos relacionados com variáveis aleatórias que representam direções e atenção especial é dada à distribuição von Mises. Essa distribuição é apropriada para variáveis aleatórias que assumem valores no círculo unitário, ou seja, representam direções em um espaço de dimensão dois. No Capítulo 4 discutimos modelos de regressão von Mises, propostos por Fisher e Lee (1992) e fazemos um estudo de influência local nesses modelos. O Capítulo 5 trata de análise de resíduos para dados circulares. São

propostos dois resíduos padronizados, obtidos a partir dos componentes da função desvio. Apresentamos o resultado de um estudo de simulação desenvolvido com o objetivo de investigar o comportamento de algumas características distribucionais desses dois resíduos. No Capítulo 6 fazemos aplicações dos resultados obtidos a dois conjuntos de dados reais. O Capítulo 7 finaliza esta tese com nossas conclusões e alguns direcionamentos para estudos subsequentes.

Capítulo 2

Análise de Influência

2.1 Motivação

Numa análise estatística é importante levar em consideração a estabilidade dos resultados obtidos e das inferências de um modo geral, com relação a possíveis perturbações nos dados ou no modelo estatístico.

Enquanto as análises de resíduos são feitas para investigar problemas com o modelo ajustado, uma análise de diagnóstico é feita assumindo o modelo como correto, e investiga-se a robustez das conclusões a pequenas perturbações nos dados.

Os elementos do conjunto de dados que efetivamente controlam aspectos da análise são ditos *influentes*. Em particular, uma observação é dita influente se ela produzir alterações relevantes no resultado da análise, quando a mesma for excluída, ou submetida a uma pequena perturbação. É importante salientar o fato de que os modelos teóricos

são descrições aproximadas de processos reais bem mais complexos.

Um dos métodos de diagnóstico empregados em modelos de regressão utiliza a estratégia da exclusão de casos, e consiste em identificar observações que influenciam, de maneira desproporcional, um particular aspecto da análise.

As medidas D -Cook e $DFFITs$ são utilizadas para diagnosticar globalmente possíveis influências e, em geral, podem ser calculadas com as informações usualmente fornecidas pelos programas computacionais desenvolvidos para análises estatísticas, tais como *SAS*, *SPSS*, *S-Plus*, *Minitab*, *BMDP*, etc. Alguns desses aplicativos já fornecem automaticamente essas medidas.

2.2 Diagnóstico em modelos de regressão

Apresentaremos a seguir as duas medidas de diagnóstico mais usuais, considerando o modelo normal linear. Para tanto vamos admitir que dispomos de uma amostra aleatória $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$, com a suposição de que a variável resposta y_i tem distribuição normal $N(\mu_i, \sigma^2)$ e que sua média μ_i segue uma estrutura linear do tipo $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$ onde $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ é um vetor de covariáveis e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é um vetor de p parâmetros desconhecidos, a serem estimados, onde $p \leq n$.

Esse modelo pode ser escrito numa forma matricial como $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, onde $\mathbf{y} = (y_1, \dots, y_n)^T$, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$, também conhecida como matriz de especificação do modelo, e $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ é um vetor de erros não observáveis com distribuição conjunta normal n -variada $N_n(\mathbf{0}, \sigma^2 \mathbf{I})$.

O estimador de mínimos quadrados do vetor de parâmetros $\boldsymbol{\beta}$ é dado pela solução do

sistema de equações normais

$$\mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0. \quad (2.1)$$

Quando a matriz \mathbf{X} for de posto completo, então o estimador de mínimos quadrados do vetor $\boldsymbol{\beta}$ fica dado por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (2.2)$$

Caso contrário, quando $\text{posto}(\mathbf{X}) < p$, a solução do sistema de equações normais (2.1) não é única, e nesse caso, substituímos $(\mathbf{X}^T \mathbf{X})^{-1}$ em (2.2) por uma inversa generalizada (g -inversa), com a qual são obtidos resultados análogos ao caso em que se tem o posto completo, ou seja, $\text{posto}(\mathbf{X}) = p$. Maiores detalhes podem ser encontrados, por exemplo, em Rao (1973). Assumiremos aqui que a matriz \mathbf{X} é de posto completo, $\text{posto}(\mathbf{X}) = p$.

O estimador de mínimos quadrados do vetor de médias $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ é dado por

$$\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}} \implies \hat{\boldsymbol{\mu}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = \mathbf{H}\mathbf{y}, \quad (2.3)$$

onde $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ é a matriz de projeção ortogonal sobre o espaço vetorial gerado pelas colunas da matriz \mathbf{X} . Na literatura, \mathbf{H} é conhecida como matriz “*hat*”, uma vez que ela transforma o vetor de valores observados \mathbf{y} no vetor de valores ajustados, Weisberg (1985).

O vetor de resíduos ordinários é dado por

$$\mathbf{r} = \mathbf{y} - \hat{\boldsymbol{\mu}} = (\mathbf{I} - \mathbf{H})\mathbf{y} \quad (2.4)$$

e satisfaz $E(\mathbf{r}) = \mathbf{0}$, $\text{Var}(\mathbf{r}) = \sigma^2(\mathbf{I} - \mathbf{H})$. Uma estimativa não viesada para a variância σ^2 é dada por $s^2 = \mathbf{r}^T \mathbf{r} / (n - p)$, e resíduos “*Studentizados*” são obtidos por

$r_i/\{s^2(1-h_{ii})\}^{1/2}$, onde h_{11}, \dots, h_{nn} são os elementos da diagonal principal da matriz “hat”.

A título de ilustração, mostraremos um exemplo fictício, apresentado por Anscombe (1973) e Weisberg (1985), onde 4 conjuntos de dados (**a**, **b**, **c** e **d**) fornecem as mesmas estimativas de mínimos quadrados para os parâmetros da reta ajustada.

Tabela 2.1: Dados hipotéticos.

Obs.	Conjunto de Dados							
	a		b		c		d	
	x	y	x	y	x	y	x	y
1	10	8,04	10	9,14	10	7,46	8	6,58
2	8	6,95	8	8,14	8	6,77	8	5,76
3	13	7,58	13	8,74	13	12,74	8	7,71
4	9	8,81	9	8,77	9	7,11	8	8,84
5	11	8,33	11	9,26	11	7,81	8	8,47
6	14	9,96	14	8,10	14	8,84	8	7,04
7	6	7,24	6	6,13	6	6,08	8	5,25
8	4	4,26	4	3,10	4	5,39	19	12,50
9	12	10,84	12	9,13	12	8,15	8	5,56
10	7	4,82	7	7,26	7	6,42	8	7,91
11	5	5,68	5	4,74	5	5,73	8	6,89

Apesar de uma análise visual desses conjuntos de dados sugerir comportamentos diferentes, se fizermos a suposição de um relacionamento linear $y = \beta_0 + \beta_1 x$ entre as variáveis x e y , um ajuste por mínimos quadrados fornece as mesmas estimativas para os parâmetros β_0 e β_1 nos 4 casos, ou seja, $\hat{y} = 3 + 0,5x$, com uma estimativa do desvio

padão $s = \sqrt{1,53}$ e $R^2 = 67\%$. Baseando-se apenas nesses resultados, tudo leva a crer que esse modelo de regressão linear é apropriado para todas as 4 situações. Entretanto, com um diagrama de dispersão percebe-se que o ajuste de um modelo de regressão linear simples não é adequado em todos os casos apresentados, como pode ser visto nos gráficos da Figura 2.1. Essa figura mostra o diagrama de dispersão, para cada um dos

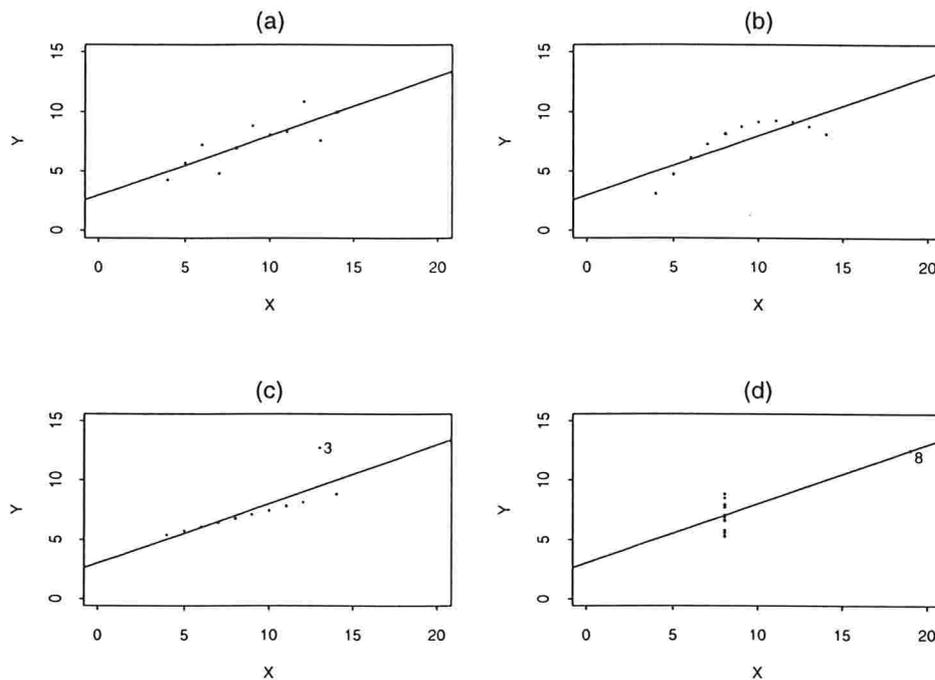


Figura 2.1: Gráficos de dispersão e retas ajustadas.

conjuntos de dados, como também a correspondente reta ajustada. Analisando esses gráficos podemos concluir o seguinte:

1. Para o conjunto de dados (a) não há indícios de problemas com o ajuste;
2. Para o conjunto (b), falta uma componente quadrática em x , ou seja, o modelo linear simples não é adequado;

3. A 3ª observação do conjunto (c) não está bem ajustada, havendo indícios de ser uma observação aberrante;
4. Por fim, as estimativas dos parâmetros para o conjunto (d) foram altamente influenciadas pela 8ª observação.

Situações como estas mostram a necessidade de se fazer uma análise mais minuciosa sobre os vários aspectos do ajuste de um modelo de regressão.

Sabemos que o estimador de mínimos quadrados é bastante sensível a observações cujos resíduos ordinários são elevados – as quais são ditas pobremente ajustadas. Da mesma forma, o estimador de mínimos quadrados é também sensível a pontos extremos no espaço gerado pela matriz de especificação do modelo, Pregibon (1981). Para o i -ésimo resíduo ordinário definido na equação (2.4), tem-se que $\text{Var}(r_i) = \sigma^2(1 - h_{ii})$, e é possível mostrar que $\frac{1}{n} \leq h_{ii} \leq \frac{1}{c}$, ver Cook e Weisberg (1982, p. 12), onde c representa o número de linhas da matriz \mathbf{X} que são iguais a \mathbf{x}_i^T . Assim, quando h_{ii} se aproxima de seu limite superior, tem-se que $\text{Var}(r_i)$ torna-se pequena, e portanto o resíduo r_i fica próximo de seu valor esperado, que é zero, com isto, segue-se que $\hat{\mu}_i \simeq y_i$. Nesse caso, já se sabe de antemão que a i -ésima observação deverá ter um resíduo pequeno, independentemente do valor observado para a variável resposta.

Essa exposição acima pode também ser feita algebricamente, usando as propriedades da matriz “hat”. Basta lembrar que \mathbf{H} só depende de \mathbf{X} e que é uma matriz de projeção, sendo portanto, simétrica e idempotente. Assim,

$$\begin{aligned} h_{ii} &= \sum_{j=1}^n h_{ij}h_{ji} \\ &= \sum_{j=1}^n h_{ij}^2 \end{aligned}$$

$$= h_{ii}^2 + \sum_{j \neq i} h_{ij}^2.$$

E segue-se daí que $h_{ii}(1 - h_{ii}) = \sum_{j \neq i} h_{ij}^2 \geq 0$. Por conseguinte, se o elemento h_{ii} for próximo de 1, os demais elementos de sua linha serão próximos de zero. Fazendo uma decomposição na expressão do valor ajustado, equação (2.3) $\hat{\mu}_i = h_{ii}y_i + \sum_{j \neq i} h_{ij}y_j$, vê-se que, se h_{ii} próximo de 1, o valor observado y_i irá influenciar fortemente a estimativa do próprio valor ajustado $\hat{\mu}_i$.

Por representar o poder de impacto que uma observação exerce sobre o próprio valor ajustado, o elemento h_{ii} é conhecido na literatura como “*leverage*” do i -ésimo caso.

Para uma matriz \mathbf{X} de posto completo tem-se que $\sum_{i=1}^n h_{ii} = p$. Baseado nisso, Hoaglin e Welsch (1978) sugerem que sejam considerados pontos com alto “*leverage*” aquelas observações cujos h_{ii} excedam o valor $2p/n$.

Com relação aos elementos que estão fora da diagonal principal da matriz \mathbf{H} , eles trazem informação sobre a influência que dois casos exercem conjuntamente sobre o ajuste do modelo. Em particular, h_{ij} pode ser interpretado como a influência que y_i exerce sobre $\hat{\mu}_j$, e também a influência que y_j exerce sobre $\hat{\mu}_i$.

Pelo que vimos, os elementos h_{11}, \dots, h_{nn} são úteis para detectar os pontos extremos e a influência exercida por cada observação sobre a respectiva predição. Dispõe-se atualmente de várias medidas de diagnóstico para detectar observações que exercem influência sobre outros aspectos do ajuste, tais como estimativas dos parâmetros, e, em geral, essas medidas combinam informações contidas tanto em h_{ii} quanto no resíduo ordinário r_i . Apresentaremos aqui apenas duas dessas medidas, que são conhecidas como D_i , sugerida por Cook (1977), e $DFFITs_i$, que foi proposta por Belsley, Kuh e Welsch (1980). Essas medidas se destinam a avaliar a influência individual de cada

observação sobre a estimativa do vetor de coeficientes β . A medida $DFFITs_i$ também avalia a influência sobre a estimativa da variância σ^2 .

Na análise de um modelo normal linear, quando a i -ésima observação é excluída do ajuste, a distância de Cook é definida como

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(-i)})^T (\mathbf{X}^T \mathbf{X})(\hat{\beta} - \hat{\beta}_{(-i)})}{ps^2}, \quad (2.5)$$

onde o índice $(-i)$ significa que a estimativa do vetor β está sendo feita sem a i -ésima observação. Essa medida está baseada em elipsóides de confiança, e sua análise é feita geralmente através de um "index plot". Os valores de D_i são comparados com o quantil da distribuição $F_{p,(n-p)}$, numa analogia com regiões de confiança.

Usando álgebra matricial, podemos expressar a nova estimativa de β como $\hat{\beta}_{(-i)} = \hat{\beta} - \frac{r_i}{1 - h_{ii}} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_i$, e com esta relação a distância de Cook pode ainda ser escrita na forma

$$D_i = \left\{ \frac{r_i}{s(1 - h_{ii})^{1/2}} \right\}^2 \left(\frac{h_{ii}}{1 - h_{ii}} \right) \frac{1}{p}. \quad (2.6)$$

Da equação (2.6), podemos observar que D_i será grande quando o resíduo "Studentizado" $t_i = \frac{r_i}{s(1 - h_{ii})^{1/2}}$ for grande e/ou h_{ii} for próximo de um. Nesse caso, dizemos que o ponto é ao mesmo tempo aberrante e influente. Para uma observação com r_i grande e h_{ii} pequeno (próximo de $1/n$), pode acontecer que D_i não detecte esse ponto. Por essa razão, a distância de Cook pode não ser adequada em situações como essa. Esse problema não existe com a medida $DFFITs_i$, que também é dada por uma distância entre as estimativas $\hat{\beta}$ e $\hat{\beta}_{(-i)}$, na métrica de $(\mathbf{X}^T \mathbf{X})$, da mesma forma que D_i , mas usa uma estimativa para a variância σ^2 calculada com os dados incompletos, ou seja, excluindo a i -ésima observação. Denotando essa estimativa por $s_{(-i)}^2$, a medida é definida

como

$$\begin{aligned} DFFITS_i &= \sqrt{\frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)})^T (\mathbf{X}^T \mathbf{X}) (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)})}{s_{(-i)}^2}} \\ &= \frac{|r_i|}{s_{(-i)}(1 - h_{ii})^{1/2}} \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2}. \end{aligned}$$

Em geral, essa medida também é analisada graficamente, e atenção especial deve ser dada para as observações com $DFFITS_i \geq 2\sqrt{p/(n-p)}$.

Diversos autores, dentre os quais Atkinson (1981), Belsley, Kuh, e Welsch (1980), Cook e Weisberg (1982), Hoaglin e Welsch (1978), defendem o uso de uma ou outra medida. Cook, Peña e Weisberg (1988) mostram que é possível expressar D_i e $DFFITS_i$ através de uma medida unificada, o afastamento pela verossimilhança (ou *likelihood displacement*), que veremos na próxima secção. Cook, Peña e Weisberg concluem que podemos empregar simultaneamente D_i e $DFFITS_i$, pois elas não medem necessariamente as mesmas coisas.

2.3 Afastamento pela verossimilhança

Uma outra maneira de avaliar a influência que uma observação exerce sobre as estimativas do modelo é através de uma medida construída a partir da função de log-verossimilhança. Para motivar, vamos admitir, por enquanto, que o parâmetro σ^2 seja conhecido. Se escrevermos $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\mu}}_{(-i)} = \mathbf{X}\hat{\boldsymbol{\beta}}_{(-i)}$, então a distância de Cook, equação (2.5), fica dada por

$$D_i = \frac{(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_{(-i)})^T (\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}_{(-i)})}{p\sigma^2}. \quad (2.7)$$

Ou seja, podemos interpretar D_i como uma distância entre as estimativas do vetor de médias, quando o ajuste é feito sem a i -ésima observação, com relação ao ajuste feito

com todo o conjunto de dados. Numa análise de diagnóstico, assumimos que o modelo postulado é correto, e desejamos comparar as estimativas obtidas através desse modelo com as respectivas estimativas decorrentes de uma pequena perturbação no mesmo. Em outras palavras, a análise é feita comparando-se as estimativas calculadas com um modelo perturbado em relação a um modelo sem perturbação.

Vamos agora pensar em perturbar o modelo normal linear $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, de uma forma infinitesimal. Podemos propor outros tipos de perturbação, em vez de considerar o esquema que consiste na exclusão de casos, como por exemplo:

1. Em vez de supormos que as respostas possuem a mesma variância, podemos admitir a possibilidade de se ter um modelo heterocedástico, isto é, supor $\text{Var}(y_i) = \sigma^2/\omega_i$;
2. Podemos ainda introduzir pequenas perturbações aditivas nas respostas, substituindo y_i por $y_i + \sigma\omega_i$, onde σ é o desvio padrão de y_i ;
3. Uma generalização do item anterior, seria perturbar simultaneamente todas as componentes do i -ésimo caso, substituindo o vetor (\mathbf{x}_i, y_i) por $(x_{i1} + s_1\omega_1, \dots, x_{ip} + s_p\omega_p, y_i + \sigma\omega_{p+1})$, onde s_j é um fator de escala, que pode ser a norma da j -ésima coluna da matriz \mathbf{X} , para $j = 1, \dots, p$.

Existem inúmeras formas alternativas de se perturbar o modelo proposto, cuja escolha deve levar em consideração quais os aspectos da análise que se deseja monitorar. Obviamente, deve-se pensar em esquemas de perturbação que sejam interpretáveis.

De uma maneira mais formal, seja $l(\boldsymbol{\beta})$ a função de log-verossimilhança no modelo postulado. Vamos considerar um vetor de perturbação $\boldsymbol{\omega} \in \Omega \subseteq \mathbf{R}^q$, Ω um aberto, e seja $l(\boldsymbol{\beta}, \boldsymbol{\omega})$ a função de log-verossimilhança no modelo perturbado. Assumiremos

que $l(\boldsymbol{\beta}, \boldsymbol{\omega})$ é duas vezes continuamente diferenciável em $(\boldsymbol{\beta}^T, \boldsymbol{\omega}^T)^T$, e que o modelo postulado está encaixado no modelo perturbado. Isto é, estaremos supondo que existe $\boldsymbol{\omega}_0 \in \Omega$ tal que $l(\boldsymbol{\beta}, \boldsymbol{\omega}_0) = l(\boldsymbol{\beta}), \forall \boldsymbol{\beta} \in \mathbf{R}^p$.

Em geral, a dimensão q (do vetor de perturbação) está relacionada com a dimensão do vetor $\boldsymbol{\beta}$ ou com o tamanho da amostra. Para os exemplos citados anteriormente, tem-se $q = n$ nas duas primeiras situações. No primeiro caso, $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$ e $\boldsymbol{\omega}_0 = (0, \dots, 0)^T$, no segundo caso. Para o terceiro exemplo, tem-se $q = p + 1$ e $\boldsymbol{\omega}_0 = (0, \dots, 0)^T \in \mathbf{R}^{p+1}$.

Ainda com relação ao primeiro esquema de perturbação anterior, se considerarmos o vetor $\boldsymbol{\omega} = (1, \dots, 1, \omega_i, 1, \dots, 1)^T$ e fizermos o limite $\omega_i \rightarrow 0$, segue-se que $\text{Var}(y_i) \rightarrow \infty$ e este esquema de perturbação corresponde à exclusão da i -ésima observação. Isto significa que a estratégia da exclusão de casos pode ser vista como um tipo de perturbação que é um caso particular do que estamos tratando agora.

No esquema de exclusão de casos, uma medida bastante geral é motivada pela estatística do teste da razão de máxima verossimilhança e definida como $LD_i = 2 \{l(\hat{\boldsymbol{\beta}}) - l(\hat{\boldsymbol{\beta}}_{(-i)})\}$, que denominaremos afastamento pela verossimilhança. Essa medida foi apresentada por Cook e Weisberg (1982) como *likelihood displacement* e pode também ser interpretada em termos da região de confiança assintótica $\{\boldsymbol{\beta} \in \mathbf{R}^p : 2 \{l(\hat{\boldsymbol{\beta}}) - l(\boldsymbol{\beta})\} \leq \chi_{(\alpha, p)}^2\}$, onde $\chi_{(\alpha, p)}^2$ é o percentil superior de tamanho α da distribuição qui-quadrado com p graus de liberdade, e p é a dimensão do vetor $\boldsymbol{\beta}$, ver Cox e Hinkley (1974, Cap. 9).

No modelo perturbado, vamos denotar o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ por $\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}$, e o nosso objetivo agora é comparar $\hat{\boldsymbol{\beta}}$ com $\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}$ quando $\boldsymbol{\omega}$ varia em Ω . Se a distância entre eles permanecer "pequena" quando $\boldsymbol{\omega}$ varia em Ω , isto é uma indicação de que existe uma estabilidade do modelo ajustado no que diz respeito ao particular

esquema de perturbação que foi utilizado, e portanto ao correspondente aspecto da análise que está sendo monitorado.

Uma comparação direta entre $\hat{\beta}$ e $\hat{\beta}_\omega$ pode não ser simples, devido a diversos fatores tais como diferença de escala, unidade de medida, erro de estimativa, etc. Cook (1986) sugere que a comparação entre $\hat{\beta}$ e $\hat{\beta}_\omega$ seja feita através da função

$$LD(\omega) = 2\{l(\hat{\beta}) - l(\hat{\beta}_\omega)\}, \quad \omega \in \Omega, \quad (2.8)$$

que é uma generalização do afastamento pela verossimilhança LD_i . Desta forma, o sentido de distância entre $\hat{\beta}$ e $\hat{\beta}_\omega$ passa a depender da concavidade da função de log-verossimilhança, Cook (1987). Se a função $l(\beta)$ tiver um grande achatamento, então poderemos julgar que $\hat{\beta}$ e $\hat{\beta}_\omega$ estão próximos entre si. Por outro lado, se $l(\beta)$ for suficientemente concentrada em torno de $\hat{\beta}$, então a estimativa de máxima verossimilhança no modelo perturbado $\hat{\beta}_\omega$ deverá ficar distante de $\hat{\beta}$.

Como $\hat{\beta}$ é o estimador de máxima verossimilhança para β no modelo sem perturbação, segue-se que $LD(\omega) \geq 0, \forall \omega \in \Omega$. Além disso, como $LD(\omega_0) = 0$, podemos concluir que ω_0 é um ponto de mínimo local da função afastamento pela verossimilhança. Uma análise sobre o comportamento geométrico da função $LD(\omega)$, quando ω varia em Ω , pode fornecer informações bastante relevantes a respeito de características importantes do modelo sob investigação. Isto será tratado na Seção 2.4.

Quando o parâmetro σ^2 for desconhecido, então o vetor paramétrico no modelo normal linear será dado por $\theta = (\beta^T, \sigma^2)^T$, que tem $p+1$ coordenadas. Nesse caso, se estivermos interessados apenas na parte correspondente ao β , Cook (1986 e 1987) mostra que o afastamento pela verossimilhança pode ser expresso na forma

$$LD_\beta(\omega) = 2\{l(\hat{\beta}, \hat{\sigma}^2) - l\{\hat{\beta}_\omega, g(\hat{\beta}_\omega)\}\}, \quad (2.9)$$

onde $g: \mathbb{R}^p \rightarrow \mathbb{R}$ é uma função que maximiza $l(\beta, \sigma^2)$, para cada β fixo.

2.4 Influência local

A idéia do estudo da influência local foi introduzida por Cook (1986) e consiste em investigar o comportamento da função $LD(\boldsymbol{\omega})$ numa vizinhança de $\boldsymbol{\omega}_0$, que é o ponto onde as duas log-verossimilhanças são iguais. Para tanto, considera-se a superfície geométrica $(q + 1)$ -dimensional formada pelos valores do vetor

$$\boldsymbol{\alpha}(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega}) \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_q \\ LD(\boldsymbol{\omega}) \end{pmatrix},$$

quando $\boldsymbol{\omega}$ varia em Ω . Essa superfície é denominada de *gráfico de influência*, Cook e Weisberg (1982), e o estudo de influência local consiste em analisar como a superfície $\boldsymbol{\alpha}(\boldsymbol{\omega})$ desvia-se de seu plano tangente em $\boldsymbol{\omega}_0$. Essa análise pode ser feita estudando-se as curvaturas das secções normais da superfície $\boldsymbol{\alpha}(\boldsymbol{\omega})$ em $\boldsymbol{\omega}_0$ – que são as intersecções de $\boldsymbol{\alpha}(\boldsymbol{\omega})$ com planos contendo o vetor normal ao seu plano tangente, em $\boldsymbol{\omega}_0$. Na literatura especializada em Geometria Diferencial, as curvaturas dessas secções normais são denominadas *curvaturas normais*.

Consideremos uma direção arbitrária em \mathbf{R}^q , representada por um vetor unitário \mathbf{d} , e uma linha reta em Ω , passando por $\boldsymbol{\omega}_0$, na mesma direção de \mathbf{d} . Essa reta é dada por

$$\boldsymbol{\omega}(a) = \boldsymbol{\omega}_0 + a \mathbf{d}, \text{ com } a \in \mathbf{R}.$$

Computando-se $LD\{\boldsymbol{\omega}(a)\}$ com $a \in \mathbf{R}$, podemos obter uma linha projetada (tradução livre de *lifted line*), sobre a superfície $\boldsymbol{\alpha}(\boldsymbol{\omega})$. Essa linha projetada consiste da curva plana dada por $\boldsymbol{\rho}(a) = (\rho_1, \rho_2)^T = [a, LD\{\boldsymbol{\omega}(a)\}]^T$ e pode ser vista como o conjunto dos pontos de $\boldsymbol{\alpha}(\boldsymbol{\omega})$ que são comuns ao plano gerado pelos vetores \mathbf{b}_{q+1} e $(\mathbf{d}^T, 0)^T$, onde \mathbf{b}_{q+1} é um vetor que tem o valor 1 na coordenada $q + 1$ e zero nas demais coordenadas.

