



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Regressão Beta Geograficamente Ponderada

por

Andreza de Oliveira Lima

Orientador: Prof. Dr. Alan Ricardo da Silva

Dezembro de 2015

Andreza de Oliveira Lima

Regressão Beta Geograficamente Ponderada

Dissertação apresentada ao Departamento de Estatística do Instituto de Ciências Exatas da Universidade de Brasília como requisito parcial à obtenção do título de Mestre em Estatística.

Universidade de Brasília

Brasília, Dezembro de 2015

*“Trata de saborear a vida; e fica sabendo,
que a pior filosofia é a do choramingas
que se deita à margem do rio
para o fim de lastimar o curso incessante das águas.*

*O ofício delas é não parar nunca;
acomoda-te com a lei, e trata de aproveitá-la.”*

Memórias Póstumas de Brás Cubas, Machado de Assis

Agradecimentos

Primeiramente, agradeço a Deus, que me guia em todos os caminhos.

Pela dedicação e atenção a este trabalho dispensadas, agradeço ao meu professor orientador, Alan Ricardo da Silva, também ao professor André Luiz Fernandes Cançado, e demais professores que direta ou indiretamente me auxiliaram nessa jornada.

Pela idealização do projeto e suporte, agradeço ao amigo Paulo Henrique Dourado da Silva.

Pelo companheirismo e amizade em todos esses anos de universidade e fora dela, agradeço à querida Ilka Oliveira Torres.

Pela compreensão, agradeço aos colegas de trabalho na Conab, no INEP, no CNJ e na Anatel.

Por fim, mas não menos importante, agradeço ao meu estimado marido, pelo amor e apoio incondicional, e à minha mãe e padrasto. Sem eles, eu não teria chegado onde cheguei.

Sumário

Agradecimentos	ii
Lista de Figuras	5
Lista de Tabelas	6
Resumo	7
Abstract	8
Introdução	9
1 Regressão Beta	14
1.1 Introdução	14
1.2 O Modelo de Regressão Beta	15
1.3 Estimação dos parâmetros	19
1.4 Diagnóstico e inferência	24
1.5 Simulações	27
1.6 Adaptações ao modelo de Regressão Beta	30
2 Regressão Geograficamente Ponderada	31
2.1 Introdução	31
2.2 O modelo de Regressão Geograficamente Ponderada	32
2.3 Considerações sobre a Matriz de Ponderação Espacial	38
2.4 Otimização por Seção Áurea	43
2.5 Inferência em dados espaciais	46

2.6	Diagnósticos	51
2.7	Extensões do modelo de Regressão Geograficamente Ponderada	52
3	Regressão Beta Geograficamente Ponderada	55
3.1	Introdução	55
3.2	Especificação do Modelo RBGP	56
3.3	Especificação do Modelo RBGPg	62
4	Estudo de Caso: proporção de telefones fixos no Espírito Santo	65
4.1	Introdução	65
4.2	Base de dados	65
4.3	Ajuste dos modelos globais	69
4.4	Ajuste dos modelos locais	75
5	Conclusões	83
5.1	Limitações do trabalho	83
5.2	Trabalhos futuros	84
	Referências Bibliográficas	86
A	Apêndice	93
B	Apêndice	97

Lista de Figuras

1	Exemplo do Paradoxo de Simpson	12
1.1	Distribuição beta para diferentes (μ, ϕ)	18
2.1	Exemplo de inclusão de variável <i>dummy</i>	33
2.2	Exemplo de superfície de parâmetros estimada	37
2.3	Função <i>Kernel</i> Espacial	39
2.4	<i>Kernel</i> (a) Fixo e (b) Variável	41
2.5	Ilustração do algoritmo da Seção Áurea	44
2.6	Exemplo de diagrama de Moran	47
4.1	Distribuição espacial da variável proporção de telefone fixo	66
4.2	Distribuição espacial da variável proporção de população urbana . . .	67
4.3	Distribuição espacial da variável proporção de pobres	67
4.4	Diagrama de Moran da variável proporção de telefone fixo	68
4.5	Diagrama de Moran da variável proporção de população urbana . . .	68
4.6	Diagrama de Moran da variável proporção de pobres	68
4.7	Mapa de espalhamento de Moran para a proporção de telefone fixo .	69
4.8	Mapa de Moran (95%) para a proporção de telefone fixo	69
4.9	Distribuição da variável proporção de telefone fixo	70
4.10	Distribuição dos resíduos do modelo clássico	72
4.11	Distribuição dos resíduos do modelo beta	72
4.12	Resíduos pelos valores preditos do modelo clássico	73
4.13	Resíduos pelos valores preditos do modelo beta	73
4.14	Disposição espacial dos resíduos do modelo clássico	73

4.15	Disposição espacial dos resíduos do modelo beta	73
4.16	Resíduos padronizados pelos valores preditos do modelo clássico . . .	74
4.17	Resíduos padronizados pelos valores preditos do modelo beta	74
4.18	Disposição espacial dos resíduos padronizados do modelo clássico . . .	75
4.19	Disposição espacial dos resíduos padronizados do modelo beta	75
4.20	Otimização do <i>bandwidth</i> para o modelo RGP	76
4.21	Otimização do <i>bandwidth</i> para o modelo RBGP	76
4.22	Otimização do <i>bandwidth</i> para o modelo RBGPg	76
4.23	Distribuição dos resíduos do modelo RGP	77
4.24	Distribuição dos resíduos do modelo RBGP	77
4.25	Distribuição dos resíduos do modelo RBGPg	77
4.26	Resíduos pelos valores preditos do modelo RGP	78
4.27	Resíduos pelos valores preditos do modelo RBGP	78
4.28	Resíduos pelos valores preditos do modelo RBGPg	78
4.29	Disposição espacial dos resíduos da RGP	78
4.30	Disposição espacial dos resíduos da RBGP	78
4.31	Disposição espacial dos resíduos da RBGPg	78
4.32	Resíduos padronizados pelos valores preditos do modelo RGP	79
4.33	Resíduos padronizados pelos valores preditos do modelo RBGP	79
4.34	Resíduos padronizados pelos valores preditos do modelo RBGPg	79
4.35	Disposição espacial dos resíduos padronizados da RGP	79
4.36	Disposição espacial dos resíduos padronizados da RBGP	79
4.37	Disposição espacial dos resíduos padronizados da RBGPg	79
4.38	Interceptos significativos a 5% na RGP	80
4.39	Parâmetros <i>prop_pobres</i> significativos a 5% na RGP	80
4.40	Parâmetros <i>prop_urbana</i> significativos a 5% na RGP	80
4.41	Interceptos significativos a 5% na RBGP	81
4.42	Parâmetros <i>prop_pobres</i> significativos a 5% na RBGP	81
4.43	Parâmetros <i>prop_urbana</i> significativos a 5% na RBGP	81
4.44	Interceptos significativos a 5% na RBGPg	81
4.45	Parâmetros <i>prop_pobres</i> significativos a 5% na RBGPg	81

4.46	Parâmetros prop_urbana significativos a 5% na RBGPg	81
4.47	Parâmetros ϕ significativos a 5% na RBGP	82

Lista de Tabelas

1	Características dos modelos globais e locais	11
1.1	Algoritmo de estimação da Regressão Beta com função de ligação logito	23
1.2	Condições para simulação da regressão beta	27
1.3	Algoritmo de construção dos dados simulados para a regressão beta com ligação logito	28
1.4	Resultados Simulação - Proporção de predições fora do intervalo . . .	29
2.1	Algoritmo da Seção Áurea	46
3.1	Algoritmo parcial de estimação da RBGP e da RBGPg com função de ligação logito	64
4.1	Resultados dos modelos globais	70
4.2	Resultados dos modelos locais	76
4.3	Teste de pseudo-significância para estacionariedade dos parâmetros .	82

Resumo

Modelos de Regressão Linear Clássica são comumente utilizados para descrever a relação existente entre uma variável resposta e um conjunto de variáveis explicativas. Entretanto, esse modelo assume que a variável resposta segue distribuição normal com variância constante e que as relações se dão de forma igualitária no espaço. Dessa forma, esse modelo não é o mais adequado para ajustar dados provenientes de taxas ou proporções que possuam variações pelo espaço.

Para lidar com as taxas e proporções, o modelo de Regressão Beta tem se mostrado como uma boa alternativa, já que acomoda naturalmente variáveis restritas a algum intervalo da reta real e que possuem heterocedasticidade, característica comum em dados de taxas e proporções. E para atender à heterogeneidade espacial, o modelo de Regressão Geograficamente Ponderada, por meio de uma extensão do modelo de regressão linear, permite que haja variações locais nos parâmetros e com isso proporciona melhor conhecimento do fenômeno espacial em estudo.

Neste trabalho propomos o modelo de Regressão Beta Geograficamente Ponderada, de modo que, ao unir as características dos dois modelos supracitados, possa fornecer melhor ajuste no estudo de dados restritos a algum intervalo da reta real e que possuem heterogeneidade espacial. Esse modelo foi aplicado à proporção de domicílios que possuem telefone fixo no estado do Espírito Santo e apresentou resultados mais apropriados que os resultados dos modelos globais e do modelo de Regressão Geograficamente Ponderada.

Palavras-chave: taxas; proporções; heterogeneidade espacial; distribuição beta; análise espacial; regressão beta; regressão geograficamente ponderada.

Abstract

Linear Regression Models are often used to describe the relationship between a dependent variable and a set of independent variables. However, these models are based on the assumption that the response variable is normally distributed with constant variance and the relations are equal through space. Thus, these models may not be the most appropriate to adjust spatially varying rates and proportions.

Beta Regression model deals with rates and proportions and has shown to be a good approach to model these kind of data, since it naturally adapts to variables constrained to a interval of the real line and exhibit heteroskedasticity, common characteristic in these type of data. And to deal with spatial nonstationarity, Geographically Weighted Regression allows variability in the parameters by an extension of the linear regression model, providing a better understanding of the spatial phenomena.

Therefore, we propose Geographically Weighted Beta Regression model which puts together the features of the above models such that a better fit is provided in the study of spatial varying continuous variables that are restricted to a interval of the real line. We applied this model to the proportion of households that own telephones in the state of Espírito Santo, Brazil. The results were more appropriate than the global models and Geographically Weighted Regression model.

Keywords: rates; proportions; spatial nonstationarity; beta distribution; spatial analysis; beta regression; geographically weighted regression.

Introdução

Modelos de regressão linear (Neter *et al.*, 1996) são a forma mais comum para descrever a relação existente entre uma variável resposta ou dependente e um conjunto de variáveis explicativas ou independentes. Esses modelos geralmente possuem a forma:

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \quad (1)$$

em que y_i é a i -ésima observação da variável resposta, $i = 1, \dots, n$, x_{ij} é o valor da j -ésima variável explicativa para a i -ésima observação, $j = 1, \dots, k$, e ε_i é o erro relacionado à i -ésima observação. Supõe-se nesse modelo que os erros são independentes e normalmente distribuídos, com média 0 e variância σ^2 ($\varepsilon_i \sim N(0, \sigma^2)$).

Os coeficientes β_j são chamados de parâmetros do modelo e representam a taxa de mudança que uma alteração na variável explicativa acarreta na média condicional da variável resposta. Cada β_j é determinado por meio de uma amostra de $i = 1, \dots, n$ observações, geralmente utilizando a técnica de Mínimos Quadrados Ordinários - MQO (ou do inglês, *Ordinary Least Squares* - OLS), cujas estimativas que minimizam a soma dos erros ao quadrado são obtidas matricialmente da seguinte forma:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}, \quad (2)$$

em que \mathbf{X} é uma matriz $n \times (k + 1)$ com o valor das variáveis explicativas para cada observação e o intercepto, \mathbf{y} é um vetor de tamanho n com o valor da variável resposta para cada observação e $\hat{\boldsymbol{\beta}}$ é um vetor de tamanho $k + 1$ que contém as estimativas para os parâmetros.

Para casos em que a variável resposta está restrita a um intervalo da reta real,

como por exemplo para taxas e proporções, restritas ao intervalo $(0, 1)$, esse modelo não é o mais adequado, pois pode levar a valores ajustados que ultrapassam esses limites. Uma alternativa seria realizar uma transformação dos dados para a reta real e modelar a média da variável transformada, como sugestão de Dyke e Patterson (1952). Contudo, surgem alguns inconvenientes, a depender dos objetivos do pesquisador: a interpretação dos parâmetros não ocorre de forma direta e, pela natureza dos dados de taxas e proporções, espera-se que exista uma certa assimetria (Fleiss *et al.*, 2013) e heterocedasticidade (Cribari-Neto e Zeileis, 2010), o que poderia violar as suposições de normalidade e de variância constante dos erros e levar a inferências errôneas.

Tendo em vista essa dificuldade, foi proposto por Ferrari e Cribari-Neto (2004) um modelo de regressão beta com inspiração na abordagem de Modelos Lineares Generalizados (Nelder e Wedderburn, 1972), em que a variável resposta segue uma distribuição beta e acomoda naturalmente os dados restritos a um intervalo qualquer. Uma vantagem em relação à alternativa proposta acima é a interpretação direta dos parâmetros. Por exemplo, quando se utiliza a função de ligação logito, é possível calcular as razões de chances. Também, como a distribuição beta é conhecida pela sua versatilidade (Gupta e Nadarajah, 2004), a depender da escolha dos parâmetros, ela pode assumir várias formas e assim, pode ser utilizada para modelar inúmeras relações. Inclusive, de acordo com Meaney e Moineddin (2014), quando a distribuição é simétrica e unimodal, semelhante à distribuição normal, o modelo de regressão linear clássico pode ter performance igual ou melhor que a regressão beta, visto que não há grande fuga às suposições do modelo, embora possa ainda apresentar valores preditos ligeiramente fora do intervalo.

Esse modelo, entretanto, assume que os parâmetros são constantes no espaço, o que pode não ser verdade em alguns casos, como, por exemplo, o preço de um mesmo produto pode ser influenciado por uma variável positivamente em determinado lugar e negativamente em outro, ao contrário de um experimento físico, que sempre trará o mesmo resultado independentemente do lugar onde o experimento for realizado, desde que respeitados os mesmos requisitos (Fotheringham *et al.*, 2002). Essa característica dos dados espaciais é chamada de não estacionariedade, que pode

ser ocasionada pela simples aleatoriedade das amostras ou por alguma outra razão intrínseca dos dados.

Para lidar com essa problemática da não estacionariedade com dados espaciais, Fotheringham *et al.* (1996), Brunson *et al.* (1996) e Brunson *et al.* (1998) sugeriram a técnica de Regressão Geograficamente Ponderada - RGP (ou do inglês, *Geographically Weighted Regression* - GWR), que, segundo os autores, nada mais é do que uma extensão do modelo de regressão linear tradicional, que permite que haja variações locais nos parâmetros e apresenta uma medida dessa variação. Por essa razão, o modelo de RGP é classificado como um modelo local, e o modelo (1) como global, pois, dentre outros aspectos, fornece apenas estimativas gerais dos dados. A Tabela 1, adaptada de Fotheringham *et al.* (2002), resume as características que diferem esses dois tipos de modelo, em se tratando de dados espaciais.

Tabela 1: Características dos modelos globais e locais

Modelos Globais	Modelos Locais
- Resumem os dados para toda a região em estudo	- Desagregam localmente as estatísticas globais
- Geram uma única estatística	- Geram uma estatística para cada local
- Estatísticas não podem ser mapeadas ou analisadas por Sistema Geográfico de Informação	- Estatísticas podem e devem ser mapeadas ou analisadas por Sistema Geográfico de Informação
- Ênfase na similaridade no espaço	- Ênfase nas diferenças no espaço
- Procuram por regularidades ou padrões	- Procuram por exceções ou regiões em destaque

Fonte: adaptado de Fotheringham *et al.* (2002)

A Figura 1 ilustra um caso não raro em que a utilização de um modelo global, quando na verdade as relações são locais, podem levar a um resultado bastante equivocado conhecido como Paradoxo de Simpson (Simpson, 1951; Appleton *et al.*, 1996), em que, ao se desconsiderar a localidade dos dados, o resultado é contrário ao que se teria se a localidade fosse considerada.

Observe que, em um exemplo bastante simplificado, foi possível mostrar que a utilização de modelos locais em dados espaciais pode trazer resultados que retratem melhor a realidade. Dessa forma, o propósito deste trabalho é apresentar uma

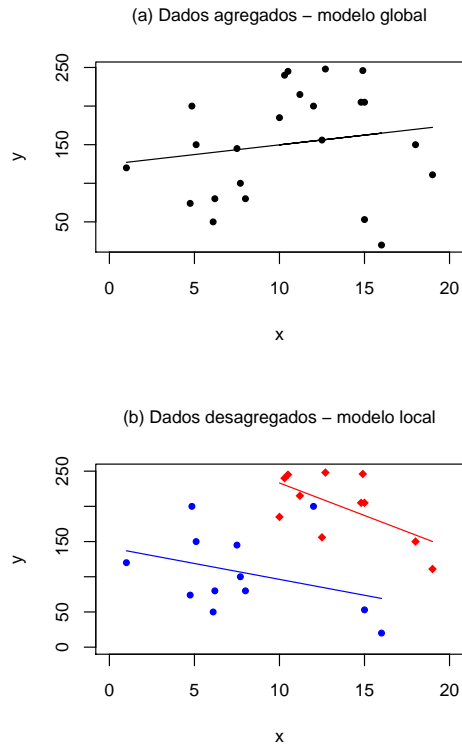


Figura 1: Exemplo do Paradoxo de Simpson

Fonte: adaptado de Fotheringham *et al.* (2002)

técnica que permita modelar dados de taxas ou proporções e que apresentem não estacionariedade espacial, por meio da junção das duas técnicas supracitadas.