Na Figura 2.2 apresentamos um exemplo de um gráfico de influência onde o espaço de

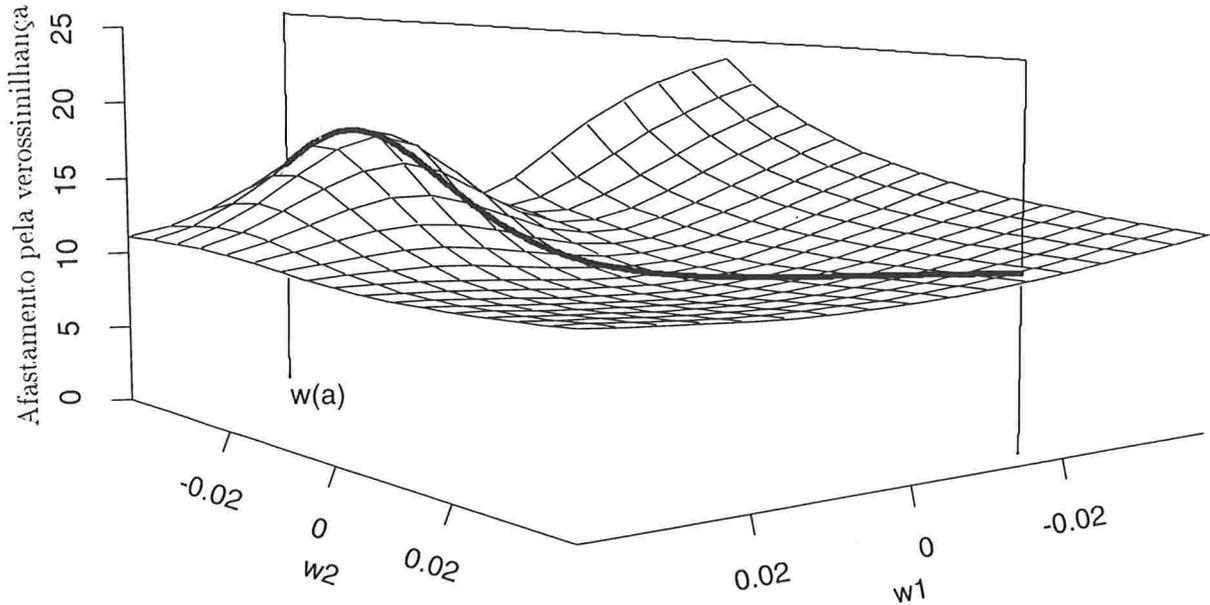


Figura 2.2: Gráfico de influência tridimensional.

perturbações é $\Omega = (-0,04, 0,04) \times (-0,04, 0,04) \subseteq \mathbf{R}^2$. Nessa figura podemos ver um plano gerado pela direção $\mathbf{d} = (-0,44, 0,90)^T$, cuja intersecção com a superfície fornece a curva plana em destaque, que é a linha projetada. Nesse exemplo a variável resposta é a percentagem de uma droga retida no fígado de ratos, ω_1 e ω_2 representam perturbações aditivas na dose administrada da droga, em dois ratos apenas. Esse exemplo está discutido em Cook (1987).

A curvatura normal da curva plana $\rho(a) = (\rho_1, \rho_2)^T$, na direção \mathbf{d} , é dada por

$$C_{\mathbf{d}} = \frac{|\dot{\rho}_1 \ddot{\rho}_2 - \dot{\rho}_2 \ddot{\rho}_1|}{(\dot{\rho}_1^2 + \dot{\rho}_2^2)^{3/2}}, \quad (2.10)$$

avaliada em ω_0 , Stoker (1969, p. 26).

Para avaliarmos essa expressão vamos lembrar da suposição de que a função de log-verossimilhança perturbada $l(\boldsymbol{\beta}, \boldsymbol{\omega})$ é duas vezes continuamente diferenciável em $(\boldsymbol{\beta}^T, \boldsymbol{\omega}^T)^T$. e que, para $\boldsymbol{\rho}(a) = [a, LD\{\boldsymbol{\omega}(a)\}]^T$, temos o seguinte:

$$\text{i. } \dot{\rho}_1 = \frac{da}{da} = 1, \text{ logo } \ddot{\rho}_1 = \frac{d^2a}{da^2} = 0;$$

$$\text{ii. } \frac{d}{da}\boldsymbol{\omega}(a) = \frac{d}{da}(\boldsymbol{\omega}_0 + a \mathbf{d}) = \mathbf{d} \text{ e portanto } \frac{d^2}{da^2}\boldsymbol{\omega}(a) = 0;$$

iii. Pela regra da cadeia,

$$\begin{aligned} \frac{d}{da}LD\{\boldsymbol{\omega}(a)\} &= [\text{gradiente}(LD)]^T \left[\frac{d}{da}\boldsymbol{\omega}(a) \right] \\ &= \left[\frac{\partial}{\partial \boldsymbol{\omega}} LD(\boldsymbol{\omega}) \right]^T \mathbf{d}; \end{aligned}$$

$$\text{iv. } \dot{\rho}_2(0) = \frac{d}{da}LD\{\boldsymbol{\omega}(a)\} \Big|_{a=0} = \left(\left[\frac{\partial}{\partial \boldsymbol{\omega}} LD(\boldsymbol{\omega}) \right]^T \mathbf{d} \right) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = 0, \text{ uma vez que } \boldsymbol{\omega}_0 \text{ é ponto de mínimo local da função } LD.$$

Substituindo essas derivadas na equação (2.10) tem-se que

$$\begin{aligned} C_{\mathbf{d}} &= |\ddot{\rho}_2| \\ &= |\dot{LD}\{\boldsymbol{\omega}(a)\}|, \text{ avaliada em } a = 0. \end{aligned}$$

Usando novamente a regra da cadeia, segue-se que

$$\begin{aligned} \dot{LD}\{\boldsymbol{\omega}(a)\} &= \frac{d^2}{da^2}LD\{\boldsymbol{\omega}(a)\} \\ &= \frac{d}{da} \left[\frac{\partial}{\partial \boldsymbol{\omega}} LD\{\boldsymbol{\omega}(a)\} \right]^T \mathbf{d} \\ &= \frac{d}{da} \frac{\partial}{\partial \boldsymbol{\omega}^T} LD\{\boldsymbol{\omega}(a)\} \mathbf{d} \\ &= \frac{d}{da} \left(\frac{\partial}{\partial \omega_1} LD\{\boldsymbol{\omega}(a)\}, \dots, \frac{\partial}{\partial \omega_q} LD\{\boldsymbol{\omega}(a)\} \right) \mathbf{d} \end{aligned}$$

$$\begin{aligned}
&= \left(\left[\frac{\partial^2}{\partial \omega_1 \partial \omega} LD\{\omega(a)\} \right]^T \frac{d}{da} \omega(a), \dots, \left[\frac{\partial^2}{\partial \omega_q \partial \omega} LD\{\omega(a)\} \right]^T \frac{d}{da} \omega(a) \right) \mathbf{d} \\
&= \left[\begin{pmatrix} \left[\frac{\partial^2}{\partial \omega_1 \partial \omega} LD\{\omega(a)\} \right]^T \\ \vdots \\ \left[\frac{\partial^2}{\partial \omega_q \partial \omega} LD\{\omega(a)\} \right]^T \end{pmatrix} \mathbf{d} \right]^T \mathbf{d} \\
&= \left[\begin{pmatrix} \frac{\partial^2}{\partial \omega_1 \partial \omega^T} LD\{\omega(a)\} \\ \vdots \\ \frac{\partial^2}{\partial \omega_q \partial \omega^T} LD\{\omega(a)\} \end{pmatrix} \mathbf{d} \right]^T \mathbf{d} \\
&= \left[\left(\frac{\partial^2}{\partial \omega \partial \omega^T} LD\{\omega(a)\} \right) \mathbf{d} \right]^T \mathbf{d} \\
&= \mathbf{d}^T \left[\frac{\partial^2}{\partial \omega \partial \omega^T} LD\{\omega(a)\} \right]^T \mathbf{d} \\
&= \mathbf{d}^T \frac{\partial^2}{\partial \omega \partial \omega^T} LD\{\omega(a)\} \mathbf{d}, \text{ pois a matriz Hessiana é simétrica.}
\end{aligned}$$

Quando essa derivada de segunda ordem é avaliada em $a = 0$, obtém-se a curvatura normal

$$C_{\mathbf{d}} = \left| \mathbf{d}^T \dot{L}\dot{D}(\omega_0) \mathbf{d} \right|. \quad (2.11)$$

Mas, pela definição do afastamento pela verossimilhança,

$$\dot{L}\dot{D}(\omega_0) = \frac{\partial^2}{\partial \omega \partial \omega^T} \left[2\{l(\hat{\beta}) - l(\hat{\beta}_\omega)\} \right] \Big|_{\omega=\omega_0} = -2 \frac{\partial^2}{\partial \omega \partial \omega^T} l(\hat{\beta}_\omega) \Big|_{\omega=\omega_0},$$

e portanto, a curvatura normal fica dada por

$$C_{\mathbf{d}} = 2|\mathbf{d}^T \ddot{\mathbf{F}} \mathbf{d}|, \quad (2.12)$$

onde $\ddot{\mathbf{F}}$ é uma matriz quadrada de dimensão $q \times q$, cujos elementos são dados pelas derivadas $\frac{\partial^2 l(\hat{\beta}_\omega)}{\partial \omega_r \partial \omega_s}$, avaliadas em $\omega = \omega_0$.

O elemento (r, s) da matriz \ddot{F} pode ser obtido através da regra da cadeia,

$$\begin{aligned}
\ddot{F}_{rs} &= \frac{\partial^2}{\partial \omega_r \partial \omega_s} l(\hat{\beta}_\omega) \Big|_{\omega=\omega_0} \\
&= \frac{\partial}{\partial \omega_r} \left\{ \frac{\partial}{\partial \omega_s} l(\hat{\beta}_\omega) \right\} \Big|_{\omega=\omega_0} \\
&= \frac{\partial}{\partial \omega_r} \left[\left\{ \frac{\partial}{\partial \hat{\beta}_\omega} l(\hat{\beta}_\omega) \right\}^T \frac{\partial \hat{\beta}_\omega}{\partial \omega_s} \right] \Big|_{\omega=\omega_0} \\
&= \left[\frac{\partial}{\partial \omega_r} \left\{ \frac{\partial}{\partial \hat{\beta}_\omega} l(\hat{\beta}_\omega) \right\} \frac{\partial \hat{\beta}_\omega}{\partial \omega_s} \right] \Big|_{\omega=\omega_0} + \left[\left\{ \frac{\partial}{\partial \hat{\beta}_\omega} l(\hat{\beta}_\omega) \right\} \frac{\partial}{\partial \omega_r} \left(\frac{\partial \hat{\beta}_\omega}{\partial \omega_s} \right) \right] \Big|_{\omega=\omega_0} \\
&= \left(\left[\left\{ \frac{\partial^2}{\partial \hat{\beta}_\omega \partial \hat{\beta}_\omega} l(\hat{\beta}_\omega) \right\}^T \frac{\partial \hat{\beta}_\omega}{\partial \omega_r} \right]^T \frac{\partial \hat{\beta}_\omega}{\partial \omega_s} \right) \Big|_{\omega=\omega_0} + \left\{ \frac{\partial}{\partial \hat{\beta}_\omega} l(\hat{\beta}_\omega) \frac{\partial^2 \hat{\beta}_\omega}{\partial \omega_r \partial \omega_s} \right\} \Big|_{\omega=\omega_0} \\
&= \left[\left(\frac{\partial \hat{\beta}_\omega}{\partial \omega_r} \right)^T \left\{ \frac{\partial^2}{\partial \hat{\beta}_\omega \partial \hat{\beta}_\omega} l(\hat{\beta}_\omega) \right\} \frac{\partial \hat{\beta}_\omega}{\partial \omega_s} \right] \Big|_{\omega=\omega_0} + \left\{ \frac{\partial}{\partial \hat{\beta}_\omega} l(\hat{\beta}_\omega) \frac{\partial^2 \hat{\beta}_\omega}{\partial \omega_r \partial \omega_s} \right\} \Big|_{\omega=\omega_0} \quad (2.13)
\end{aligned}$$

Como, por hipótese, $l(\beta, \omega_0) = l(\beta)$, $\forall \beta \in \mathbf{R}^p$, tem-se $\hat{\beta}_{\omega_0} = \hat{\beta}$, e portanto,

$$\frac{\partial}{\partial \hat{\beta}_\omega} l(\hat{\beta}_\omega) \Big|_{\omega=\omega_0} = \frac{\partial}{\partial \beta} l(\beta, \omega_0) \Big|_{\beta=\hat{\beta}} = 0,$$

já que $\hat{\beta}$ é o ponto de máximo da função de log-verossimilhança, tanto no modelo postulado $l(\beta)$, quanto no modelo perturbado $l(\beta, \omega_0)$. Com isso, a segunda parcela da derivada \ddot{F}_{rs} , equação (2.13), é nula. Portanto $\ddot{F}_{rs} = \left(\frac{\partial \hat{\beta}_\omega}{\partial \omega_r} \right)^T \left(\frac{\partial^2 l(\hat{\beta}_\omega)}{\partial \hat{\beta}_\omega \partial \hat{\beta}_\omega} \right) \left(\frac{\partial \hat{\beta}_\omega}{\partial \omega_s} \right)$, avaliada em $\omega = \omega_0$.

Em notação matricial, podemos escrever

$$\ddot{F} = \mathbf{J}^T \ddot{L} \mathbf{J}, \quad (2.14)$$

onde \mathbf{J} é uma matriz $p \times q$ cujos elementos são dados por $J_{ks} = \frac{\partial \hat{\beta}_{k\omega}}{\partial \omega_s} \Big|_{\omega=\omega_0}$, sendo $\hat{\beta}_{k\omega}$ a k -ésima componente do vetor $\hat{\beta}_\omega$, enquanto que $(-\ddot{L})$ é a matriz de informação observada de Fisher.

A fim de obtermos uma expressão para a matriz \mathbf{J} , vamos usar o fato de que $\hat{\beta}_\omega$ é um ponto de máximo local da log-verossimilhança perturbada $l(\beta, \omega)$. Assim, para cada $\omega \in \Omega$,

$$\left. \frac{\partial l(\beta, \omega)}{\partial \beta_i} \right|_{\beta = \hat{\beta}_\omega} = 0, \quad i = 1, \dots, p. \quad (2.15)$$

O primeiro membro dessa expressão é uma função das $p+q$ variáveis $\hat{\beta}_{1\omega}, \dots, \hat{\beta}_{p\omega}$ e $\omega_1, \dots, \omega_q$. Já o segundo membro é uma função constante, a função identicamente nula. Vamos agora derivar os dois membros da equação (2.15), com relação a ω , e em seguida avaliar a derivada no ponto ω_0 . A derivada do primeiro membro é dada por

$$\begin{aligned} \frac{\partial}{\partial \omega_s} \left(\left. \frac{\partial l(\hat{\beta}_{\omega_0}, \omega_0)}{\partial \beta_i} \right) \right) &= \sum_{k=1}^p \left(\left. \frac{\partial^2 l(\hat{\beta}_\omega, \omega)}{\partial \hat{\beta}_{k\omega} \partial \beta_i} \frac{\partial \hat{\beta}_{k\omega}}{\partial \omega_s} \right) \right) \Big|_{\omega = \omega_0} + \sum_{r=1}^q \left(\left. \frac{\partial^2 l(\hat{\beta}_\omega, \omega)}{\partial \omega_r \partial \beta_i} \frac{\partial \omega_r}{\partial \omega_s} \right) \right) \Big|_{\omega = \omega_0} \\ &= \sum_{k=1}^p \left(\left. \frac{\partial^2 l(\hat{\beta}_\omega, \omega)}{\partial \hat{\beta}_{k\omega} \partial \beta_i} \frac{\partial \hat{\beta}_{k\omega}}{\partial \omega_s} \right) \right) \Big|_{\omega = \omega_0} + \left(\left. \frac{\partial^2 l(\hat{\beta}_\omega, \omega)}{\partial \omega_s \partial \beta_i} \right) \right) \Big|_{\omega = \omega_0} \\ &= \sum_{k=1}^p \left(\left. \frac{\partial^2 l(\hat{\beta}_\omega, \omega)}{\partial \hat{\beta}_{k\omega} \partial \beta_i} \frac{\partial \hat{\beta}_{k\omega}}{\partial \omega_s} \right) \right) \Big|_{\omega = \omega_0} + \frac{\partial^2 l(\hat{\beta}_{\omega_0}, \omega_0)}{\partial \omega_s \partial \beta_i}. \end{aligned}$$

Quanto ao segundo membro, sua derivada é nula, por se tratar de uma função constante.

Assim, para $i = 1, \dots, p$, temos a seguinte relação

$$\sum_{k=1}^p \left(\left. \frac{\partial^2 l(\hat{\beta}_\omega, \omega)}{\partial \hat{\beta}_{k\omega} \partial \beta_i} \frac{\partial \hat{\beta}_{k\omega}}{\partial \omega_s} \right) \right) \Big|_{\omega = \omega_0} + \frac{\partial^2 l(\hat{\beta}_{\omega_0}, \omega_0)}{\partial \omega_s \partial \beta_i} = 0. \quad (2.16)$$

Como as duas funções de log-verossimilhança são iguais em ω_0 , para todo β ,

$$\left. \frac{\partial^2 l(\hat{\beta}_\omega, \omega)}{\partial \hat{\beta}_{k\omega} \partial \beta_i} \right) \Big|_{\omega = \omega_0} = \frac{\partial^2 l(\hat{\beta}_{\omega_0}, \omega_0)}{\partial \hat{\beta}_{k\omega_0} \partial \beta_i} = \frac{\partial^2 l(\hat{\beta})}{\partial \beta_k \partial \beta_i} = \left. \frac{\partial^2 l(\beta)}{\partial \beta_k \partial \beta_i} \right) \Big|_{\beta = \hat{\beta}}$$

Com isto, da equação (2.16) segue-se que, para $i = 1, \dots, p$, e $s = 1, \dots, q$,

$$\sum_{k=1}^p \left\{ \left(\left. \frac{\partial^2 l(\beta)}{\partial \beta_k \partial \beta_i} \right) \right) \Big|_{\beta = \hat{\beta}} \left(\left. \frac{\partial \hat{\beta}_{k\omega}}{\partial \omega_s} \right) \right) \Big|_{\omega = \omega_0} \right\} + \frac{\partial^2 l(\hat{\beta}_{\omega_0}, \omega_0)}{\partial \omega_s \partial \beta_i} = 0. \quad (2.17)$$

Por fim, vamos definir uma matriz Δ , de dimensão $p \times q$, dada pela derivada $\Delta = \frac{\partial^2 l(\beta, \omega)}{\partial \beta \partial \omega^T}$, avaliada em $\beta = \hat{\beta}$ e $\omega = \omega_0$.

Assim, a equação (2.17) pode ser escrita na forma $\sum_{k=1}^p \ddot{L}_{ki} J_{ks} + \Delta_{is} = 0$, ou ainda, $\sum_{k=1}^p \ddot{L}_{ik} J_{ks} + \Delta_{is} = 0$, usando a simetria da matriz \ddot{L} . Portanto, $(\ddot{L}J)_{is} = -\Delta_{is}$, para $i = 1, \dots, p$ e $s = 1, \dots, q$. Com isto segue-se que

$$\mathbf{J} = -\ddot{L}^{-1} \Delta. \quad (2.18)$$

Substituindo \mathbf{J} na equação (2.14) obtém-se

$$\ddot{\mathbf{F}} = \mathbf{J}^T \ddot{L} \mathbf{J} = \left(-\Delta^T \ddot{L}^{-1} \right) \ddot{L} \left(-\ddot{L}^{-1} \Delta \right) \implies$$

$$\ddot{\mathbf{F}} = \Delta^T \ddot{L}^{-1} \Delta.$$

Com isto, a curvatura normal da superfície $\alpha(\omega)$, na direção de um vetor unitário \mathbf{d} , fica dada por

$$C_{\mathbf{d}} = 2 \left| \mathbf{d}^T \Delta^T \ddot{L}^{-1} \Delta \mathbf{d} \right|. \quad (2.19)$$

Com a equação (2.19) podemos avaliar a influência que pequenas perturbações podem exercer sobre os componentes do modelo – tais como estimativas dos parâmetros, e outros resultados da análise. Interesse particular está na direção que produz a maior influência local. Essa direção \mathbf{d}_{\max} é o autovetor normalizado correspondente ao maior autovalor C_{\max} da matriz $\ddot{\mathbf{F}} = \Delta^T \ddot{L}^{-1} \Delta$. Com o vetor \mathbf{d}_{\max} identificamos os fatores mais influentes para o esquema de perturbação em análise.

2.4.1 Outras expressões para $C_{\mathbf{d}}$

Vamos supor agora uma partição $\beta^T = (\beta_1^T, \beta_2^T)^T$, onde β_1 é de dimensão $p_1 \times 1$ e β_2 é de dimensão $p_2 \times 1$, com $p_1 + p_2 = p$, e vamos admitir que o nosso interesse reside

particularmente em β_1^T . Nesse caso, o gráfico de influência fica dado por

$$\alpha_s(\omega) = \begin{pmatrix} \omega \\ LD_s(\omega) \end{pmatrix},$$

onde

$$LD_s(\omega) = 2[l(\hat{\beta}_1, \hat{\beta}_2) - l\{\hat{\beta}_{1\omega}, g(\hat{\beta}_{1\omega})\}]. \quad (2.20)$$

Aqui, $g : \mathbf{R}^{p_1} \rightarrow \mathbf{R}^{p_2}$ é uma função que maximiza $l(\beta_1, \beta_2)$, para cada β_1 fixo, e $l\{\beta_1, g(\beta_1)\}$ representa a log-verossimilhança perfilada para β_1 . O estimador $\hat{\beta}_{1\omega}$ é determinado a partir de $\hat{\beta}_\omega^T = (\hat{\beta}_{1\omega}^T, \hat{\beta}_{2\omega}^T)^T$.

Vamos considerar agora a partição

$$\ddot{\mathbf{L}} = \begin{pmatrix} \ddot{\mathbf{L}}_{11} & \ddot{\mathbf{L}}_{12} \\ \ddot{\mathbf{L}}_{21} & \ddot{\mathbf{L}}_{22} \end{pmatrix},$$

onde $(-\ddot{\mathbf{L}})$ é a matriz de informação observada de Fisher, no modelo modelo postulado.

Com isso, a curvatura normal da superfície $\alpha_s(\omega)$ na direção de um vetor unitário \mathbf{d} fica dada pela expressão

$$C_{\mathbf{d}}(\beta_1) = 2 \left| \mathbf{d}^T \Delta^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_{22}) \Delta \mathbf{d} \right|, \quad (2.21)$$

onde

$$\mathbf{B}_{22} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddot{\mathbf{L}}_{22}^{-1} \end{pmatrix}.$$

Atenção especial deve ser dada à direção \mathbf{d}_{\max} , que produz a maior influência local, a qual pode ser calculada a partir da equação (2.21), e que é dada pelo autovetor normalizado correspondente ao maior autovalor da matriz $\Delta^T (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_{22}) \Delta$.

Diversas aplicações e resultados adicionais sobre o método de influência local podem ser encontrados em Lawrence (1988), Thomas e Cook (1989, 1990), Paula (1993, 1995, 1996), Galea, Paula e Bolfarine (1997), entre outros.

Thomas e Cook (1990) empregam esse método para fazer diagnóstico de influência sobre predições nos modelos lineares generalizados, utilizando uma função objetivo baseada em diferença de predições. Paula (1996) faz uma extensão desse estudo para os modelos próprios de dispersão, propostos por Jørgensen (1997).

Fung e Kwan (1997) apresentam uma discussão interessante sobre a aplicação de influência local para outras medidas de influência diferentes do afastamento pela verossimilhança. Fung e Kwan mostram que uma medida de influência, digamos \hat{T}_ω , é invariante de escala se $\dot{\Gamma} = \left. \frac{\partial \hat{T}_\omega}{\partial \omega} \right|_{\omega=\omega_0} = \mathbf{0}$. Quando esta derivada é diferente de zero a ordem entre os componentes de \mathbf{d}_{\max} não é necessariamente preservada sob mudanças de escala. Em particular, para o afastamento pela verossimilhança, tem-se que $\dot{\Gamma} = \left. \frac{\partial LD(\omega)}{\partial \omega} \right|_{\omega=\omega_0} = \mathbf{0}$. Esta propriedade segue também, por exemplo, para as medidas de influência propostas por Thomas e Cook (1990) e Paula (1993). Porém ela não vale para outras medidas de influência discutidas por Fung e Kwan (1997).

Cadigan e Farrel (1999) sugerem o uso do valor esperado da influência local como uma ferramenta útil na interpretação da análise de diagnóstico.

Capítulo 3

Introdução a Dados Circulares

3.1 Preliminares

De uma forma geral, quando se deseja estudar dados que representam *direções*, eles podem ser tratados em um espaço de qualquer dimensão, entretanto nas situações práticas geralmente o interesse está nos espaços de dimensão 2 ou 3 (bi- ou tri-dimensional), onde as direções podem ser representadas geometricamente por pontos sobre a circunferência de um círculo, ou sobre a superfície de uma esfera, respectivamente. Num espaço de dimensão maior, as direções são representadas por pontos sobre a superfície de uma hiperesfera.

Existe interesse por medidas na forma de ângulos em diversas áreas do conhecimento. De modo particular, encontramos exemplos de dados direcionais em Geologia, Biologia, Meteorologia, Oceanografia, Medicina e em Engenharia. Alguns exemplos típicos de dados direcionais nessas áreas são:

1. Direção do vento e de correntes marinhas;
2. Horário de internação na UTI de um determinado hospital;
3. Orientação da fratura em uma rocha, quando da sua formação geológica;
4. O vectorcardiograma (um exame cujo resultado consiste de um registro gráfico tri-dimensional) análogo ao eletrocardiograma (unidimensional);
5. Direção escolhida por um pássaro, após o mesmo ter sido submetido a um determinado tratamento.

Dados como esses devem ser tratados com técnicas diferentes daquelas utilizadas para variáveis definidas em um espaço euclidiano usual, isto é, dados que assumem valores em \mathbf{R}^n . Isto porque as superfícies esféricas de um modo geral, e as circunferências em particular, são espaços limitados e fechados, e além do mais, o conceito de uma origem não é bem definido, Gould (1969). As medidas resumo que são usadas para variáveis lineares, não fazem sentido quando se trata de uma variável direcional, conforme mostraremos mais adiante.

3.2 Dados circulares

No caso bi-dimensional, uma observação de uma variável direcional pode ser vista como um ponto P sobre a circunferência do círculo unitário, centrado na origem O do sistema de coordenadas cartesianas e, portanto, tem-se que \overrightarrow{OP} é um vetor unitário em \mathbf{R}^2 . Essa observação direcional pode ainda ser vista como um ângulo y , formado pelo vetor

\overrightarrow{OP} e o semi-eixo positivo Ox , medido no sentido anti-horário. Assim, uma observação de uma variável circular pode ser representada pelo par ordenado $(\cos y, \operatorname{sen} y)$, em coordenadas cartesianas, ou pelo par $(1, y)$, em coordenadas polares.

Embora os dados circulares possam ser coletados em diferentes tipos de unidade, como por exemplo hora do dia em que houve a ocorrência do evento de interesse, a conversão desses dados para graus é feita por uma simples transformação linear, obtendo-se valores no intervalo $[0^\circ, 360^\circ)$ ou $[-180^\circ, 180^\circ)$, dependendo do interesse na análise. Em princípio, podemos trabalhar com dados circulares expressos em qualquer medida de ângulo – tais como graus, grado, radiano, etc. Mas, por razões teóricas, adota-se trabalhar em radianos, com intervalo de variação $[0, 2\pi)$ ou $[-\pi, \pi)$, dependendo do interesse na análise.

Em alguns casos, como por exemplo no estudo de fratura em rochas, não há razão para distinguir qual é o início e qual é o fim dessa fratura. Um dado desse tipo é denominado de *axial*, diferente do anterior, denominado de *vetorial* (Fisher, 1993).

Para trabalharmos com distribuições de probabilidade referentes a variáveis aleatórias circulares precisaremos estabelecer algumas medidas descritivas relacionadas com os conceitos de posição, dispersão, etc. Em princípio, é bastante tentador considerar as mesmas medidas usualmente empregadas para dados lineares, tais como média e variância, e estendê-las também para dados circulares. Entretanto, um simples exemplo fictício mostra a inadequação dessas medidas quando se trabalha no círculo.

Exemplo 3.1 *Vamos imaginar uma amostra de tamanho 2 em que os ângulos observados tenham sido $y_1 = 1^\circ$ e $y_2 = 359^\circ$, que são “próximos entre si” e equidistantes da direção 0° . Nesse exemplo, a média aritmética amostral resulta em $\bar{y} = (1^\circ + 359^\circ)/2 = 180^\circ$ que corresponde ao ponto no círculo diametralmente oposto ao que sugere a nossa*

intuição. Por outro lado, o desvio padrão amostral fica dado por $s_{n-1} = 253^\circ$, ou seja, um valor que não descreve o fato de y_1 e y_2 estarem “próximos entre si”.