Assim, à luz dos trabalhos de Atkinson *et al.* (2003), Nakaya *et al.* (2005) e Silva e Rodrigues (2014), que introduziram o modelo de RGP com as distribuições Binomial (logística), Poisson e Binomial Negativa, respectivamente, o objetivo deste trabalho é propor a Regressão Beta Geograficamente Ponderada (RBGP), já que, até o momento, essas adaptações ao modelo RGP beneficiam apenas variáveis respostas discretas, ou seja, não são adequadas para o caso em que a variável resposta está restrita a um intervalo contínuo, como é o caso das taxas e proporções.

Com isso, pretende-se proporcionar um melhor ajuste a trabalhos como o de Comber *et al.* (2012) e Lawson *et al.* (2015), que analisam taxas pelo modelo de Regressão Geograficamente Ponderada clássico, ou seja, assumindo que a variável resposta contínua e restrita ao intervalo $(0, 1)$ segue uma distribuição normal, que, como elucidado acima, nem sempre é uma suposição verdadeira.

Com esse objetivo, nos Capítulos 1 e 2 serão especificados o modelo de Regressão Beta e o modelo RGP tradicionais, bem como suas respectivas peculiaridades. O modelo de Regressão Beta Geograficamente Ponderada será exposto no Capítulo 3. O Capítulo 4 traz uma aplicação com dados sobre a proporção de domicílios que possuem telefone fixo no estado do Espírito Santo. Por fim, no Capítulo 5 serão apresentadas as considerações finais e os possíveis estudos futuros. Nos Apêndices podem ser encontradas partes dos algoritmos que foram utilizados nas análises.

Capítulo 1

Regressão Beta

1.1 Introdução

Vem de longa data a procura por métodos para modelar apropriadamente dados provenientes de taxas ou proporções. Como já salientavam Dyke e Patterson (1952), o modelo de regressão linear não é o mais adequado para esses dados, pois podem levar a valores ajustados fora do intervalo $(0, 1)$. Então, a proposta dos autores era utilizar esse modelo, mas com uma transformação da variável resposta. Porém, dessa forma a interpretação dos parâmetros não seria direta e, além disso, quando o número de observações não é grande o suficiente ou quando a proporção p está muito próxima de 0 ou 1, a distribuição dos dados de taxas e proporções podem apresentar uma certa assimetria, e assim a suposição de normalidade do modelo de regressão linear tradicional (Equação (1)) não seria satisfeita. Também, esses tipos de dados estão sujeitos a maior variação em volta do valor médio e menor variação nas extremidades do intervalo, caracterizando a heterocedasticidade, o que também violaria a suposição de que os erros do modelo possuem a mesma variância.

Mais tarde, como citam Kieschnick e McCullough (2003), outros artifícios sugeridos foram, por exemplo, manter a distribuição normal para a variável resposta, mas com uma estrutura de regressão não linear (Cox, 1996), ou então, em uma analogia ao modelo de regressão Tobit (Tobin, 1958) para modelar dados contínuos não negativos, Barclay *et al.* (1997) utilizaram o modelo de regressão clássica, mas censurada aos casos em que o preditor linear está dentro do intervalo $(0, 1)$. Jorgensen (1997)

sugere ainda um modelo de dispersão que tem por base a utilização da distribuição simplex.

A ideia de utilizar um modelo de regressão fundamentado na distribuição beta possui sustentáculo em diversos trabalhos aplicados, como por exemplo, Falls (1974) e Sulaiman *et al.* (1999), os quais apontam que essa distribuição se ajusta bem a dados de proporções, e tem seus primórdios nos trabalhos de Brehm e Gates (1993) e McDonald e Xu (1995). O primeiro assume que a variável segue uma distribuição beta cujos parâmetros são funções das variáveis explicativas, já o segundo emprega uma distribuição beta generalizada que depende de cinco parâmetros.

Mais tarde, Ferrari e Cribari-Neto (2004) propuseram o modelo de regressão beta adotado neste trabalho, o qual tem por objetivo modelar variáveis contínuas, porém restritas ao intervalo $(0, 1)$, como seria o caso de taxas, proporções ou índices, em função de um conjunto de variáveis explicativas por meio de uma estrutura de regressão. O modelo proposto pelos autores em tela tem por base uma reparametrização da distribuição beta, que depende de apenas dois parâmetros e, principalmente devido a sua versatilidade, utilidade e implementações computacionais disponíveis (Cribari-Neto e Zeileis, 2010; Swearingen *et al.*, 2011), vem sendo aplicado em diversas áreas do conhecimento, como na medicina (Peplonska *et al.*, 2012), economia (Castellani *et al.*, 2012), gestão pública (Pereira *et al.*, 2014), engenharia florestal (Eskelson *et al.*, 2011), dentre outras, nas quais tem provado ser uma boa alternativa na modelagem desses dados.

Nas próximas seções, o modelo será abordado, bem como os métodos de estimação e de inferência e diagnóstico a ele relacionados. A última seção apresenta ainda um estudo com dados simulados, para melhor compreensão do desempenho do modelo.

1.2 O Modelo de Regressão Beta

A fim de contornar os impasses citados anteriormente para descrever relações que envolvem variáveis restritas a um intervalo da reta real, Ferrari e Cribari-Neto (2004) partiram do modelo tradicional (1), mas considerando que a variável resposta

segue uma distribuição beta, tendo por base os Modelos Lineares Generalizados - MLG (Nelder e Wedderburn, 1972; McCullagh *et al.*, 1989). A distribuição beta foi escolhida pela sua versatilidade, já que pode assumir diversas formas a depender dos valores dos parâmetros, e pela própria definição da função beta, que naturalmente já está restrita ao intervalo em questão:

$$B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt. \quad (1.1)$$

A densidade da distribuição beta é dada por

$$f(y) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1-y)^{b-1} \quad (1.2)$$

em que $0 < y < 1$, $a > 0$, $b > 0$ e $\Gamma(\cdot)$ é a função gamma.

Vale citar que na abordagem dos modelos lineares generalizados, formulados para as distribuições da família exponencial, existe uma estrutura que relaciona a média da distribuição aos preditores lineares. Embora a distribuição beta não pertença de fato à família exponencial de distribuições, uma abordagem semelhante foi utilizada. Assim, para se obter uma estrutura de regressão para a média da variável resposta, seguindo a abordagem de MLG do modelo de Ferrari e Cribari-Neto (2004), é necessário que haja uma reparametrização dessa densidade apresentada em (1.2). Porém, cumpre citar que Vasconcellos e Cribari-Neto (2005) sugeriram um modelo alternativo que utiliza a própria distribuição, sem reparametrizar.

Para a parametrização (1.2), sabe-se que a média e a variância da distribuição beta são dadas, respectivamente, por:

$$E(y) = \frac{a}{a+b} \quad (1.3)$$

$$var(y) = \frac{ab}{(a+b)^2(a+b+1)}. \quad (1.4)$$

Fazendo $\mu = E(y)$ e $\phi = a + b$, tem-se que $a = \mu\phi$ e $b = (1 - \mu)\phi$. Segue então,

a partir das Equações (1.3) e (1.4), que

$$E(y) = \mu \quad (1.5)$$

e

$$var(y) = \frac{V(\mu)}{1 + \phi}, \quad (1.6)$$

em que $V(\mu) = \mu(1 - \mu)$. Tem-se então a reparametrização da distribuição beta em função da média μ e do parâmetro de precisão ϕ (no sentido de que, para μ fixo, a variância de y decresce quando ϕ cresce):

$$f(y) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}, \quad (1.7)$$

em que $0 < \mu < 1$ e $\phi > 0$.

Com essa nova parametrização, foram construídos os painéis na Figura 1.1, os quais mostram quão flexível pode ser a distribuição beta pela combinação dos valores (μ, ϕ) , podendo ser simétrica ($\mu = 1/2$) ou assimétrica ($\mu \neq 1/2$), assumir “forma de J”, “J invertido” e até “forma de U”. É possível observar ainda que para μ fixo, a dispersão da distribuição diminui se ϕ aumenta, e vice versa. Adicionalmente, para $\mu = 1/2$ e $\phi = 2$, tem-se o caso da distribuição uniforme.

Para a definição do modelo, sejam então y_1, y_2, \dots, y_n variáveis aleatórias independentes que seguem a distribuição beta dada pela parametrização em (1.7), com média μ_t e precisão ϕ , para $t = 1, 2, \dots, n$. O modelo de regressão beta de Ferrari e Cribari-Neto (2004) é então obtido ao se assumir que a média de y_t pode ser representada por

$$g(\mu_t) = \sum_{i=1}^k x_{ti}\beta_i = \eta_t, \quad (1.8)$$

ou seja, $\mu_t = g^{-1}(\eta_t)$, em que x_{t1}, \dots, x_{tk} são os valores para as k variáveis explicativas com seus respectivos parâmetros β_k ($k < n$), lembrando que para o intercepto $x_{t1} = 1 \forall t$, e $g(\cdot)$ é uma função estritamente monótona e diferenciável até a segunda ordem que associa o intervalo $(0, 1)$ na reta real, é a chamada função de ligação. Como os valores extremos do intervalo não são suportados pelo modelo, frequente-

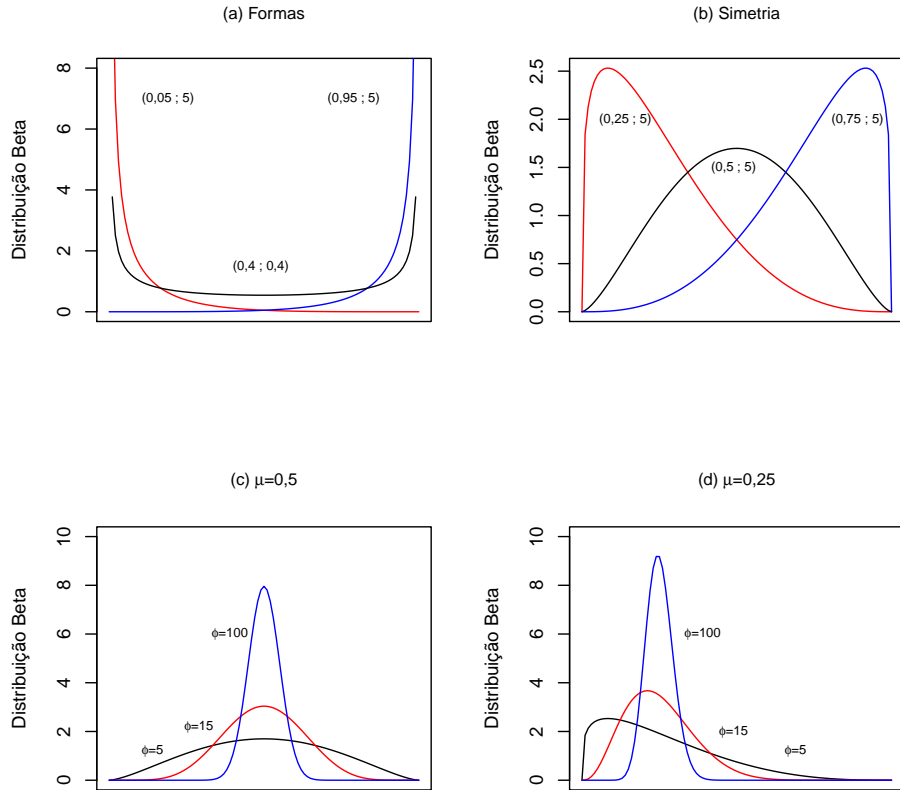


Figura 1.1: Distribuição beta para diferentes (μ, ϕ)

mente se faz a transformação $(y_t(n-1) + 0,5)/n$ na variável resposta para eliminar problemas na estimação, entretanto, caso haja muitas ocorrências desses valores, o mais indicado seria a aplicação do modelo inflado (Cook *et al.*, 2008).

Observe que, a partir de (1.6), pode-se verificar que a variância dos y_t está definida em função das médias μ_t , logo, também depende do valor das variáveis explicativas pela relação (1.8). Dessa forma, o modelo acomoda naturalmente variáveis respostas com variâncias não-constantes, resolvendo o problema da heterocedasticidade existente no modelo de regressão tradicional:

$$var(y_t) = \frac{\mu_t(1 - \mu_t)}{1 + \phi} = \frac{g^{-1}(\mathbf{x}_t\boldsymbol{\beta})[1 - g^{-1}(\mathbf{x}_t\boldsymbol{\beta})]}{1 + \phi} \quad (1.9)$$

em que \mathbf{x}_t corresponde à t -ésima linha da matriz \mathbf{X} de variáveis explicativas e $\boldsymbol{\beta}$ é o vetor de parâmetros do modelo.

Também vale citar que o modelo proposto pode ser utilizado para dados que são restritos a qualquer outro intervalo (c, d) para c e d conhecidos, bastando para isso modelar $(y_t - c)/(d - c)$ no lugar de y_t , com $c < d$.

Sobre as funções de ligação, é importante destacar que existem várias opções de escolha. As mais comuns são:

- logito: $g(\mu_t) = \log\left(\frac{\mu_t}{1-\mu_t}\right)$;
- probito: $g(\mu_t) = \Phi^{-1}(\mu_t)$, em que $\Phi(\cdot)$ é a função de distribuição acumulada normal padrão;
- log-log: $g(\mu_t) = -\log(-\log(\mu_t))$;
- complemento log-log: $\log(-\log(1 - \mu_t))$.

Dentre essas funções de ligação, no contexto da regressão beta, a logito é bastante útil, pois permite que sejam calculadas as razões de chances ($\exp(\beta_i)$), que representam em quanto a mudança em determinada variável modifica a chance, mantendo as demais variáveis constantes. Mas independentemente da função de ligação escolhida, o método de estimação é o mesmo, que segue na próxima Seção.

1.3 Estimação dos parâmetros

Para estimar os parâmetros do modelo de regressão beta via máxima verossimilhança, utiliza-se a função de log-verossimilhança, que tem por base uma amostra de n observações independentes, como segue:

$$\ell(\boldsymbol{\beta}, \phi) = \sum_{t=1}^n \ell_t(\mu_t, \phi), \quad (1.10)$$

em que $\boldsymbol{\beta}' = (\beta_1, \beta_2, \dots, \beta_k)$ é o vetor de tamanho k com os parâmetros do modelo e, pela definição,

$$\ell_t(\mu_t, \phi) = \log \Gamma(\phi) - \log \Gamma(\mu_t \phi) - \log \Gamma((1-\mu_t)\phi) + (\mu_t \phi - 1) \log y_t + ((1-\mu_t)\phi - 1) \log(1-y_t) \quad (1.11)$$

desde que as médias μ_t satisfaçam a relação (1.8).

A função escore $U^T(\boldsymbol{\beta}, \phi) = (U_\beta(\boldsymbol{\beta}, \phi), U_\phi(\boldsymbol{\beta}, \phi))$ é obtida pela diferenciação da função de log-verossimilhança com respeito aos β_i ($U_\beta(\boldsymbol{\beta}, \phi)$) e a ϕ ($U_\phi(\boldsymbol{\beta}, \phi)$). As diferenciações são iguais a

$$U_\beta(\boldsymbol{\beta}, \phi) = \phi \mathbf{X}^T \mathbf{T}(\mathbf{y}^* - \boldsymbol{\mu}^*) \quad (1.12)$$

e

$$U_\phi(\boldsymbol{\beta}, \phi) = \sum_{t=1}^n \{\mu_t(y_t^* - \mu_t^*) + \log(1 - y_t) - \psi((1 - \mu_t)\phi) + \psi(\phi)\}, \quad (1.13)$$

em que \mathbf{X} é uma matriz $n \times k$ que contém os valores das variáveis explicativas para cada uma das observações, \mathbf{y}^* e $\boldsymbol{\mu}^*$ são os vetores de tamanho n com elementos iguais, respectivamente, a

$$y_t^* = \log(y_t/(1 - y_t)) \quad (1.14)$$

e

$$\mu_t^* = \psi(\mu_t\phi) - \psi((1 - \mu_t)\phi), \quad (1.15)$$

sendo $\psi(\cdot)$ a função digamma, e

$$\mathbf{T} = \text{diag}(1/g'(\mu_1), \dots, 1/g'(\mu_n)), \quad (1.16)$$

sendo $g'(\mu_t)$ a derivada de primeira ordem da função de ligação escolhida.

Para obter a matriz de informação de Fisher \mathbf{K} , é preciso diferenciar a função escore com respeito aos parâmetros, obtendo

$$\mathbf{K} = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix} \quad (1.17)$$

com $K_{\beta\beta} = \phi \mathbf{X}^T \mathbf{Z} \mathbf{X}$, $K_{\beta\phi} = K_{\phi\beta}^T = \mathbf{X}^T \mathbf{T} \mathbf{c}$ e $K_{\phi\phi} = \text{tr}(\mathbf{D})$, em que \mathbf{Z} e \mathbf{D} são matrizes diagonais de tamanho n com elementos:

$$z_t = \phi \{\psi'(\mu_t\phi) + \psi'((1 - \mu_t)\phi)\} \frac{1}{g'(\mu_t)^2} \quad (1.18)$$

e

$$d_t = \psi'(\mu_t\phi)\mu_t^2 + \psi'((1-\mu_t)\phi)(1-\mu_t)^2 - \psi'(\phi), \quad (1.19)$$

sendo que $\psi'(\cdot)$ é a função trigamma e \mathbf{c} é um vetor de tamanho n cujos elementos são $c_t = \phi\{\psi'(\mu_t\phi)\mu_t - \psi'((1-\mu_t)\phi)(1-\mu_t)\}$.

Os estimadores de máxima verossimilhança para $\boldsymbol{\beta}$ e ϕ são obtidos ao se igualar as Equações (1.12) e (1.13) a zero. Observe que, ao contrário do que acontece nos Modelos Lineares Generalizados, $\boldsymbol{\beta}$ e ϕ não são ortogonais. Como não possuem forma fechada, os estimadores $\hat{\boldsymbol{\beta}}$ e $\hat{\phi}$ podem ser encontrados ao se maximizar numericamente a função de log-verossimilhança a partir de algum algoritmo não-linear de otimização, como Newton ou Quasi-Newton, por exemplo. Em geral, esses algoritmos envolvem a especificação de um valor de partida ou valor inicial para o processo iterativo, para os quais Ferrari e Cribari-Neto (2004) sugerem que sejam utilizados os valores da regressão linear tradicional (1) dos valores transformados $\check{y}_t = g(y_t)$ em \mathbf{X} para os β_k , ou seja, $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\check{\mathbf{y}}$, e para ϕ , o valor

$$\frac{1}{n} \sum_{t=1}^n \frac{\check{\mu}_t(1-\check{\mu}_t)}{\check{\sigma}_t^2} - 1, \quad (1.20)$$

em que $\check{\mu}_t$ é obtido ao se aplicar $g^{-1}(\cdot)$ aos valores ajustados pela regressão linear tradicional de \check{y}_t em \mathbf{X} , isto é, $\check{\mu}_t = g^{-1}(\mathbf{x}_t(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\check{\mathbf{y}})$, sendo \mathbf{x}_t a t -ésima linha da matriz \mathbf{X} , e

$$\check{\sigma}_t^2 = \frac{\check{\mathbf{e}}^T\check{\mathbf{e}}}{(n-k)g'(\check{\mu}_t)^2}, \quad (1.21)$$

sendo $\check{\mathbf{e}}$ o vetor de resíduos da regressão linear, ou seja, $\check{\mathbf{e}} = \check{\mathbf{y}} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\check{\mathbf{y}}$.

Assim, para um número de observações grande e sob as devidas condições de regularidade para os estimadores de máxima verossimilhança, tem-se que, assintoticamente,

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\phi} \end{pmatrix} \sim N_{k+1} \left(\begin{pmatrix} \boldsymbol{\beta} \\ \phi \end{pmatrix}, \mathbf{K}^{-1} \right),$$

em que \mathbf{K}^{-1} pode ser obtido utilizando inversão de matrizes particionadas, como

segue:

$$\mathbf{K}^{-1} = \begin{pmatrix} K^{\beta\beta} & K^{\beta\phi} \\ K^{\phi\beta} & K^{\phi\phi} \end{pmatrix} \quad (1.22)$$

com

$$K^{\beta\beta} = \frac{1}{\phi} (\mathbf{X}^T \mathbf{Z} \mathbf{X})^{-1} \left[\mathbf{I}_k + \frac{\mathbf{X}^T \mathbf{T} \mathbf{c} \mathbf{c}^T \mathbf{T}^T \mathbf{X} (\mathbf{X}^T \mathbf{Z} \mathbf{X})^{-1}}{\gamma \phi} \right],$$

$$K^{\beta\phi} = (k^{\phi\beta})^T = -\frac{1}{\gamma \phi} (\mathbf{X}^T \mathbf{Z} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{T} \mathbf{c}$$

e $K^{\phi\phi} = \gamma^{-1}$, considerando que $\gamma = \text{tr}(\mathbf{D}) - \phi^{-1} \mathbf{c}^T \mathbf{T}^T \mathbf{X} (\mathbf{X}^T \mathbf{Z} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{T} \mathbf{c}$ e \mathbf{I}_k é a matriz identidade de tamanho k . A fim de facilitar a implementação em qualquer plataforma computacional, construiu-se a Tabela 1.1 que apresenta o algoritmo de estimação para a regressão beta utilizando a função de ligação logito. Lembrando que, para outras funções de ligação, deve-se alterar os valores \tilde{y} , $\tilde{\mu}$ e $g'(\cdot)$ de acordo com a função escolhida. O algoritmo completo encontra-se no Apêndice A, incluindo inferência e algumas medidas de diagnóstico.

Vale lembrar que, de acordo com o trabalho de Ospina *et al.* (2006), esses estimadores de máxima verossimilhança estão sujeitos a viés quando a amostra é pequena. Nesse sentido, os autores apresentam correções para os estimadores, porém, ponderam que a magnitude do viés a que esses estimadores de máxima verossimilhança estão sujeitos é desprezível quando o tamanho da amostra é grande. Ainda sobre pequenas amostras, Bayer e Cribari-Neto (2013) abordam o problema da inferência nesses casos.

Na próxima seção, serão discutidos métodos de inferência e medidas de diagnóstico para o modelo de regressão beta, considerando os estimadores de máxima verossimilhança expostos nesta seção, cujas demonstrações podem ser encontradas no apêndice do trabalho de Ferrari e Cribari-Neto (2004).

Tabela 1.1: Algoritmo de estimação da Regressão Beta com função de ligação logito

n = número de linhas de \mathbf{y}
 $\mathbf{y}_{transformado} = \hat{\mathbf{y}} = g(\mathbf{y}) = \log(\mathbf{y}/(1 - \mathbf{y}))$
 $\boldsymbol{\beta}_{inicial} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{y}}$
 $\hat{\mathbf{e}} = \hat{\mathbf{y}} - \mathbf{X} \boldsymbol{\beta}_{inicial}$
 $\hat{\boldsymbol{\mu}} = g^{-1}(\mathbf{X} \boldsymbol{\beta}_{inicial}) = \exp(\mathbf{X} \boldsymbol{\beta}_{inicial}) / (1 + \exp(\mathbf{X} \boldsymbol{\beta}_{inicial}))$
 $g'(\hat{\boldsymbol{\mu}}) = (1/\hat{\boldsymbol{\mu}}) + (1/(1 - \hat{\boldsymbol{\mu}}))$
 $\hat{\sigma}^2 = \frac{\hat{\mathbf{e}}^T \hat{\mathbf{e}}}{(n-k)g'(\hat{\boldsymbol{\mu}})\#g'(\hat{\boldsymbol{\mu}})}$
 $\phi_{inicial} = 0$
 Para $i = 1$ até n faça {
 $\phi_{inicial} = \phi_{inicial} + \hat{\boldsymbol{\mu}}[i](1 - \hat{\boldsymbol{\mu}}[i]) / (n\hat{\sigma}^2[i])$
 }
 $\boldsymbol{\beta} = \boldsymbol{\beta}_{inicial}$
 $\phi = \phi_{inicial} - 1$
 Função de Máxima Verossimilhança:
 $\max_like(\boldsymbol{\beta}, \phi) = \{$
 $logver1 = lgamma(\phi \boldsymbol{\mu})$
 $logver2 = lgamma((1 - \boldsymbol{\mu})\phi)$
 $logver3 = (\phi \boldsymbol{\mu} - 1)\#log(\mathbf{y})$
 $logver4 = ((1 - \boldsymbol{\mu})\phi - 1)\#log(1 - \mathbf{y})$
 $logver = 0$

 Para $i = 1$ até n faça {
 $max_like = logver + lgamma(\phi) - logver1[i] - logver2[i] + logver3[i] + logver4[i]$
 }
 }
 $it=0$
 $dif=1$
 Enquanto ($|dif| > 10^{-8}$ e $it < \maxit$) faça {
 $\boldsymbol{\mu} = g^{-1}(\mathbf{X} \boldsymbol{\beta}) = \exp(\mathbf{X} \boldsymbol{\beta}) / (1 + \exp(\mathbf{X} \boldsymbol{\beta}))$
 $g'(\boldsymbol{\mu}) = (1/\boldsymbol{\mu}) + (1/(1 - \boldsymbol{\mu}))$
 $\mathbf{y}^* = \log(\mathbf{y}/(1 - \mathbf{y}))$
 $\boldsymbol{\mu}^* = \psi'(\boldsymbol{\mu}\phi) - \psi'((1 - \boldsymbol{\mu})\phi)$
 $\mathbf{T} = \text{diag}(1/g'(\boldsymbol{\mu}))$
 $\mathbf{c} = \phi(\psi''(\boldsymbol{\mu}\phi)\#\boldsymbol{\mu} - \psi''((1 - \boldsymbol{\mu})\phi)\#(1 - \boldsymbol{\mu}))$
 $\mathbf{Z} = \text{diag}((\phi\psi''(\boldsymbol{\mu}\phi) + \phi\psi''((1 - \boldsymbol{\mu})\phi)) / (g'(\boldsymbol{\mu})\#g'(\boldsymbol{\mu})))$
 $\mathbf{D} = \text{diag}(\psi''(g'(\boldsymbol{\mu})\phi)\#g'(\boldsymbol{\mu})\#g'(\boldsymbol{\mu}) + \psi''((1 - g'(\boldsymbol{\mu})\phi)\#(1 - g'(\boldsymbol{\mu})\phi)) - \psi''(\phi))$
 $\mathbf{k}_{\beta\beta} = \phi \mathbf{X}^T \mathbf{Z} \mathbf{X}$
 $\mathbf{k}_{\beta\phi} = \mathbf{X}^T \mathbf{T} \mathbf{c}$
 $\mathbf{k}_{\phi\phi} = \text{tr}(\mathbf{D})$
 $\mathbf{km} = (\mathbf{k}_{\beta\beta} \|\mathbf{k}_{\beta\phi}\|) / (\mathbf{k}_{\beta\phi}^T \|\mathbf{k}_{\phi\phi}\|)$
 $\mathbf{k}^{-1} = \mathbf{km}^{-1}$
 $U_{\beta}(\boldsymbol{\beta}, \phi) = \phi(\mathbf{X}^T \mathbf{T}(\mathbf{y}^* - \boldsymbol{\mu}^*))$
 $U_{\phi}(\boldsymbol{\beta}, \phi) = 0$
 Para $i = 1$ até n faça {
 $U_{\phi}(\boldsymbol{\beta}, \phi) = U_{\phi}(\boldsymbol{\beta}, \phi) + \mu[i](y^*[i] - \mu^*[i]) + \log(1 - y[i]) - \psi'((1 - \mu[i])\phi) + \psi'(\phi)$
 }
 $\mathbf{U} = U_{\beta}(\boldsymbol{\beta}, \phi) \|\mathbf{U}_{\phi}(\boldsymbol{\beta}, \phi)$
 $param = \boldsymbol{\beta} \|\phi$
 $param^{n+1} = param + \mathbf{k}^{-1} \mathbf{U}$
 $dif = \max_like(param) - \max_like(param^{n+1})$
 $\boldsymbol{\beta} = \boldsymbol{\beta}^{n+1}$
 $\phi = \phi^{n+1}$
 $it = it + 1$
 }
 }

$lgamma(.) = \log(\Gamma(.))$

$\#$: multiplicação elemento por elemento

1.4 Diagnóstico e inferência

Após a estimação do modelo, boas práticas em análise de regressão incluem a verificação da qualidade do ajuste e a análise dos resíduos como um requisito para a validade das inferências realizadas a partir dele. Com esse foco, Ferrari e Cribari-Neto (2004) sugerem a utilização de uma medida semelhante ao coeficiente de determinação, chamada de pseudo R^2 (R_p^2), que é calculada como a raiz quadrada da correlação amostral entre $\hat{\eta}$ e $g(y)$ e está definida no intervalo $0 \leq R_p^2 \leq 1$. Quanto mais próximo de 1, maior é a variação dos dados explicada pelo modelo.

Também para a qualidade do ajuste, os autores utilizam o conceito de Deviance ($D(\mathbf{y}; \boldsymbol{\mu}, \phi)$), definida como duas vezes a diferença entre a máxima log-verossimilhança que pode ser atingida pelo modelo saturado e a log-verossimilhança do modelo sob análise, ou seja, $D(\mathbf{y}; \boldsymbol{\mu}, \phi) = \sum_{t=1}^n 2(\ell_t(\tilde{\mu}_t, \phi) - \ell(\mu_t, \phi))$, em que $\tilde{\mu}_t$ é o valor tal que $\partial \ell_t / \partial \mu_t = 0$, isto é, $\phi(y_t^* - \mu_t^*) = 0$, com y_t^* dado por (1.14) e μ_t^* por (1.15). Se ϕ é grande o suficiente, os autores provaram que $\mu_t^* \approx \log(\mu_t / (1 - \mu_t))$ e então $\tilde{\mu}_t \approx y_t$.

Se ϕ é conhecido, então a deviance é $D(\mathbf{y}; \bar{\boldsymbol{\mu}}, \phi)$, sendo $\bar{\boldsymbol{\mu}}$ o estimador de máxima verossimilhança de $\boldsymbol{\mu}$ obtido pelo modelo em análise. Já se o valor de ϕ não é conhecido, pode ser utilizada a aproximação $D(\mathbf{y}; \hat{\boldsymbol{\mu}}, \hat{\phi}) = \sum_{t=1}^n (r_t^d)^2$, em que

$$r_t^d = \text{sgn}(y_t - \hat{\mu}_t) [2(\ell_t(\tilde{\mu}_t, \hat{\phi}) - \ell(\hat{\mu}_t, \hat{\phi}))]^{1/2} \quad (1.23)$$

é chamado de “deviance residual”, que é uma medida que também pode ser utilizada para identificar observações discrepantes, tendo em vista que a t -ésima observação contribui em $(r_t^d)^2$ na deviance do modelo.

Para a análise de resíduos, Cribari-Neto e Zeileis (2010) não aconselham a utilização dos resíduos puros, devido à heterocedasticidade inerente ao modelo (1.9). Nesse sentido, o trabalho de Ferrari e Cribari-Neto (2004) traz também os resíduos padronizados do modelo:

$$r_t = \frac{y_t - \hat{\mu}_t}{\sqrt{\widehat{\text{var}}(y_t)}}, \quad (1.24)$$

com $\hat{\mu}_t = g^{-1}(\mathbf{x}_t \hat{\boldsymbol{\beta}})$ e $\widehat{\text{var}}(y_t) = (\hat{\mu}_t(1 - \hat{\mu}_t)) / (1 + \hat{\phi})$. O ideal para um bom modelo é

que o gráfico desses resíduos pelo índice t não mostre tendências ou padrões. Também pode ser utilizado o gráfico dos resíduos pelo preditor estimado $\hat{\eta}_t$ para verificar se a função de ligação foi bem especificada, caso em que não haverá padrões. Como a distribuição dos resíduos não é conhecida, os autores sugerem ainda a utilização de *half normal plots* com envelope simulado. Espinheira *et al.* (2008b) propõem ainda mais dois resíduos para o modelo de regressão beta.

Para a identificação de observações influentes na estimação dos parâmetros, Ferrari e Cribari-Neto (2004) sugerem a aproximação para a distância de Cook, dada por

$$Cook_t = \frac{h_{tt}r_t^2}{k(1 - h_{tt})^2}, \quad (1.25)$$

que é uma combinação dos resíduos padronizados e da matriz de influência (ou “matriz *hat*”), tal que $\mathbf{y} = \mathbf{H}\hat{\mathbf{y}}$, isto é:

$$\mathbf{H} = \mathbf{Z}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{Z}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}^{1/2} \quad (1.26)$$

em que \mathbf{Z} é a matriz diagonal com elementos dados por (1.18).

Os autores sugerem ainda a utilização da medida de influência generalizada (*generalized leverage*), definida para cada observação pela diagonal da matriz

$$GL(\boldsymbol{\beta}, \phi) = \mathbf{TX}(\mathbf{X}^T\mathbf{ZX})^{-1}\mathbf{X}^T\mathbf{TM} + \frac{1}{\gamma\phi}\mathbf{TX}(\mathbf{X}^T\mathbf{QX})^{-1}\mathbf{X}^T\mathbf{Tf}(\mathbf{f}^T\mathbf{TX}(\mathbf{X}^T\mathbf{QX})^{-1}\mathbf{X}^T\mathbf{TM} - \mathbf{b}^T), \quad (1.27)$$

em que $\mathbf{Q} = \text{diag}(q_1, \dots, q_n)$ com

$$q_t = \left[\phi\psi'(\mu_t\phi) + \psi'((1 - \mu_t)\phi) + (y_t^* - \mu_y^*)\frac{g''(\mu_t)}{g'(\mu_t)} \right] \frac{1}{[g'(\mu_t)]^2},$$

$\mathbf{M} = \text{diag}(m_1, \dots, m_n)$ com $m_t = 1/y_t(1 - y_t)$, \mathbf{f} é o vetor com elementos $f_t = c_t - (y_t^* - \mu_t^*)$ e \mathbf{b} é o vetor com elementos $b_t = -(y_t - \mu_t)/(y_t(1 - y_t))$. Espinheira *et al.* (2008a) sugerem ainda outras medidas de influência.

Se a análise dos resíduos e da qualidade do ajuste for satisfatória, então os resultados da inferência para grandes amostras são confiáveis. Nesse sentido, o

trabalho de Ferrari e Cribari-Neto (2004) traz algumas opções para a construção de intervalos de confiança e três testes sobre os parâmetros, com hipótese $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{(0)}$ contra $H_1 : \boldsymbol{\beta}_1 \neq \boldsymbol{\beta}_1^{(0)}$, em que $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_m)'$ e $\boldsymbol{\beta}_1^{(0)} = (\beta_1^{(0)}, \dots, \beta_m^{(0)})'$ são dados, para $m < k$.

O primeiro teste é o da razão de verossimilhanças, com estatística $\omega_1 = 2[\ell(\hat{\boldsymbol{\beta}}, \hat{\phi}) - \ell(\tilde{\boldsymbol{\beta}}, \tilde{\phi})]$, em que $\ell(\tilde{\boldsymbol{\beta}}, \tilde{\phi})$ é a máxima log-verossimilhança restrita aos valores da hipótese nula. Sob H_0 , $\omega_1 \xrightarrow{D} \chi_m^2$, ou seja, a estatística do teste segue assintoticamente uma distribuição qui-quadrado com m graus de liberdade.