Uma outra particularidade de dados circulares diz respeito à escolha da orientação do sistema de eixos ortogonais, e pode ser exemplificada da seguinte forma. Nesse exemplo que acabamos de ver, se escolhermos a direção zero como o eixo das ordenadas, em vez do eixo das abcissas, então as observações y_1 e y_2 passarão a ter os valores $y'_1 = 271^\circ$ e $y'_2 = 269^\circ$, respectivamente. Neste caso, a média amostral seria igual a 270° , que é um valor compatível com a nossa intuição, nesse novo sistema de coordenadas. Isso mostra que uma análise descritiva de dados circulares, através das medidas utilizadas para variáveis definidas na reta, é bastante sensível à escolha da origem no sistema de eixos ortogonais. Isso é decorrente do fato do círculo possuir uma métrica bastante peculiar. Por exemplo, a diferença entre dois arcos y e y^* no círculo unitário é dada por $\min\{\theta_1, \theta_2\}$, onde $\theta_1 = |y - y^*| \bmod(2\pi)$ e $\theta_2 = |y + 2\pi - y^*| \bmod(2\pi)$, para $y < y^*$.

3.3 Medidas descritivas no círculo

Consideremos uma amostra aleatória de n observações circulares y_1, \dots, y_n , às quais podemos associar n vetores unitários $\overrightarrow{OP_1}, \dots, \overrightarrow{OP_n}$, conforme foi mencionado anteriormente. Uma maneira natural e intuitiva de resumir esses vetores unitários é através da sua soma, conforme ilustra a Figura 3.1. Do ponto de vista algébrico, calculemos as quantidades

$$S = \sum_{i=1}^n \text{sen}(y_i), \quad C = \sum_{i=1}^n \text{cos}(y_i), \quad R^2 = C^2 + S^2.$$

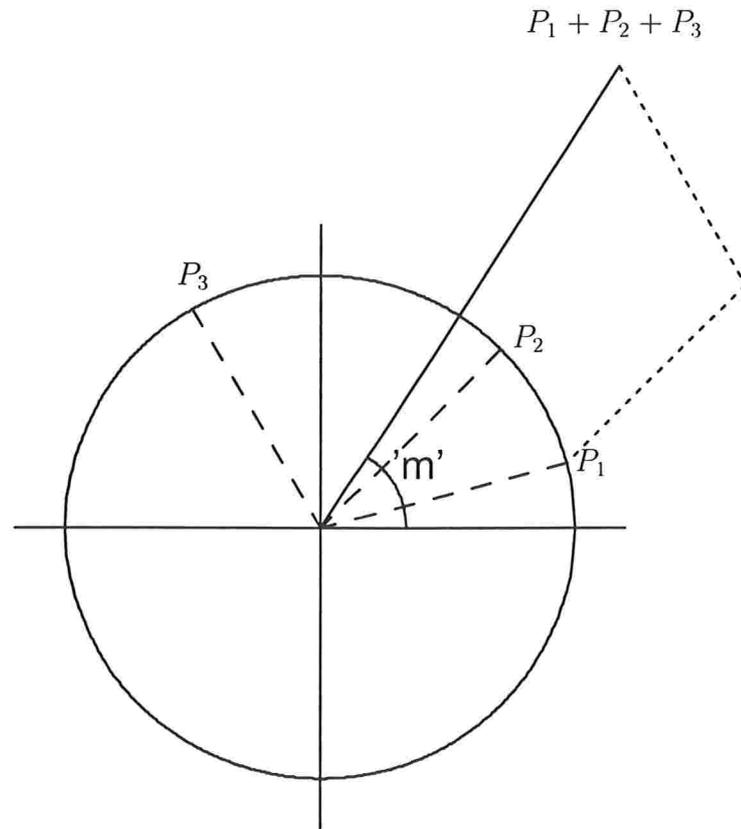


Figura 3.1: Representação gráfica de uma amostra de 3 observações circulares.

A *média circular*¹ de y_1, \dots, y_n é definida como sendo o ângulo $\bar{\mu}$ correspondente ao vetor resultante da soma $\overrightarrow{OP_1} + \dots + \overrightarrow{OP_n}$. Esse ângulo $\bar{\mu}$ deve satisfazer as condições $\cos(\bar{\mu}) = C/R$ e $\sin(\bar{\mu}) = S/R$, ou seja, $\bar{\mu} = \arctan(S/C)$. Assim,

$$\bar{\mu} = \begin{cases} \arctan(S/C), & \text{se } S \geq 0 \text{ e } C > 0 \\ \arctan(S/C) + \pi, & \text{se } C < 0 \\ \arctan(S/C) + 2\pi, & \text{se } S < 0 \text{ e } C > 0. \end{cases} \quad (3.1)$$

¹Também denominada *média direcional*, *direção média* ou *direção de referência* de y_1, \dots, y_n .

Alguns casos excepcionais são

$$\bar{\mu} = \begin{cases} 90^\circ, & \text{se } S > 0 \text{ e } C = 0 \\ 270^\circ, & \text{se } S < 0 \text{ e } C = 0 \\ \text{indeterminado,} & \text{se } S = 0 \text{ e } C = 0. \end{cases}$$

No exemplo fictício da página 29, tem-se $n = 2$ e

$$\left. \begin{array}{l} S = \text{sen}(y_1) + \text{sen}(y_2) = 0 \\ C = \text{cos}(y_1) + \text{cos}(y_2) = 1,996 \end{array} \right\} \Rightarrow \bar{\mu} = \arctan(0) \Rightarrow \bar{\mu} = 0^\circ.$$

Por outro lado,

$$\left. \begin{array}{l} S' = \text{sen}(y'_1) + \text{sen}(y'_2) = -1,996 \\ C' = \text{cos}(y'_1) + \text{cos}(y'_2) = 0 \end{array} \right\} \Rightarrow \bar{\mu}' = 270^\circ.$$

A quantidade $R = \sqrt{C^2 + S^2}$ representa o comprimento do vetor resultante e fornece informação sobre o grau de concentração dos ângulos observados. Isto porque, se todas as n observações circulares forem iguais a um mesmo ângulo, então $R = n$. Esse caso representa uma concentração máxima possível, e corresponde a uma variabilidade mínima. A outra situação extrema é quando os dados encontram-se uniformemente distribuídos sobre o círculo. Nesse caso tem-se uma concentração mínima e variabilidade máxima. Vale salientar que o conceito de variabilidade máxima não existe para dados lineares em geral. No caso de concentração mínima, que significa $R = 0$, a média circular não está definida. É importante salientar também que $R = 0$ não significa necessariamente uma distribuição uniforme em torno do círculo, pois pode acontecer algum tipo de estrutura de agrupamentos, fornecendo $R = 0$. Um exemplo é dado em Fisher (1993, p. 32).

Em vez de R , é mais comum usar o comprimento médio do vetor resultante, definido por $\bar{R} = R/n$, o qual tem a vantagem de variar no intervalo $[0, 1]$. A variância circular

amostral é definida por $V = 1 - \bar{R}$, enquanto que o desvio padrão circular amostral é definido por $\hat{\sigma} = \{-2 \log(1 - V)\}^{\frac{1}{2}}$, diferentemente da definição usual para dados lineares, onde o desvio padrão corresponde à raiz quadrada da variância. Uma aproximação que pode ser usada para o desvio padrão circular amostral é dada por $\hat{\sigma} \simeq \{2(1 - \bar{R})\}^{\frac{1}{2}}$, válida para \bar{R} grande. Maiores detalhes sobre medidas de concentração para dados circulares podem ser encontrados em Mardia (1972, p. 74) e em Fisher (1993, p. 54).

3.4 Momentos trigonométricos

No que se segue, entenderemos por uma variável aleatória circular como sendo uma variável aleatória já convertida em ângulos e assumindo valores no intervalo $[0, 2\pi)$.

Definição 3.1 *Definimos a função de distribuição de uma variável aleatória circular Y como sendo uma função F satisfazendo*

$$(i) \quad F(y) = P(0 < Y \leq y), \quad \forall y \in [0, 2\pi);$$

$$(ii) \quad F(y + 2\pi) - F(y) = 1, \quad \forall y \in (-\infty, \infty).$$

Para $-\infty < y_1 < y_2 < \infty$, com $y_2 - y_1 < 2\pi$, a probabilidade de Y pertencer ao arco de círculo (y_1, y_2) é dada por

$$P\{Y \in \text{arc}(y_1, y_2)\} = F(y_2) - F(y_1) = \int_{y_1}^{y_2} dF(y).$$

Vale salientar que, embora $F(0) = 0$ e $F(2\pi) = 1$, a função de distribuição no círculo **não satisfaz** as condições $\lim_{y \rightarrow -\infty} F(y) = 0$ e $\lim_{y \rightarrow \infty} F(y) = 1$, que são satisfeitas por funções de distribuição de variáveis aleatórias definidas na reta.

Definição 3.2 *Seja F a função de distribuição de uma variável aleatória circular Y . Se F for absolutamente contínua, definimos a função densidade de probabilidade de Y como sendo uma função f tal que*

$$\int_{y_1}^{y_2} f(y)dy = F(y_2) - F(y_1), \quad -\infty < y_1 < y_2 < \infty.$$

Definição 3.3 *O p -ésimo momento trigonométrico não central de Y é definido como $\mu'_p = E\{e^{ipY}\} = \alpha_p + i\beta_p$, $p = 0, 1, 2, \dots$, onde $i = \sqrt{-1}$ é a unidade imaginária. α_p e β_p são denominados momentos circulares de Y e são dados por*

$$\begin{aligned} \alpha_p &= \int_0^{2\pi} \cos(py) dF(y) \text{ e} \\ \beta_p &= \int_0^{2\pi} \text{sen}(py) dF(y). \end{aligned}$$

Definição 3.4 *Definimos a função característica de uma variável aleatória circular Y como sendo $\phi_Y(p) = E\{\cos(pY)\} + iE\{\text{sen}(pY)\}$, $p \in \mathbf{R}$.*

Usando a teoria de séries de Fourier, mostra-se que é suficiente considerar a função $\phi_Y(p)$ apenas para valores inteiros de p . Como consequência das propriedades das funções trigonométricas tem-se

$$\alpha_{-p} = \alpha_p, \quad \beta_{-p} = -\beta_p, \quad |\alpha_p| \leq 1 \quad \text{e} \quad |\beta_p| \leq 1.$$

Desta forma, os momentos trigonométricos não centrais de Y identificam a sua função característica, e escrevemos simplesmente $\phi_Y(p) = \phi_p = \alpha_p + i\beta_p$. Para $p = 1$, tem-se $\mu'_1 = (\rho, \mu)$, em coordenadas polares, onde μ é a média circular de Y e a distância polar ρ corresponde ao comprimento do vetor médio, $\rho = (\alpha_1^2 + \beta_1^2)^{1/2}$. Pode-se ver que o primeiro momento trigonométrico (não central) μ'_1 contém mais informação do que o primeiro momento ordinário de uma variável aleatória linear. O p -ésimo momento trigonométrico central de Y é definido por $\mu_p = E\{e^{ip(Y-\mu)}\}$, $p = 0, 1, 2, \dots$

Definimos ainda a variância circular de Y por $V_0 = 1 - \rho = 1 - E\{\cos(Y - \mu)\}$ e o desvio padrão circular por $\sigma = \{-2 \log(1 - V_0)\}^{\frac{1}{2}}$. Tem-se que $0 \leq V_0 \leq 1$, enquanto que $\sigma > 0$. A variância circular em torno de um ângulo qualquer ν é definida como $V'_0 = 1 - E\{\cos(Y - \nu)\}$, e é possível provar que $V'_0 = V_0 + 2\rho \sin^2\{\frac{1}{2}(\mu - \nu)\}$, de onde se tem que V'_0 assume um mínimo em $\nu = \mu$.

Uma outra medida de dispersão para variáveis circulares também utilizada é a dispersão circular, Fisher (1993, p. 42), que é definida por $\delta = \frac{1 - \rho_2}{2\rho^2}$, onde $\rho_2 = (\alpha_2^2 + \beta_2^2)^{1/2}$.

3.5 Modelos probabilísticos para dados circulares

Muitos modelos probabilísticos para dados circulares têm sido derivados a partir de transformações de modelos probabilísticos lineares usuais, ou mesmo de vetores aleatórios bi-dimensionais, dando origem às distribuições arqueadas – do inglês *wrapped distributions*.

De uma maneira geral, dada uma variável aleatória linear Z , podemos “construir” uma variável aleatória circular Y , assumindo valores no intervalo $[0, 2\pi)$ através da transformação $Y = Z \{\text{mod } (2\pi)\}$, ou seja, tomando o resto da divisão da variável aleatória Z por 2π . Essas são as chamadas **distribuições arqueadas**. Se Z for absolutamente contínua, com função densidade de probabilidade $f_Z(z)$, então a correspondente função densidade de probabilidade de Y será dada por

$$f_Y(y) = \sum_{k=-\infty}^{\infty} f_Z(y + 2k\pi), \quad \text{para } y \in [0, 2\pi).$$

3.5.1 A Distribuição uniforme circular - \mathcal{U}_c

Dizemos que uma variável aleatória Y tem distribuição uniforme sobre círculo unitário se sua função densidade de probabilidade for dada por

$$f_Y(y) = \frac{1}{2\pi}, \quad 0 \leq y < 2\pi.$$

A função característica da distribuição uniforme é dada por

$$\begin{aligned} \phi_p &= \frac{(e^{2ip\pi} - 1)}{2ip\pi} \\ &= \frac{\cos(2p\pi) + i\operatorname{sen}(2p\pi) - 1}{2ip\pi}, \quad p = \pm 1, \pm 2, \dots \\ \implies \phi_p &= \begin{cases} 0, & \text{se } p \neq 0 \\ 1, & \text{se } p = 0. \end{cases} \end{aligned}$$

Vale salientar que a média angular não está definida para este modelo e que $V = 1$. Isto significa que a concentração é nula, em torno de qualquer direção.

3.5.2 A Distribuição normal arqueada - $\mathcal{WN}(\mu, \sigma^2)$

Um caso especial entre as distribuições arqueadas é aquela que se obtém a partir de uma distribuição Gaussiana. Se uma variável aleatória Z tem distribuição normal $N(\mu_Z, \sigma_Z^2)$, então a variável aleatória $Y = Z \{\bmod(2\pi)\}$ tem distribuição normal arqueada com parâmetros μ_Z e σ_Z^2 . Sua função densidade de probabilidade é dada por

$$\begin{aligned} f_Y(y, \mu_Z, \sigma_Z^2) &= \sum_{k=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_Z^2}} \exp\left\{-\frac{(y + 2k\pi - \mu_Z)^2}{2\sigma_Z^2}\right\} \\ &= \left\{1 + 2 \sum_{p=1}^{\infty} \rho^{p^2} \cos(py)\right\} / (2\pi), \quad y \in [0, 2\pi), \end{aligned}$$

onde $\rho = e^{-\frac{1}{2}\sigma^2}$ é o comprimento médio resultante. A média circular de Y fica dada por $\mu = \mu_Z \{\text{mod } (2\pi)\}$, e o desvio padrão circular é $\sigma = \{-2 \log(1 - V_0)\}^{\frac{1}{2}}$, onde $V_0 = 1 - \rho$ é a variância circular de Y , Mardia (1972, p. 55) e Fisher (1993, p. 47).

3.5.3 A Distribuição von Mises - $\mathcal{VM}(\mu, \lambda)$

Outros modelos probabilísticos são obtidos através de analogias no círculo de importantes caracterizações univariadas, dando origem a igualmente importantes distribuições de probabilidade para variáveis aleatórias direcionais. A distribuição von Mises é tão importante para dados circulares, quanto o modelo Gaussiano é para dados lineares em geral. Esse modelo foi proposto pelo físico alemão R. von Mises, em 1918, ao investigar o desvio de pesos atômicos de valores inteiros. Seu propósito era mostrar que os pesos atômicos são números inteiros sujeitos a erros. Transformando a parte fracionária dos pesos atômicos em desvios angulares, de forma que

$$8,25 \mapsto 90^0, \quad 8,75 \mapsto 270^0, \quad 9,00 \mapsto 0^0, \quad 9,25 \mapsto 90^0, \quad \text{etc.}$$

ele passou do problema físico original para um problema estatístico, a partir do qual originou essa distribuição.

Dizemos que uma variável aleatória circular Y tem distribuição von Mises com parâmetros μ e λ , se sua função densidade de probabilidade for dada por

$$f_Y(y; \mu, \lambda) = \frac{1}{2\pi I_0(\lambda)} e^{\lambda \cos(y-\mu)}, \quad 0 \leq y < 2\pi, \quad 0 \leq \mu < 2\pi \quad \text{e} \quad \lambda > 0,$$

onde $I_0(\lambda)$ é a função de Bessel modificada do 1º tipo e ordem zero, que é dada pela série

$$I_0(\lambda) = \sum_{r=0}^{\infty} \frac{1}{r!^2} \left(\frac{1}{2}\lambda\right)^{2r}.$$

As Figuras 3.2 e 3.3 apresentam gráficos dessa função densidade para diversos valores dos parâmetros μ e λ . A função de distribuição de Y é dada por

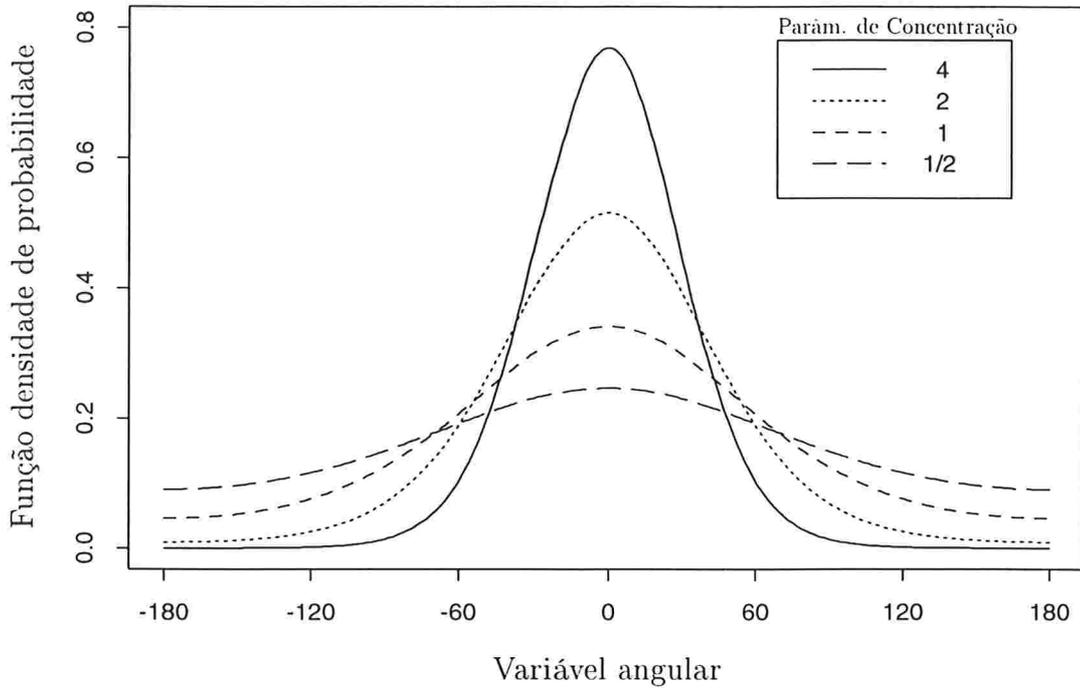


Figura 3.2: Densidade da distribuição von Mises para $\mu = 0^\circ$ e $\lambda = 1/2, 1, 2$ e 4 .

$$F_Y(y) = \{2\pi I_0(\lambda)\}^{-1} \int_0^y e^{\lambda \cos(t-\mu)} dt,$$

para a qual não existe uma forma fechada.

Algumas características da distribuição von Mises envolvem a função $\mathcal{A}_p(\lambda) = \frac{I_p(\lambda)}{I_0(\lambda)}$, onde $I_p(\lambda)$ é a função de Bessel modificada do 1º tipo e ordem p , que é definida pela série

$$I_p(\lambda) = \sum_{r=0}^{\infty} \frac{1}{(p+r)! r!} \left(\frac{1}{2}\lambda\right)^{2r+p}, \quad p = 0, 1, 2, \dots$$

Essas funções e suas propriedades, como também tabulações, podem ser encontradas em Abramowitz e Stegun (1970).

O parâmetro μ representa a média circular de Y enquanto que λ é denominado parâmetro de concentração, e é tal que o comprimento resultante médio $\rho = \mathcal{A}_1(\lambda)$. A dispersão circular de Y é dada por $\delta = [\lambda \mathcal{A}_1(\lambda)]^{-1}$ e as componentes dos seus momentos trigonométricos são $\alpha_p = \mathcal{A}_p(\lambda)$ e $\beta_p = 0$ para $p \geq 1$. Essa distribuição goza de

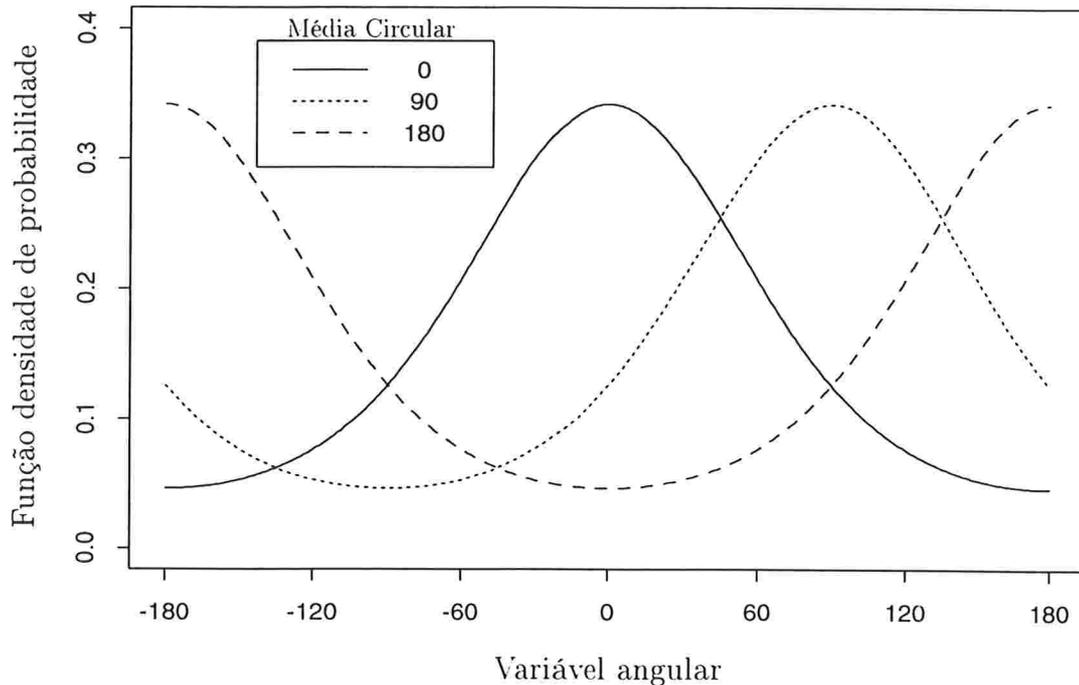


Figura 3.3: Densidade da distribuição von Mises para $\lambda = 1$ e $\mu = 0^\circ, 90^\circ$ e 180° .

propriedades bastante interessantes, algumas das quais são:

1. Ela é unimodal e simétrica em torno da média direcional μ ;
2. Quando $\lambda \rightarrow 0$, a distribuição de von Mises converge para a uniforme circular;
3. Quando $\lambda \rightarrow \infty$, a distribuição converge para a distribuição degenerada no ponto $y = \mu$;

4. Para valores não muito pequenos de λ , a distribuição von Mises $\mathcal{VM}(\mu, \lambda)$ é próxima de uma distribuição normal arqueada $\mathcal{WN}(\mu, 1/\lambda)$, no sentido de que

$$\sup_{y \in [0, 2\pi)} |F_{\mathcal{VM}}(y) - F_{\mathcal{WN}}(y)| \leq 0,0125,$$

onde $F_{\mathcal{VM}}(y)$ e $F_{\mathcal{WN}}(y)$ representam as funções de distribuição acumuladas da distribuição von Mises e da distribuição normal arqueada, respectivamente. A distância máxima de 0,0125 é atingida quando $\lambda = 1, 4$, Stephens (1963).

Simulação da Distribuição von Mises

Best e Fisher (1979) apresentam um algoritmo bastante eficiente para a simulação da distribuição von Mises, o qual está disponível na biblioteca de subrotinas do IMSL (1991). Esse algoritmo será tratado no Capítulo 5.

Dispõe-se também de um outro algoritmo, sugerido por Dagpunar (1990), que pode ser mais rápido do que esse, quando o parâmetro λ não for constante, isto é, varia a cada entrada no processo iterativo, vide Fisher (1993, p. 49).

3.5.4 Estimação dos parâmetros da distribuição von Mises

Consideremos uma amostra aleatória Y_1, \dots, Y_n da distribuição von Mises $\mathcal{VM}(\mu, \lambda)$, com valores observados y_1, \dots, y_n . A função de log-verossimilhança fica dada por

$$l(\mu, \lambda; y_1, \dots, y_n) = -n \log I_0(\lambda) + \lambda \sum_{i=1}^n \cos(y_i - \mu).$$

Uma vez calculada a derivada parcial de $l(., .; .)$ com relação a μ e efetuadas algumas operações algébricas, vamos concluir que o estimador de máxima verossimilhança de μ deve satisfazer as equações $\cos(\hat{\mu}) = C/R$ e $\sin(\hat{\mu}) = S/R$, onde $C = \sum_{i=1}^n \cos(y_i)$, $S = \sum_{i=1}^n \sin(y_i)$ e $R^2 = C^2 + S^2$. Portanto,

$$\hat{\mu} = \begin{cases} \arctan(S/C), & \text{se } S \geq 0 \text{ e } C > 0 \\ \arctan(S/C) + \pi, & \text{se } C < 0 \\ \arctan(S/C) + 2\pi, & \text{se } S < 0 \text{ e } C > 0, \end{cases} \quad (3.2)$$

ou seja, o estimador de máxima verossimilhança da média direcional μ é exatamente a média circular amostral, definida pela equação (3.1). A propósito, Bingham e Mardia (1975) mostram que a distribuição von Mises é a única distribuição circular onde a direção média amostral é o estimador de máxima verossimilhança para a direção média populacional.

Alguns casos excepcionais para esse estimador são

$$\bar{\mu} = \begin{cases} 90^\circ, & \text{se } S > 0 \text{ e } C = 0 \\ 270^\circ, & \text{se } S < 0 \text{ e } C = 0 \\ \text{indeterminado,} & \text{se } S = 0 \text{ e } C = 0. \end{cases}$$

Por sua vez, o estimador de máxima verossimilhança de λ deve satisfazer a equação

$$-n \frac{I_1(\hat{\lambda})}{I_0(\hat{\lambda})} + \cos(\hat{\mu}) C + \sin(\hat{\mu}) S = 0,$$

cuja solução é $\mathcal{A}_1(\hat{\lambda}) = \bar{R}$, onde $\bar{R} = \frac{R}{n}$, e $\mathcal{A}_1(\hat{\lambda}) = \frac{I_1(\hat{\lambda})}{I_0(\hat{\lambda})}$. Assim, o estimador de máxima verossimilhança de λ fica dado por $\hat{\lambda} = \mathcal{A}_1^{-1}(\bar{R})$.

Mostra-se que o estimador $\hat{\lambda}$ é viciado e uma aproximação para o vício foi obtida por Best e Fisher (1981). Eles concluem que, para amostras pequenas e valores pequenos de

λ a distribuição de $\hat{\lambda}$ tem uma grande assimetria, com uma longa cauda à direita. Consequentemente, as inferências baseadas em valores pequenos de $\hat{\lambda}$ podem ser suspeitas, a menos que n seja suficientemente grande.

Embora a função \mathcal{A}_1 esteja implementada de forma eficiente em rotinas do IMSL (1991), pode-se também calcular a sua inversa \mathcal{A}_1^{-1} , com uma razoável precisão, usando a seguinte aproximação, sugerida por Best e Fisher (1981),

$$\mathcal{A}_1^{-1}(r) \simeq \begin{cases} 2r + r^3 + (5r^5)/6 & , & r < 0,53 \\ -0,4 + 1,39r + 0,43/(1-r) & , & 0,53 \leq r \leq 0,85 \\ (r^3 - 4r^2 + 3r)^{-1} & , & r > 0,85. \end{cases} \quad (3.3)$$

Para pequenas amostras, $n \leq 15$, Fisher (1993, p. 88) sugere o seguinte estimador

$$\hat{\lambda}^* = \begin{cases} \max\{\hat{\lambda} - 2(n\hat{\lambda})^{-1}, 0\}, & \text{se } \hat{\lambda} < 2 \\ \frac{(n-1)^3 \hat{\lambda}}{n^3 + n}, & \text{se } \hat{\lambda} \geq 2, \end{cases}$$

onde $\hat{\lambda}$ é o estimador de máxima verossimilhança, $\hat{\lambda} = \mathcal{A}_1^{-1}(\bar{R})$.

3.5.5 Outras distribuições circulares

Mostramos aqui as distribuições mais importantes para variáveis aleatórias que assumem valores no círculo. Além dessas, existem diversas outras distribuições, tais como Cardióide, Poisson Arqueada, Cauchy Arqueada, Lattice, entre outras. Maiores detalhes podem ser encontrados em Batschelet (1981), Fisher (1993) e Mardia (1972).

Capítulo 4

Influência Local em Dados Circulares

4.1 Modelos de regressão para dados circulares

Quando se investiga medidas de associação envolvendo variáveis direcionais e variáveis lineares (isto é, que assumem valores num espaço euclidiano), é necessário levar em consideração os seguintes tipos de associação:

circular-circular que mede o grau de associação entre duas variáveis direcionais;

circular-linear que pode ser utilizada para prever o valor esperado de uma variável direcional dado o valor de uma variável linear;

linear-circular que pode ser utilizada para prever o valor esperado de uma variável linear dado o valor de uma variável direcional.

Considerando uma amostra aleatória $(x_1, Y_1), \dots, (x_n, Y_n)$, onde x_1, \dots, x_n são números reais conhecidos e cada Y_i segue uma distribuição von Mises $\mathcal{VM}(\mu_i, \lambda)$, Gould (1969)

faz uma analogia à teoria de regressão normal linear e propõe um modelo admitindo que as médias angulares seguem uma estrutura linear do tipo $\mu_i = \mu + \beta x_i$, $i = 1, \dots, n$, onde μ , β e λ são parâmetros desconhecidos, a serem estimados. Gould obtém equações de estimação referentes aos estimadores de máxima verossimilhança para os parâmetros desse modelo, sendo que $\hat{\beta}$ é calculado através de um procedimento iterativo, o qual é fortemente dependente do ponto de inicialização do processo. Não nos deteremos aqui nesse modelo, mas o leitor interessado pode encontrar maiores detalhes em Gould (1969) e Mardia (1972).

Fisher e Lee (1983) discutem um coeficiente de correlação para medir associação entre duas variáveis circulares, digamos Y e W , fazendo uma analogia ao coeficiente de correlação linear, ou seja, ao caso de associação entre variáveis lineares. O coeficiente de correlação proposto é motivado pelo que se denomina de **associação toroidal** (ou T -associação), que pode ser definida da seguinte maneira.

Dizemos que variáveis Y e W possuem uma T -associação perfeita se

$$Y = \begin{cases} W + \mu_0 \{\text{mod } (2\pi)\}, & \text{associação positiva,} \\ -W + \mu_0 \{\text{mod } (2\pi)\}, & \text{associação negativa,} \end{cases}$$

onde μ_0 é um ângulo constante.

O conceito de associação toroidal e a definição acima proporcionaram a obtenção de uma medida, o *coeficiente de correlação toroidal*, que é definido como

$$\rho_T = \frac{E\{\text{sen}(Y_1 - Y_2) \text{sen}(W_1 - W_2)\}}{[E\{\text{sen}^2(Y_1 - Y_2)\} E\{\text{sen}^2(W_1 - W_2)\}]^{1/2}},$$

onde (Y_1, W_1) e (Y_2, W_2) são independentes com a mesma distribuição de (Y, W) . Essa expressão é bastante semelhante à que define o coeficiente de correlação linear usual.

Podemos destacar as seguintes propriedades do coeficiente de correlação toroidal ρ_T :

- i. $-1 \leq \rho_T \leq 1$;
- ii. $|\rho_T| = 1 \iff$ a T -associação entre Y e W for perfeita;
- iii. se Y e W forem independentes, então $\rho_T = 0$.

Um estimador para ρ_T , baseado numa amostra aleatória $(y_1, w_1), \dots, (y_n, w_n)$ tem a seguinte expressão:

$$\hat{\rho}_T = \frac{\sum_{1 \leq i < j \leq n} \text{sen}(y_i - y_j) \text{sen}(w_i - w_j)}{\left\{ \sum_{1 \leq i < j \leq n} \text{sen}^2(y_i - y_j) \sum_{1 \leq i < j \leq n} \text{sen}^2(w_i - w_j) \right\}^{1/2}}.$$

Não nos deteremos aqui em associação do tipo circular-circular. Para maiores detalhes, e outras propriedades de ρ_T , como por exemplo invariância sob transformação de rotação e seu comportamento assintótico podem ser encontradas em Fisher e Lee (1983) e Fisher (1993).

4.2 Modelos von Mises de regressão

Fisher e Lee (1992) apresentam uma classe de modelos para respostas direcionais com distribuição von Mises, onde é possível modelar tanto a média direcional da distribuição, como o parâmetro de concentração. Consideremos n observações independentes y_1, \dots, y_n , e vamos admitir que elas estão relacionadas com $\mathbf{x}_1, \dots, \mathbf{x}_n$, onde \mathbf{x}_i é um vetor de p variáveis explicativas, com $p \leq n$. Cada y_i é o valor observado de uma variável aleatória circular com distribuição von Mises $\mathcal{VM}(\mu_i, \lambda_i)$ e a dependência dos parâmetros μ_i e λ_i sobre \mathbf{x}_i é feita através de funções de ligação apropriadas. A abordagem que veremos a seguir guarda uma estreita analogia com a teoria dos modelos lineares generalizados, vide Nelder e Wedderburn (1972) e McCullagh e Nelder (1989).

4.2.1 O Modelo von Mises de médias

Nesse modelo, admite-se que os parâmetros de concentração $\lambda_1, \dots, \lambda_n$ são todos iguais a um valor λ e que cada média direcional μ_i está relacionada com as p variáveis explicativas por $\mu_i = \mu + g(\boldsymbol{\beta}^T \mathbf{x}_i)$, onde $g : \mathbf{R} \rightarrow [-\pi, \pi)$ é uma função de ligação, monótona, duas vezes continuamente diferenciável e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ é um vetor de parâmetros desconhecidos a serem estimados. A função de ligação g deve satisfazer ainda a condição $g(0) = 0$, que tem a finalidade de dar uma interpretação para o parâmetro μ como sendo uma “origem”, uma vez que a variável resposta assume valores no círculo, onde o conceito de origem não é bem definido. Fisher e Lee (1992) discutem algumas possíveis escolhas para a função g e uma função de ligação bastante usual é $g(t) = 2 \arctan(t)$.

Paula (1996) discute a aplicação de diagnóstico de influência local nos modelos próprios de dispersão, propostos por Jørgensen (1997). A classe dos modelos próprios de dispersão inclui a distribuição von Mises, quando a média direcional μ_i é modelada por $\mu_i = g(\boldsymbol{\beta}^T \mathbf{x}_i)$, ou seja, quando a “origem” μ está, por assim dizer, incorporada ao preditor linear $\eta = \boldsymbol{\beta}^T \mathbf{x}$.

Estimação dos parâmetros no modelo von Mises de médias

No modelo von Mises de médias, o vetor paramétrico é dado por $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}^T, \lambda)^T$ e a função de log-verossimilhança fica expressa como

$$l(\mu, \boldsymbol{\beta}, \lambda; y_1, \dots, y_n) = -n \log I_0(\lambda) + \lambda \sum_{i=1}^n \cos \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}.$$

Os estimadores de máxima verossimilhança de μ , $\boldsymbol{\beta}$ e λ são obtidos a partir da solução do sistema de equações

$$\mathbf{X}^T \dot{\mathbf{G}} \mathbf{u} = \mathbf{0}, \quad (4.1)$$

$$\bar{R} \operatorname{sen} \hat{\mu} = S, \quad (4.2)$$

$$\bar{R} \operatorname{cos} \hat{\mu} = C, \quad (4.3)$$

$$\mathcal{A}_1(\hat{\lambda}) = \bar{R}, \quad (4.4)$$

onde

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}, \\ \dot{\mathbf{G}} &= \operatorname{diag}[g'(\boldsymbol{\beta}^T \mathbf{x}_1), \dots, g'(\boldsymbol{\beta}^T \mathbf{x}_n)], \\ \mathbf{u} &= (u_1, \dots, u_n)^T, \text{ com} \\ u_i &= \operatorname{sen}\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}, \\ S &= \frac{1}{n} \sum_{i=1}^n \operatorname{sen}\{y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}, \\ C &= \frac{1}{n} \sum_{i=1}^n \operatorname{cos}\{y_i - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}, \\ \bar{R} &= (S^2 + C^2)^{1/2} \text{ e} \\ \mathcal{A}_1(\lambda) &= \frac{I_1(\lambda)}{I_0(\lambda)}. \end{aligned}$$

A equação (4.1) não tem uma solução explícita, e deve ser resolvida utilizando-se algum procedimento iterativo. Uma solução pode ser obtida através do algoritmo de mínimos quadrados reponderados – *iteratively reweighted least squares* (IRLS), Green (1984). A equação de atualização para $\hat{\boldsymbol{\beta}}$ é

$$\mathbf{X}^T \dot{\mathbf{G}}^2 \mathbf{X} (\hat{\boldsymbol{\beta}}^{(m+1)} - \hat{\boldsymbol{\beta}}^{(m)}) = \mathbf{X}^T \dot{\mathbf{G}}^2 \mathbf{y}^*, \quad (4.5)$$

onde $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^T$, com $y_i^* = u_i / \{\mathcal{A}_1(\lambda) g'(\boldsymbol{\beta}^T \mathbf{x}_i)\}$.

Processo iterativo

A obtenção das estimativas de máxima verossimilhança $\hat{\mu}$, $\hat{\beta}$ e $\hat{\lambda}$, segue o seguinte processo iterativo:

1. Obter uma estimativa inicial $\hat{\beta}_0$;
2. Calcular S , C e R , pelas expressões acima;
3. Calcular estimativas $\hat{\mu}$ e $\hat{\lambda}$, pelas equações (4.2) - (4.4);
4. Obter uma estimativa atualizada $\hat{\beta}$, usando a equação (4.5);
5. Repetir as etapas 2 a 4 até obter a convergência.

Os estimadores $\hat{\mu}$ e $\hat{\beta}$ são assintoticamente não correlacionados com $\hat{\lambda}$ e a variância assintótica de $\hat{\beta}$ é dada por

$$Var_a(\hat{\beta}) = \frac{1}{\lambda \mathcal{A}_1(\lambda)} \left\{ (\mathbf{X}^T \dot{\mathbf{G}}^2 \mathbf{X})^{-1} + \frac{(\mathbf{X}^T \dot{\mathbf{G}}^2 \mathbf{X})^{-1} \mathbf{X}^T \mathbf{g} \mathbf{g}^T \mathbf{X} (\mathbf{X}^T \dot{\mathbf{G}}^2 \mathbf{X})^{-1}}{n - \mathbf{g}^T \mathbf{X} (\mathbf{X}^T \dot{\mathbf{G}}^2 \mathbf{X})^{-1} \mathbf{X}^T \mathbf{g}} \right\}, \quad (4.6)$$

onde \mathbf{g} é um vetor contendo os elementos da diagonal principal da matriz $\dot{\mathbf{G}}$.

A variância assintótica de $\hat{\lambda}$ é dada por $Var_a(\hat{\lambda}) = \frac{1}{n \mathcal{A}'_1(\lambda)}$, enquanto que a variância circular assintótica de $\hat{\mu}$ é $\{2(n-p)\lambda \mathcal{A}_1(\lambda)\}^{-1}$ e uma estimativa para seu desvio padrão circular é dado por $\hat{\sigma}_a(\hat{\mu}) = \{(n-p)\hat{\lambda} \mathcal{A}_1(\hat{\lambda})\}^{-1/2}$.

4.2.2 O Modelo von Mises de dispersão

Em algumas situações práticas podemos estar interessados em modelar o parâmetro de concentração λ da distribuição von Mises, em vez da sua média direcional, como

por exemplo, nos experimentos de acompanhar a direção seguida por animais soltos após um estímulo, ou após terem sido submetidos a um tratamento. Observa-se que nos primeiros momentos ocorre uma grande variabilidade, indicando uma espécie de "indecisão inicial" dos animais, entretanto com o passar do tempo a variabilidade tende a diminuir. Ou seja, há indícios de que o parâmetro de concentração é função de uma covariável, que pode ser, por exemplo, o tempo t decorrido após a soltura do animal.

No modelo de dispersão, consideram-se n vetores $\mathbf{z}_1, \dots, \mathbf{z}_n$, com $\mathbf{z}_i = (1, z_{i1}, \dots, z_{iq})^T$ e admite-se que o parâmetro de concentração λ_i pode ser modelado por $\lambda_i = h(\boldsymbol{\gamma}^T \mathbf{z}_i)$, onde $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_q)^T$ é um vetor de parâmetros desconhecidos, a serem estimados, e $h : \mathbf{R} \rightarrow [0, +\infty)$ é uma função conhecida, duas vezes continuamente diferenciável, que atua como uma função de ligação. Em geral, as covariáveis z_{i1}, \dots, z_{iq} correspondem a um subconjunto de x_{i1}, \dots, x_{ip} .

Estimação dos parâmetros no modelo von Mises de dispersão

Nesse modelo o vetor paramétrico é dado por $\boldsymbol{\theta} = (\mu, \boldsymbol{\gamma}^T)^T$ e a log-verossimilhança fica expressa como

$$l(\mu, \boldsymbol{\gamma}; y_1, \dots, y_n) = - \sum_{i=1}^n \log I_0\{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} + \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos(y_i - \mu).$$

Os estimadores de máxima verossimilhança dos parâmetros $\mu, \gamma_0, \gamma_1, \dots, \gamma_q$ devem satisfazer o sistema de $(q + 3)$ equações,

$$\sum_{i=1}^n \{\cos(y_i - \hat{\mu}) - \mathcal{A}_1(\hat{\lambda}_i)\} h'(\boldsymbol{\gamma}^T \mathbf{z}_i) = 0, \quad (4.7)$$

$$\sum_{i=1}^n \{\cos(y_i - \hat{\mu}) - \mathcal{A}_1(\hat{\lambda}_i)\} h'(\boldsymbol{\gamma}^T \mathbf{z}_i) z_{ij} = 0, \text{ para } j = 1, \dots, q, \quad (4.8)$$

$$R \operatorname{sen} \hat{\mu} = S, \quad (4.9)$$

$$R \operatorname{cos} \hat{\mu} = C, \quad (4.10)$$

onde

$$\begin{aligned} \mathbf{Z} &= \begin{bmatrix} \mathbf{z}_1^T \\ \vdots \\ \mathbf{z}_n^T \end{bmatrix}, \\ S &= \sum_{i=1}^n \lambda_i \text{sen}(y_i), \\ C &= \sum_{i=1}^n \lambda_i \cos(y_i) \text{ e} \\ R &= (S^2 + C^2)^{1/2}. \end{aligned}$$

As equações (4.7) e (4.8) não admitem solução explícita e, assim como no modelo de médias, elas devem ser resolvidas utilizando-se algum procedimento iterativo. A solução através do algoritmo de mínimos quadrados ponderados – IRLS, Green (1984), usa a seguinte equação de atualização para $\hat{\gamma}$:

$$\mathbf{Z}^T \mathbf{W} \mathbf{Z} (\hat{\gamma}^{(m+1)} - \hat{\gamma}^{(m)}) = \mathbf{Z}^T \mathbf{W} \mathbf{y}^*, \quad (4.11)$$

onde $\mathbf{y}^* = (y_1^*, \dots, y_n^*)^T$, com $y_i^* = \frac{\cos(y_i - \mu) - \mathcal{A}_1(\lambda_i)}{h'(\boldsymbol{\gamma}^T \mathbf{z}_i) \mathcal{A}'_1(\lambda_i)}$, e \mathbf{W} é uma matriz diagonal com elementos $w_i = \{h'(\boldsymbol{\gamma}^T \mathbf{z}_i)\}^2 \mathcal{A}'_1(\lambda_i)$.

Processo iterativo

A obtenção das estimativas de máxima verossimilhança $\hat{\mu}$ e $\hat{\gamma}$, segue o seguinte processo iterativo:

1. Obter uma estimativa inicial $\hat{\gamma}_0$;
2. Calcular S e C pelas expressões acima;
3. Calcular uma estimativa $\hat{\mu}$ a partir das equações (4.9) e (4.10);

4. Obter uma estimativa atualizada $\hat{\gamma}$, usando a equação (4.11);
5. Repetir as etapas 2 a 4 até obter a convergência.

A variância assintótica de $\hat{\gamma}$ é dada por

$$\text{Var}_a(\hat{\gamma}) = (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1}, \quad (4.12)$$

enquanto que o comprimento médio resultante assintótico de $\hat{\mu}$ é

$$1 - \frac{1}{2 \sum_{i=1}^n \lambda_i \mathcal{A}_1(\lambda_i)}. \quad (4.13)$$

4.2.3 O Modelo von Mises misto

Segundo Fisher (1993) argumenta, não é incomum encontrar conjuntos de dados envolvendo variáveis direcionais onde tanto a média direcional quanto o parâmetro de concentração dependem de covariáveis.

Ao se combinar o modelo de médias com o modelo de dispersão, obtém-se uma classe de modelos de regressão que apresenta uma grande flexibilidade na modelagem envolvendo respostas angulares. Referimo-nos a essa classe simplesmente como *modelo von Mises misto*. A função de log-verossimilhança para $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ é dada por

$$l(\mu, \boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y}) = - \sum_{i=1}^n \log I_0\{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} + \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}. \quad (4.14)$$

Essa função é maximizada através da utilização simultânea dos dois métodos citados nas Secções 4.2.1 e 4.2.2. Entretanto, são necessárias as seguintes alterações:

1. A matriz $\dot{\mathbf{G}}^2$ da equação de atualização para $\boldsymbol{\beta}$, equação (4.5) deve ser substituída por $\dot{\mathbf{G}} \boldsymbol{\Lambda} \dot{\mathbf{G}}$, onde $\boldsymbol{\Lambda}$ é uma matriz diagonal com elementos $\lambda_i \mathcal{A}_1(\lambda_i)$;

2. Na definição do vetor \mathbf{y}^* , equação (4.5), substituir $\mathcal{A}_1(\lambda)$ por $\mathcal{A}_1(\lambda_i)$;
3. Na equação de atualização para $\boldsymbol{\gamma}$, equação (4.11), a i -ésima coordenada do vetor \mathbf{y}^* fica dada por

$$y_i^* = \frac{\cos \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{z}_i)\} - \mathcal{A}_1(\lambda_i)}{h'(\boldsymbol{\gamma}^T \mathbf{z}_i) \mathcal{A}'_1(\lambda_i)}.$$

Processo iterativo

A obtenção das estimativas de máxima verossimilhança $\hat{\mu}$, $\hat{\boldsymbol{\beta}}$ e $\hat{\boldsymbol{\gamma}}$, segue o seguinte processo iterativo:

1. Obter estimativas iniciais $\hat{\boldsymbol{\beta}}_0$ e $\hat{\boldsymbol{\gamma}}_0$;
2. Obter uma estimativa atualizada $\hat{\boldsymbol{\beta}}$, usando a equação (4.5), levando-se em consideração a alteração descrita acima;
3. Obter uma estimativa atualizada $\hat{\boldsymbol{\gamma}}$, usando a equação (4.11), também considerando a alteração correspondente acima;
4. Calcular a estimativa $\hat{\mu}$ pelas equações (4.9) e (4.10);
5. Repetir as etapas 2 a 4 até obter a convergência.

As variâncias assintóticas de $\hat{\boldsymbol{\beta}}$ e de $\hat{\boldsymbol{\gamma}}$ são dadas por

$$\text{Var}_a(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \dot{\mathbf{G}} \boldsymbol{\Lambda} \dot{\mathbf{G}} \mathbf{X})^{-1}, \quad (4.15)$$

$$\text{Var}_a(\hat{\boldsymbol{\gamma}}) = (\mathbf{Z}^T \mathbf{W} \mathbf{Z})^{-1}. \quad (4.16)$$

O comprimento assintótico da resultante média é

$$1 - \frac{1}{2 \sum_{i=1}^n \lambda_i \mathcal{A}_1(\lambda_i)}. \quad (4.17)$$

Além disso, $\hat{\mu}$ e $\hat{\boldsymbol{\beta}}$ são assintoticamente não correlacionados com $\hat{\boldsymbol{\gamma}}$, Fisher e Lee (1992).

4.3 Influência local no modelo von Mises misto

No modelo von Mises misto são modeladas simultaneamente a média direcional e o parâmetro de concentração. Neste caso, assumimos que a função densidade de probabilidade, para a observação y_i , é dada por

$$f(y_i; \mu_i, \lambda_i) = \{2\pi I_0(\lambda_i)\}^{-1} e^{\lambda_i \cos(y_i - \mu_i)}, \quad 0 \leq y < 2\pi, 0 \leq \mu < 2\pi \text{ e } \lambda > 0, \quad (4.18)$$

onde $\mu_i = \mu + g(\boldsymbol{\beta}^T \mathbf{z}_i)$ e $\lambda_i = h(\boldsymbol{\gamma}^T \mathbf{z}_i)$.

As funções $g: \mathbf{R} \rightarrow [-\pi, \pi)$ e $h: \mathbf{R} \rightarrow [0, +\infty)$, que atuam como funções de ligação, devem ser duas vezes continuamente diferenciáveis e, além disso, $g(0) = 0$.

Nesse modelo, o vetor paramétrico fica dado por $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$, onde μ é uma média direcional de referência, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ e $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \dots, \gamma_q)^T$. A função de log-verossimilhança para $\boldsymbol{\theta}$ tem a forma

$$l(\boldsymbol{\theta}; \mathbf{y}) = - \sum_{i=1}^n \log I_0\{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} + \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}.$$

A fim de usarmos os resultados do Capítulo 2, que trata de influência local, precisaremos das derivadas parciais de primeira e segunda ordens da função $l(\cdot; \cdot)$, em relação aos vários parâmetros. Vamos obter inicialmente as derivadas parciais de primeira ordem. Assim,

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta}; \mathbf{y})}{\partial \mu} &= \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \text{sen}\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}, \\ \frac{\partial l(\boldsymbol{\theta}; \mathbf{y})}{\partial \beta_r} &= \sum_{i=1}^n x_{ir} h(\boldsymbol{\gamma}^T \mathbf{z}_i) g'(\boldsymbol{\beta}^T \mathbf{x}_i) \text{sen}\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}, \end{aligned}$$

para $1 \leq r \leq p$. Com relação aos parâmetros referentes à dispersão tem-se que

$$\frac{\partial l(\boldsymbol{\theta}; \mathbf{y})}{\partial \gamma_s} = \sum_{i=1}^n [-\mathcal{A}_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} + \cos\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}] h'(\boldsymbol{\gamma}^T \mathbf{z}_i) z_{is}.$$

para $0 \leq s \leq q$, com $z_{i0} = 1$.

Vamos agora às derivadas parciais de segunda ordem.

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{y})}{\partial \mu^2} &= - \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}, \\
\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{y})}{\partial \mu \partial \beta_s} &= - \sum_{i=1}^n x_{is} g'(\boldsymbol{\beta}^T \mathbf{x}_i) \{h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}, \\
\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{y})}{\partial \mu \partial \gamma_s} &= \sum_{i=1}^n h'(\boldsymbol{\gamma}^T \mathbf{z}_i) \text{sen} \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\} z_{is}, \\
\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{y})}{\partial \beta_r \partial \beta_s} &= \sum_{i=1}^n x_{ir} h(\boldsymbol{\gamma}^T \mathbf{z}_i) \left[g''(\boldsymbol{\beta}^T \mathbf{x}_i) x_{is} \text{sen} \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\} + \right. \\
&\quad \left. g'(\boldsymbol{\beta}^T \mathbf{x}_i) \cos \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\} \{-g'(\boldsymbol{\beta}^T \mathbf{x}_i)\} x_{is} \right], \\
\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{y})}{\partial \beta_r \partial \gamma_s} &= \sum_{i=1}^n x_{ir} h'(\boldsymbol{\gamma}^T \mathbf{z}_i) z_{is} g'(\boldsymbol{\beta}^T \mathbf{x}_i) \text{sen} \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}, \\
\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{y})}{\partial \gamma_r \partial \gamma_s} &= \sum_{i=1}^n z_{ir} \left[-\mathcal{A}'_1 \{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} \{h'(\boldsymbol{\gamma}^T \mathbf{z}_i)\}^2 z_{is} z_{ir} \right. \\
&\quad \left. -\mathcal{A} \{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} h''(\boldsymbol{\gamma}^T \mathbf{z}_i) z_{is} z_{ir} \right] + \\
&\quad h''(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\} z_{is}.
\end{aligned}$$

Com essas derivadas podemos então obter uma expressão para a matriz de informação de Fisher observada $\ddot{L} = -\frac{\partial^2 l(\boldsymbol{\theta}; y_1, \dots, y_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$, de dimensão $(p+q+2) \times (p+q+2)$, onde $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\gamma}}^T)^T$ é o estimador de máxima verossimilhança de $\boldsymbol{\theta}$. Considerando de uma forma particionada, podemos escrever

$$\ddot{L} = \frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = \begin{bmatrix} L_{11} & L_{12} & L_{13} \\ & L_{22} & L_{23} \\ & & L_{33} \end{bmatrix}$$

sendo que a parte triangular inferior é obtida por simetria.

Vamos introduzir uma notação matricial, considerando, para cada $i = 1, \dots, n$, os elementos

1. $t_i = \cos\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}$
2. $u_i = \text{sen}\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\},$

e, a partir destes, vamos definir as matrizes

$$\begin{aligned} \mathcal{M}_1 &= \text{diag}\{h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}\} \\ \mathcal{M}_2 &= \text{diag}\{h(\boldsymbol{\gamma}^T \mathbf{z}_i) \text{sen}\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}\} \\ \mathcal{M}_3 &= \text{diag}\{h'(\boldsymbol{\gamma}^T \mathbf{z}_i) \text{sen}\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\} \cos\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}\} \\ \mathcal{M}_4 &= \text{diag}\{h''(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\} - \mathcal{A}_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_1)\} h''(\boldsymbol{\gamma}^T \mathbf{z}_i) - \\ &\quad \mathcal{A}'_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_1)\} \{h'(\boldsymbol{\gamma}^T \mathbf{z}_i)\}^2\}, \end{aligned}$$

e com isto, a matriz de segundas derivadas da função de log-verossimilhança, com relação a $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$, fica dada por

$$\ddot{L} = \begin{bmatrix} -\mathbf{1}^T \mathcal{M}_1 \mathbf{1} & -\mathbf{1}^T \mathcal{M}_1 \dot{G} X & -\mathbf{1}^T \mathcal{M}_3 Z \\ -X^T \dot{G} \mathcal{M}_1 \mathbf{1} & X^T \mathcal{M}_2 [\ddot{G} - \mathcal{M}_1 \dot{G}^2] X & -X^T \mathcal{M}_3 Z \\ -Z^T \mathcal{M}_3 \mathbf{1} & -Z^T \mathcal{M}_3 X & Z^T \mathcal{M}_4 Z \end{bmatrix}.$$

Vamos agora introduzir alguns tipos de perturbação no modelo von Mises misto, a fim de estudarmos a influência local. Consideraremos os seguintes tipos de perturbação:

1. Ponderação de Casos: $l(\boldsymbol{\theta}; y_i) \mapsto \omega_i l(\boldsymbol{\theta}; y_i)$;
2. Perturbando a resposta: $y_i \mapsto y_i + \sigma_i \omega_i$; , onde σ_i é um fator de escala;
3. Perturbando uma variável explanatória individualmente.

4.3.1 Ponderação de casos

A log-verossimilhança perturbada tem a forma

$$l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y}) = - \sum_{i=1}^n \omega_i \log I_0\{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} + \sum_{i=1}^n \omega_i h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\},$$

e para este esquema de perturbação, o vetor correspondente à não perturbação é o vetor n -dimensional $\boldsymbol{\omega}_0 = (1, \dots, 1)^T$. Vamos calcular as derivadas necessárias a obtenção da $\ddot{\mathbf{F}}$, vide equação (2.19), a partir da qual poderemos obter a direção de maior influência local.