O segundo teste é o de escore, com estatística $\omega_2 = \tilde{\mathbf{U}}_{1\beta}^T \tilde{\mathbf{K}}_{11}^{\beta\beta} \tilde{\mathbf{U}}_{1\beta}$, em que $\tilde{\mathbf{U}}_{1\beta} = \phi \mathbf{X}_1^T \mathbf{T}(\mathbf{y}^* - \boldsymbol{\mu}^*)$ é o vetor que contém os m primeiros elementos da função escore para $\boldsymbol{\beta}$, com a mesma partição de $\boldsymbol{\beta}_1$ para $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2]$ e $\mathbf{K}_{11}^{\beta\beta}$ é a matriz formada pelas m primeiras linhas e m primeiras colunas da matriz \mathbf{K}^{-1} (1.22). O til indica que os valores são avaliados sob a hipótese nula. Considerando H_0 , tem-se que $\omega_2 \xrightarrow{D} \chi_m^2$.

Por fim, o trabalho traz a estatística para o teste Wald sob a mesma hipótese H_0 , $\omega_3 = (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)})^T (\hat{\mathbf{K}}_{11}^{\beta\beta})^{-1} (\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)})$, em que $\hat{\mathbf{K}}_{11}^{\beta\beta}$ é o valor para $\mathbf{K}_{11}^{\beta\beta}$ irrestrito e $\hat{\boldsymbol{\beta}}_1$ é a estimativa de máxima verossimilhança de $\boldsymbol{\beta}_1$. Tem-se que sob H_0 , $\omega_3 \xrightarrow{D} \chi_m^2$. Para testar a significância de apenas um parâmetro, pode-se utilizar a adaptação da estatística de Wald $\hat{\beta}_i / se(\hat{\beta}_i)$, em que $se(\hat{\beta}_i)$ é o erro padrão do estimador de máxima verossimilhança obtido pela inversa da matriz de informação de Fisher, nos valores estimados, com distribuição aproximadamente normal padrão sob a hipótese nula.

Para a construção de intervalos de confiança para os β_i , Ferrari e Cribari-Neto (2004) sugerem a inversão dos testes citados acima ou a aproximação $\hat{\beta}_i \pm \Phi^{-1}(1 - \alpha/2) se(\hat{\beta}_i)$, com confiança de $(1 - \alpha)100\%$. Da mesma forma, um intervalo aproximado para o parâmetro de precisão ϕ pode ser $\hat{\phi} \pm \Phi^{-1}(1 - \alpha/2) se(\hat{\phi})$, em que $\hat{\phi} = \hat{\gamma}^{-1/2}$. Para o valor médio da variável resposta $\boldsymbol{\mu}$ para uma dada combinação de variáveis explicativas \mathbf{X} , tem-se o intervalo

$$[g^{-1}(\hat{\boldsymbol{\eta}} - \Phi^{-1}(1 - \alpha/2) se(\hat{\boldsymbol{\eta}})); g^{-1}(\hat{\boldsymbol{\eta}} + \Phi^{-1}(1 - \alpha/2) se(\hat{\boldsymbol{\eta}}))],$$

com $\hat{\boldsymbol{\eta}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ e $se(\hat{\boldsymbol{\eta}}) = \sqrt{\mathbf{X}^T \widehat{cov}(\hat{\boldsymbol{\beta}}) \mathbf{X}}$, sendo a covariância estimada obtida da matriz de informação de Fisher avaliada no valor estimado, excluindo-se a linha e a

coluna correspondentes ao parâmetro de precisão.

1.5 Simulações

De acordo com Meaney e Moineddin (2014), a performance do modelo de regressão clássica para dados restritos e contínuos é razoável quando os dados são simétricos no meio do intervalo em análise. A fim de melhor compreender a vantagem da utilização do modelo de regressão beta em relação ao modelo clássico, elaborou-se um estudo com dados simulados nas diversas formas que a distribuição beta pode assumir.

Para tal, Gupta e Nadarajah (2004) apontaram as condições em que a distribuição assume suas diversas formas em termos da parametrização usual da beta (1.2). Com essas condições transformadas para a reparametrização (1.7) utilizada no modelo de regressão beta, construiu-se a Tabela 1.2, que apresenta os valores aproximados de referência para obter cada uma das formas da distribuição beta.

Tabela 1.2: Condições para simulação da regressão beta

Forma	Condições
Unimodal:	$a > 1$ e $b > 1 \Leftrightarrow 1/\phi < \mu < (\phi - 1)/\phi$
Simétrica:	$a = b \Leftrightarrow \mu = 0,5$
J:	$0 < b < 1 \Leftrightarrow \mu > (\phi - 1)/\phi$
J invertido:	$0 < a < 1 \Leftrightarrow \mu < 1/\phi$

Ainda, considerando a função de ligação logito, observou-se que para um dado valor de μ que se deseja atingir com ϕ fixo, pode-se utilizar qualquer valor para os parâmetros das variáveis explicativas desde que $\beta_0 = \log(\mu_{alvo}/(1 - \mu_{alvo})) - \sum \beta_i x_i$, sendo que para valores $|\beta_i| > 1$ apurou-se que o valor de μ real tende a fugir do valor de μ_{alvo} pretendido, de forma que foi possível simular a distribuição em forma de “U”. Assim, os dados foram simulados para a regressão beta em seus distintos formatos, como descreve o algoritmo da Tabela 1.3, para duas variáveis explicativas.

Para $n = 200, 500$ e 1000 observações, repetiu-se o algoritmo de simulação 100 vezes para cada uma das formas da distribuição e utilizou-se como uma medida de conformidade do modelo a porcentagem de vezes em que pelo menos um valor

Tabela 1.3: Algoritmo de construção dos dados simulados para a regressão beta com ligação logito

Defina o número de observações n
Defina o valor de μ_{alvo} e de ϕ
Assuma uma distribuição para os parâmetros:
$\beta_1 \sim N(0,5;0,1)$
$\beta_2 \sim N(-0,2;0,05)$
Assuma uma distribuição para as n variáveis explicativas:
$x_1 \sim N(4,1)$
$x_2 \sim \text{Poisson}(2)$
Faça $\beta_0 = \log(\mu_{alvo}/(1 - \mu_{alvo})) - \beta_1\bar{x}_1 - \beta_2\bar{x}_2$
Para cada observação, faça:
$\eta = \beta_0 + \beta_1x_1 + \beta_2x_2$
$\mu = e^\eta/(1 + e^\eta)$
$a = \mu\phi$
$b = (1 - \mu)\phi$
$y \sim \text{beta}(a, b)$

\bar{x}_i é a média de x_i

estimado caiu fora do intervalo (0, 1). A Tabela 1.4 apresenta os resultados dessas simulações. Cumpre informar que, devido às condições da Tabela 1.2, na simulação das distribuições em forma de “J” e “J invertido” optou-se por utilizar um valor de ϕ menor, de modo que o valor necessário de μ_{alvo} para atingir essa forma não tivesse muitas casas decimais.

Analisando os resultados, note que na regressão beta não existe a possibilidade de o valor estimado cair fora do intervalo. Já para a regressão clássica, constatou-se que, de modo geral, quanto maior o número n de observações, maior é a chance de os modelos oferecerem previsões fora do intervalo (0, 1); e mesmo no caso em que ela teria seu melhor desempenho para dados contínuos e restritos, ou seja, quando é simétrica em torno de $\mu = 0,5$ (Meaney e Moineddin, 2014), os resultados obtidos foram contrários à sua utilização, a depender do rigor do pesquisador, tendo em vista que, na simulação realizada, entre 11% e 25% dos modelos forneceram pelo menos um valor predito fora do intervalo. Para as situações extremas, em forma de “J” ou de “U”, a grande maioria dos modelos clássicos não apresentaram conformidade com as características dos dados (91% a 100% dos modelos estimaram dados fora do intervalo). Nos casos em que os valores estimados permaneceram dentro do intervalo,

Tabela 1.4: Resultados Simulação - Proporção de predições fora do intervalo

Forma	Regressão Clássica	Regressão Beta
<i>n</i> = 200		
Levemente assimétrica à direita ($\mu = 0,25; \phi = 200$)	72%	0%
Simétrica ($\mu = 0,50; \phi = 200$)	11%	0%
Levemente assimétrica à esquerda ($\mu = 0,75; \phi = 200$)	60%	0%
J ($\mu = 0,97; \phi = 20$)	92%	0%
J invertido ($\mu = 0,04; \phi = 20$)	91%	0%
U ($\mu = 0,50; \phi = 200$), com $\beta_1 = 3$ e $\beta_2 = -1$	100%	0%
<i>n</i> = 500		
Levemente assimétrica à direita ($\mu = 0,25; \phi = 200$)	92%	0%
Simétrica ($\mu = 0,50; \phi = 200$)	23%	0%
Levemente assimétrica à esquerda ($\mu = 0,75; \phi = 200$)	77%	0%
J ($\mu = 0,97; \phi = 20$)	97%	0%
J invertido ($\mu = 0,04; \phi = 20$)	96%	0%
U ($\mu = 0,50; \phi = 200$), com $\beta_1 = 3$ e $\beta_2 = -1$	100%	0%
<i>n</i> = 1000		
Levemente assimétrica à direita ($\mu = 0,25; \phi = 200$)	90%	0%
Simétrica ($\mu = 0,50; \phi = 200$)	25%	0%
Levemente assimétrica à esquerda ($\mu = 0,75; \phi = 200$)	86%	0%
J ($\mu = 0,97; \phi = 20$)	98%	0%
J invertido ($\mu = 0,04; \phi = 20$)	100%	0%
U ($\mu = 0,50; \phi = 200$), com $\beta_1 = 3$ e $\beta_2 = -1$	100%	0%

a análise dos resíduos mostrou-se seriamente inadequada.

Evidentemente, essa comparação não é a mais justa, posto que os dados analisados pela regressão beta são transformados enquanto que na regressão clássica não o são. Todavia, julga-se relevante a análise a fim de verificar que quanto maior for a fuga à normalidade, maior será o distanciamento do modelo clássico à realidade desses dados provenientes de taxas ou proporções.

Destarte, o modelo de regressão beta mostra-se mais apropriado tanto conceitualmente, considerando a natureza dos dados, como na prática, dadas as evidências fornecidas pelo estudo exposto.

1.6 Adaptações ao modelo de Regressão Beta

As Seções anteriores abordaram a base da regressão beta. Mas apesar de ser relativamente recente, esse modelo já foi bastante explorado e várias adaptações além do que foi exposto foram propostas, como o modelo bayesiano (Branscum *et al.*, 2007); o modelo inflado, que aceita os valores extremos 0 e 1 (Cook *et al.*, 2008); o modelo misto, que inclui efeitos aleatórios (Zimprich, 2010); modelos adaptados para séries temporais (Rydlewski, 2007; Rocha e Cribari-Neto, 2009) e para dados longitudinais ou em painel (Hunger *et al.*, 2012), dentre outros estudos.

Para lidar com a não estacionariedade e/ou dependência espacial, foi proposto por Simas *et al.* (2010) um modelo generalizado que permite que o parâmetro de precisão seja diferente para cada observação, cuja estrutura poderia ser utilizada de alguma forma para incorporar o padrão de dependência espacial. Nesse caso, assume-se que cada variável resposta y_i segue a reparametrização (1.7) com média μ_i e precisão ϕ_i , e assim o modelo é obtido ao se assumir que a média e parâmetro de precisão satisfazem, respectivamente,

$$g_1(\mu_i) = \eta_{1i} \tag{1.28}$$

e

$$g_2(\phi_i) = \eta_{2i} \tag{1.29}$$

em que g_1 é uma função de ligação que mapeia o intervalo $(0, 1)$ na reta real, g_2 também é uma função de ligação, mas que mapeia o intervalo $(0, \infty)$ na reta real e os η_{1i} e η_{2i} são os preditores em função das variáveis explicativas, podendo ser lineares ou não. A estimação é análoga à do modelo de Ferrari e Cribari-Neto (2004). Observe, entretanto, que esse modelo não fornece estatísticas locais. Igualmente, um modelo com as características dos modelos espaciais globais de Anselin (1988) foi proposto por Cepeda-Cuervo e Núñez-Antón (2013). Além desses, encontrou-se ainda um modelo com apelo local que foi proposto por Schmid *et al.* (2013), porém depende da definição de uma função de suavização da superfície via *P-spline*, o que é desnecessário na Regressão Geograficamente Ponderada, tema do próximo Capítulo.

Capítulo 2

Regressão Geograficamente Ponderada

2.1 Introdução

Como já foi discutido anteriormente, o objetivo deste trabalho é modelar dados restritos ao intervalo $(0,1)$ que apresentam dependência espacial. O modelo de Regressão Beta, apresentado no Capítulo anterior, cuida da parte da modelagem que geralmente envolve taxas, proporções e índices. Porém, ele assume que os parâmetros são constantes no espaço, sendo que, na verdade, é razoável supor que exista alguma relação intrínseca dos dados que se configurem como uma variação espacial dos parâmetros, como por exemplo o comportamento das pessoas, a cultura ou a política de cada local ou a falta de alguma variável no modelo como simplificador da realidade (Fotheringham *et al.*, 2002).

Nesse contexto, surge o modelo de Regressão Geograficamente Ponderada (RGP), proposto por Fotheringham *et al.* (1996), sendo definido como uma técnica não paramétrica descritiva para explorar a não estacionariedade e/ou a dependência espacial dos dados, a qual se configura quando locais próximos apresentam valores semelhantes para alguma variável de interesse (dependência positiva) ou, menos comum, quando locais próximos apresentam valores discrepantes (dependência negativa).

No contexto da regressão, a condição de não estacionariedade espacial significa que um modelo global, como é o caso da regressão beta, não é capaz de refletir a vari-

ação intrínseca dos dados. O ideal então seria a utilização de modelos que levassem em consideração essa estrutura local da relação entre as variáveis. Fotheringham *et al.* (2002) consideram que os modelos locais atuam como microscópios, ampliando o conhecimento que se tem sobre áreas menores, corroborando as informações da Tabela 1.

A Regressão Geograficamente Ponderada então tenta contornar essa limitação dos modelos globais ao captar essa variação por meio da calibragem de um modelo de regressão múltipla que permite que existam diferentes relações entre a variável resposta e o conjunto de variáveis explicativas para diferentes pontos no espaço. Esse modelo tem se mostrado bastante eficiente, conforme os estudos de Mulatu *et al.* (2013) em demografia, de de Sousa *et al.* (2013) em geografia, de Bitter *et al.* (2007) na determinação de preços de casas e de Zhang e Shi (2004) em engenharia florestal, dentre outros.

Nas próximas seções serão discutidas a especificação do modelo e de seus métodos de estimação e as questões relativas à inferência, à dependência e à não estacionariedade espacial. Ainda, serão tratados assuntos ligados ao diagnóstico do ajuste e às extensões já propostas ao modelo.

2.2 O modelo de Regressão Geograficamente Ponderada

De acordo com Brunson *et al.* (1996), o principal objetivo da análise espacial é identificar a natureza da relação existente entre um conjunto de variáveis entre si e com o espaço. Com o intuito de descrever e mapear essa relação quando há variação espacial, foi proposto o modelo de Regressão Geograficamente Ponderada - RGP.

Tendo por base o modelo de Regressão Linear (1), algumas técnicas para modelar dados espaciais não estacionários e/ou com dependência espacial já existiam antes da proposição do modelo em questão, porém com algumas desvantagens em relação à RGP. Em resumo, algumas dessas técnicas são:

- Inclusão de variáveis *dummy*: consiste em incluir no modelo de regressão linear

tradicional variáveis *dummy* ou indicadoras que informam o espaço geográfico onde a observação foi coletada. Essa variável geralmente é do tipo

$$I_{t\ell} = \begin{cases} 1, & \text{se a observação } t \text{ foi coletada no local } \ell \\ 0, & \text{caso contrário.} \end{cases} \quad (2.1)$$

Note que nesse caso há apenas um deslocamento do intercepto para cada local distinto, o que causa uma variação na média global do modelo para cada localidade diferente, sem contudo haver variação nos demais parâmetros do modelo (inclinação), como exemplificado na Figura 2.1.

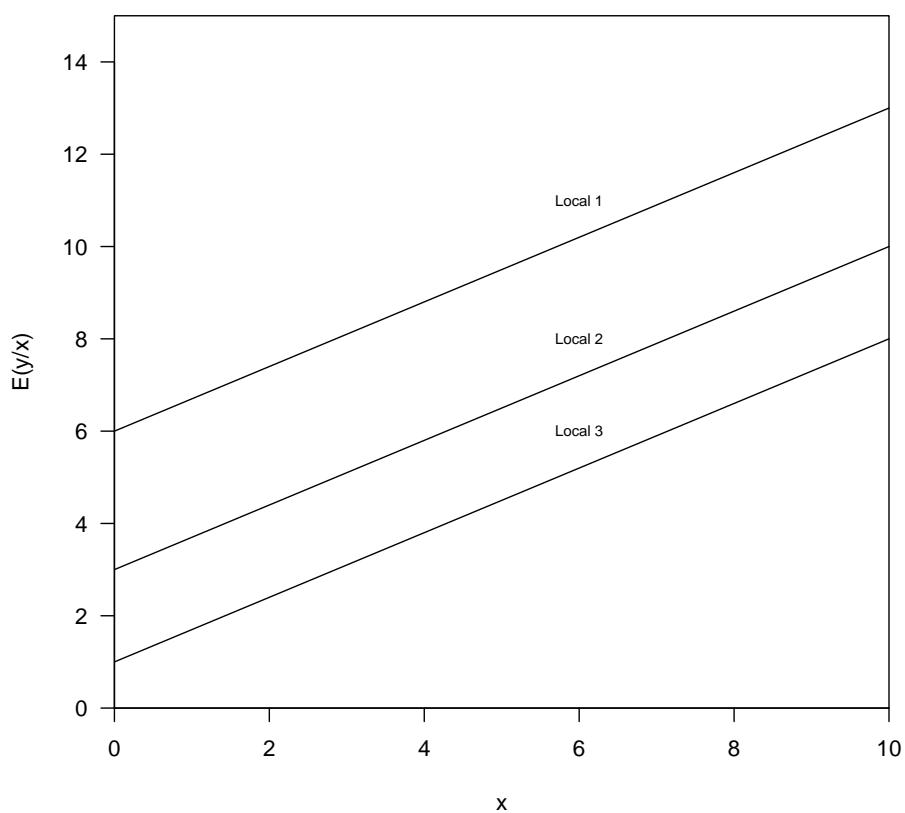


Figura 2.1: Exemplo de inclusão de variável *dummy*

- Um modelo global para cada local: Apesar de apresentar um indicativo da existência ou não da heterogeneidade espacial, se os parâmetros apresentarem valores muito distintos, ajustar um modelo para cada localidade do estudo pode

não ser uma boa estratégia, pois, além de ser um trabalho árduo, cada região pode ter quantidades diferentes de observações, resultando em estimativas com muita dispersão e podendo haver inclusive casos em que não seja nem possível ajustar o modelo, a depender do valor das variáveis. Outro inconveniente é que dessa forma, assume-se que dentro das regiões os parâmetros são constantes e que pode haver variações muito drásticas de uma região para outra, quando na verdade se sabe que dados espaciais geralmente apresentam correlação suave e que não obedecem necessariamente às fronteiras geográficas.