Inicialmente, temos a derivada

$$\frac{\partial l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \omega_r} = -\log I_0\{h(\boldsymbol{\gamma}^T \mathbf{z}_r)\} + h(\boldsymbol{\gamma}^T \mathbf{z}_r) \cos\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r)\}.$$

Agora,

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \mu \partial \omega_r} &= h(\boldsymbol{\gamma}^T \mathbf{z}_r) \operatorname{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r)\}, \\ \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \beta_s \partial \omega_r} &= h(\boldsymbol{\gamma}^T \mathbf{z}_r) \operatorname{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r)\} g'(\boldsymbol{\beta}^T \mathbf{x}_r) x_{rs}, \\ \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \gamma_s \partial \omega_r} &= [-\mathcal{A}_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_r)\} + \cos\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r)\}] h'(\boldsymbol{\gamma}^T \mathbf{z}_r) z_{rs}. \end{aligned}$$

Definindo

$$\mathcal{D}_1 = \operatorname{diag}\{h'(\boldsymbol{\gamma}^T \mathbf{z}_i)[- \mathcal{A}_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} + \cos\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}]\}$$

e usando a matriz \mathcal{M}_2 , tem-se que

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \mu \partial \boldsymbol{\omega}^T} &= \mathbf{1}^T \mathcal{M}_2, \\ \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T} &= \mathbf{X}^T \dot{\mathcal{G}} \mathcal{M}_2 \text{ e} \\ \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\omega}^T} &= \mathbf{Z}^T \mathcal{D}_1. \end{aligned}$$

Portanto, a matriz $\Delta = \frac{\partial^2 l(\hat{\boldsymbol{\beta}}, \boldsymbol{\omega}; \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T}$, fica dada por

$$\Delta = \begin{bmatrix} \mathbf{1}^T \mathcal{M}_2 \\ \mathbf{X}^T \dot{\mathbf{G}} \mathcal{M}_2 \\ \mathbf{Z}^T \mathcal{D}_1 \end{bmatrix}_{(p+q+2) \times n},$$

onde as derivadas são avaliadas em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = (1, \dots, 1)^T$. É importante observar que, para esse esquema de perturbação em particular, a matriz Δ não depende do vetor $\boldsymbol{\omega}$.

O vetor d_{\max} é então obtido, que é o autovetor normalizado correspondente ao maior autovalor da matriz $\Delta^T \ddot{\mathbf{L}}^{-1} \Delta$. Um gráfico de $|d_{\max}|$ versus a ordem das observações pode revelar quais são os casos que exercem mais influência sobre as estimativas dos parâmetros.

4.3.2 Perturbando a variável resposta y

No modelo misto, que estamos considerando, é assumido que cada Y_i tem uma concentração λ_i diferente das demais observações. Desta forma, é mais conveniente atribuir um peso para o elemento da perturbação ω_i . Como peso, usaremos uma estimativa do desvio padrão circular para a distribuição von Mises, dado por $\sigma_i = \{-2 \log \mathcal{A}_1(\lambda_i)\}^{1/2}$, Mardia (1972). Uma outra possibilidade seria ponderar ω_i por uma estimativa da dispersão circular que, no caso da distribuição von Mises, tem a forma $\delta_i = \{\lambda_i A(\lambda_i)\}^{-1}$.

Usando a ponderação com $\hat{\sigma}_i$, a função de log-verossimilhança perturbada fica dada por

$$l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y}) = - \sum_{i=1}^n \log I_0\{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} + \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos \{y_i + \hat{\sigma}_i \omega_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\},$$

sendo que o vetor correspondente à não perturbação é $\boldsymbol{\omega}_0 = \mathbf{0}$, o vetor nulo n -dimensional.

Vamos obter inicialmente a derivada

$$\frac{\partial l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \omega_r} = -\hat{\sigma}_r h(\boldsymbol{\gamma}^T \mathbf{z}_r) \text{sen}\{y_r + \hat{\sigma}_r \omega_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r)\}.$$

Com isso,

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \mu \partial \omega_r} &= \hat{\sigma}_r h(\boldsymbol{\gamma}^T \mathbf{z}_r) \cos\{y_r + \hat{\sigma}_r \omega_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r)\}, \\ \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \beta_s \partial \omega_r} &= \hat{\sigma}_r h(\boldsymbol{\gamma}^T \mathbf{z}_r) \cos\{y_r + \hat{\sigma}_r \omega_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r)\} g'(\boldsymbol{\beta}^T \mathbf{x}_r) x_{rs}, \\ \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \gamma_s \partial \omega_r} &= -\hat{\sigma}_r h'(\boldsymbol{\gamma}^T \mathbf{z}_r) z_{rs} \text{sen}\{y_r + \hat{\sigma}_r \omega_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r)\}. \end{aligned}$$

Vamos agora avaliar essas derivadas de segunda ordem em $\boldsymbol{\omega}_0 = \mathbf{0}$ e $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, o estimador de máxima verossimilhança. Definindo $\mathbf{S} = \text{diag}\{\hat{\sigma}_1, \dots, \hat{\sigma}_n\}$ e usando as matrizes \mathcal{M}_1 e \mathcal{M}_2 tem-se que

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \mu \partial \boldsymbol{\omega}^T} &= \mathbf{1}^T \mathcal{M}_1 \mathbf{S}, \\ \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T} &= \mathbf{X}^T \dot{\mathbf{G}} \mathcal{M}_1 \mathbf{S}, \\ \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\omega}^T} &= \mathbf{Z}^T \mathcal{M}_2 \mathbf{S}. \end{aligned}$$

Portanto a matriz $\Delta = \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^T}$, fica dada por

$$\Delta = \begin{bmatrix} \mathbf{1}^T \mathcal{M}_1 \mathbf{S} \\ \mathbf{X}^T \dot{\mathbf{G}} \mathcal{M}_1 \mathbf{S} \\ \mathbf{Z}^T \mathcal{M}_2 \mathbf{S} \end{bmatrix}_{(p+q+2) \times n},$$

sendo que todas essas submatrizes são computadas a partir da estimativa de máxima verossimilhança $\hat{\boldsymbol{\theta}}$.

O autovetor normalizado correspondente ao maior autovalor da matriz $\Delta^T \ddot{\mathbf{L}}^{-1} \Delta$ fornece a direção de maior influência local, que é o vetor $d_{1\text{max}}$. Esse vetor contém informação

sobre quais são as respostas que exercem maior influência sobre as estimativas dos parâmetros.

4.3.3 Perturbando uma covariável individualmente

Seguindo Thomas e Cook (1990), vamos modificar a j -ésima coluna da matriz \mathbf{X} , adicionando um vetor $\boldsymbol{\omega}$ de pequenas perturbações multiplicadas por um fator de escala s , por exemplo, o inverso do desvio padrão amostral de x_j . O vetor de não perturbação é $\boldsymbol{\omega}_0 = \mathbf{0}$, o vetor $\mathbf{0}$ n -dimensional.

Sem perda de generalidade podemos considerar $j = 1$ pois, caso contrário, basta permutar as colunas de \mathbf{X} e trabalhar com a coluna j como sendo a primeira coluna da nova matriz de especificação do modelo. Da mesma forma, assumiremos que a covariável que está sendo perturbada corresponde também à segunda coluna da matriz \mathbf{Z} . Lembremos que a primeira coluna de \mathbf{Z} é uma coluna de uns.

As linhas da nova matriz \mathbf{X} serão constituídas por $\mathbf{x}_i^* = (x_{i1} + s\omega_i, x_{i2}, \dots, x_{ip})^T$ e analogamente, as linhas da nova matriz \mathbf{Z} serão $\mathbf{z}_i^* = (1, z_{i1} + s\omega_i, z_{i2}, \dots, z_{iq})^T$, para $i = 1, \dots, n$. A função de log-verossimilhança do modelo perturbado é

$$l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y}) = - \sum_{i=1}^n \log I_0\{h(\boldsymbol{\gamma}^T \mathbf{z}_i^*)\} + \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i^*) \cos\{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i^*)\}.$$

Assim,

$$\begin{aligned} \frac{\partial l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \omega_r} &= -s\gamma_1 \mathcal{A}_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_r^*)\} h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \\ &\quad + s\gamma_1 h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \cos\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} \\ &\quad + s\beta_1 h(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \text{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} g'(\boldsymbol{\beta}^T \mathbf{x}_r^*), \end{aligned}$$

para $r = 1, \dots, n$. Com isso,

$$\begin{aligned}
\frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \mu \partial \omega_r} &= s\gamma_1 h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \text{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} - \\
&\quad s\beta_1 h(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \cos\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} g'(\boldsymbol{\beta}^T \mathbf{x}_r^*), \\
\frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \beta_1 \partial \omega_r} &= s\gamma_1 h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \text{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} g'(\boldsymbol{\beta}^T \mathbf{x}_r^*) (x_{r1} + s\omega_r) + \\
&\quad sh(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \left[\text{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} g'(\boldsymbol{\beta}^T \mathbf{x}_r^*) - \right. \\
&\quad \left. \beta_1 \cos\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} \{g'(\boldsymbol{\beta}^T \mathbf{x}_r^*)\}^2 (x_{r1} + s\omega_r) + \right. \\
&\quad \left. \beta_1 \text{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} g''(\boldsymbol{\beta}^T \mathbf{x}_r^*) (x_{r1} + s\omega_r) \right], \\
\frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \beta_s \partial \omega_r} &= s\gamma_1 h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \text{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} g'(\boldsymbol{\beta}^T \mathbf{x}_r^*) x_{rs} + \\
&\quad s\beta_1 h(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \left[-\cos\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} \{g'(\boldsymbol{\beta}^T \mathbf{x}_r^*)\}^2 x_{rs} + \right. \\
&\quad \left. \text{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} g''(\boldsymbol{\beta}^T \mathbf{x}_r^*) x_{rs} \right], \text{ para } s = 2, \dots, p, \\
\frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \gamma_1 \partial \omega_r} &= -s\mathcal{A}_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_r^*)\} h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \\
&\quad -s\gamma_1 \mathcal{A}'_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_r^*)\} \{h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*)\}^2 (z_{r1} + s\omega_r) \\
&\quad -s\gamma_1 \mathcal{A}_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_r^*)\} h''(\boldsymbol{\gamma}^T \mathbf{z}_r^*) (z_{r1} + s\omega_r) \\
&\quad +sh'(\boldsymbol{\gamma}^T \mathbf{z}_r^*) \cos\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} \\
&\quad +s\gamma_1 h''(\boldsymbol{\gamma}^T \mathbf{z}_r^*) (z_{r1} + s\omega_r) \cos\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} \\
&\quad +s\beta_1 h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*) (z_{r1} + s\omega_r) \text{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} g'(\boldsymbol{\beta}^T \mathbf{x}_r^*), \\
\frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega}; \mathbf{y})}{\partial \gamma_s \partial \omega_r} &= -s\gamma_1 \mathcal{A}'_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_r^*)\} \{h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*)\}^2 z_{rs} \\
&\quad -s\gamma_1 \mathcal{A}_1\{h(\boldsymbol{\gamma}^T \mathbf{z}_r^*)\} h''(\boldsymbol{\gamma}^T \mathbf{z}_r^*) z_{rs} \\
&\quad +s\gamma_1 h''(\boldsymbol{\gamma}^T \mathbf{z}_r^*) z_{rs} \cos\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} \\
&\quad +s\beta_1 h'(\boldsymbol{\gamma}^T \mathbf{z}_r^*) z_{rs} \text{sen}\{y_r - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_r^*)\} g'(\boldsymbol{\beta}^T \mathbf{x}_r^*), \text{ para } s \neq 1.
\end{aligned}$$

Como antes, nessa última derivada tem-se $z_{r0} = 1$, para $1 \leq r \leq n$.

A matriz Δ pode ser escrita na forma de blocos como

$$\Delta = \begin{bmatrix} (\Delta\mu) \\ (\Delta\beta)_1 + (\Delta\beta)_2 \\ (\Delta\gamma)_1 + (\Delta\gamma)_2 \end{bmatrix}_{(p+q+2) \times n}.$$

A linha correspondente ao parâmetro μ é dada pelo vetor

$$(\Delta\mu) = \begin{bmatrix} s\gamma_1 h'(\gamma^T z_1) u_1 - s\beta_1 h(\gamma^T z_1) t_1 g'(\beta^T x_1), \\ \dots, s\gamma_1 h'(\gamma^T z_n) u_n - s\beta_1 h(\gamma^T z_n) t_n g'(\beta^T x_n) \end{bmatrix}.$$

Para o bloco correspondente ao vetor β , tem-se $(\Delta\beta)_1 = \mathbf{X}^T \mathcal{D}_2$, onde

$$\mathcal{D}_2 = \text{diag} \left\{ s\gamma_1 h'(\gamma^T z_r) u_r g'(\beta^T x_r) \right. \\ \left. - s\beta_1 h(\gamma^T z_r) t_1 \{g'(\beta^T x_r)\}^2 \right. \\ \left. + s\beta_1 h(\gamma^T z_r) u_r g''(\beta^T x_r) \right\},$$

enquanto,

$$(\Delta\beta)_2 = \begin{bmatrix} sh(\gamma^T z_1) u_1 & \dots & sh(\gamma^T z_n) u_n \\ 0 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & 0 \end{bmatrix}_{p \times n}.$$

Para o bloco correspondente ao vetor γ , tem-se $(\Delta\gamma)_1 = \mathbf{Z}^T \mathcal{D}_3$, onde

$$\mathcal{D}_3 = \text{diag} \left\{ -s\gamma_1 \mathcal{A}_1 \{h(\gamma^T z_r)\} \{h'(\gamma^T z_r)\}^2 \right. \\ \left. - s\gamma_1 \mathcal{A}_1 \{h(\gamma^T z_r)\} h''(\gamma^T z_r) \right. \\ \left. + s\gamma_1 h''(\gamma^T z_r) t_r \right. \\ \left. + s\beta_1 h'(\gamma^T z_r) u_r \right\}.$$

Por fim, a matriz $(\Delta\gamma)_2$ é uma matriz nula de dimensão $\{(q+1) \times n\}$, com exceção da segunda linha, em que os elementos são dados por

$$\{sh'(\gamma^T z_r) t_r - \varepsilon \mathcal{A}_1 \{h(\gamma^T z_r)\} h'(\gamma^T z_r), \text{ com } r = 1, \dots, n\}.$$

O autovetor normalizado d_{\max} (que corresponde ao maior autovalor da matriz $\Delta^T \ddot{L}^{-1} \Delta$) indica quais são os valores da covariável sob análise que exercem maior influência sobre as estimativas dos parâmetros.

4.4 Influência local sobre predições

Com a finalidade de fazer diagnóstico de influência sobre predições de modelos lineares generalizados, empregando o método de influência local, Thomas e Cook (1990) discutem o uso de outra função objetivo, diferente do afastamento pela verossimilhança.

Vamos considerar o modelo von Mises misto,

$$\begin{aligned} y_i &\sim \mathcal{VM}(\mu_i, \lambda_i) \\ \mu_i &= \mu + g(\mathbf{x}_i^T \boldsymbol{\beta}), \\ \lambda_i &= h(\mathbf{z}_i^T \boldsymbol{\gamma}) \text{ com } \mathbf{z}_i^T = (1, \mathbf{x}_i^T). \end{aligned}$$

A função de log-verossimilhança para $\boldsymbol{\theta} = (\mu, \boldsymbol{\beta}^T, \boldsymbol{\gamma}^T)^T$ fica expressa como

$$l(\boldsymbol{\theta}; \mathbf{y}) = - \sum_{i=1}^n \log I_0\{h(\boldsymbol{\gamma}^T \mathbf{z}_i)\} + \sum_{i=1}^n h(\boldsymbol{\gamma}^T \mathbf{z}_i) \cos \{y_i - \mu - g(\boldsymbol{\beta}^T \mathbf{x}_i)\}.$$

Seja $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_p)^T$ um vetor contendo valores (não necessariamente observados) das covariáveis X_1, \dots, X_p . Com o modelo proposto, a predição no ponto $\tilde{\mathbf{x}}$, é dada por $\hat{y}(\tilde{\mathbf{x}}) = \hat{\mu} + g(\tilde{\mathbf{x}}^T \hat{\boldsymbol{\beta}})$. Para um vetor de perturbação $\boldsymbol{\omega}$, a predição calculada com o modelo perturbado fica expressa como $\hat{y}(\tilde{\mathbf{x}}; \boldsymbol{\omega}) = \hat{\mu}_{\boldsymbol{\omega}} + g(\tilde{\mathbf{x}}^T \hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})$. Novamente estamos supondo que existe $\boldsymbol{\omega}_0 \in \Omega$ tal que $\hat{y}(\tilde{\mathbf{x}}) = \hat{y}(\tilde{\mathbf{x}}; \boldsymbol{\omega}_0)$, $\forall \boldsymbol{\theta} \in \mathbf{R}^{p+q+2}$.

A fim de avaliar influência sobre predições, vamos usar a função

$$f(\tilde{\mathbf{x}}, \boldsymbol{\omega}) = \{\hat{y}(\tilde{\mathbf{x}}) - \hat{y}(\tilde{\mathbf{x}}; \boldsymbol{\omega})\}^2,$$

como função objetivo, proposta por Thomas e Cook (1990). Novamente, o estudo da influência local consiste em investigar o comportamento da função $f(\tilde{\mathbf{x}}, \boldsymbol{\omega})$ numa vizinhança de $\boldsymbol{\omega}_0$, que é o ponto onde as duas predições são idênticas.

Após cálculos semelhantes aos da Secção 2.4, a curvatura normal da superfície $(\boldsymbol{\omega}^T, f(\tilde{\mathbf{x}}, \boldsymbol{\omega}))^T$, na direção \mathbf{d} , fica dada por

$$C_{\mathbf{d}}(\tilde{\mathbf{x}}) = 2 |\mathbf{d}^T \ddot{\mathbf{f}} \mathbf{d}|,$$

onde $\ddot{\mathbf{f}} = \frac{\partial^2 f(\tilde{\mathbf{x}}, \boldsymbol{\omega})}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^T}$, avaliada em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

Nesse caso, a matriz $\ddot{\mathbf{f}}$ pode ser decomposta na forma $\ddot{\mathbf{f}} = \mathbf{U}\mathbf{U}^T$, para $\mathbf{U} = \mathbf{J}_\mu + \dot{g}\mathbf{J}_\beta\tilde{\mathbf{x}}$, onde $\dot{g} = g'(\tilde{\mathbf{x}}^T\hat{\boldsymbol{\beta}})$ e \mathbf{J} é uma matriz em blocos dada por

$$\mathbf{J} = \begin{bmatrix} \partial\hat{\mu}_\omega/\partial\boldsymbol{\omega} \\ \partial\hat{\boldsymbol{\beta}}_\omega/\partial\boldsymbol{\omega} \\ \partial\hat{\boldsymbol{\gamma}}_\omega/\partial\boldsymbol{\omega} \end{bmatrix} = \begin{bmatrix} \mathbf{J}_\mu^T \\ \cdots \\ \mathbf{J}_\beta^T \\ \cdots \\ \mathbf{J}_\gamma^T \end{bmatrix} \begin{matrix} \} 1 \\ \\ \} p \\ \\ \} q+1 \end{matrix},$$

avaliada em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Definindo $\mathbf{e}_1 = (1, 0, \dots, 0)^T$ e $\mathbf{E}_\beta = [\mathbf{0}:\mathbf{I}_p:\mathbf{0}]^T$, a direção de maior influência local pode ser expressa como

$$\mathbf{d}_{\max} \propto \boldsymbol{\Delta}^T \ddot{\mathbf{L}}^{-1} (\mathbf{e}_1 + \dot{g}\mathbf{E}_\beta\tilde{\mathbf{x}}), \quad (4.19)$$

onde $(-\ddot{\mathbf{L}})$ é a matriz de informação observada de Fisher e $\boldsymbol{\Delta} = \frac{\partial^2 l(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T}$, avaliada em $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

A idéia agora é fazer $\tilde{\mathbf{x}}$ assumir os valores $\mathbf{x}_1, \dots, \mathbf{x}_n$ e avaliar a influência de cada ponto sobre a própria predição. Essa análise pode ser feita através de um gráfico de \mathbf{d}_{\max} contra o índice das observações.

Assim, uma vez definido o tipo de perturbação (ponderação de casos, perturbação aditiva na resposta y , etc.), a direção \mathbf{d}_{\max} pode ser obtida através da expressão (4.19) com o uso da correspondente matriz Δ e da matriz $(-\ddot{\mathbf{L}})$ obtidas na Secção 4.3.

Outras medidas de influência têm sido propostas por alguns autores e, a propósito disso, Fung e Kwan (1997) discutem a aplicação de influência local para outras medidas de influência diferentes do afastamento pela verossimilhança. Fung e Kwan mostram que uma medida de influência, digamos $\hat{T}_{\boldsymbol{\omega}}$, é invariante de escala se $\dot{\Gamma} = \frac{\partial \hat{T}_{\boldsymbol{\omega}}}{\partial \boldsymbol{\omega}} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \mathbf{0}$. Quando esta derivada é diferente de zero a ordem entre os componentes de \mathbf{d}_{\max} não é necessariamente preservada sob mudanças de escala. No presente caso, esta propriedade é satisfeita pela medida $f(\tilde{\mathbf{x}}, \boldsymbol{\omega}) = \{\hat{y}(\tilde{\mathbf{x}}) - \hat{y}(\tilde{\mathbf{x}}; \boldsymbol{\omega})\}^2$, proposta por Thomas e Cook (1990), entretanto ela não vale para outras medidas de influência discutidas por Fung e Kwan (1997).

Capítulo 5

Estudo de Resíduos em Dados Circulares

5.1 Introdução

Como já destacamos anteriormente, observações que representam direções devem ser analisadas por métodos estatísticos especiais. A metodologia empregada para dados lineares não fazem sentido para informações circulares.

Esta particularidade está presente também quando se analisa os resíduos referentes a uma variável direcional. Para variáveis lineares, muitos resíduos são obtidos a partir de uma padronização de $Y - \mu$, onde μ é a esperança matemática de Y . Devido à periodicidade existente para dados direcionais, um resíduo como este não é apropriado, pois o ângulo correspondente à diferença $(Y - \mu) \{\text{mod}(2\pi)\}$ cresce no intervalo $[0, \frac{\pi}{2})$ e é decrescente em $[\frac{\pi}{2}, 2\pi)$. Por conseguinte, o significado de resíduo “grande” ou resíduo

“pequeno” não é tão claro para uma resposta circular, o quanto é para uma variável linear.

Vários resíduos são discutidos em Cox e Snell (1968), McCullagh e Nelder (1989), Jørgensen (1997), entre outros. De uma maneira geral, podemos definir o resíduo referente à i -ésima observação através de uma função $r_i = r(Y_i, \hat{\mu}_i)$ que objetiva identificar discrepâncias entre o modelo e os dados. Cox e Snell (1968) fornecem aproximações para os dois primeiros momentos do resíduo r_i , admitindo que $Y_i = Y(\hat{\mu}_i, \epsilon_i)$, onde ϵ_i é uma variável aleatória possuindo uma distribuição de probabilidade completamente especificada.

5.2 Componente da função desvio

Quando se procura ajustar um modelo a um conjunto de dados, a validação desse ajuste passa pela análise de uma estatística especial, que tem a finalidade de “medir” a qualidade do modelo ajustado. Uma vez que os resíduos são usados para identificar discrepâncias entre um modelo ajustado e o conjunto de dados, é conveniente buscar uma definição para resíduo que leve em consideração a contribuição de cada observação sobre essa medida de qualidade de ajuste.

Uma medida de qualidade de ajuste bastante utilizada nos modelos lineares generalizados é a função desvio (*deviance*), definida por $D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \sum_{i=1}^n d_i^2$, com

$$d_i = d(y_i; \mu_i) = \pm\sqrt{2}\{l_i(y_i; \tilde{\mu}_i) - l_i(y_i; \hat{\mu}_i)\}^{1/2},$$

onde $l_i(y_i; \cdot)$ é a contribuição de y_i para a log-verossimilhança total, $\tilde{\mu}_i$ é o estimador de máxima verossimilhança de μ_i baseado apenas em y_i , e $\hat{\mu}_i$ é o estimador de máxima verossimilhança de μ_i baseado em toda a amostra y_1, \dots, y_n .

Para a distribuição von Mises, $y_i \sim \mathcal{VM}(\mu_i, \lambda)$ o componente da função desvio é dado por $d_i = d(y_i; \hat{\mu}_i, \lambda) = \pm\sqrt{2\lambda} \{1 - \cos(y_i - \hat{\mu}_i)\}^{1/2}$, onde o sinal de d_i é o mesmo de $y_i - \hat{\mu}_i$. Da relação trigonométrica $\cos(\theta) = \cos^2(\theta/2) - \text{sen}^2(\theta/2)$, segue-se que $\text{sen}(\theta/2) = \left\{ \frac{1 - \cos(\theta)}{2} \right\}^{1/2}$, e portanto uma expressão mais apropriada para d_i é dada por

$$d(y_i; \hat{\mu}_i, \lambda) = 2\sqrt{\lambda} \text{sen}\left(\frac{y_i - \hat{\mu}_i}{2}\right). \quad (5.1)$$

Nosso objetivo é estudar a distribuição de $d(y_i; \hat{\mu}_i, \lambda)$, quando a média direcional μ_i é modelada por $\mu_i = \mu + g(\beta^T \mathbf{x}_i)$, ou seja, nas condições do modelo de médias von Mises. apresentado na Secção 4.2.1.

É importante notar que estamos tratando de um resíduo que está definido em \mathbf{R} , e não mais de uma variável angular. Considerando o grande uso do gráfico normal de probabilidade, bem como das bandas de confiança construídas a partir da geração de envelopes, Atkinson (1985), concentramos nossa atenção na busca de uma transformação em $d(y_i; \hat{\mu}_i, \lambda)$ de forma a padronizá-lo e com isto obter um novo resíduo cuja distribuição de probabilidade seja próxima da distribuição normal $N(0, 1)$.

5.2.1 Resíduo d_i^*

McCullagh (1987, p. 214) demonstra que, para os modelos lineares generalizados, a distribuição de probabilidade de

$$\frac{d(Y_i; \eta_i) + \rho_{3i}/6}{\sqrt{1 + (14\rho_{3i}^2 - 9\rho_{4i})/36}}$$

é aproximadamente uma normal padrão. Na expressão acima, η_i representa o preditor linear de Y_i , ρ_{3i} e ρ_{4i} são os coeficientes de assimetria e de curtose de $\frac{\partial l_i}{\partial \eta_i}$, respectivamente, onde l_i é a contribuição de Y_i para a verossimilhança total. Assim, empregando

as equações (23), (26) e (27) de Cox e Snell (1968), segue-se que $E\{d(Y_i; \eta_i)\} = 0$ e $\text{Var}\{d(Y_i; \eta_i)\} = 1 - h_{ii}$, onde os termos negligenciados são $O(n^{-1})$, Hinkley et al. (1991, p. 94). Nessa expressão da variância, h_{ii} representa o elemento da diagonal principal da matriz $\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$, onde \mathbf{W} é uma matriz diagonal com os pesos $\frac{\partial^2 l_i}{\partial \eta_i^2}$, calculados na estimativa de máxima verossimilhança dos parâmetros. A matriz \mathbf{H} faz o papel da matriz “hat” nos modelos lineares generalizados.

A partir dessa característica da matriz “hat”, consideramos uma correção para o componente da função desvio, equação (5.1), definindo o resíduo

$$d_i^* = \pm 2\sqrt{\lambda} \frac{\text{sen}\{(y_i - \hat{\mu}_i)/2\}}{(1 - h_{ii}^*)^{1/2}}, \quad (5.2)$$

onde h_{ii}^* representa o elemento da diagonal principal da matriz “hat” \mathbf{H}^* que, no caso do modelo de médias von Mises, fica dada por $\mathbf{H}^* = \dot{\mathbf{G}} \mathbf{X} (\mathbf{X}^T \dot{\mathbf{G}}^2 \mathbf{X})^{-1} \mathbf{X}^T \dot{\mathbf{G}}$ com $\dot{\mathbf{G}} = \text{diag}\{g'(\mathbf{x}_i^T \boldsymbol{\beta})\}$, avaliada na estimativa de máxima verossimilhança $(\hat{\mu}, \hat{\boldsymbol{\beta}}^T, \hat{\lambda})^T$.

A fim de investigarmos a distribuição de probabilidade do resíduo d_i^* , equação (5.2), foi feito um estudo de simulação, de forma semelhante aos procedimentos aplicados por Williams (1984), para os modelos lineares generalizados, que passaremos a descrever e discutir em seguida.

5.2.2 Resíduo r_i

Uma outra transformação que fizemos, buscando uma padronização no componente da função desvio, baseia-se na abordagem empregada por Davison e Gigli (1989), que consiste fundamentalmente em expressar o componente do desvio como função de uma variável aleatória com distribuição conhecida, como por exemplo a normal padrão, e em seguida fazer uma expansão em série de Taylor dessa transformação, até termos de

segunda ordem.

Lembremos inicialmente que a função densidade de probabilidade de uma variável aleatória Y com distribuição von Mises $\mathcal{VM}(\mu, \lambda)$ é dada por

$$f_Y(y; \mu, \lambda) = \frac{1}{2\pi I_0(\lambda)} \exp\{\lambda \cos(y - \mu)\}, \quad 0 \leq y < 2\pi, \quad 0 \leq \mu < 2\pi, \quad \lambda > 0. \quad (5.3)$$

Se fizermos a transformação $y - \mu = z/\lambda^{1/2}$ em (5.3) e usarmos o fato de que $\cos(t) = 1 - t^2/2 + o(t^2)$, quando $t \rightarrow 0$, então temos

$$\begin{aligned} \exp\{\lambda \cos(y - \mu)\} &= \exp\{\lambda \cos(z/\lambda^{1/2})\} \\ &= \exp\left[\lambda \left\{1 - \frac{z^2}{2\lambda} + o(z^2)\right\}\right] \\ &= e^{(\lambda - z^2/2)} e^{o(z^2)}, \quad \text{quando } z \rightarrow 0. \end{aligned} \quad (5.4)$$

Mas, por outro lado,

$$\begin{aligned} e^{o(z^2)} &= 1 + o(z^2) + o\{o(z^2)\} \\ &= 1 + o(z^2). \end{aligned} \quad (5.5)$$

Substituindo (5.4) e (5.5) na equação (5.3), segue-se que

$$f_Y(y; \mu, \lambda) = \frac{e^\lambda}{2\pi I_0(\lambda)} e^{-\frac{z^2}{2}} \{1 + o(z^2)\}, \quad \text{quando } z \rightarrow 0. \quad (5.6)$$

Desta forma, vemos que a função densidade de probabilidade da distribuição von Mises $\mathcal{VM}(\mu, \lambda)$ é aproximadamente proporcional à função densidade de probabilidade da distribuição normal $N(0, 1)$.

Essa relação entre a distribuição von Mises e a distribuição normal pode ainda ser vista através do resultado devido a Stephens (1963), apresentado na Secção 3.5.3. Para valores não muito pequenos de λ , $\sup_{y \in [0, 2\pi)} |F_{\mathcal{VM}}(y) - F_{\mathcal{VN}}(y)| \leq 0,0125$. Além disso, quando $\lambda \rightarrow 0$ e quando $\lambda \rightarrow \infty$, isto é, para valores “extremos” de λ , tem-se

uma correspondência exata entre as distribuições von Mises e normal arqueada, Mardia (1972, p. 66). Como a distribuição normal arqueada é definida pela transformação $Z \pmod{2\pi}$, onde Z é uma variável aleatória com distribuição normal $N(\mu_Z, \sigma_Z^2)$, temos assim justificado um relacionamento entre a distribuição von Mises e a distribuição normal.

Para usarmos a abordagem utilizada por Davison e Gigli (1989), precisaremos conhecer a mediana da distribuição von Mises $\mathcal{VM}(\mu, \lambda)$. Nós sabemos que essa distribuição é simétrica em torno da média direcional μ , quando o gráfico da sua função densidade de probabilidade é considerado sobre o círculo unitário. Entretanto, se fizermos uma “linearização” através de um “corte” no círculo e passarmos a considerar esse gráfico na reta, então a distribuição só poderá ser simétrica se seu suporte for $(\mu - \frac{\pi}{2}, \mu + \frac{\pi}{2})$, ou seja, um intervalo que depende do parâmetro μ .

Esse problema pode ser contornado se fizermos uma transformação da seguinte maneira. Para $Y \sim \mathcal{VM}(\mu, \lambda)$, no intervalo $[-\pi, \pi)$, consideremos inicialmente $W_0 = (Y - \mu) \pmod{2\pi}$, e em seguida

$$W = \begin{cases} W_0 + 2\pi, & \text{se } -2\pi < W_0 < -\pi \\ W_0, & \text{se } -\pi \leq W_0 < \pi \\ W_0 - 2\pi, & \text{se } \pi \leq W_0 < 2\pi. \end{cases} \quad (5.7)$$

Desta forma, segue-se que $W \sim \mathcal{VM}(0, \lambda)$, no intervalo $[-\pi, \pi)$ e portanto a direção mediana de W é $w_m = 0$, ou seja, é igual à sua média direcional.

Seguindo a abordagem de Davison e Gigli (1989), e baseando-se na relação existente entre a distribuição von Mises e a distribuição normal, considerando $F_W(w; \nu, \lambda)$ a função de distribuição acumulada da variável aleatória W , podemos escrever o componente da função desvio $d(W; \nu, \lambda) = d[F_W^{-1}\{\Phi(Z); \nu, \lambda\}; \nu, \lambda] = T(Z; \nu, \lambda)$, onde $Z \sim N(0, 1)$ e T

é uma função desconhecida, possivelmente não linear.

Em particular, para $W \sim \mathcal{VM}(\nu, \lambda)$, a função de distribuição acumulada de W é dada por

$$F_W(w; \nu, \lambda) = \{2\pi I_0(\lambda)\}^{-1} \int_0^w e^{\lambda \cos(t-\nu)} dt.$$

Como não temos uma forma fechada para $F_W(w; \nu, \lambda)$, não é possível obtermos uma expressão analítica para a função T . Entretanto, podemos obter uma expansão em série de Taylor para $T(z; \nu, \lambda)$, em torno de $z = 0$, a saber,

$$T(z; \nu, \lambda) = T(0; \nu, \lambda) + T'(0; \nu, \lambda)z + \frac{T''(0; \nu, \lambda)}{2}z^2 + R_2(z, 0). \quad (5.8)$$

Considerando w_m como sendo a mediana de $F_W(w; \nu, \lambda)$, Davison e Gigli (1989) mostram que

$$\begin{aligned} T(0; \nu, \lambda) &= d(w_m; \nu, \lambda), \\ T'(0; \nu, \lambda) &= \frac{d'(w_m; \nu, \lambda)}{(2\pi)^{1/2} f(w_m)} \text{ onde } f \text{ é a função densidade de probabilidade de } W, \\ T''(0; \nu, \lambda) &= \frac{1}{(2\pi) f^2(w_m)} \left\{ d''(w_m; \nu, \lambda) - \frac{f'(w_m) d'(w_m; \nu, \lambda)}{f(w_m)} \right\}. \end{aligned}$$

O componente da função desvio para a von Mises é,

$$d(w; \nu, \lambda) = 2\sqrt{\lambda} \operatorname{sen}\left(\frac{w - \nu}{2}\right).$$

Assim,

$$\begin{aligned} d'(w; \nu, \lambda) &= \sqrt{\lambda} \cos\left(\frac{w - \nu}{2}\right) e \\ d''(w; \nu, \lambda) &= -\frac{\sqrt{\lambda}}{2} \operatorname{sen}\left(\frac{w - \nu}{2}\right). \end{aligned}$$

Por outro lado, de (5.3),

$$\begin{aligned} f(w) &= \frac{1}{2\pi I_0(\lambda)} e^{\lambda \cos(w)} e \\ f'(w) &= -\frac{\lambda \operatorname{sen}(w)}{2\pi I_0(\lambda)} e^{\lambda \cos(w)}. \end{aligned}$$

Substituindo essas expressões em (5.8), e baseando-se em (5.6), segue-se que $d(w; 0, \lambda) = T(z; 0, \lambda) = (2\lambda\pi)^{1/2} I_0(\lambda) e^{-\lambda} z + o(w^2)$, quando $w \rightarrow 0$. Assim, o componente da função desvio da distribuição von Mises $\mathcal{VM}(0, \lambda)$ pode ser aproximado por uma função linear da distribuição normal $N(0, 1)$.

Agora, para o caso de uma distribuição von Mises com média direcional $\mu \neq 0$, segue-se de (5.7) que

$$\begin{aligned} d(w_i; \nu_i, \lambda) &= \sqrt{2\lambda} \left\{ 1 - \cos \left((y_i - \mu_i) \{ \text{mod}(2\pi) \} \right) \right\}^{1/2} \\ &= \sqrt{2\lambda} \left\{ 1 - \cos(y_i - \mu_i) \right\}^{1/2} \\ &= d(y_i; \mu_i, \lambda). \end{aligned}$$

Ou seja, o componente da função desvio de uma von Mises $\mathcal{VM}(\mu, \lambda)$, com $\mu \neq 0$, segue a mesma distribuição que o componente da função desvio de uma von Mises $\mathcal{VM}(0, \lambda)$.

Com isso, podemos definir um segundo resíduo padronizado que é dado por

$$r_i = \pm (2/\pi)^{1/2} \frac{\text{sen} \{ (y_i - \hat{\mu}_i)/2 \}}{I_0(\lambda) e^{-\lambda}}, \quad (5.9)$$

cuja distribuição de probabilidade pretendemos comparar com a distribuição normal padrão $N(0, 1)$.

5.3 Simulação

Com o objetivo de investigar o comportamento da distribuição de probabilidade dos resíduos d_i^* e r_i , definidos pelas equações (5.2) e (5.9), respectivamente, foi feito um estudo de simulação, que passaremos a descrever e comentar os resultados obtidos.

Nesse estudo de simulação foram geradas amostras de tamanhos $n = 25, 50$ e 100

observações da distribuição von Mises $\mathcal{VM}(\mu_i, \lambda)$. O parâmetro μ_i foi considerado como no modelo de médias von Mises, com apenas uma covariável x e assumindo-se a ligação $\mu_i = \mu + 2 \arctan(x\beta)$. Para a covariável x foram gerados números pseudo-aleatórios no intervalo $(0, 6)$, sendo que, para cada amostra y_1, \dots, y_n a ser simulada, eram gerados novos valores de x e em seguida centralizados, a fim de acelerar a convergência do algoritmo de estimação. Os valores considerados para o parâmetro de concentração λ foram 1, 2, 4 e 7.

Para gerar a amostra y_1, \dots, y_n foi feito um programa em S-Plus a fim de implementar o algoritmo proposto por Best e Fisher (1979). Esse algoritmo simula cada observação von Mises $\mathcal{VM}(\mu, \lambda)$ a partir de 3 uniformes no intervalo $[0, 1]$ e consiste em:

Passo 1. Definir $a = 1 + \sqrt{(1 + 4\lambda^2)}$, $b = (a - \sqrt{2a})/2\lambda$ e $r = (1 + b^2)/(2b)$;

Passo 2. Considerar U_1, U_2 e U_3 números pseudo-aleatórios no intervalo $[0, 1]$;

Passo 3. Calcular $z = \cos(\pi U_1)$, $f = (1 + rz)/(r + z)$ e $c = \lambda(r - f)$;

Passo 4. Se $c(2 - c) - U_2 > 0$ ir para o Passo 6;

Passo 5. Se $\log(c/U_2) + 1 - c < 0$ considerar novos números U_1, U_2 e U_3 , pseudo-aleatórios no intervalo $[0, 1]$, e voltar para o Passo 3;

Passo 6. $y = \text{sinal}(U_3 - 0,5) \arccos(f) + \mu \{\text{mod}(2\pi)\}$.

Para cada par (n, λ) , ou seja, para cada combinação dos valores do tamanho da amostra n e do parâmetro de concentração λ , foram geradas 1000 amostras de tamanho n da distribuição von Mises $\mathcal{VM}(\mu + 2 \arctan(x\beta), \lambda)$. Em cada amostra gerada, foi aplicado o processo de estimação descrito pela equação (4.5) a fim de obter uma solução do sistema de equações (4.1) - (4.4). Uma vez obtidas as estimativas dos parâmetros $(\hat{\mu}, \hat{\beta}^T, \hat{\lambda})^T$, foram calculados os componentes da função desvio d_1, \dots, d_n , pela equação

(5.1), e em seguida esses resíduos foram padronizados de acordo com (5.2), obtendo-se então d_1^*, \dots, d_n^* . Analogamente, obtivemos r_1, \dots, r_n , utilizando (5.9). De acordo com o que mostramos, d_1^*, \dots, d_n^* e r_1, \dots, r_n devem ser vistas como duas amostras da distribuição normal $N(0, 1)$.

Em seguida calculamos a média amostral, o desvio padrão amostral e os coeficientes de assimetria e curtose de cada uma dessas amostras de n resíduos, tanto para d_i^* , como para r_i . O passo seguinte foi calcular a média aritmética dessas 4 estatísticas, referente a todas as 1000 amostras geradas, com o objetivo de compará-las com as medidas correspondentes à distribuição normal padrão. Os resultados estão apresentados na Tabela 5.1. Podemos ver que tanto a média quanto o coeficiente de assimetria não diferem substancialmente do valor correspondente da distribuição normal padrão. Na maioria dos casos o desvio padrão apresenta-se ligeiramente inferior a um, sugerindo que as distribuições desses dois resíduos são mais concentradas em torno de suas médias do que a distribuição normal $N(0, 1)$. Por sua vez, o coeficiente de curtose é sistematicamente negativo, indicando que as distribuições dos resíduos r_i e d_i^* são mais achatadas do que a distribuição normal. De uma forma geral, os dois resíduos mostraram comportamentos bem similares entre si, e a concordância de suas distribuições com a distribuição normal padrão é mais acentuada para valores grandes do parâmetro de concentração λ .

Tabela 5.1: Medidas de locação e escala para os resíduos padronizados r_i e d_i^* .

Parâmetro de Concentração	Medida	Resíduo r_i			Resíduo d_i^*		
		$n = 25$	$n = 50$	$n = 100$	$n = 25$	$n = 50$	$n = 100$
$\lambda = 1$	média	0,0094	0,0052	0,0014	0,0113	0,0063	0,0017
	d.padrão	0,8796	0,8316	0,8020	1,0285	0,9639	0,8213
	assimetria	-0,0009	0,0024	0,0010	-0,0006	0,0024	0,0010
	curtose	-1,1783	-1,1904	-1,2049	-1,1828	-1,1931	-1,2065
$\lambda = 2$	média	-0,0152	-0,0080	0,0072	-0,0179	-0,0099	0,0069
	d.padrão	0,7445	0,6830	0,6480	0,8606	0,7524	0,6692
	assimetria	0,0185	-0,0070	-0,0117	0,0191	0,0072	-0,0118
	curtose	-1,1689	-1,1581	-1,1933	-1,1620	-1,1546	-1,1908
$\lambda = 4$	média	0,0050	0,0008	-0,0047	0,0062	0,0010	-0,0057
	d.padrão	0,9678	0,9121	0,8800	1,1087	1,0439	1,0064
	assimetria	-0,0034	0,0060	0,0035	-0,0034	0,0061	0,0036
	curtose	-1,2316	-1,2330	-1,2553	-1,2371	-1,2376	-1,2574
$\lambda = 7$	média	0,0006	0,0006	-0,0059	0,0012	0,0007	-0,0072
	d.padrão	1,0335	0,9625	0,9084	1,1670	1,0968	1,0479
	assimetria	-0,0031	0,0056	0,0009	-0,0032	0,0056	0,0010
	curtose	-1,1460	-1,1072	-1,0667	-1,1602	-1,1173	-1,0729

Capítulo 6

Aplicações

6.1 Aplicação 1

A primeira aplicação que faremos consiste de uma análise dos dados obtidos pelo Prof. Dr. R. Ranvaud, do Instituto de Ciências Biológicas da USP, tendo sido analisado, numa outra abordagem, em Artes (1997) e em Artes, Paula e Ranvaud (1998a, 1998b).

Descrição do conjunto de dados

O conjunto de dados corresponde à direção seguida por pombos-correio, medida após 30 segundos, 60 segundos e 90 segundos da soltura como também a direção do pombo no momento do seu desaparecimento no horizonte. Outra informação coletada foi o tempo decorrido até o desaparecimento de cada pombo.

Um dos objetivos do estudo era investigar o mecanismo de orientação dos pombos-

correio que, segundo uma teoria existente, é influenciada tanto pelo campo magnético da Terra quanto pelo Sol. Os dados foram coletados na cidade de Camocim (CE), localizada a cerca de 300 km da capital Fortaleza, e fica próxima à linha do equador geográfico. As aves foram soltas em dois pontos distintos, (CMC-1) e (CMC-2), que foram escolhidos em função de seus relevos, favorecendo a observação dos pombos. Tendo em vista a influência da luz solar, a soltura foi feita em três períodos do dia, caracterizados como MANHÃ, MEIO-DIA e TARDE. Essa escolha do local possibilitou um controle experimental com relação à posição do Sol (proximidade com o equador geográfico), como também com relação ao campo magnético, uma vez que o equador magnético da Terra estava próximo à cidade de Camocim na época em que o experimento foi realizado.

Definindo y_{ijk_r} como sendo a diferença entre a direção do pombo i após j segundos da soltura e a direção de desaparecimento, na localidade r no período k do dia, Artes (1997) adotou um modelo, que o denominou *modelo completo*, para as médias circulares μ_{ijk_r} e parâmetros de concentração λ_{ijk_r} . Nesse modelo completo foi explorada a estrutura longitudinal dos dados, e após o ajuste do mesmo, bem como a realização de testes de hipóteses, foi possível fazer o ajuste de um modelo mais reduzido.

No modelo reduzido, a média circular ficou constante e não foi detectada diferença entre os parâmetros de concentração referentes aos pombos soltos no período da manhã e os que foram soltos à tarde. A média circular referente a esses dois períodos foi estimada em $5,07^\circ$.

Nosso objetivo aqui é analisar a dependência entre a direção de cada pombo no momento do seu desaparecimento e o tempo decorrido desde a soltura do mesmo. A Figura 6.1 contém um gráfico de dispersão de todos os pombos. Cada ponto nesse gráfico caracteriza a “direção escolhida” e o tempo decorrido até o desaparecimento. Tendo em vista

a disposição dos pontos nesse gráfico, aplicamos o teste de Rayleigh, ver Mardia (1972, p. 133), para testar a hipótese da distribuição dos dados ser uma uniforme circular, ou seja, testar se a direção escolhida pelos pombos ocorria aleatoriamente no círculo. Com os 176 pombos, observamos $\bar{R} = 0,3241$, o que dá $2n\bar{R}^2 = 36,9972$, comparando-se com a distribuição assintótica de referência, χ_2^2 , a hipótese de uniformidade é rejeitada. Nas Figuras 6.2 e 6.3 representamos essas mesmas direções, destacando o local (CMC-1

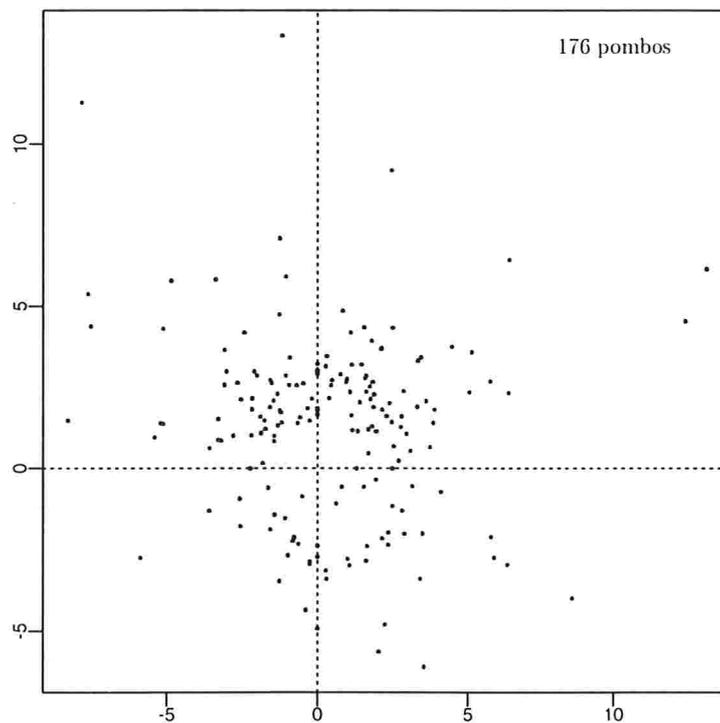


Figura 6.1: Localização dos pombos no desaparecimento.

e CMC-2) e o período de soltura (MANHÃ, MEIO-DIA e TARDE). Podemos observar que a posição de cada pombo no momento do seu desaparecimento parece depender de outros fatores, além do tempo de voo até a ocorrência desse evento. Tendo em vista que no momento o nosso objetivo não é explorar a estrutura longitudinal desse conjunto de

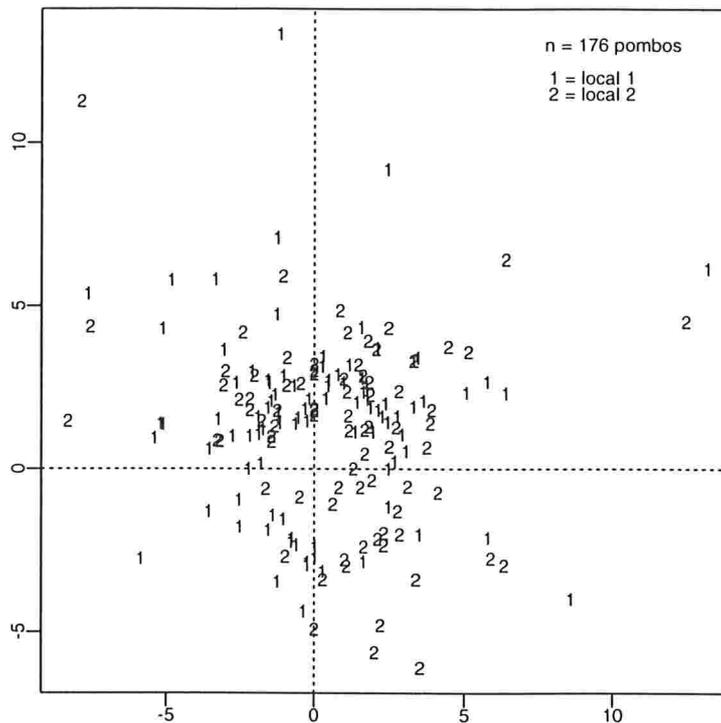


Figura 6.2: Localização dos pombos: Local de soltura.

dados e considerando as conclusões obtidas por Artes (1997), de não ter sido detectada diferença entre os parâmetros de concentração referentes aos pombos soltos no período da manhã e os que foram soltos à tarde, fizemos uma aplicação utilizando apenas o subconjunto dos pombos correspondentes a esses dois períodos de soltura, consistindo de 112 pombos. Como variável resposta, consideramos y_i como sendo a distância angular entre os valores observados aos 30 segundos, e no momento do desaparecimento, do i -ésimo pombo.

A Figura 6.4 apresenta um gráfico sugerido por Fisher (1993). Nesse gráfico, o eixo das abscissas corresponde ao *tempo* de vôo até o desaparecimento, e no eixo das ordenadas estamos considerando y_i e $y_i + 2\pi$, que ajuda a visualizar comportamentos sistemáticos em observações circulares. Analisando esse gráfico, não se percebe nenhuma tendência

da direção média. Por outro lado, observa-se uma variabilidade maior para os pombos que demoraram pouco tempo para o desaparecimento do que para os pombos que demoraram mais tempo. Isto sugere um modelo de dispersão. Considerando $y_i \in [-\frac{\pi}{2}, \frac{\pi}{2})$, foi assumido inicialmente que:

1. y_i segue uma distribuição von Mises $\mathcal{VM}(\mu_i, \lambda_i)$, para $i = 1, \dots, 112$;
2. A direção de referência, ou média direcional, de y_i foi inicialmente modelada por $\mu_i = \mu + 2 \arctan(t_i \beta)$, onde t_i representa o tempo decorrido até o desaparecimento do pombo i ;
3. O parâmetro de concentração foi modelado por $\lambda_i = \exp\{\gamma_0 + t_i \gamma_1\}$.

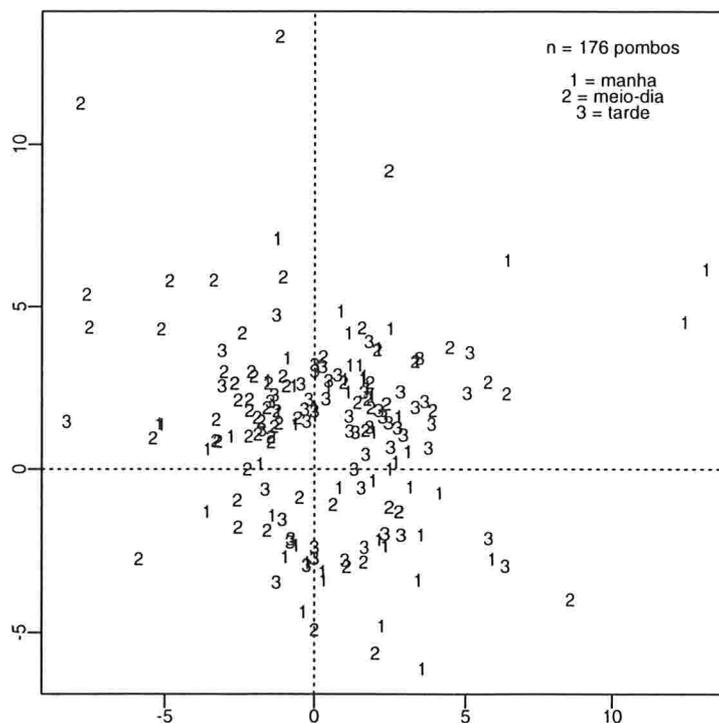


Figura 6.3: Localização dos pombos: Período da soltura.

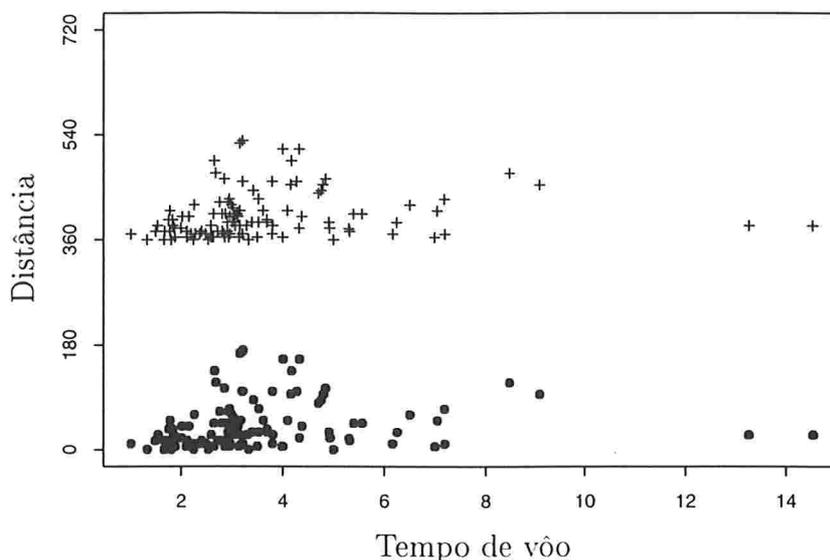


Figura 6.4: Distância entre as direções dos pombos (MANHÃ e TARDE).

Isto significa que estamos partindo de um modelo von Mises misto. Aplicando-se o procedimento iterativo descrito na página 52, para estimar os parâmetros desse modelo, o ajuste obtido apresentou um desvio de $D = 124,6$ com 108 graus de liberdade, indicando que o ajuste é aceitável¹. As estimativas para os parâmetros com respectivos erros padrão apresentados na Tabela 6.1. Analisando os níveis descritivos, há indícios de que o modelo se ajusta bem aos dados. Em seguida fizemos uma análise gráfica de resíduos. Na Figura 6.5 apresentamos os gráficos normal de probabilidade para os resíduos padronizados r e d^* , respectivamente. Em cada gráfico apresentamos bandas de confiança contruídas através da técnica de envelopes. Os limites de cada envelope foram determinados a partir de 19 simulações, conforme sugere Atkinson (1985). Desta forma, a probabilidade do resíduo observado r_i (ou d_i^*) exceder o limite superior do envelope é aproximadamente $1/20$. Esses gráficos mostram que algumas pressuposições do modelo foram violadas. Além de um possível problema de superdispersão, a causa mais provável

¹Estamos usando o resultado: $D(y; \hat{\mu}) \stackrel{\text{modelo}}{\sim} \chi_{n-p}^2$, quando $\lambda \rightarrow \infty$, Jørgensen (1987).

Tabela 6.1: Estimativas dos parâmetros do modelo von Mises misto.

Parâmetro	Estimativa	D. padrão (aprox.)	Estim/D. padrão	Nível descritivo
μ	-1,9900	0,0548	-36,3240	$< 0,0001$
β	1,3111	0,1261	10,3955	$< 0,0001$
γ_0	3,0352	0,5954	5,0974	$< 0,0001$
γ_1	-1,0123	0,3221	-3,1423	$< 0,0001$

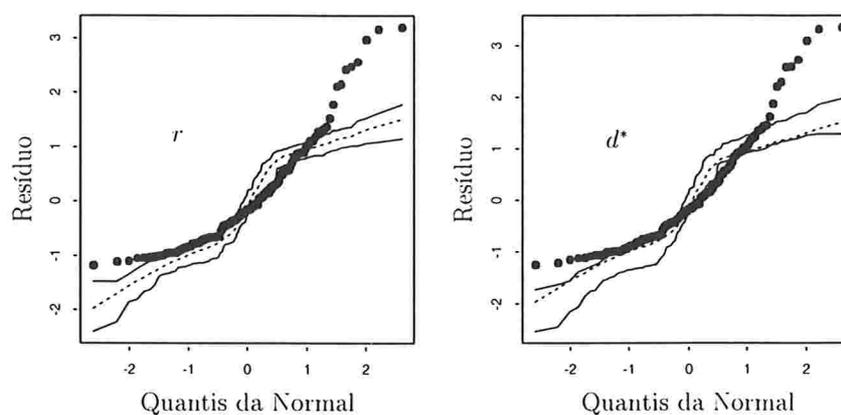


Figura 6.5: Gráfico dos envelopes para os resíduos r e d^* .

dessa violação pode ter sido a distribuição assumida para a variável resposta, uma vez que já que estamos modelando o parâmetro de concentração. Ressaltamos que a distância entre duas observações direcionais y e y^* , é dada por $\min\{\theta_1, \theta_2\}$, onde $\theta_1 = |y - y^*| \bmod\{2\pi\}$ e $\theta_2 = |y + 2\pi - y^*| \bmod\{2\pi\}$, para $y < y^*$.