- Método de expansão (Casetti, 1972): assume que os parâmetros do modelo global são funções do espaço geográfico ($y_t = \sum_j x_{tj}\beta_j(p_j) + \varepsilon_t$, $t = 1, \dots, n$, em que p_j é a localização geográfica da observação j). Dessa forma, é possível mapear a variação dos parâmetros, porém, a desvantagem deste método em relação à RGP é que ele exige do pesquisador um elevado conhecimento sobre a relação determinística dos parâmetros em cada uma das localidades, o que nem sempre é simples, podendo envolver relações bastante complexas. Já a RGP permite que os parâmetros para cada local possam ser estimados e mapeados sem a necessidade de se ajustar uma superfície.
- Filtro espacial adaptativo, ou do inglês *Spatial Adaptive Filtering - SAF* (Foster e Gorr, 1986): com inspiração nas séries temporais, incorpora a variação nos parâmetros de maneira *ad-hoc* para “corrigir” as estimativas devido ao *lag*, de tal maneira que permite o mapeamento dos parâmetro mas não é possível aplicar testes estatísticos (Brunsdon *et al.*, 1996; Fotheringham *et al.*, 2002).
- Modelo de coeficientes aleatórios (Aitkin, 1996) e modelagem multinível (Goldstein, 1987): segundo Brunsdon *et al.* (1996), esses métodos assumem que os parâmetros são variáveis aleatórias, em uma abordagem bayesiana. Neste, a distribuição normal é utilizada, naquele, uma mistura finita de distribuições. A principal desvantagem é que nenhum deles assume um padrão de associação espacial e também necessitam da definição de distribuições *a priori*, ao passo que a RGP assume uma associação espacial do tipo que decresce conforme a distância aumenta (*distance-decay*), que parece ser mais apropriado por deixar

que os próprios dados deem forma a essa associação, sem a necessidade de uma intervenção subjetiva do pesquisador.

- Modelos de regressão espacial: de acordo com Fotheringham *et al.* (2002), são modelos que tentam incorporar a dependência espacial na estrutura de regressão, por meio, por exemplo, da flexibilização das suposições de erros independentes, utilizando uma matriz de covariância que representa a distância entre os elementos. Entretanto, esses modelos não são considerados modelos locais por originarem apenas parâmetros globais (vide Tabela 1). Como exemplo, tem-se o modelo de Krigagem (Krige, 1966) e o modelo autorregressivo espacial (Ord, 1975).
- Regressão por janelas móveis: é uma forma rudimentar da RGP (Fotheringham *et al.*, 2002), aplicando uma regressão tradicional a cada região do estudo, sem considerar a área geográfica como limite das relações espaciais, mas sim um *grid* ou janelas independentes. Porém, devido a essa metodologia, os resultados dependem em certo grau do tamanho da “janela” e apresentam certa descontinuidade entre as regiões, onde se espera que o processo de correlação espacial seja contínuo.

O modelo de Regressão Geograficamente Ponderada procura contornar as limitações dos modelos citados acima partindo de uma extensão do próprio modelo de regressão tradicional (1), permitindo que os parâmetros da regressão variem de acordo com a localidade de forma contínua, sem ajustar uma superfície a eles (não-paramétrico) e de modo que possam ser estatisticamente testados, fazendo o uso de uma associação espacial do tipo *distance-decay*, em que pontos mais próximos possuem um peso maior na estimação do que pontos mais distantes. O fato de ter por base o modelo de regressão tradicional e ainda assim incorporar as relações espaciais locais torna o modelo RGP mais intuitivo (Fotheringham *et al.*, 2002), sendo uma vantagem em relação aos modelos listados. Por outro lado, a principal dificuldade se encontra na calibragem do modelo, já que geralmente envolve um esforço computacional maior.

Para a definição do modelo, considere que existam n observações e k variáveis

explicativas. O modelo de RGP é dado pela expressão:

$$y_i = \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad (2.2)$$

em que (u_i, v_i) representa as coordenadas do i -ésimo ponto no espaço, para $i = 1, \dots, n$, $\beta_k(u_i, v_i)$ é o parâmetro para a k -ésima variável explicativa, em função do local da i -ésima observação, x_{ik} é o valor da k -ésima variável explicativa para o local i e ε_i é o erro associado à i -ésima observação, com suposição de que são independentes e identicamente distribuídos $\varepsilon_i \sim N(0, \sigma^2)$.

Observe que, apesar de os coeficientes da regressão variarem geograficamente, assume-se que a variância dos erros é fixa. Também, é importante frisar que os parâmetros β_{ik} , como uma simplificação da notação $\beta_k(u_i, v_i)$, são obtidos por meio das observações coletadas próximas do ponto i , ou seja, o ponto i onde as estimativas dos parâmetros são calculadas (pontos de regressão) não necessariamente precisam se referir aos pontos de coleta das informações (pontos amostrais), embora caso sejam pontos coincidentes, há a possibilidade de cálculo de resíduos e de predições de novas observações.

O principal produto deste modelo é um conjunto de parâmetros calibrados para cada localidade, o que permite a visualização da variação espacial dos mesmos a partir de mapas e ajuda a compreender melhor o fenômeno em estudo. Os parâmetros β_{ik} podem ser estimados via mínimos quadrados ponderados, uma adaptação da técnica de mínimos quadrados ordinários (2), que atribui pesos w_i em diferentes observações para gerar as estimativas. Matricialmente, tem-se

$$\hat{\beta}_i = (\mathbf{X}^T \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_i \mathbf{y}, \quad (2.3)$$

em que $\hat{\beta}_i$ é o vetor de tamanho k com as estimativas para os parâmetros na localidade i , \mathbf{X} é a matriz $n \times k$ com o valor das k variáveis explicativas para as n observações, \mathbf{W}_i é a matriz de ponderação para a localidade i , de dimensão $n \times n$, e \mathbf{y} é o vetor de tamanho n que contém o valor das variáveis resposta nesse ponto. Note que essas estimativas são viesadas, devido aos pesos.

A interpretação dos parâmetros é análoga à interpretação da regressão clássica (1), só que para uma superfície de parâmetros. A variação na magnitude das estimativas indica as mudanças locais na influência da variável explicativa sobre a variável resposta (Charlton *et al.*, 2009). Por exemplo, na Figura 2.2, nas áreas mais escuras a influência da variável explicativa é mais forte que nas mais claras.

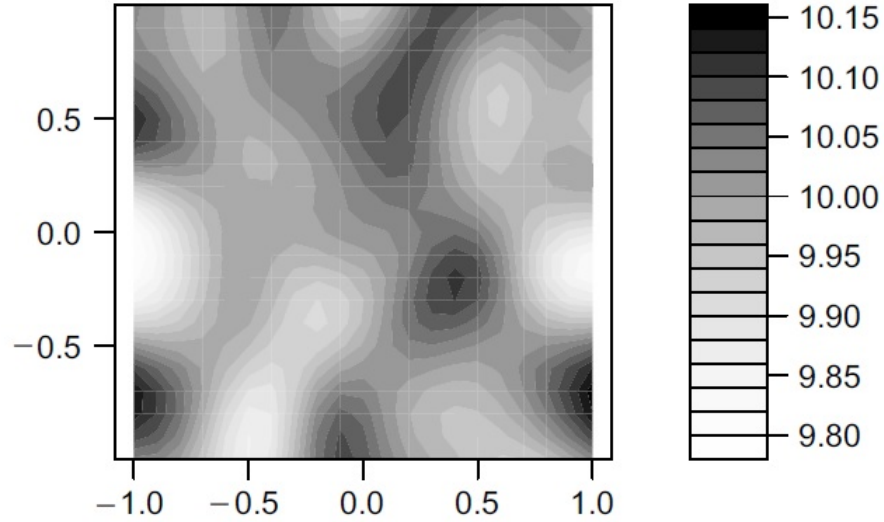


Figura 2.2: Exemplo de superfície de parâmetros estimada

Fonte: Fotheringham *et al.* (2002)

Dadas as estatísticas locais, também é importante o cálculo dos erros-padrão locais. Ao se reescrever a Equação (2.3) da seguinte maneira:

$$\hat{\beta}_i = \mathbf{C}y, \quad (2.4)$$

com $\mathbf{C} = (\mathbf{X}^T \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_i$, tem-se, analogamente à regressão clássica, que a variância das estimativas dos parâmetros é dada por:

$$\text{var}(\hat{\beta}_i) = \mathbf{C} \mathbf{C}^T \sigma^2, \quad (2.5)$$

em que σ^2 é estimado pela soma de quadrados dos resíduos normalizados para a regressão local

$$\hat{\sigma}^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n - 2\nu_1 + \nu_2}, \quad (2.6)$$

com $\nu_1 = \text{tr}(\mathbf{S})$ e $\nu_2 = \text{tr}(\mathbf{S}^T\mathbf{S})$, em que \mathbf{S} é a matriz conhecida como “matriz *hat*”, tal que $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$, ou seja, $\mathbf{S} = \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}$, válida para quando os pontos amostrais e de regressão são coincidentes. O termo $n - 2\nu_1 + \nu_2$ é o número efetivo de graus de liberdade dos resíduos e o termo $2\nu_1 - \nu_2$ é o número efetivo de parâmetros estimados pelo modelo. Como os valores $\text{tr}(\mathbf{S})$ e $\text{tr}(\mathbf{S}^T\mathbf{S})$ são muito próximos, é comum utilizar apenas ν_1 como uma aproximação para o número efetivo de parâmetros no modelo (Fotheringham *et al.*, 2002). Para encontrar o erro-padrão, basta aplicar a raiz quadrada sobre os valores estimados pela Equação (2.5). Esses valores também podem ser colocados em um mapa, para verificar a diferença na dispersão dos parâmetros para cada região.

Por fim, vale citar ainda que a RGP também produz localmente outras medidas de diagnóstico e ajuste dos modelos, que podem ser encontradas no Capítulo 9 de Fotheringham *et al.* (2002), como por exemplo o coeficiente de determinação local R_i^2 , que fornece uma noção de quão bem os modelos locais foram ajustados. Para tanto, utiliza-se a fórmula

$$R_i^2 = \frac{TSS_i^w - RSS_i^w}{TSS_i^w}, \quad (2.7)$$

em que $TSS_i^w = \sum_j w_{ij}(y_j - \bar{y})^2$ e $RSS_i^w = \sum_j w_{ij}(y_j - \hat{y}_j)^2$, $j = 1, \dots, n$, são a soma de quadrados total e a soma de quadrados residual geograficamente ponderadas, respectivamente.

2.3 Considerações sobre a Matriz de Ponderação Espacial

Os valores da matriz de ponderação \mathbf{W} devem ser calculados para cada ponto i . Observe que, dessa forma, cada ponto amostral recebe um peso diferente na estimação em i , variando de acordo com cada localidade, ou seja, exige-se que seja feita uma calibragem para cada ponto de interesse. No contexto dos dados com correlação espacial, a ideia é que os pesos sejam uma medida de proximidade da observação ao ponto de estimação i , e esse é um aspecto bastante importante na estimação do modelo RGP: a determinação da função de ponderação.

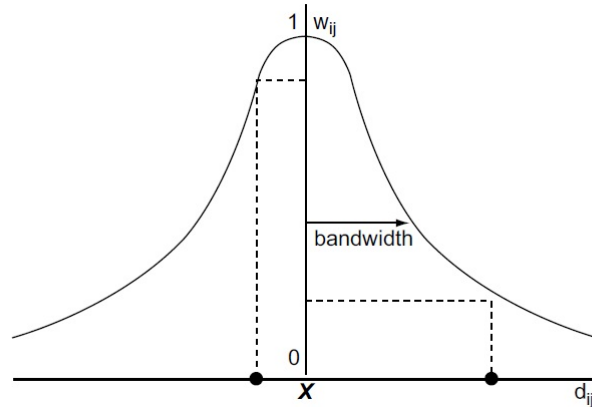


Figura 2.3: Função *Kernel* Espacial

Fonte: Fotheringham *et al.* (2002)

- : ponto amostral
- x: ponto de regressão

Na literatura são encontradas diversas formas para a escolha da função de ponderação espacial, a qual fornece os componentes da matriz \mathbf{W} , mas há consenso de que elas devem possuir as características de uma função *Kernel*, como $\lim_{d \rightarrow \infty} K(d) = 0$ e ser uma função monótona decrescente para números reais positivos, conforme exemplificado na Figura 2.3. Um exemplo de função de ponderação espacial que considera a distância d_{ij} entre o ponto de estimação i e o ponto amostral j é:

$$w_{ij} = \begin{cases} 1, & d_{ij} < d \\ 0, & \text{caso contrário} \end{cases} \quad (2.8)$$

em que d é fixado arbitrariamente. Observe que nesse caso, se d for muito grande, de modo que todas as observações sejam utilizadas na estimação de β_{ik} com peso unitário, então tem-se na verdade o estimador de Mínimos Quadrados Ordinários. Da mesma forma, não é desejável que d seja muito pequeno, já que pode ocorrer de apenas o próprio ponto ser utilizado na estimação. A aplicação dessa função remete à abordagem de janelas móveis citadas anteriormente, com uma descontinuidade abrupta ao se aproximar da distância d , já que não há decaimento da função para distâncias maiores. Uma forma de não haver descontinuidade seria utilizar uma

função contínua exponencial quadrática do tipo

$$w_{ij} = e^{-\frac{d_{ij}^2}{2b^2}}, \quad (2.9)$$

que assume valores decrescentes quanto maior for a distância d_{ij} , na forma da distribuição normal. Porém, desse modo, a função de ponderação associará um peso para cada ponto de interesse, o que aumenta o esforço computacional. Sugere-se então a utilização de uma mistura das estruturas em (2.8) e (2.9), exemplificada com a função bi-quadrática:

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2, & d_{ij} < b \\ 0, & \text{caso contrário.} \end{cases} \quad (2.10)$$

Em vários trabalhos, $b > 0$ é comumente chamado de *bandwidth* e funciona como um parâmetro de suavização da função de ponderação: quanto menor for o seu valor, maior será o decaimento da função e menor importância será designada à informação j na estimação de i , e vice-versa. Observe que (2.9) e (2.10) seguem as propriedades da função *Kernel* (vide Figura 2.3), de modo que as observações mais próximas da localidade i exercem maior influência sobre a estimação neste ponto, em comparação com aquelas observações que se encontram mais distantes.

As funções citadas acima são chamadas de *Kernel* espacial fixas, uma vez que assumem apenas um valor para o limite b . De acordo com Fotheringham *et al.* (2002), a utilização desse tipo de função pode ser equívoca quando os dados são esparsos em uma região e mais densos em outra, já que no primeiro caso a estimação se dará com menos dados do que no segundo caso, podendo ocasionar estimativas com alta dispersão (vide Figura 2.4).

Nesse caso, Fotheringham *et al.* (2002) sugerem a utilização de *Kernels* em que o parâmetro b se adapta a essas diferenças de densidade utilizando a ideia de R vizinhos próximos. Para tal, ordena-se a distância entre os pontos i e j e assume-se

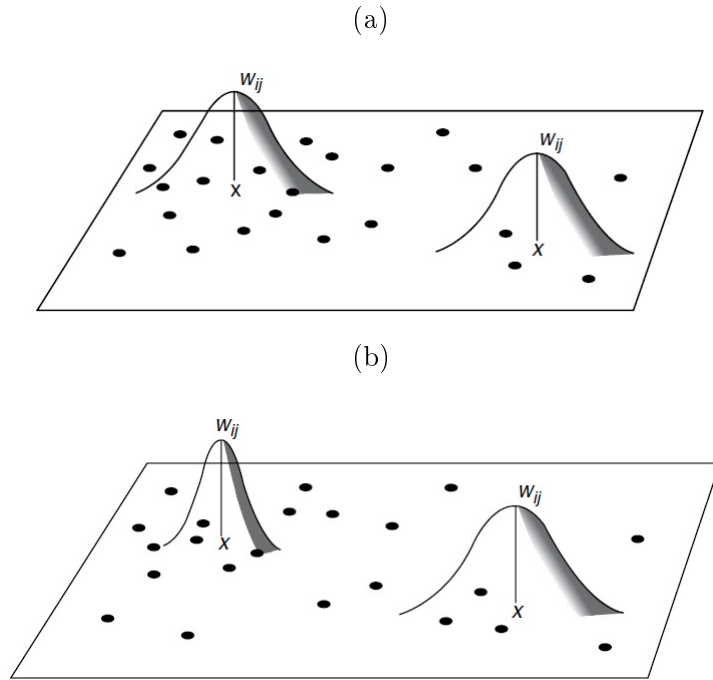


Figura 2.4: *Kernel* (a) Fixo e (b) Variável

Fonte: Fotheringham *et al.* (2002)

que R_{ij} represente essa ordem. Algumas funções que utilizam essa ideia são:

$$w_{ij} = \begin{cases} 1, & \text{se } j \text{ é um dos } R \text{ vizinhos mais próximos de } i \\ 0, & \text{caso contrário} \end{cases} \quad (2.11)$$

ou

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2, & \text{se } j \text{ é um dos } R \text{ vizinhos mais próximos de } i \\ 0, & \text{caso contrário.} \end{cases} \quad (2.12)$$

em que b é a distância de i até o R -ésimo vizinho. Ou ainda

$$w_{ij} = e^{-\frac{R_{ij}}{b}}. \quad (2.13)$$

Um outro modo para que o parâmetro de suavização seja variável seria garantir que a soma dos pesos w_{ij} na função (2.9) ou (2.10) seja igual a alguma constante C arbitrária.