Redefinimos a variável resposta e passamos a considerar simplesmente a diferença da direção no momento do desaparecimento de cada pombo, em relação à sua posição aos 30 segundos de voo. O gráfico da Figura 6.6 mostra um diagrama de dispersão desses dados, e na Figura 6.7 temos plotado simultaneamente os pontos y e $y + 2\pi$ versus

a variável *tempo*. Nota-se uma grande variabilidade entre os pombos que demoraram pouco tempo para sumir no horizonte, em contrapartida aos pombos que ficaram visíveis por um período maior. Ajustamos inicialmente um modelo von Mises de médias. O

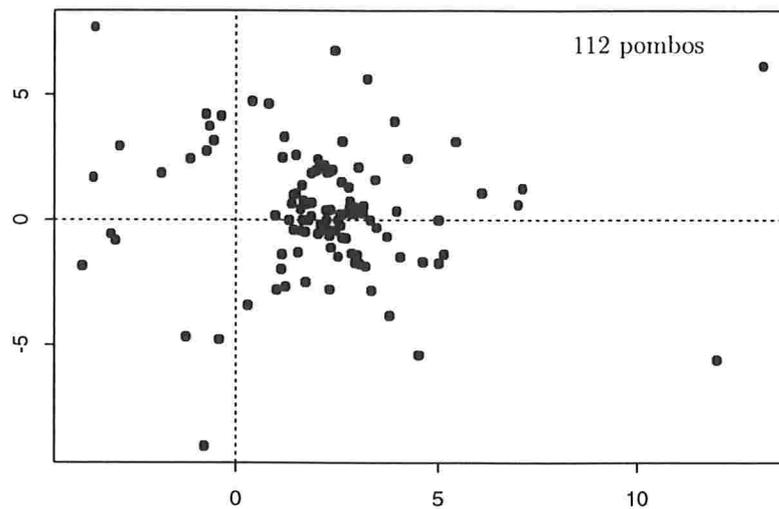


Figura 6.6: Gráfico de dispersão da diferença entre as direções dos pombos.

ajuste forneceu um desvio $D = 135,755$ com 109 graus de liberdade, e as estimativas dos parâmetros estão na Tabela 6.2. Os níveis descritivos sugerem que o parâmetro β pode ser excluído do modelo, corroborando assim a indicação da Figura 6.7, de que a média direcional de y_i não depende do tempo de vôo.

Análise dos Resíduos

Na Figura 6.8 apresentamos os gráficos normal de probabilidade para os resíduos padronizados r e d^* . Em cada gráfico apresentamos bandas de confiança contruídas através

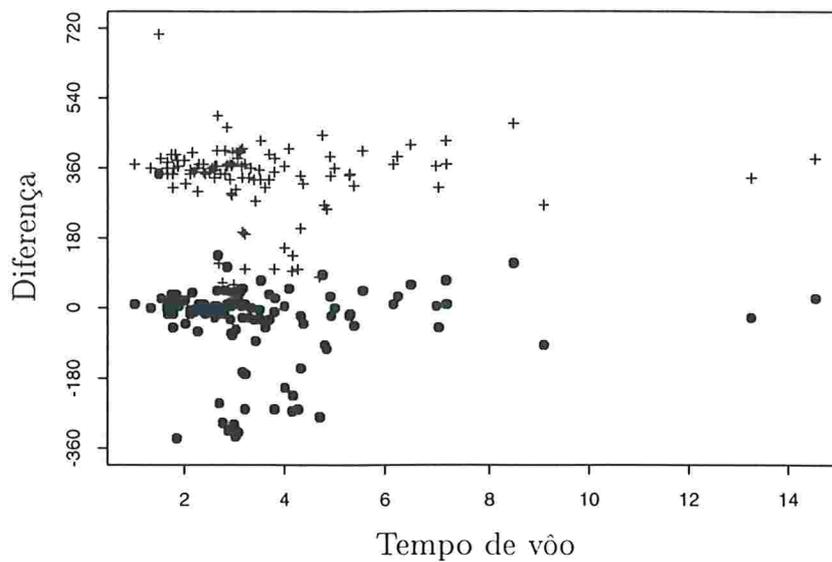


Figura 6.7: Diferença entre as direções dos pombos.

da técnica de envelopes, cujos limites foram obtidos a partir de 19 simulações, Atkinson (1985). Os dois gráficos mostram indícios de superdispersão. Com isto, passamos a

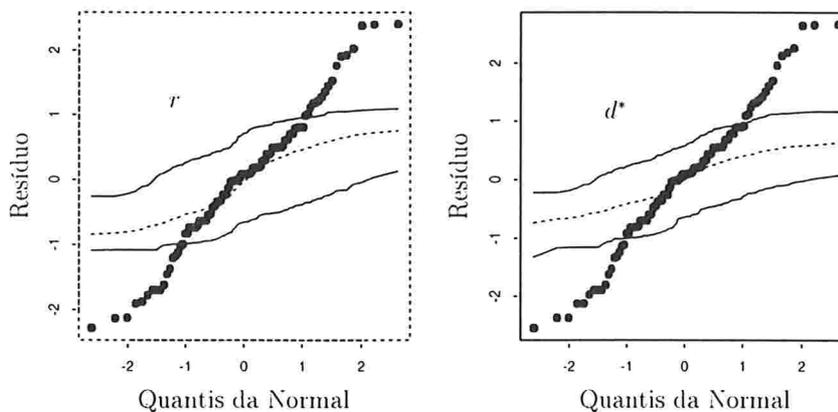


Figura 6.8: Gráfico dos envelopes para os resíduos r e d^* .

considerar um modelo de dispersão, em que todas as médias direcionais são iguais entre si, e os parâmetros de concentração são modelados por $\lambda_i = \exp\{\gamma_0 + t_i\gamma_1\}$. O ajuste desse modelo apresentou um desvio $D = 113,1811$, com 109 graus de liberdade. As estimativas para os parâmetros μ , γ_0 e γ_1 , com respectivos erros padrão estão apre-

Tabela 6.2: Estimativas dos parâmetros do modelo von Mises de médias.

Parâmetro	Estimativa	D. padrão (aprox.)	Estim./D. padrão	Nível descritivo
μ	-0,1394	0,0882	-1,5811	0,1139
β	-0,0038	0,0215	-0,1749	0,8612
λ	1,7649	0,2130	8,2867	< 0,0001

sentados na Tabela 6.3. Observando-se os níveis descritivos, optamos por assumir esse modelo de dispersão.

A média circular estimada pelo modelo, corresponde a uma diferença de aproximadamente -7° entre a direção inicial (aos 30 segundos) e a direção final (no desaparecimento). O parâmetro de concentração para o i -ésimo pombo foi estimado por $\lambda_i = \exp\{5,38 - 2,72t_i\}$.

Tabela 6.3: Estimativas dos parâmetros do modelo von Mises de dispersão.

Parâmetro	Estimativa	D. padrão (aprox.)	Estim./D. padrão	Nível descritivo
μ	-0,1149	0,0685	-1,6770	0,0935
γ_0	5,3761	0,9921	5,4190	< 0,0001
γ_1	-2,7195	0,5944	-4,5752	< 0,0001

Análise dos Resíduos

Nas Figuras 6.9 e 6.10 apresentamos os gráficos normal de probabilidade para os resíduos padronizados r e d^* , respectivamente. Em cada gráfico apresentamos bandas de confiança contruídas através da técnica de envelopes, cujos limites foram obtidos a partir de 19 simulações, Atkinson (1985). Esses gráficos não indicam afastamentos das suposições iniciais para o ajuste desse modelo de dispersão.

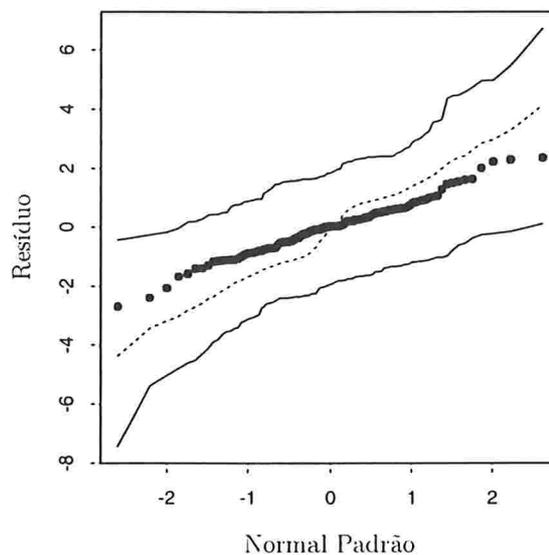


Figura 6.9: Gráfico dos envelopes para o resíduo r .

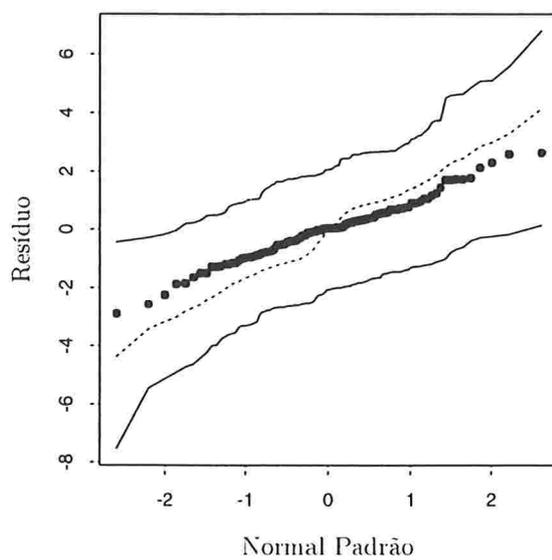


Figura 6.10: Gráfico dos envelopes para o resíduo d^* .

Influência Local

Assumindo o modelo von Mises de dispersão, temos na Figura 6.11 um gráfico mostrando a influência das observações sobre as estimativas dos parâmetros e na Figura 6.12 temos esses mesmos pontos plotados contra a variável *tempo* de voo. Foi empregado aqui o esquema de perturbação que consiste na ponderação dos casos e vemos que 2 pontos se destacam, que correspondem aos pombos 17 e 158, e outros 3 deles apresentam influência moderada, que são os pombos 61, 113 e 170. Esses pombos estão entre os que desapareceram mais rapidamente. Isso mostra que as estimativas dos parâmetros de concentração foram influenciadas pelos pombos “mais decididos”, ou seja, os pombos com os *menores valores na covariável tempo de voo*.

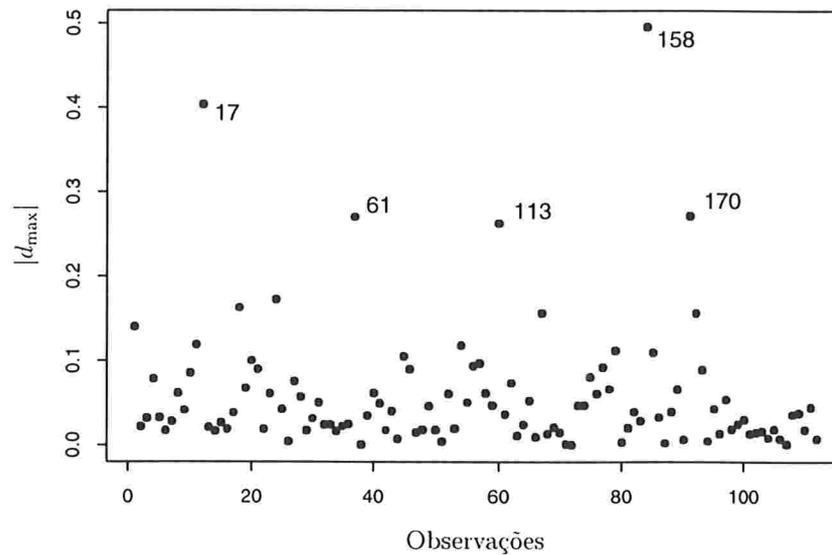


Figura 6.11: Influência das observações sobre as estimativas dos parâmetros.

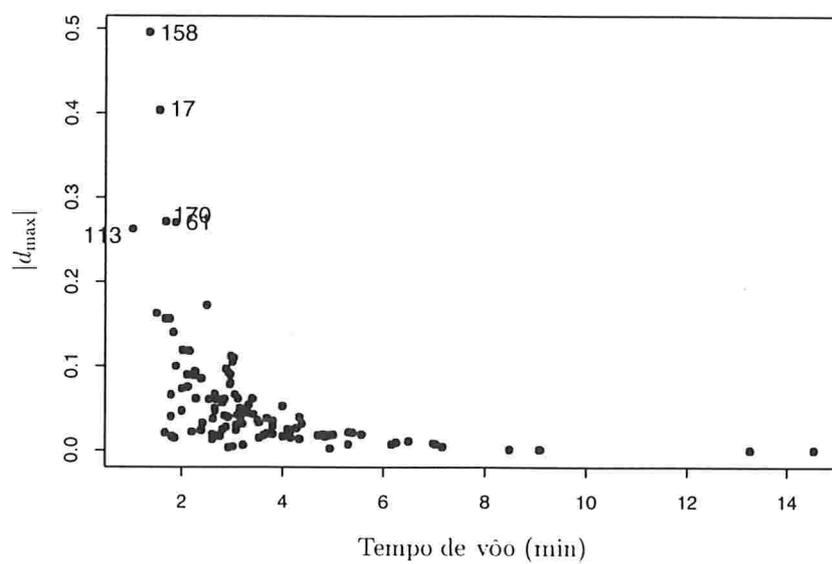


Figura 6.12: Influência das observações sobre as estimativas de γ .

6.2 Aplicação 2

Uma segunda aplicação foi feita considerando o conjunto de dados descrito em Fisher e Lee (1992, p. 136), representando a direção tomada e a distância percorrida por 31 caramujos após os mesmos terem sido transportados do seu *habitat* natural.

Tabela 6.4: Medidas de distância percorrida e direção escolhida por 31 caramujos (*Nodilittorina unifasciata*) após o deslocamento do seu *habitat* natural.

Obs.	x	y									
1	107	67	9	12	171	17	21	165	25	38	83
2	46	66	10	25	166	18	1	133	26	70	71
3	33	74	11	37	98	19	71	101	27	7	74
4	67	61	12	69	60	20	60	105	28	48	91
5	122	58	13	5	197	21	71	71	29	7	38
6	69	60	14	83	98	22	71	84	30	21	200
7	43	100	15	68	86	23	57	75	31	27	56
8	30	89	16	38	123	24	53	98			

x = distância percorrida (em *cm*) e y = direção escolhida (em graus).

Foi assumido que a direção escolhida por cada caramujo segue uma distribuição de probabilidade von Mises $\mathcal{VM}(\mu_i, \lambda_i)$. A média direcional foi modelada por $\mu_i = \mu + 2 \arctan(\beta x_i)$, onde x_i representa a distância (em *cm*) percorrida pelo respectivo caramujo. Inicialmente considerou-se que os parâmetros de concentração são todos iguais entre si, $\lambda_1 = \dots = \lambda_{31}$. Os ângulos originais foram convertidos em radiano e transformados para o intervalo $[-\pi, \pi)$, e as distâncias x_1, \dots, x_{31} foram centralizadas.

Após aplicar o processo iterativo descrito pela equação (4.5) a fim de resolver o sistema de equações (4.1) - (4.4), obtivemos as seguintes estimativas de máxima verossimilhança, com respectivos desvios padrão assintóticos, $\hat{\mu} = 97^\circ$ ($\hat{\sigma}_{\hat{\mu}} = 0,853$), $\hat{\beta} = -0,0065$ (0,0004) e $\hat{\lambda} = 3,187$ (0,133). O desvio total obtido foi de $D(y; \hat{\mu}) = 34,796$ com 29 graus de liberdade. Fazendo-se uma comparação com a distribuição de referência χ_{n-p}^2 , não há indícios de que o modelo ajustado seja inadequado ².

Foi feita em seguida uma análise de resíduos através do gráfico de envelopes, considerando ambos os resíduos definidos em (5.2) e (5.9). Cada envelope foi obtido a partir de 19 simulações, conforme sugere Atkinson (1985). Os gráficos estão nas Figura 6.13 e 6.14. A disposição dos resíduos indica que o ajuste do modelo não está bom, sendo que alguns resíduos apontam para pontos mal ajustados.

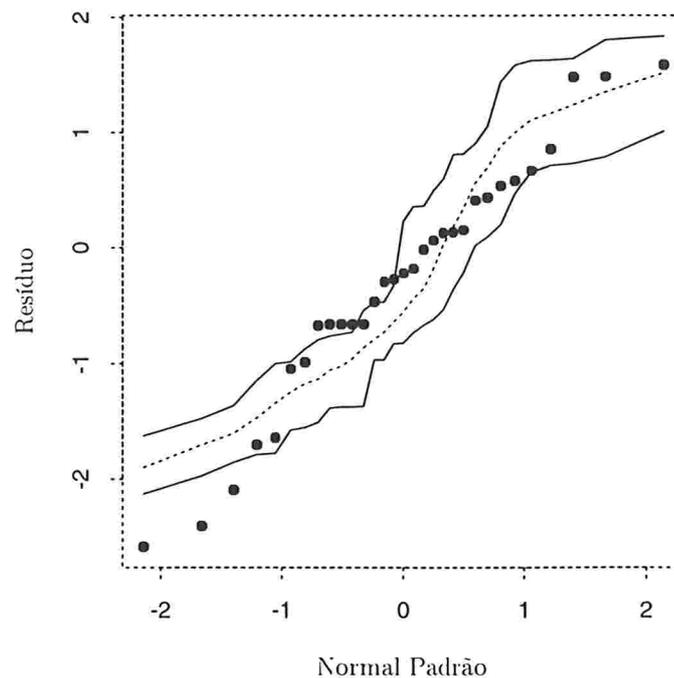


Figura 6.13: Gráfico dos envelopes para o resíduo d^* (modelo de médias).

²Nós usamos aqui o resultado de Jørgensen (1987) que $D(y; \hat{\mu}) \stackrel{\text{modelo}}{\sim} \chi_{n-p}^2$, quando $\lambda \rightarrow \infty$.

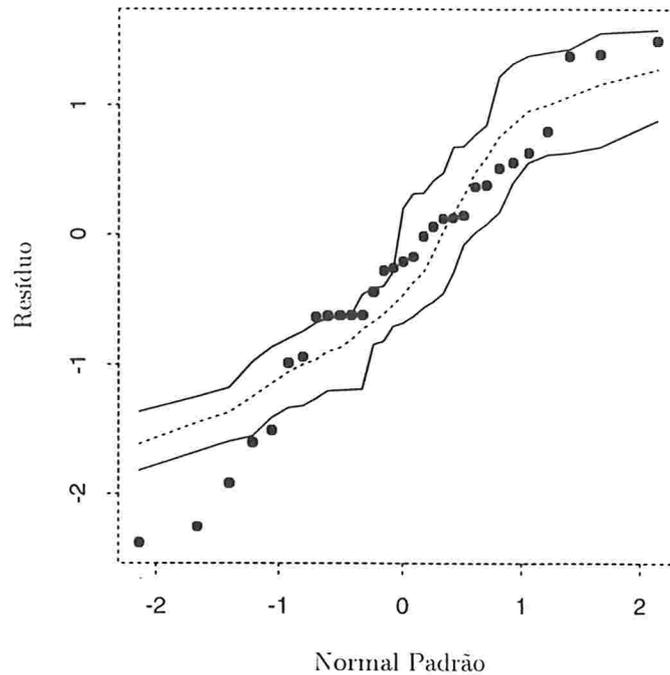
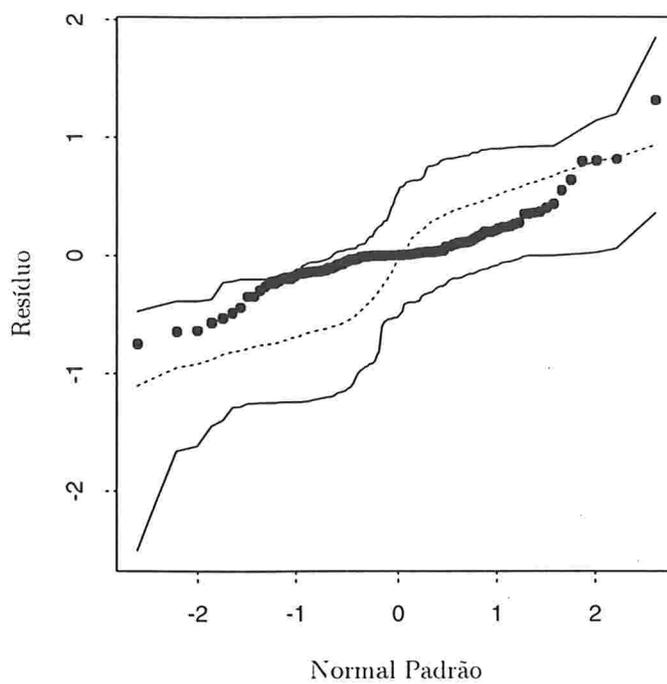
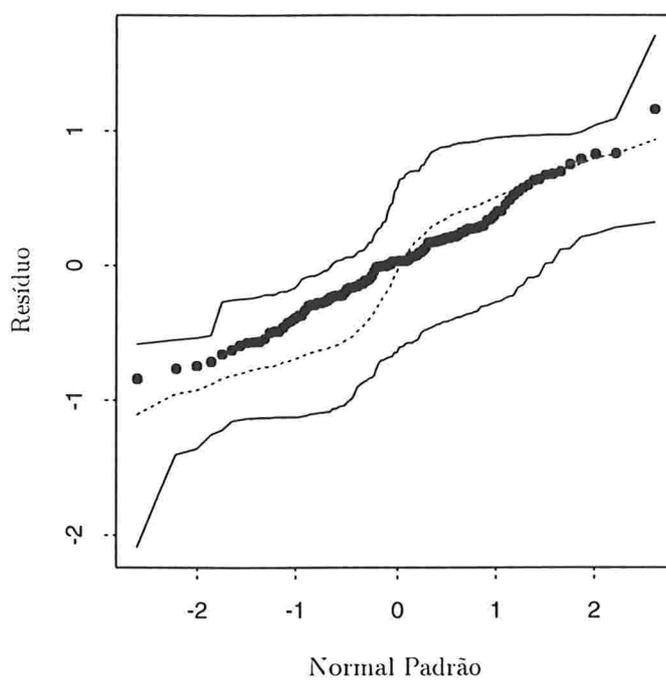


Figura 6.14: Gráfico dos envelopes para o resíduo r (modelo de médias).

Segundo Fisher (1993), não é incomum encontrarmos conjuntos de dados onde tanto a direção média quanto o parâmetro de concentração têm a influência de covariáveis. Para esse conjunto de dados, uma exploração gráfica sugere que a variabilidade parece ser maior para os animais que percorreram uma pequena distância, vide Fisher e Lee (1992, Figura 2(b)). De acordo com Paula (1996), para esse conjunto de dados, parece ser mais apropriado procurar ajustar um modelo misto $\mathcal{VM}(\mu_i, \lambda_i)$ com $\mu_i = \mu + 2 \arctan(\beta x_i)$ e $\lambda_i = \exp\{\gamma_0 + \gamma_1 x_i\}$.

Empregamos então o processo iterativo descrito pela equação (4.11) a fim de resolver o sistema de equações (4.7) - (4.10) e obtivemos as seguintes estimativas de máxima verossimilhança, com respectivos desvios padrão assintóticos, $\hat{\mu} = 117,1^\circ$ ($\hat{\sigma}_{\hat{\mu}} = 0,0458$), $\hat{\beta} = -0,009(0,0025)$, $\hat{\gamma}_0 = 1,78(0,25)$ e $\hat{\gamma}_1 = 0,045(0,009)$. As Figuras (6.15) e (6.16) apresentam gráficos dos envelopes para os resíduos d^* e r , respectivamente. Os dois gráficos indicam que o modelo misto foi ajustado de forma satisfatória

Figura 6.15: Gráfico dos envelopes para o resíduo d^* (modelo misto).Figura 6.16: Gráfico dos envelopes para o resíduo r (modelo misto).

Influência Local

Perturbando o modelo com o esquema de ponderação dos casos, apresentamos na Figura 6.17 um gráfico de $|d_{\max}|$ contra a covariável x , mostrando a influência local das observações sobre as estimativas dos parâmetros $\hat{\mu}$, $\hat{\beta}$ e $\hat{\gamma}$. A Figura 6.18 mostra a influência local das observações apenas sobre as estimativas de γ_0 e γ_1 . Podemos ver que as estimativas tiveram influência dos caramujos 14 e 19 e que os caramujos 1, 5 e 14 foram os que exerceram maior influência sobre as estimativas de γ_0 e γ_1 , que são os parâmetros referentes à concentração (ou dispersão). Observando as distâncias percorridas pelos mesmos, vemos que as estimativas dos parâmetros foram influenciadas por aqueles caramujos que se deslocaram mais do que os outros.

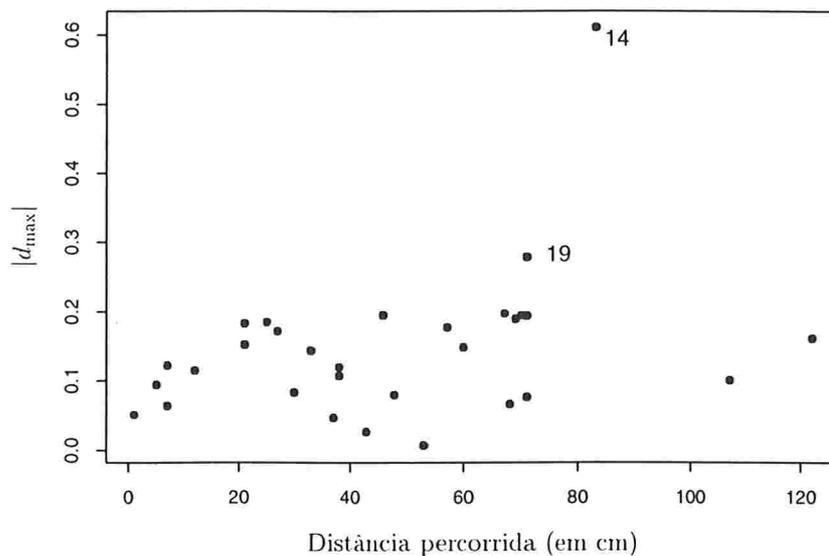


Figura 6.17: Influência dos casos sobre as estimativas dos parâmetros.

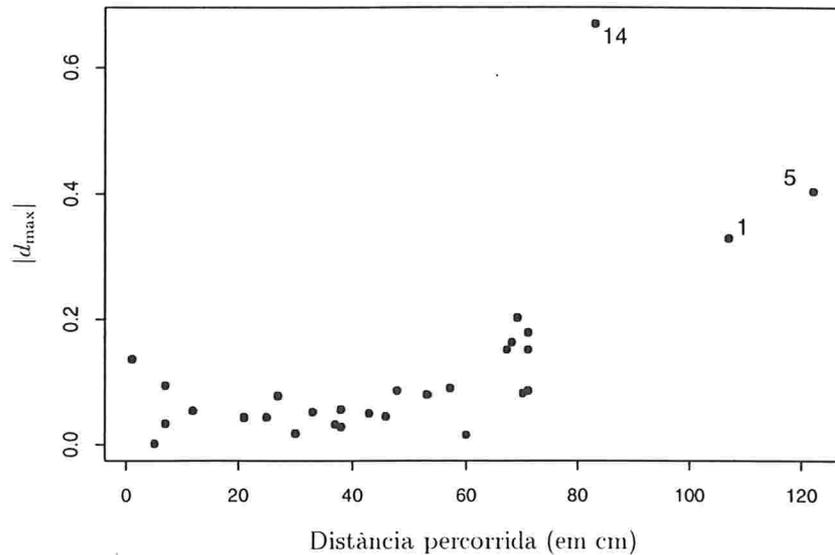


Figura 6.18: Influência das observações sobre as estimativas de γ .

Fazendo o diagnóstico de influência sobre as predições do modelo, tratado na Secção 4.4 e que usa a função objetivo proposta por Thomas e Cook (1990), apresentamos na Figura 6.19 o gráfico de influência das observações sobre as predições, onde foi empregada a perturbação de ponderação dos casos. Fazendo uma perturbação aditiva na variável resposta y , temos na Figura 6.20 um gráfico mostrando a influência dos valores observados para esta variável sobre as próprias predições. Nesses dois gráficos podemos observar que a influência maior sobre as predições é exercida também pelos caramujos 1, 5 e 14, ou seja, os caramujos com os *maiores valores na covariável distância percorrida*.

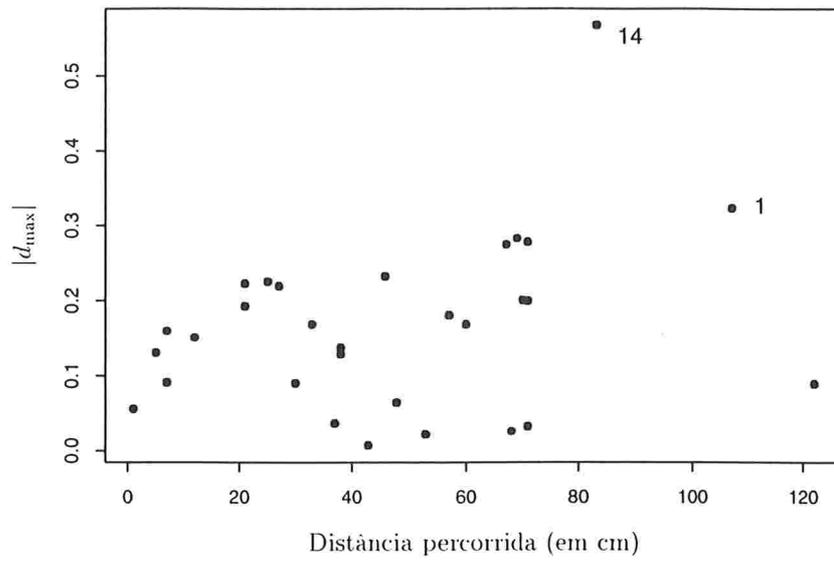


Figura 6.19: Influência dos casos sobre as predições.

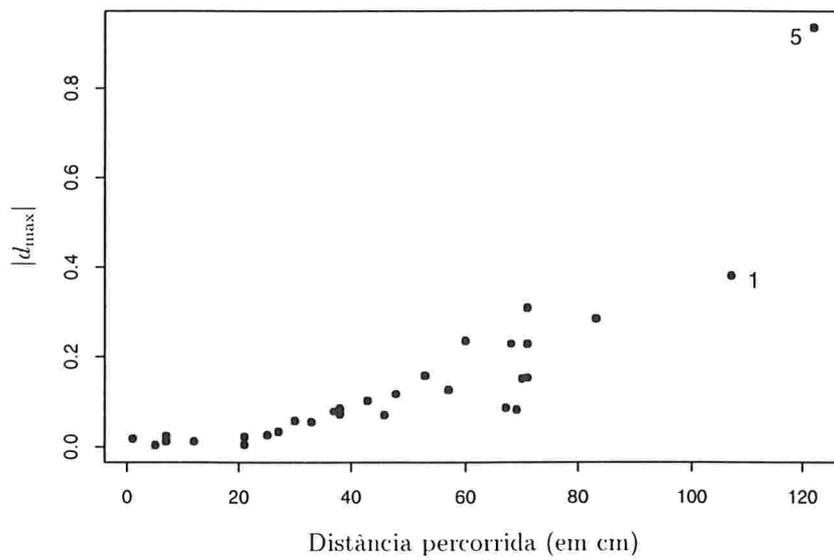


Figura 6.20: Influência das respostas sobre as predições.

Nessas duas aplicações, observamos uma inversão na influência exercida pelas observações, no que diz respeito aos valores para as respectivas covariáveis. Na Aplicação 1 os pombos influentes foram aqueles mais negativamente afastados, no espaço da covariável que representa o tempo de vôo. Por outro lado, na Aplicação 2 os caramujos influentes foram aqueles mais positivamente afastados no espaço da covariável que indica a distância percorrida.

Capítulo 7

Considerações Finais

7.1 Conclusões

O objetivo inicial deste trabalho visava um estudo de diagnóstico em modelos de regressão onde a variável resposta representa uma *direção* e está definida na circunferência do círculo unitário. Discutimos no Capítulo 2 a abordagem proposta por Cook (1986) para diagnóstico de influência local, que se apresenta como técnica moderna e que pode ser empregada em diversas situações onde se disponha da verossimilhança.

Alguns dos conceitos e resultados mais importantes sobre dados circulares foram apresentados no Capítulo 3, e atenção especial foi dada à distribuição von Mises. No Capítulo 4 apresentamos uma classe de modelos de regressão para respostas circulares y_1, \dots, y_n , a saber:

y_1, \dots, y_n são independentes;

$$y_i \sim \mathcal{VM}(\mu_i, \lambda_i);$$

$$\mu_i = \mu + g(\mathbf{x}_i^T \boldsymbol{\beta}) \text{ e } \lambda_i = h(\mathbf{z}_i^T \boldsymbol{\gamma}).$$

Aplicamos a metodologia de influência local a esses modelos, estando os resultados apresentados ainda no Capítulo 4.

Foi dada essa ênfase aos modelos de regressão von Mises tendo em vista que:

- i. A importância da distribuição von Mises para dados circulares é comparável à importância da distribuição normal para dados lineares;
- ii. Não se dispõe atualmente de uma família de locação/escala para dados circulares que possibilite uma transformação com vistas a se obter uma dispersão unitária. Entretanto o inverso do parâmetro de concentração da distribuição von Mises $\mathcal{VM}(\mu, \lambda)$ pode ser tratado, embora que de forma aproximada, como uma variância;
- iii. Os modelos de regressão von Mises propostos por Fisher e Lee (1992) apresentam uma grande flexibilidade na análise de dados direcionais.

Outras contribuições estão no Capítulo 5. Tendo em vista a periodicidade existente no círculo, o entendimento do resíduo ordinário $y_i - \hat{\mu}_i$ assumir um valor “grande” ou “pequeno” não é imediato. Transformando os componentes da função desvio, propomos dois resíduos que podem assumir valores em toda a reta \mathbf{R} . Mostramos que esses resíduos possuem distribuições de probabilidade aproximadamente normal padrão. Desta forma, podemos empregar, por exemplo, as diversas técnicas gráficas para análise de resíduos. Em particular, podemos construir bandas de confiança através do uso de envelopes, Atkinson (1981).

7.2 Futuros Estudos

Em artigo recente Cadigan e Farrel (1999) empregam o valor esperado da curvatura normal C_d , vide equação (2.19), e mostram como essa curvatura esperada pode auxiliar na interpretação do diagnóstico. Em particular, dos elementos influentes, eles distinguem quais são devido à má especificação do modelo e quais se devem ao desenho do experimento. Uma linha que desejamos prosseguir consiste em adotar a estratégia de procurar classificar os elementos influentes ao se fazer diagnóstico nos modelos de regressão von Mises.

Preisser e Qaqish (1996) utilizam diagnóstico em equações de estimação generalizadas, através da estratégia de deleção de casos. Nessa linha, pretendemos avançar nosso estudo de influência local numa abordagem ligada à teoria de equações de estimação generalizadas.

Outros tópicos que pretendemos também desenvolver são:

1. Fazer outros estudos de simulação com os resíduos d^* e r , considerando modelos de análise de variância;
2. Investigar a distribuição desses resíduos em outras situações;
3. Fazer estudos adicionais sobre os resíduos propostos, considerando modelos com erros nas variáveis.

Apêndice A

Intervalos de Confiança para os parâmetros da distribuição von Mises

A.1 Intervalos de confiança para μ

Consideremos y_1, \dots, y_n uma amostra aleatória da distribuição von Mises $\mathcal{VM}(\mu, \lambda)$ e seja $\hat{\mu}$ o estimador de máxima verossimilhança para μ , que é dado por

$$\hat{\mu} = \begin{cases} \arctan(S/C), & \text{se } S \geq 0 \text{ e } C > 0 \\ \arctan(S/C) + \pi, & \text{se } C < 0 \\ \arctan(S/C) + 2\pi, & \text{se } S < 0 \text{ e } C > 0, \end{cases}$$

onde $C = \sum_{i=1}^n \cos(y_i)$ e $S = \sum_{i=1}^n \sin(y_i)$, conforme mostramos na página 41. O erro padrão circular de $\hat{\mu}$ é estimado por

$$\hat{\sigma}_{\mathcal{VM}} = \frac{1}{(n\bar{R}\hat{\lambda})^{1/2}},$$

onde $\bar{R} = \frac{(C^2 + S^2)^{1/2}}{n}$ e $\hat{\lambda} = \mathcal{A}_1^{-1}(\bar{R})$. Um intervalo de confiança aproximado para μ , com um nível de confiança de $100(1 - \alpha)\%$, é dado por

$$\hat{\mu} \pm \arcsen(z_{\frac{\alpha}{2}} \hat{\sigma}_{\nu, \mu}),$$

onde $z_{\frac{\alpha}{2}}$ corresponde ao percentil de ordem $100(1 - \frac{\alpha}{2})$ da distribuição normal padrão. A função \mathcal{A}_1^{-1} , pode ser calculada com uma razoável precisão, usando a aproximação (3.3) da página 42, sugerida por Best & Fisher (1981).

Além dessa abordagem, também é possível obter intervalos de confiança para μ através de *bootstrap*. Os detalhes podem ser obtidos em Fisher (1993, p.88)

A.2 Intervalos de confiança para λ

Quando a estimativa $\hat{\lambda} < 2$ é aconselhável usar o método de *bootstrap* para se construir intervalos de confiança para λ . Por outro lado, quando $\hat{\lambda} \geq 2$ é possível obter intervalos de confiança para λ , através de cálculos diretos. Sejam

$$a = \frac{n - R}{\chi_{n-1}^2(1 - \frac{\alpha}{2})} \quad \text{e} \quad b = \frac{n - R}{\chi_{n-1}^2(\frac{\alpha}{2})},$$

onde $\chi_{n-1}^2(p)$ corresponde ao percentil de ordem $100(1 - p)$ da distribuição qui-quadrado com $(n - 1)$ graus de liberdade.

Um intervalo de confiança aproximado para λ , com um nível de confiança de $100(1 - \alpha)\%$, é dado por

$$\left(\frac{1 + (1 + 3a)^{1/2}}{4a}, \frac{1 + (1 + 3b)^{1/2}}{4b} \right).$$

Apêndice B

Alguns resultados sobre as funções de Bessel

Neste apêndice apresentamos alguns resultados relacionados com a função de Bessel $I_p(t)$, e também com a função $\mathcal{A}_p(t) = \frac{I_p(t)}{I_0(t)}$. Essas funções são de extrema importância para a distribuição von Mises, e os resultados aqui apresentados foram compilados a partir de Abramowits & Stegun (1970), Mardia (1972) e Fisher (1993).

B.1 Derivadas

A função de Bessel modificada de 1º tipo e ordem p é definida como

$$I_p(t) = \sum_{r=0}^{\infty} \frac{1}{(p+r)! r!} \left(\frac{1}{2}t\right)^{2r+p}, \quad p = 0, 1, 2, \dots$$

A seguir serão enumeradas expressões que correspondem às derivadas de maior interesse

para nós. de funções envolvendo $I_0(t)$, $I_1(t)$.

1. $\frac{d}{dt}I_0(t) = I_1(t)$
2. $\frac{d}{dt}I_1(t) = I_0(t) - \frac{I_1(t)}{t} = \frac{1}{2}\{I_0(t) + I_2(t)\}$
3. $\frac{d}{dt}\{tI_1(t)\} = tI_0(t)$
4. $\frac{d}{dt}\mathcal{A}_1(t) = 1 - \frac{\mathcal{A}_1(t)}{t} - \mathcal{A}_1^2(t)$
5. $\frac{d^2}{dt^2}\mathcal{A}_1(t) = \frac{\mathcal{A}_1(t)}{t^2} - \left\{2\mathcal{A}_1(t) + \frac{1}{t}\right\}\frac{d}{dt}\mathcal{A}_1(t)$

B.2 Avaliação de I_0 e I_1

As funções $I_0(x)$ e $I_1(x)$ podem ser calculadas a partir das seguintes aproximações polinomiais, propostas por Abramowitz & Stegun (1970, p. 378). Considerando $t = \frac{x}{3,75}$, então temos o seguinte:

i. Se $0 \leq x \leq 3,75$,

$$\begin{aligned} I_0(x) &= 1 + 3,5156229t^2 + 3,0899424t^4 + 1,2067492t^6 \\ &+ 0,2659732t^8 + 0,0360768t^{10} + 0,0045813t^{12} + \epsilon \end{aligned} \quad (\text{B.1})$$

$$|\epsilon| < 1,6 \times 10^{-7}$$

$$\begin{aligned} x^{-1}I_1(x) &= 0,5 + 0,87890594t^2 + 0,51498869t^4 + 0,15084934t^6 \\ &+ 0,02658733t^8 + 0,00301532t^{10} + 0,00032411t^{12} + \epsilon \end{aligned} \quad (\text{B.2})$$

$$|\epsilon| < 8,0 \times 10^{-9}$$

ii. Se $x > 3,75$,

$$\begin{aligned}
 x^{\frac{1}{2}}e^{-x}I_0(x) &= 0,39894228 + 0,01328592/t + 0,00225319/t^2 \\
 &- 0,00157565/t^3 + 0,00916281/t^4 - 0,02057706/t^5 \\
 &+ 0,02635537/t^6 - 0,01647633/t^7 + 0,00392377/t^8 + \epsilon \quad (\text{B.3}) \\
 |\epsilon| &< 1,9 \times 10^{-7}
 \end{aligned}$$

$$\begin{aligned}
 x^{\frac{1}{2}}e^{-x}I_1(x) &= 0,39894228 - 0,03988024/t - 0,00362018/t^2 \\
 &+ 0,00163801/t^3 - 0,01031555/t^4 + 0,02282967/t^5 \\
 &- 0,02895312/t^6 + 0,01787654/t^7 - 0,00420059/t^8 + \epsilon \quad (\text{B.4}) \\
 |\epsilon| &< 2,2 \times 10^{-7}
 \end{aligned}$$

B.3 Fórmulas de recorrência

Para $p = 2, 3, \dots$ e $t > 0$ dispomos das seguintes fórmulas de recorrência:

1. $I_p(t) = I_{p-2}(t) - \frac{2(p-1)}{t}I_{p-1}(t);$
2. $\mathcal{A}_p(t) = \mathcal{A}_{p-2}(t) - \frac{2(p-1)}{t}\mathcal{A}_{p-1}(t).$

Apêndice C

Geração de Envelopes

No modelo de regressão normal linear $y_i \sim N(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$, $i = 1, \dots, n$, independentes, o vetor de resíduos ordinários é dado por $\mathbf{r} = \mathbf{y} - \mathbf{X}^T \hat{\boldsymbol{\beta}} = (\mathbf{I} - \mathbf{H})\mathbf{y}$ e satisfaz $E(\mathbf{r}) = \mathbf{0}$, $\text{Var}(\mathbf{r}) = \sigma^2(\mathbf{I} - \mathbf{H})$, onde $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ é a matriz “hat”.

Em geral, a análise desses resíduos é feita através de técnicas gráficas. Uma dessas técnicas é o gráfico normal de probabilidades, que é bastante informativo sobre existência de pontos discrepantes, afastamento de normalidade ou ainda falta de homogeneidade das variâncias. Tendo em vista a dificuldade de avaliar se o gráfico normal de probabilidades se afasta efetivamente do comportamento esperado (uma reta com coeficiente angular igual a 1), Atkinson (1981, 1985) sugere a construção de bandas de confiança através de simulações, que denominou de *envelopes*. No caso do modelo normal linear, o procedimento consiste em gerar resíduos que tenham média zero e estrutura de covariâncias $(\mathbf{I} - \mathbf{H})$.

A construção das bandas de confiança segue os seguintes passos:

1. Gerar n observações $N(0, 1)$ e armazená-las em $\mathbf{y} = (y_1, \dots, y_n)^T$;
2. Fazer o ajuste do modelo, obter $r_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}$, e em seguida calcular $t_i^* = r_i / \{1 - h_{ii}\}^{1/2}$, $i = 1, \dots, n$;
3. Repetir o passo anterior m vezes, obtendo t_{ij}^* , $i = 1, \dots, n$ e $j = 1, \dots, m$;
4. Colocar cada grupo de n resíduos em ordem crescente, obtendo $t_{(i)j}^*$, $i = 1, \dots, n$ e $j = 1, \dots, m$;
5. Obter $t_{(i)I}^* = \min_j t_{(i)j}^*$ e $t_{(i)S}^* = \max_j t_{(i)j}^*$, que fornecem os limites correspondentes ao i -ésimo resíduo.

A sugestão de Atkinson (1985) é considerar $m=19$. Dessa forma, a probabilidade do maior resíduo de um particular envelope exceder o limite superior é aproximadamente igual a $1/20$.

Bibliografia

- [1] Abeyasekera, S. e Collet, D. (1982). On the estimation of the parameters of the von Mises distribution. *Communications in Statistics - Theory and Methods* **11**, 18, 2083-2090.
- [2] Abramowitz, M. e Stegun, I. A. (1970). *Handbook of Mathematical Functions*. New York: Dover Pub.
- [3] Andrews D. F. e Pregibon D. (1978). Finding the outliers that matter. *Journal of the Royal Statistical Society B* **40**, 85-93.
- [4] Anscombe F. J. (1961). Examination of residuals. *Proc. Fourth Berkeley Symposium*. 1-36.
- [5] Anscombe F. J. (1973). Graphs in statistical analysis. *The American Statistician* **27**, 17-21.
- [6] Artes, R. (1997). *Extensões da Teoria das Equações de Estimação Generalizadas a Dados Circulares e Modelos de Dispersão*. Tese de Doutorado. Instituto de Matemática e Estatística da Universidade de São Paulo.

- [7] Artes, R., Paula, G. A. e Ranvaud, R. (1998a). *Estudo da direção tomada por pombos através de equações de estimação para dados circulares longitudinais*. Relatório Técnico. São Paulo, IME-USP. (RT-MAE 9810).
- [8] Artes, R., Paula, G. A. e Ranvaud, R. (1998b). *Strategy for the analysis of circular longitudinal data based on generalized estimating equations*. Relatório Técnico. São Paulo, IME-USP. (RT-MAE 9816).
- [9] Atkinson, A. C. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika* **68**, 13-20.
- [10] Atkinson, A. C. (1982). Regression diagnostics, transformations and constructed variables. *Journal of the Royal Statistical Society B* **44**, 1-36.
- [11] Atkinson, A. C. (1985). *Plots, Transformations and Regressions*. Oxford Statistical Science Series, Oxford: Clarendon Press.
- [12] Bagchi, P. e Guttman, I. (1990). Spuriousity and outliers in directional data. *Journal of Applied Statistics* **17**, 3, 341-350.
- [13] Barret, Bruce E. e Ling, Robert F. (1992). General classes of influence measures for multivariate regression. *Journal of the American Statistical Association* **87**, 184-191.
- [14] Batschelet, E. (1981). *Circular Statistics in Biology*. London: Academic Press.
- [15] Belsley, D. A., Kuh, E. e Welsch, R. E. (1980). *Regression Diagnostics*. New York: John Wiley.
- [16] Best, D. J. e Fisher, N. I. (1979). Efficient simulation of the von Mises distribution. *Applied Statistics* **28**, 152-157.

- [17] Best, D. J. e Fisher, N. I. (1981). The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. *Communications in Statistics - Simulation and Computation* **10**, 493-502.
- [18] Billor, N. e Loynes, R. M. (1993). Local influence: A new approach. *Communications in Statistics - Theory and Methods* **22**, 1595-1611.
- [19] Bingham, M. S. e Mardia, K. V. (1975). Maximum likelihood characterization of the von Mises distribution. *In: Statistical Distributions in Scientific Work* (C. P. Patil et al., eds), Vol. 3, pp. 387-398. Reidel, Dordrecht.
- [20] Cadigan, N. G. e Farrell, P. J. (1999). Expected local influence in the normal linear regression model. *Statistics e Probability Letters* **41**, 25-30.
- [21] Collet, D. (1980). Outliers in circular data. *Applied Statistics* **29**, 50-57.
- [22] Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics* **19**, 15-18.
- [23] Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association* **74**, 169-174.
- [24] Cook, R. D. (1986). Assessment of local influence (with discussion). *Journal of the Royal Statistical Society B* **48**, 2, 133-169.
- [25] Cook, R. D. (1987). Influence assessment. *Journal of Applied Statistics* **14**, 2, 117-131.
- [26] Cook, R. D. e Weisberg, S. (1982). *Residuals and Influence in Regression*. New York: Chapman and Hall.
- [27] Cook, R. D. e Tsai, C.-L. (1985). Residuals in nonlinear regression. *Biometrika* **72**, 1, 23-29.

- [28] Cook, R. D., Peña, D. e Weisberg, S. (1988). The Likelihood displacement: A unifying principle for influence measures. *Communications in Statistics - Theory and Methods* **17**, 3, 623-640.
- [29] Cordeiro, G. M. e Paula, G. A. (1989). Fitting non-exponential family nonlinear models in GLIM by using the offset-facility. *Lecture Notes in Statistics* **57**, 105-114.
- [30] Cordeiro, G. M. e Paula, G. A. (1992). Estimation, large-sample parametric tests and diagnostics for non exponential family nonlinear models. *Communications in Statistics - Simulation and Computation* **21**, 149-172.
- [31] Cox, D. R. e Snell, E. J. (1968). A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B* **30**, 248-275.
- [32] Cox, D. R. e Snell, E. J. (1971). On test statistics calculated from residuals. *Biometrika* **58**, 589-594.
- [33] Cox, D. R. e Hinkley, D. V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- [34] Dagpunar, J. S. (1990). Sampling from the von Mises distribution via a comparison of random numbers. *Journal of Applied Statistics* **17**, 165-168.
- [35] Davison, A. C. e Gigli, A. (1989). Deviance residuals and normal scores plots. *Biometrika* **76**, 211-221.
- [36] Davison, A. C. e Tsai, C.-L. (1992). Regression model diagnostics. *International Statistical Review* **60**, 3, 337-353.
- [37] Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge: University Press.

- [38] Fisher, N. I. e Lee, A. J. (1983). A correlation coefficient for circular data. *Biometrika* **70**, 327-332.
- [39] Fisher, N. I. e Lewis, T. (1983). Estimating the common mean directional of several circular or spherical distributions with differing dispersions. *Biometrika* **70**, 333-341.
- [40] Fisher, N. I. e Lee, A. J. (1992). Regression models for an angular response. *Biometrics* **48**, 665-677.
- [41] Fung, W. K. e Kwan, C. W. (1997). A Note on local influence based on normal curvature. *Journal of the Royal Statistical Society B* **59**, 839-843.
- [42] Galea, M., Paula, G. A. e Bolfarine, H. (1997). Local influence in elliptical linear regression models. *The Statistician* **46**, 71-79.
- [43] Gauss, C. F. (1823). *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae*. Göttingen: Dieterich.
- [44] Gould, A. L. (1969). A regression technique for angular variates. *Biometrics* **25**, 683-700.
- [45] Gray, J. B. (1989). On the use of regression diagnostics. *The Statistician* **38**, 97-105.
- [46] Green, P. J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society B* **46**, 2, 149-192.
- [47] Hastie, T. (1987). A Closer look at the deviance. *The American Statistician* **41**, 16-20.

- [48] Hinkley, D. V., Reid, N. e Snell, E. J. (1991). *Statistical Theory and Modelling - In honour of Sir David Cox, FRS*. London: Chapman and Hall.
- [49] Hoaglin, D. C. e Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician* **32**, 17-22.
- [50] Hossain, A. e Naik, D. N. (1989). Detection of influential observations in multivariate regression. *Journal of Applied Statistics* **16**, 25-37.
- [51] IMSL (1991). *The IMSL Libraries*. 2nd edition. Sugar Land. Texas: IMSL Inc.
- [52] Jørgensen, B. (1983). Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika* **70**, 19-28.
- [53] Jørgensen, B. (1984). The Delta Algorithm and GLIM. *International Statistical Review* **52**, 3, 283-300.
- [54] Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society B* **49**, 127-162.
- [55] Jørgensen, B. (1997a). Proper dispersion models (with discussion). *Brazilian Journal of Probability and Statistics* **11**, 2, 89-140.
- [56] Jørgensen, B. (1997b). *The Theory of Dispersion Models*. London: Chapman and Hall.
- [57] Kim, M. G. (1995). Local influence in multivariate regression. *Communications in Statistics - Theory and Methods* **20**, 1271-1278.
- [58] Lawrance, A. J. (1988). Regression transformation diagnostic using local influence. *Journal of the American Statistical Association* **83**, 1067-1079.

- [59] Lee, A. H. e Hui, Y. V. (1994). On influence assessment in generalized linear models. *Computational Statistics* **9**, 277-285.
- [60] Lee, A. H. e Zhao, Y. (1996). Sensitivity of Pearson's goodness-of-fit statistic in generalized linear models. *Communications in Statistics - Theory and Methods* **25**, 143-157.
- [61] Mardia, K. V. (1972). *Statistics of Directional Data*. London: Academic Press.
- [62] Mardia, K. V. (1975). Statistics of directional data (with discussion). *Journal of the Royal Statistical Society B* **37**, 349-393.
- [63] Mardia, K. V. e El-Atoum, S. A. M. (1976). Bayesian inference for the von Mises-Fisher distribution. *Biometrika* **63**, 203-206.
- [64] McCullagh, P. (1987). *Tensor Methods in Statistics*. New York: Chapman and Hall.
- [65] McCullagh, P. e Nelder, J. A. (1989). *Generalized Linear Models*. 2nd edition. London: Chapman and Hall.
- [66] Pan, J.X., Fang, K.T. e von Rosen, D. (1997). Local influence assessment in the growth curve model with unstructured covariance. *Journal of Statistical Planning and Inference* **62**, 263-278.
- [67] Paula, G. A. (1993). Assessing local influence in restricted regressions models. *Computational Statistics and Data Analysis* **16**, 63-79.
- [68] Paula, G. A. (1995). Influence and residuals in restricted generalized linear models. *Journal of Statistical Computation and Simulation* **51**, 315-331.
- [69] Paula, G. A. (1996). Influence diagnostics in proper dispersion models. *Australian Journal of Statistics* **38**, 307-316.

- [70] Pierce, D. A. e Schafer, D. W. (1982). Residuals in generalized linear models. *Journal of the American Statistical Association* **81**, 977-986.
- [71] Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics* **9**, 705-724.
- [72] Preisser, J. S. e Qaqish, B. F. (1996). Deletion diagnostics for generalised estimating equations. *Biometrika* **83**, 551-562.
- [73] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. 2nd edition. New York: John Wiley.
- [74] Rivest, L.-P. (1982). Some statistical methods for bivariate circular data. *Journal of the Royal Statistical Society B* **44**, 81-90.
- [75] Souza, F. A. M. e Paula, G. A. (1998). *Deviance Residuals for an Angular Response*. Relatório Técnico. São Paulo, IME-USP. (RT-MAE 9832).
- [76] St. Laurent, R. T. e Cook, R. D. (1993). Leverage, local influence and curvature in nonlinear regression. *Biometrika* **80**, 99-106.
- [77] Stephens, M. A. (1963). Random walk on a circle. *Biometrika* **50**, 385-390.
- [78] Stoker, J. J. (1969). *Differential Geometry*. New York: John Wiley.
- [79] Thomas, W. (1990). Influence on confidence regions for regression coefficients in generalized linear models. *Journal of the American Statistical Association* **85**, 393-397.
- [80] Thomas, W. (1991). Influence diagnostics for the cross-validated smoothing parameter in spline smoothing. *Journal of the American Statistical Association* **86**, 693-698.

- [81] Thomas, W. e Cook, R. D. (1989). Assessing influence on regression coefficients in generalized linear models. *Biometrika* **76**, 741-749.
- [82] Thomas, W. e Cook, R. D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics* **32**, 59-65.
- [83] Tsai, C.-L. e Wu, X. (1992). Transformation-model diagnostics. *Technometrics* **34**, 197-202.
- [84] Upton, G. (1993). Testing for outliers in circular data. *Journal of Applied Statistics* **20**, 2, 229-235.
- [85] Wehrly, T. E. e Shine, E. P. (1981). Influence curves of estimators for directional data. *Biometrika* **68**, 334-335.
- [86] Weisberg, S. (1985). *Applied Linear Regression*. 2nd edition. New York: John Wiley.
- [87] Williams, D. A. (1984). Residuals in generalized linear models. In: *Proceedings of the 12th. International Biometrics Conference, Tokyo*, pp. 59-68.
- [88] Williams, D. A. (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics* **36**, 181-191.
- [89] Wood, A. T. A. (1994). Simulation of the von Mises Fisher distribution. *Communications in Statistics - Simulation* **23**, 157-164.
- [90] Wood, F. S. (1973). The use of individual effects and residuals in fitting equations to data. *Technometrics* **15**, 4, 677-695.