Em todas essas funções, d geralmente é a distância euclidiana em relação ao centroide da região, quando são utilizadas coordenadas cartesianas, porém, nada impede que seja utilizado outro tipo de medida de distância, como coordenadas UTM¹ ou até mesmo tempo de deslocamento entre os locais.

Apesar das inúmeras opções possíveis para escolha da função de ponderação, Brunson *et al.* (1998) argumentam que devido à característica de *distance-decay* e à suavidade das funções, os resultados da RGP são pouco afetados pela escolha da função de ponderação. Nesse sentido, de acordo com Leung *et al.* (2000), a função exponencial (2.9) é a mais utilizada. Por outro lado, segundo Charlton *et al.* (2009), a escolha do *bandwidth* b é mais crítica, o que faz com que a escolha ótima desse valor seja uma etapa de grande importância na calibragem da RGP.

Para a escolha do valor de b ótimo, uma solução intuitiva seria encontrar o valor b que minimize a soma $\sum_{i=1}^n (y_i - \hat{y}_i^*(b))^2$, em que $\hat{y}_i^*(b)$ seria o valor estimado pelo modelo (2.2) com o valor b indicando a suavização da função de ponderação. O problema é que dessa forma o valor b ótimo é o valor tal que apenas o próprio elemento i permaneça na estimativa dos parâmetros no local i , ou seja, a soma se iguala a zero, gerando um modelo com um parâmetro para cada observação. Uma solução então, proposta em Brunson *et al.* (1996), seria utilizar a técnica de validação cruzada (*cross-validation*), em que na verdade o valor ótimo procurado é o que minimiza a soma

$$\sum_{i=1}^n (y_i - \hat{y}_{\neq i}^*(b))^2, \quad (2.14)$$

em que $\hat{y}_{\neq i}^*(b)$ é o valor ajustado para o local i , sem considerar o ponto i na estimação. Dessa forma, atinge-se o objetivo de selecionar apenas os pontos próximos de i . Nesse caso, observe que $w_{ij} = 1$ se $i = j$, porém, para fins computacionais, quando i e j são coincidentes, deve-se assumir que $w_{ij} = 0$ na validação cruzada.

Charlton *et al.* (2009) sugerem ainda utilizar, em vez da soma (2.14), o valor do critério de Akaike corrigido (quanto menor, melhor é o modelo), que pode ser utilizado também como uma medida de ajuste para a comparação de modelos, cuja

¹Sistema Universal Transverso de Mercator.

forma de acordo com Burnham e Anderson (2002) é

$$AIC_c = 2\nu_1 - 2\ell + 2\frac{\nu_1(\nu_1 + 1)}{n - \nu_1 - 1}, \quad (2.15)$$

sendo ℓ a log-verossimilhança do modelo, como segue:

$$\ell = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\hat{\sigma}^2) - \sum_i \frac{(y_i - \hat{y}_i)^2}{\hat{\sigma}^2}, \quad (2.16)$$

em que n é o número de observações, $\hat{\sigma}$ é o valor estimado do desvio padrão dos resíduos via máxima verossimilhança e ν_1 representa o número de parâmetros estimados pelo modelo.

Uma discussão levantada por Fotheringham *et al.* (2002) reforça a importância da escolha do *bandwidth*. Conforme citado na Seção anterior, as estimativas para β_i são viesadas. Segundo os autores, a magnitude desse viés tem uma relação muito estreita com a escolha do parâmetro de suavização e apresenta uma relação de *trade-off* com a variância das estimativas: quanto menor for o *bandwidth*, de modo que menos observações sejam utilizadas na estimação em i , maior será a variância e menor será o viés, e vice-versa.

Com relação a essa escolha, a otimização de (2.14) ou de (2.15) pode ser feita por meio de uma ferramenta visual, em que para vários valores de b são calculados os valores dos critérios, os quais são colocados em um gráfico onde é possível identificar o mínimo, ou então pela técnica de procura de mínimo de funções denominada Seção Áurea ou *Golden Section Search* (Greig, 1980), a qual é comumente utilizada no contexto da RGP e será abordada na próxima Seção.

2.4 Otimização por Seção Áurea

A técnica de procura de mínimo (ou máximo) de funções conhecida como Seção Áurea ou *Golden Section Search* (Greig, 1980) é bastante útil quando não é possível ou quando há dificuldades em empregar o artifício matemático de derivar a função que se deseja analisar e identificar em que ponto essa derivada se anula.

Supondo que a função em estudo possua apenas um ponto de mínimo (global)

e que este esteja no intervalo $[a, b]$, de tamanho $b - a$, a ideia por trás de alguns algoritmos de otimização é que a cada iteração se exclua uma parte do intervalo em que se sabe que o mínimo não se encontra, e assim ir diminuindo o intervalo total até que ele seja suficientemente pequeno para ser considerado o ponto de mínimo.

Conforme a Figura 2.5 e seguindo essa ideia, selecionam-se dois pontos x_1 e x_2 no intervalo. Se $f(x_1) > f(x_2)$, conclui-se que a função decresce na direção de x_2 , e então o mínimo é maior que x_1 e o intervalo de procura se resume a $(x_1, b]$. Caso contrário, se $f(x_1) < f(x_2)$, então o mínimo deve ser menor que x_2 , e o intervalo de procura se resume a $[a, x_2)$. Note que nesses casos não se pode concluir que o mínimo está entre x_1 e x_2 porque ambos podem estar do mesmo lado do ponto de mínimo. A única maneira de concluir que o mínimo está no intervalo (x_1, x_2) é se $f(x_1) = f(x_2)$, pois isso significa que cada ponto está de um lado do ponto de mínimo.

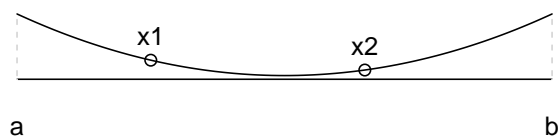


Figura 2.5: Ilustração do algoritmo da Seção Áurea

O desafio dessa otimização é encontrar os valores x_1 e x_2 que possibilitem encontrar o mínimo de maneira mais eficiente, e esse é o diferencial da Seção Áurea. Nesse algoritmo, procura-se por x_1 e x_2 tal que

$$\frac{\text{Tamanho do intervalo}}{\text{Tamanho da maior parte}} = \frac{\text{Tamanho da maior parte}}{\text{Tamanho da menor parte}}.$$

Ao se considerar o intervalo unitário, tem-se:

$$\frac{1}{r} = \frac{r}{1-r}$$

$$1-r = r^2$$

$$r^2 + r - 1 = 0 \quad (2.17)$$

cuja raiz positiva é igual a

$$r = \frac{-1 + \sqrt{5}}{2} = 0,618 \quad (2.18)$$

que representa a razão áurea, medida utilizada desde a antiguidade para definir a razão que seria a mais perfeita sob o ponto de vista estético, encontrada em muitos elementos da natureza, na geometria, no corpo humano.

Assim, escolhe-se $x_1 = b - (b - a)r$ e $x_2 = a + (b - a)r$. Dessa forma, após atualizar o intervalo de procura, pelo menos um dos pontos vai poder ser reutilizado, diminuindo o esforço computacional. Como exemplo, considerando novamente o intervalo unitário, tem-se

$$x_1 = 1 - 0,618 = 0,382 \quad (2.19)$$

e

$$x_2 = 0 + 0,618 = 0,618 \quad (2.20)$$

e se $f(x_1) < f(x_2)$ então o novo intervalo de procura é $[0; 0,618)$. Recalculando os novos valores x_1 e x_2 com esse novo intervalo de procura, obtém-se

$$x_1 = 0,618 - 0,618 \times 0,618 = 0,236 \quad (2.21)$$

e

$$x_2 = 0 + 0,618 \times 0,618 = 0,382 \quad (2.22)$$

ou seja, o novo valor encontrado x_2 é o mesmo já utilizado no passo anterior, não sendo necessária nova avaliação da função nesse ponto. Isso aconteceria também caso $f(x_1) > f(x_2)$, mas com o outro ponto. Em suma, a técnica se resume ao algoritmo da Tabela 2.1.

Para utilizar esse algoritmo na RGP, bastaria substituir a função $f(\cdot)$ pela equação da validação cruzada (2.14) ou do AIC (2.15), a depender do critério escolhido, e o x , argumento a ser otimizado, pelo *bandwidth* b .

Essa técnica de otimização pode não ser a mais eficiente, já que pode falhar na

Tabela 2.1: Algoritmo da Seção Áurea

$\begin{aligned} &\text{Enquanto } b - a < \varepsilon \text{ faça } \{ \\ &\quad x_1 = b - (b - a) \times 0,618 \\ &\quad x_2 = a + (b - a) \times 0,618 \\ &\quad f_1 = f(x_1) \\ &\quad f_2 = f(x_2) \\ &\quad \text{Se } f_1 < f_2 \text{ faça } \{ \\ &\quad\quad b = x_2 \\ &\quad\quad \} \\ &\quad \text{Se } f_1 > f_2 \text{ faça } \{ \\ &\quad\quad a = x_1 \\ &\quad\quad \} \\ &\quad \} \end{aligned}$

identificação de mínimos globais em funções que não sejam unimodais, mas justamente por ser um algoritmo muito simples tem sido uma técnica muito utilizada e geralmente possui boa performance no contexto da RGP, o que não exige o pesquisador de fazer confirmações específicas para seu banco de dados.

Com a definição desse algoritmo, tem-se todos os elementos para estimação dos parâmetros. As duas próximas Seções abordam alguns passos para inferência e diagnóstico.

2.5 Inferência em dados espaciais

O uso da Regressão Geograficamente Ponderada como uma forma de visualizar padrões em dados espaciais é bastante intuitivo, porém, considerando o que foi explanado até agora, nada se pode afirmar sobre o padrão observado ser de fato devido a alguma característica dos dados e não devido a uma variação aleatória qualquer, havendo a necessidade de se realizar uma análise inferencial. Conforme observa Fotheringham *et al.* (2002), reconhecer que dados espaciais raramente são independentes é entender que a estatística inferencial aplicada aos modelos de regressão clássica não é adequada na Regressão Geograficamente Poderada. Dessa forma, faz-se necessária a identificação da existência de correlação espacial e, em caso positivo, sua posterior incorporação à análise inferencial.

A dependência ou correlação espacial ocorre basicamente quando o valor de uma

variável depende, de alguma maneira, do local onde ela foi registrada e geralmente decorre da heterogeneidade ou não estacionariedade espacial, a qual por sua vez ocorre quando há mais de um padrão de associação espacial, ou seja, para diferentes locais, uma mesma variável pode ter influência diferente no valor de outra variável.

Graficamente, a correlação espacial pode ser visualizada ao mapear os resíduos da regressão clássica, para verificar padrões onde as estimativas foram melhor ou pior ajustadas, ou com o auxílio do diagrama de Moran, mais conhecido como *Moran Scatterplot* (Anselin, 1996), que relaciona o valor padronizado pela média da variável na localidade i ($\mathbf{z} = (\mathbf{y} - \bar{\mathbf{y}})$) com o valor padronizado pela média de seus vizinhos ($\mathbf{Wz} = \mathbf{W}(\mathbf{y} - \bar{\mathbf{y}})$, em que \mathbf{W} é a matriz de vizinhança, mais comumente binária). Pelos quadrantes do gráfico gerado é possível identificar aquelas observações que possuem valores altos e estão cercadas de outras observações com valores altos (alto-alto ou *high-high*), observações com valores baixos que estão cercadas por valores baixos (baixo-baixo ou *low-low*), observações com valores altos que estão cercadas por valores baixos (alto-baixo ou *high-low*) e observações com valores baixos que estão cercadas por observações com valores altos (baixo-alto ou *low-high*).

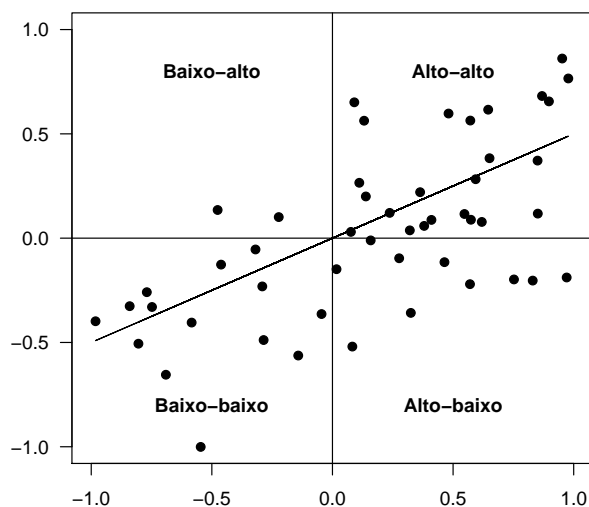


Figura 2.6: Exemplo de diagrama de Moran

A interpretação que esse diagrama suscita é que quanto mais pontos estiverem

nos quadrantes alto-alto e baixo-baixo, maior será a indicação de correlação espacial positiva; e caso os pontos estejam distribuídos predominantemente nos outros dois quadrantes, há evidência de correlação espacial negativa. Se os pontos se apresentarem de maneira aleatória, então menos evidência se tem de que os dados possuem alguma correlação espacial ou pode ser um caso em que os dados apresentam tanto a correlação positiva quanto a negativa (processo não estacionário), o que não apareceria em uma estatística global.

Analicamente, Fotheringham *et al.* (2002) à época já mostravam que são bem consolidadas as medidas para identificar correlação espacial em um único índice globalmente, como o Índice de Geary (Geary, 1954) e o Índice de Moran (Moran, 1950; Cliff e Ord, 1972). Posteriormente, surgiram trabalhos que enfatizam estatísticas locais para medir a correlação (Getis e Ord, 1992; Ord e Getis, 1995, 2001; Anselin, 1995, 1998; Rogerson, 1999), as quais permitem a identificação de diferentes comportamentos de associação no espaço, que não apareceriam na estatística global, podendo omitir um caso do paradoxo de Simpson.

Os índices globais mais utilizados são

$$\text{Índice de Moran: } I = \frac{n}{\sum_i \sum_{j \neq i} w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.23)$$

e

$$\text{Índice de Geary: } C = \frac{n-1}{2 \sum_i \sum_{j \neq i} w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - y_j)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (2.24)$$

em que $-1 < I < 1$, com -1 indicando perfeita correlação negativa, 0 indicando ausência de correlação e 1 indicando correlação positiva perfeita, e $0 < C < 2$, com 0 indicando correlação positiva perfeita, 1 indicando ausência de correlação e 2 indicando correlação negativa perfeita. Observe que ambos os índices fazem uso da matriz de vizinhança \mathbf{W} , que não deve ser confundida com a matriz de ponderação espacial, levando sempre em consideração o fenômeno em estudo.

As versões locais dos índices acima são dados, respectivamente, por

$$I_i = \frac{nz_i \sum_{j=1}^n w_{ij} z_j}{\sum_{i=1}^n z_j^2}, \quad (2.25)$$

em que $z_i = y_i - \bar{y}$, e

$$C_i = \frac{\sum_{j=1}^n w_{ij}(y_i - y_j)^2}{\sum_{j=1}^n (y_j - \bar{y})^2}. \quad (2.26)$$

Esses resultados locais podem ser visualizados pelo *Box map*, que consiste em colorir no mapa as classes definidas no diagrama de Moran para cada localidade, ou ainda pelo mapa de Moran ou *Moran Map*, que destaca apenas os índices que foram significativos, fornecendo indícios da não estacionariedade espacial.

As ferramentas gráficas são sempre de grande ajuda, mas não devem ser as únicas para o reconhecimento da correlação e/ou não estacionariedade. Em se tratando de testes estatísticos, todos os índices acima podem ter sua significância testada por meio de testes de pseudo-significância, que têm por base a abordagem de Monte Carlo (Hope, 1968). Sob a hipótese nula de independência espacial, qualquer permutação dos pares $(\mathbf{y}_i, \mathbf{y}_j)$, que representam os valores das variáveis em diferentes localidades, são igualmente prováveis. Sendo assim, os valores observados dos índices podem ser comparados com a distribuição aleatorizada de N pares permutados, e então o p-valor do teste seria o número de vezes m em que o índice calculado foi maior que o índice observado sobre o número de permutações mais o observado $m/(N + 1)$. Outra alternativa seria utilizar um teste paramétrico, aproveitando-se do fato de que ambos os índices possuem distribuições assintoticamente normais.

Segundo Brunson *et al.* (1996) e Leung *et al.* (2000), a análise inferencial na Regressão Geograficamente Ponderada pode ser utilizada como uma alternativa ou complemento às medidas acima e deve se concentrar em responder aos seguintes questionamentos:

⇒ A Regressão Geograficamente Ponderada descreve os dados significativamente melhor que a regressão linear clássica? (Há variação espacial suficiente para que o modelo local seja melhor que o modelo global para descrever os dados, prezando pelo Princípio da Parcimônia?)

⇒ Cada conjunto de parâmetros β_{ik} , $i = 1, \dots, n$, exibe variação significativa na região em estudo?

A primeira questão é na realidade um teste de adequabilidade do ajuste do modelo, ou seja, se os dados realmente podem ser descritos pela relação (2.2). Brunson

et al. (1996) argumentam que o questionamento pode se resumir em se testar o valor do parâmetro de suavização b , pois, já que ele é responsável pela inclusão da dependência espacial no modelo tradicional, pode ser entendido como uma medida de discrepância entre o modelo RGP e o modelo de regressão clássica. Por exemplo, na função de ponderação exponencial (2.9), bastaria testar a hipótese $H_0 : \theta = 0$, sendo $\theta = 1/b$. Nesse caso, qualquer valor diferente de 0 indica que há heterogeneidade espacial e que o modelo local seria o melhor candidato. Por outro lado, se a hipótese não for rejeitada, espera-se que os parâmetros resultantes da RGP sejam muito parecidos e coincidentes com o parâmetro estimado pelo modelo de regressão clássica.

Já a segunda questão envolve testar se uma determinada variável poderia ou não ser bem representada por um coeficiente constante, $H_0 : \beta_{ik} = \beta_k, \forall i$. Para essa questão, os referidos autores sugerem utilizar a ideia de variabilidade dos parâmetros. Como são estimados n valores para cada uma das k variáveis explicativas, é possível calcular o desvio padrão amostral (s_k), de acordo com Brunson *et al.* (1998):

$$s_k = \frac{\sum_i (\hat{\beta}_{ik} - \bar{\beta}_{.k})^2}{n}, \quad (2.27)$$

em que $\bar{\beta}_{.k}$ é a média das estimativas para a k -ésima variável explicativa. Assim, quanto menor for s_k , mais evidência se tem de que o parâmetro deve ser constante, e o teste se resume a testar a hipótese $H_0 : S_k = 0$.

Também aqui, ambas as hipóteses podem ser testadas por meio dos testes de pseudo-significância descritos anteriormente. Porém, Charlton *et al.* (2009) foram além e citam ainda a possibilidade de utilização do coeficiente de determinação (R^2) ou coeficiente de determinação ajustado (R_{adj}^2) e dos critérios de Akaike corrigido (AIC_c) ou Bayesiano (BIC) para medir a adequabilidade do modelo. Embora não envolvam testes estatísticos, essas medidas são eficazes na comparação de diferentes modelos.

Ainda com o objetivo de responder a essas questões, Leung *et al.* (2000) utilizam a noção de Soma de Quadrados Residual com o número efetivo de graus de liberdade dos resíduos (vide Equação (2.6)). Assumindo que os erros são normais e que os

estimadores são não viesados, sua distribuição aproximada é calculada e é aplicado o teste F, no caso da primeira questão. Para a segunda, os autores propõem uma aproximação para a distribuição de s_k .

Um outro tópico que pode ser de interesse é determinar o intervalo de confiança das estimativas para cada ponto i de interesse. Aqui, basta utilizar a mesma abordagem da regressão clássica, utilizando a variância dos estimadores definida em (2.5). Há ainda o teste múltiplo proposto por da Silva e Fotheringham (2015), o qual considera uma abordagem semelhante à correção de Bonferroni para testar em que locais os parâmetros são de fato significativos.

2.6 Diagnósticos

Conforme citado anteriormente, os resíduos do modelo de regressão clássica podem ser sobrepostos ao mapa da região em estudo. Da mesma forma, os resíduos da Regressão Geograficamente Ponderada, com fórmula matricial $\mathbf{e} = (\mathbf{I} - \mathbf{S})\mathbf{y}$, em que \mathbf{S} é a “matriz *hat*”, podem e devem ser colocados em um mapa, mas ao contrário da regressão clássica, espera-se que os resíduos de dados heterogêneos se apresentem de maneira aleatória, já que levam em consideração as especificidades de cada localidade na estimação dos parâmetros.

Na análise de observações discrepantes, podem ser utilizados os resíduos padronizados $r_i = e_i/(\hat{\sigma}\sqrt{q_{ii}})$, em que q_{ii} são os elementos diagonais da matriz $\mathbf{Q} = (\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^T$, ou padronizados externamente, cuja diferença é que na estimação de $\hat{\sigma}$ não é considerada a observação i .

Para a identificação de observações influentes, pode-se utilizar a distância de Cook: $D_i = r_i^2 s_{ii}/(1 - s_{ii})(2\nu_1 - \nu_2)$, em que a quantidade $2\nu_1 - \nu_2$ é o número efetivo de parâmetros do modelo.

De fato, percebe-se que qualquer medida aplicável aos modelos de regressão tradicional pode ser adaptada para uso nos modelos de Regressão Geograficamente Ponderada. Por ora, esse não é o foco deste trabalho.

2.7 Extensões do modelo de Regressão Geograficamente Ponderada

Nas Seções anteriores foram abordados os principais aspectos sobre a Regressão Geograficamente Ponderada. Entretanto, conforme Fotheringham *et al.* (2002), desde sua proposição em Fotheringham *et al.* (1996), muitas adaptações já foram propostas.

Uma primeira adaptação vem a ser os modelos RGP mistos ou semiparamétricos. É natural supor que, em alguns casos, a relação entre as variáveis pode ser em parte variável e em parte fixa, especialmente. Em outras palavras, tem-se que uma parte das variáveis explicativas do modelo são globais (não apresentam ou apresentam pouca variação espacial) e outra parte são locais, considerando a abordagem da Regressão Geograficamente Ponderada, como segue:

$$y_i = \sum_{j=1}^{k_f} \beta_j X_{ij}(f) + \sum_{l=1}^{k_v} \beta_l(u_i, v_i) x_{il}(v) + \varepsilon_i, \quad (2.28)$$

em que o índice f indica as variáveis que são fixas e o índice v indica as variáveis que variam no espaço. Note que se os valores dos parâmetros globais fossem conhecidos, então os parâmetros locais poderiam ser estimados ao se aplicar o modelo RGP básico (2.2) à variável y_i subtraída do valor do primeiro somatório da Equação (2.28), analogamente para o caso inverso. Partindo dessa reflexão, Fotheringham *et al.* (2002) apresentam um método para estimação do modelo.

Considere a forma matricial de (2.28):

$$\mathbf{y} = \mathbf{X}_f \boldsymbol{\beta}_f + \mathbf{m} + \boldsymbol{\varepsilon}, \quad (2.29)$$

em que \mathbf{X}_f é a matriz das variáveis globais com parâmetros $\boldsymbol{\beta}_f$ e \mathbf{m} é o vetor de tamanho n com elementos $\sum_{l=1}^{k_v} \beta_l(u_i, v_i) x_{il}(v)$. Reescrevendo (2.29), tem-se

$$\mathbf{y} - \mathbf{X}_f \boldsymbol{\beta}_f = \mathbf{m} + \boldsymbol{\varepsilon}. \quad (2.30)$$

Se os valores β_f são conhecidos e relembrando que os valores estimados pela RGP podem ser escritos como $\hat{\mathbf{y}} = \mathbf{S}\mathbf{y}$, em que \mathbf{S} é a “matriz *hat*”, então \mathbf{m} pode ser estimado por

$$\hat{\mathbf{m}} = \mathbf{S}(\mathbf{y} - \mathbf{X}_f\beta_f) \quad (2.31)$$

e substituindo esse valor na Equação (2.29) com um rearranjo, tem-se

$$\mathbf{z} = \mathbf{Q}\beta_f + \varepsilon, \quad (2.32)$$

em que $\mathbf{z} = (\mathbf{I} - \mathbf{S})\mathbf{y}$ e $\mathbf{Q} = (\mathbf{I} - \mathbf{S})\mathbf{X}_f$, que se resume a um modelo de regressão clássica, com resolução pelo método de Mínimos Quadrados Ordinários:

$$\hat{\beta}_f = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{z} = (\mathbf{X}_f^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbf{X}_f)^{-1}\mathbf{X}_f^T(\mathbf{I} - \mathbf{S})^T(\mathbf{I} - \mathbf{S})\mathbf{y}. \quad (2.33)$$

Observe que \mathbf{z} são os resíduos do modelo RGP de \mathbf{y} em relação às variáveis locais \mathbf{X}_v e as colunas de \mathbf{Q} são os resíduos obtidos ao se ajustar um modelo RGP das variáveis fixas \mathbf{X}_f em relação às variáveis locais \mathbf{X}_v . Assim, tem-se os seguintes passos para estimação:

1. Para cada coluna de \mathbf{X}_f :
 - Aplicar o modelo RGP com \mathbf{X}_v .
 - Calcular os resíduos do modelo.
2. Aplicar o modelo RGP de \mathbf{y} em \mathbf{X}_v .
3. Calcular os resíduos do modelo.
4. Aplicar MQO nos resíduos do item 2 em relação aos resíduos do item 1, o que fornece as estimativas de $\hat{\beta}_f$.
5. Subtrair $\mathbf{y} - \mathbf{X}_f\hat{\beta}_f$ e aplicar RGP a esse resultado contra \mathbf{X}_v para obter as estimativas locais.

Outra adaptação se refere à abordagem bayesiana, a qual foi proposta por LeSage (2004), utilizando a técnica de Monte Carlo via Cadeia de Markov (ou, do inglês, MCMC - *Monte Carlo Markov Chain*) para estimar as distribuições dos parâmetros

nos pontos de regressão. Nesse estudo, mostrou-se que os resultados são similares aos obtidos pela abordagem clássica, mas com a superfície para os parâmetros mais suavizada nos extremos e erros padrões razoavelmente menores (Fotheringham *et al.*, 2002).

Contudo, para este trabalho as adaptações mais relevantes são aquelas que envolvem variáveis respostas não normais, como os trabalhos já citados de Atkinson *et al.* (2003), Nakaya *et al.* (2005) e Silva e Rodrigues (2014), que introduziram o modelo de Regressão Geograficamente Ponderada para as distribuições Binomial (logística), Poisson e Binomial Negativa, respectivamente. Observe que todas as distribuições incorporadas ao modelo RGP são discretas, e para o caso de dados contínuos e restritos ao intervalo $(0, 1)$, nenhuma delas se mostra adequada, exceto talvez a RGP-Normal, quando a variável resposta for simétrica.

Para abarcar adequadamente esses casos e os casos para dados contínuos restritos a qualquer outro intervalo, conforme a problemática exposta no Capítulo 1 deste trabalho e adicionando a variação espacial, propõe-se o modelo de Regressão Beta Geograficamente Ponderada, especificado no Capítulo a seguir.

Capítulo 3

Regressão Beta Geograficamente Ponderada

3.1 Introdução

Conforme descrito por Meaney e Moineddin (2014), a regressão clássica funciona bem quando os dados de taxas ou proporções são simétricos em torno do meio do intervalo $(0, 1)$, mas as simulações no Capítulo 1 mostraram que ainda assim pode haver casos de valores estimados fora do intervalo. Ainda, de acordo com Fleiss *et al.* (2013), dados de taxas e proporções geralmente são assimétricos, e, nessa condição, a regressão clássica tende a ter um pior ajuste, com previsões fora do intervalo e fuga às suposições de normalidade e de variância constante. Sendo assim, a Regressão Beta tem se mostrado como uma boa alternativa para a modelagem de dados com essa característica.

Já para dados que possuem alguma relação com o espaço, a regressão clássica, por ser um modelo global, falha na identificação e descrição das relações locais entre as variáveis, podendo inclusive levar a conclusões espúrias, contrárias à realidade. Nessa situação, a Regressão Geograficamente Ponderada claramente se sobressai, evidenciando seu potencial.

Ao unir as características dos dois modelos supracitados, o modelo de Regressão Beta Geograficamente Ponderado (RBGP), objeto deste trabalho, vem com o intuito de melhor ajustar modelos em que a variável resposta possui estrutura de depen-

dência e/ou variação espacial, porém está restrita a um intervalo contínuo qualquer, em especial o intervalo $(0, 1)$, como taxas e proporções, a exemplo dos estudos de Comber *et al.* (2012) e Lawson *et al.* (2015).

A próxima Seção traz a especificação do modelo RBGP, bem como do processo de estimação dos parâmetros e de obtenção dos erros padrões. Logo em seguida, a outra seção traz um modelo alternativo, RBGPg, para contornar uma problemática encontrada na mensuração da quantidade de parâmetros estimados do modelo RBGP.

3.2 Especificação do Modelo RBGP

Para a definição do modelo de Regressão Beta Geograficamente Ponderada, combinando as formulações dos modelos elucidados nos Capítulos anteriores, considere que existam n observações e k variáveis explicativas e sejam y_1, y_2, \dots, y_n as variáveis resposta que seguem a distribuição beta dada pela parametrização em (1.7), ou seja, restritas ao intervalo $(0, 1)$ e com média μ_i e precisão ϕ . O modelo RBGP é então obtido ao se assumir que a média da variável resposta no local i pode ser representada por

$$g(\mu_i) = \eta_i = \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad (3.1)$$

em que $g(\cdot)$ é uma função de ligação que associa o intervalo $(0, 1)$ à reta real, (u_i, v_i) representa as coordenadas do i -ésimo ponto no espaço, $i = 1, \dots, n$, $\beta_k(u_i, v_i)$ é o parâmetro para a k -ésima variável explicativa em função do local da i -ésima observação, x_{ik} é o valor da k -ésima variável explicativa para o local i e ε_i é o erro associado à i -ésima observação, com suposição de que são independentes.

Para a estimação dos parâmetros do modelo de Regressão Beta Geograficamente Ponderada foi considerado o conceito de verossimilhança local, que em vez de estimar todos os $\beta_k(u_i, v_i)$ globalmente, o processo de estimação é realizado de ponto em ponto. Considerando o ponto (u_0, v_0) , o valor $\beta_k(u_0, v_0)$ pode ser aproximado por γ_k , como em um modelo de regressão clássico, mas considerando apenas as observações

próximas ao ponto (u_0, v_0) . Dessa maneira, tem-se

$$WL(\gamma_0, \dots, \gamma_k | D) = \sum_{i=1}^n L(\gamma_0, \dots, \gamma_k | D) w_{0i}, \quad (3.2)$$

em que w_{0i} é a ponderação espacial em função da distância entre o ponto (u_i, v_i) e o ponto (u_0, v_0) , que pode ser qualquer função como as descritas na Seção 2.3, inclusive utilizando as mesmas técnicas para definição do *bandwidth* ótimo, e $L(\gamma_0, \dots, \gamma_k | D)$ é a verossimilhança do modelo clássico:

$$y_i = \gamma_0 + \sum_{j=1}^k \gamma_j x_{ij} + \varepsilon_i. \quad (3.3)$$

Tem-se então que os estimadores globais γ_k com base nas observações próximas ao ponto (u_0, v_0) são consistentes e assintoticamente não viesados para os parâmetros locais $\beta_k(u_0, v_0)$.

Assim, como o modelo RBGP usa a distribuição beta, adaptou-se a verossimilhança local com a utilização da verossimilhança da regressão beta, e então a função de verossimilhança para o modelo RBGP é dada por:

$$\ell(\boldsymbol{\beta}(\mathbf{u}, \mathbf{v}), \phi) = \sum_{i=1}^n \ell_i(\mu(u_i, v_i), \phi), \quad (3.4)$$

ou, seguindo o conceito de log-verossimilhança local para o ponto (u_0, v_0) , com as adaptações necessárias,

$$\ell(\boldsymbol{\gamma}, \phi) = \sum_{i=1}^n \ell_i(\mu_i, \phi) w_{0i}, \quad (3.5)$$

em que, de acordo com o modelo de Regressão Beta,

$$\begin{aligned} \ell_i(\mu_i, \phi) &= \log \Gamma(\phi) - \log \Gamma(\mu_i \phi) - \log \Gamma((1 - \mu_i) \phi) + \\ &(\mu_i \phi - 1) \log y_i + ((1 - \mu_i) \phi - 1) \log(1 - y_i). \end{aligned} \quad (3.6)$$

Por conseguinte, a função score do modelo de Regressão Beta Geograficamente Ponderada, representada por $U^T(\boldsymbol{\gamma}, \phi) = (U_\gamma(\boldsymbol{\gamma}, \phi), U_\phi(\boldsymbol{\gamma}, \phi))$, aqui é obtida pela diferenciação da função de log-verossimilhança local com respeito aos γ_i ($U_\gamma(\boldsymbol{\gamma}, \phi)$)

e a $\phi(U_\phi(\boldsymbol{\gamma}, \phi))$. Com respeito aos γ_i no ponto (u_0, v_0) , tem-se

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\gamma}, \phi)}{\partial \gamma_k} &= \sum_{i=1}^n \left\{ \frac{\partial \ell_i(\mu_i, \phi)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \gamma_k} \right\} w_{0i} \\ &= \sum_{i=1}^n \underbrace{\left\{ \phi(y_i^* - \mu_i^*) \frac{1}{g'(\mu_i)} x_{ik} \right\}}_{(*)} w_{0i}\end{aligned}\quad (3.7)$$

e com respeito a ϕ :

$$\frac{\partial \ell(\boldsymbol{\gamma}, \phi)}{\partial \phi} = \sum_{i=1}^n \underbrace{\left\{ \mu_i(y_i^* - \mu_i^*) + \log(1 - y_i) - \psi((1 - \mu_i)\phi) + \psi(\phi) \right\}}_{(**)} w_{0i}, \quad (3.8)$$

em que, assim como na Regressão Beta, $g'(\mu_i)$ é a derivada de primeira ordem da função de ligação escolhida, $y_i^* = \log(y_i/(1 - y_i))$ e $\mu_i^* = \psi(\mu_i\phi) - \psi((1 - \mu_i)\phi)$, sendo $\psi(\cdot)$ a função digamma.

Novamente, seguindo o processo de estimação da Regressão Beta, os parâmetros locais γ_k e o parâmetro de precisão ϕ podem ser estimados ao se igualar a função escore a 0, mas, como não há fórmula fechada, pode-se maximizar numericamente a função de log-verossimilhança local a partir de algum algoritmo não-linear de otimização, como Newton ou Quasi-Newton, por exemplo. É interessante notar que, dessa maneira, o modelo RBGP naturalmente acomoda valores diferentes de ϕ para cada local de interesse, sem contudo considerar um valor de ϕ diferente para cada observação nas calibrações locais.

O procedimento de otimização, seguindo a Regressão Geograficamente Ponderada clássica, também deve ser realizado para cada um dos n pontos, de modo a obter as superfícies de parâmetros. Para a escolha do *bandwidth* ótimo, analogamente, pode-se utilizar a validação cruzada (2.14) ou o critério de Akaike corrigido, que para este modelo tem a mesma forma que em (2.15):

$$AIC_c = 2\nu_1 - 2\ell + 2\frac{\nu_1(\nu_1 + 1)}{n - \nu_1 - 1}, \quad (3.9)$$

mas com log-verossimilhança obtida pelos valores estimados pelo modelo RBGP, ou

seja, $\hat{\mu}_i$ e $\hat{\phi}_i$, dada por

$$\ell = \sum_i [\log \Gamma(\hat{\phi}_i) - \log \Gamma(\hat{\mu}_i \hat{\phi}_i) - \log \Gamma((1 - \hat{\mu}_i) \hat{\phi}_i) + (\hat{\mu}_i \hat{\phi}_i - 1) \log y_i + ((1 - \hat{\mu}_i) \hat{\phi}_i - 1) \log(1 - y_i)], \quad (3.10)$$

em que ν_1 representa o número de parâmetros do modelo, obtido com o auxílio da “matriz *hat*”.

Os valores iniciais para a otimização do vetor de parâmetros γ_0 , que contém os k parâmetros para o ponto (u_0, v_0) , e de ϕ , que é o vetor de tamanho n com os valores de ϕ para cada local i , foram adaptados a partir dos valores iniciais utilizados na regressão beta, mas considerando a matriz de ponderação local. Para os γ_{0k} , os valores iniciais são as estimativas da regressão geograficamente ponderada clássica aplicada na transformação dos dados pela função de ligação $\check{y}_i = g(y_i)$, ou seja, $\gamma_{0\text{inicial}} = (\mathbf{X}^T \mathbf{W}_0 \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_0 \check{\mathbf{y}}$. Ao se realizar esse cálculo para todas as localidades, obtém-se uma matriz com k linhas e n colunas, na qual cada coluna representa os parâmetros iniciais para cada localidade. Da mesma forma, como a regressão beta produz um valor μ_t para cada observação, tem-se que a Regressão Geograficamente Ponderada irá produzir um vetor de μ_t para cada local, gerando uma matriz $n \times n$, e assim, o valor inicial de ϕ para o ponto (u_0, v_0) é dado por

$$\phi_0 = \frac{1}{n} \sum_{t=1}^n \frac{\check{\mu}_{0t}(1 - \check{\mu}_{0t})}{\check{\sigma}_{0t}^2} - 1, \quad (3.11)$$

em que $\check{\mu}_{it}$ é obtido ao se aplicar $g^{-1}(\cdot)$ aos valores ajustados pela regressão geograficamente ponderada tradicional de \check{y}_t em \mathbf{X} , isto é, $\check{\mu}_{it} = g^{-1}(\mathbf{x}_t(\mathbf{X}^T \mathbf{W}_0 \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_0 \check{\mathbf{y}})$, com \mathbf{x}_t sendo a t -ésima linha da matriz \mathbf{X} , e

$$\check{\sigma}_{0t}^2 = \frac{\check{\mathbf{e}}^T \check{\mathbf{e}}}{(n - \nu_1) g'(\check{\mu}_{0t})^2} \quad (3.12)$$

sendo $\check{\mathbf{e}}$ o vetor de resíduos da RGP de $\check{\mathbf{y}}$ e ν_1 representa o número de parâmetros estimados no modelo, que é o traço da “matriz *hat*”.

De posse das estimativas dos parâmetros, para obter a matriz de informação de

Fisher para cada ponto e, conseqüentemente, os erros padrões dos estimadores, é preciso obter a esperança da derivada da função escore com respeito a cada um dos parâmetros. Assim, para o ponto (u_0, v_0) tem-se:

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\gamma}, \phi)}{\partial \gamma_k \partial \gamma_j} &= \sum_{i=1}^n \left\{ \frac{\partial(*)}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \gamma_j} \right\} \\ &= \sum_{i=1}^n \left\{ -\phi^2 \frac{1}{g'(\mu_i)^2} x_{ik} x_{ij} [\psi'(\mu_i \phi) + \psi'((1 - \mu_i) \phi)] w_{0i} \right\} \end{aligned} \quad (3.13)$$

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\gamma}, \phi)}{\partial \gamma_k \partial \phi} &= \sum_{i=1}^n \frac{\partial(*)}{\partial \phi} \\ &= \sum_{i=1}^n \left\{ \frac{1}{g'(\mu_i)} x_{ik} [y_i^* - \mu_i^* - \phi(\mu_i \psi'(\mu_i \phi) - (1 - \mu_i) \psi'((1 - \mu_i) \phi))] w_{0i} \right\} \end{aligned} \quad (3.14)$$

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\gamma}, \phi)}{\partial \phi^2} &= \sum_{i=1}^n \frac{\partial(**)}{\partial \phi} \\ &= \sum_{i=1}^n \left\{ -\mu_i^2 \psi'(\mu_i \phi) + (1 - \mu_i)^2 \psi'((1 - \mu_i) \phi) + \psi'(\phi) \right\} w_{0i}, \end{aligned} \quad (3.15)$$

em que $(*)$ e $(**)$ são como definidos em (3.7) e (3.8), respectivamente, $\psi'(\cdot)$ é a função trigamma, $g'(\mu_i)$ é tratado como constante em relação a μ_i , pois $g(\mu_i) = \eta_i$, ou seja, considera-se $g'(\mu_i)$ em função de η_i , e \mathbf{y}^* e $\boldsymbol{\mu}^*$ são os vetores de tamanho n com elementos $y_i^* = \log(y_i/(1 - y_i))$ e $\mu_i^* = \psi(\mu_i \phi) - \psi((1 - \mu_i) \phi)$.

E então, considerando que $E(y_i^*) = \mu_i^*$, a matriz de informação de Fisher do modelo é dada por

$$\mathbf{K} = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix} \quad (3.16)$$

com $K_{\beta\beta} = \mathbf{X}^T \boldsymbol{\Phi} \mathbf{W} \mathbf{Z} \mathbf{X}$, $K_{\beta\phi} = K_{\phi\beta}^T = \mathbf{X}^T \mathbf{W} \mathbf{T} \mathbf{c}$ e $K_{\phi\phi} = \text{tr}(\mathbf{W} \mathbf{D})$. É importante frisar que aqui \mathbf{W} é a matriz de ponderação espacial completa de tamanho $n \times n$, que contém todos os pesos com base nas distâncias 2 a 2 entre os pontos, ou seja, o elemento (i, j) dessa matriz contém os pesos em função das distâncias entre o

ponto (u_i, v_i) e o ponto (u_j, v_j) , $\forall i, j = 1, \dots, n$. Conseqüentemente, essa matriz possui diagonal igual a 1, diferentemente da matriz \mathbf{W}_0 , a qual considera apenas as distâncias para o ponto (u_0, v_0) .

Os demais componentes da matriz de informação de Fisher são as matrizes diagonais Φ , \mathbf{Z} , \mathbf{T} e \mathbf{D} , todas de tamanho n e cujos elementos são respectivamente iguais a:

$$\phi_i = \hat{\phi}_i \quad (3.17)$$

$$z_i = \hat{\phi}_i \{ \psi'(\hat{\mu}_i \hat{\phi}_i) + \psi'((1 - \hat{\mu}_i) \hat{\phi}_i) \} \frac{1}{g'(\hat{\mu}_i)^2} \quad (3.18)$$

$$t_i = \frac{1}{g'(\hat{\mu}_i)} \quad (3.19)$$

$$d_i = \psi'(\hat{\mu}_i \hat{\phi}_i) \hat{\mu}_i^2 + \psi'((1 - \hat{\mu}_i) \hat{\phi}_i) (1 - \hat{\mu}_i)^2 - \psi'(\hat{\phi}_i) \quad (3.20)$$

e \mathbf{c} é o vetor tamanho n com elementos iguais a

$$c_i = \hat{\phi}_i \{ \psi'(\hat{\mu}_i \hat{\phi}_i) \hat{\mu}_i - \psi'((1 - \hat{\mu}_i) \hat{\phi}_i) (1 - \hat{\mu}_i) \}. \quad (3.21)$$

Observe que tudo que depende dos valores estimados $\hat{\mu}_i$ e $\hat{\phi}_i$ deve ser tratado globalmente, enquanto o que depende de \mathbf{W} deve ser tratado localmente. Em termos computacionais, isso significa que primeiro estimam-se os parâmetros e a matriz \mathbf{W}_i para cada local i , já que o processo de estimação se dá ponto a ponto, e de posse desses valores obtém-se a matriz \mathbf{W} completa e por conseguinte a matriz de informação de Fisher.

Como esses estimadores são de máxima verossimilhança, sob as devidas condições de regularidade e para um grande número de observações, segue que os estimadores possuem distribuição aproximada dada por

$$\begin{pmatrix} \hat{\beta} \\ \hat{\phi} \end{pmatrix} \sim N_{k+1} \left(\begin{pmatrix} \beta \\ \phi \end{pmatrix}, \mathbf{K}^{-1} \right).$$

Após a estimação, presume-se que os procedimentos para inferência e diagnóstico sejam análogos aos discutidos no Capítulo 2, sendo que para a obtenção da “matriz *hat*”, importante para a mensuração do número real de parâmetros estimados pelo

modelo ($\nu_1 = \text{tr}(\mathbf{S})$), empregou-se a matriz do modelo de Regressão Beta (1.26) adaptada aos pesos espaciais:

$$\mathbf{S} = \mathbf{Z}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{Z} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Z}^{1/2} \quad (3.22)$$

em que a matriz \mathbf{Z} possui elementos como em (3.18).

Observe, porém, que nesse caso o traço da “matriz *hat*” não contabiliza a estimação dos ϕ locais, o que pode ser constatado ao se utilizar um *bandwidth* muito grande, de modo que o modelo RBGP se iguale ao modelo global de regressão beta, caso em que o valor do traço dessa matriz tenderá ao número de variáveis explicativas do modelo. Assim, as estatísticas que utilizam o número de parâmetros do modelo, como o AIC por exemplo, ficam prejudicadas, já que o traço da “matriz *hat*” subestima o real número de parâmetros. Entretanto, assumiu-se como estimativa que o número de parâmetros do modelo é $\nu_1 + 1$, mas adicionalmente propõe-se na próxima Seção um modelo alternativo para tentar controlar melhor essa questão do número de parâmetros envolvidos no modelo.

3.3 Especificação do Modelo RBGPg

Como discutido na Seção anterior, o método tradicional para aferir a quantidade de parâmetros estimados não reflete a realidade no modelo RBGP, já que não consegue captar a estimação dos ϕ locais. Assim, uma alternativa encontrada nos moldes do trabalho de Silva e Rodrigues (2014) foi o modelo de Regressão Beta Geograficamente Ponderada com ϕ global (RBGPg), cuja proposta é justamente utilizar um modelo em que o valor de ϕ é controlado, ou seja, não é estimado localmente, mas sim globalmente.

Então, o modelo de Regressão Beta Geograficamente Ponderada com ϕ global considera que existem n observações e k variáveis explicativas e que as variáveis respostas y_1, y_2, \dots, y_n seguem a distribuição beta dada pela parametrização em (1.7), ou seja, restritas ao intervalo $(0, 1)$ e com média μ_i , mas com precisão ϕ_g para todas as localidades, o qual é obtido pelo modelo global de regressão beta. O modelo

RBGPg é então obtido ao se assumir que a média da variável resposta no local i pode ser representada por

$$g(\mu_i) = \eta_i = \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad (3.23)$$

em que, analogamente ao modelo RBGP, $g(\cdot)$ é uma função de ligação que associa o intervalo $(0, 1)$ à reta real, (u_i, v_i) representa as coordenadas do i -ésimo ponto no espaço, $i = 1, \dots, n$, $\beta_k(u_i, v_i)$ é o parâmetro para a k -ésima variável explicativa, em função do local da i -ésima observação, x_{ik} é o valor da k -ésima variável explicativa para o local i e ε_i é o erro associado à i -ésima observação, com suposição de que são independentes.

A estimação segue o mesmo procedimento que o do modelo RBGP, mas impedindo que ϕ varie localmente, forçando que a cada iteração do algoritmo de estimação ϕ seja igual a ϕ_g . Dessa forma, tem-se a certeza de que o número de parâmetros é igual ao número de parâmetros das variáveis explicativas mais a contribuição do ϕ estimado globalmente, ou seja, $\nu_1 = \text{tr}(\mathbf{S}) + 1$, em que \mathbf{S} é a “matriz *hat*” (3.22).

Em termos gerais, esse modelo significa dizer que os parâmetros variam localmente, mas a precisão dos dados é fixa no espaço, o que pode não retratar a realidade, mas pode ser útil a depender dos objetivos de quem utiliza o modelo, conforme algoritmo para a função de ligação logito, descrito no Apêndice B desta dissertação, com realce para a parte de atualização do Newton Raphson na Tabela 3.1. Observe que para o caso da RBGP o algoritmo é o mesmo a menos da parte que força o valor $\phi = \phi_g$, que é o valor estimado para ϕ pelo modelo global de regressão beta.

Com isso, foram abordados os principais aspectos do modelo de Regressão Beta Geograficamente Ponderada proposto neste trabalho. O próximo Capítulo traz uma aplicação para dados reais sobre a proporção de domicílios que possuem telefone fixo no estado do Espírito Santo. Para tal, foi desenvolvido um programa na linguagem SAS/IML que executa tanto a regressão beta quanto a otimização do *bandwidth* via validação cruzada e o ajuste do modelo RBGP e RBGPg em si, com opção de função de ligação logito, probito, log-log e complemento log-log e de funções de ponderação exponencial e bi-quadrática, esta fixa e variável.

Tabela 3.1: Algoritmo parcial de estimação da RBGP e da RBGPg com função de ligação logito

(...)

$$param = \beta || \phi$$

$$param^{n+1} = param + \mathbf{k}^{-1} \mathbf{U}$$

$$\beta[, i] = param^{n+1}[1 : k]$$

Se for o modelo RBGP, então faça {

$$\phi = param^{n+1}[k + 1]$$

}

Se for o modelo RBGPg, então faça {

$$\phi = \phi_g$$

}

$$it = it + 1$$

Capítulo 4

Estudo de Caso: proporção de telefones fixos no Espírito Santo

4.1 Introdução

Este Capítulo traz uma aplicação dos modelos de Regressão Beta Geograficamente Ponderada propostos no Capítulo anterior em comparação com o modelo de regressão clássica, o modelo de regressão beta e o modelo de regressão geograficamente ponderada considerando a distribuição normal, para fins de verificar sua performance. Os resultados aqui apresentados foram obtidos com o auxílio do *software* para análise de dados SAS, e o modelo ora proposto foi ajustado por meio de programação desenvolvida em linguagem SAS/IML deste mesmo *software*.

A próxima Seção traz uma descrição das variáveis em estudo e nas Seções seguintes apresentam-se os resultados para o modelos globais e locais, respectivamente.

4.2 Base de dados

Os dados desse exemplo se referem à proporção de domicílios que possuem telefone fixo nos municípios do Espírito Santo (variável resposta) em relação à proporção de população urbana e à proporção de indivíduos pobres em cada município desse estado (variáveis explicativas). Esses dados foram obtidos a partir do Censo 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE, 2010) e do Portal Atlas Brasil

(PNUD *et al.*, 2010), que reúne algumas dessas informações.

A variável resposta, proporção de domicílios que possuem telefone fixo, foi calculada a partir da quantidade de domicílios que possuem telefone fixo dividida pelo número total de domicílios particulares permanentes para cada município. A variável explicativa proporção de população urbana corresponde à população residente na área urbana dividida pela população total do município. Já a variável proporção de indivíduos pobres, por definição, é composta pela proporção de indivíduos que vivem em domicílios particulares permanentes com renda domiciliar *per capita* igual ou inferior a R\$140,00 mensais no período base (agosto de 2010).

As Figuras 4.1 a 4.3 exibem a distribuição espacial das variáveis em estudo, acompanhadas de suas respectivas medidas descritivas. A Figura 4.1 mostra como é a relação da variável resposta com o espaço. No mapa, é possível observar concentrações maiores no litoral do estado, principalmente ao sul, em volta da capital Vitória, e concentrações menores na parte interior do estado, principalmente a sudoeste. Essa relação é semelhante à relação apresentada na Figura 4.2, mas em maior magnitude nesta. Essas áreas também coincidem parcialmente com as áreas mais claras e mais escuras, respectivamente, da Figura 4.3, a qual apresenta a variável proporção de pobres, indicando uma relação contrária entre essa variável e as demais.

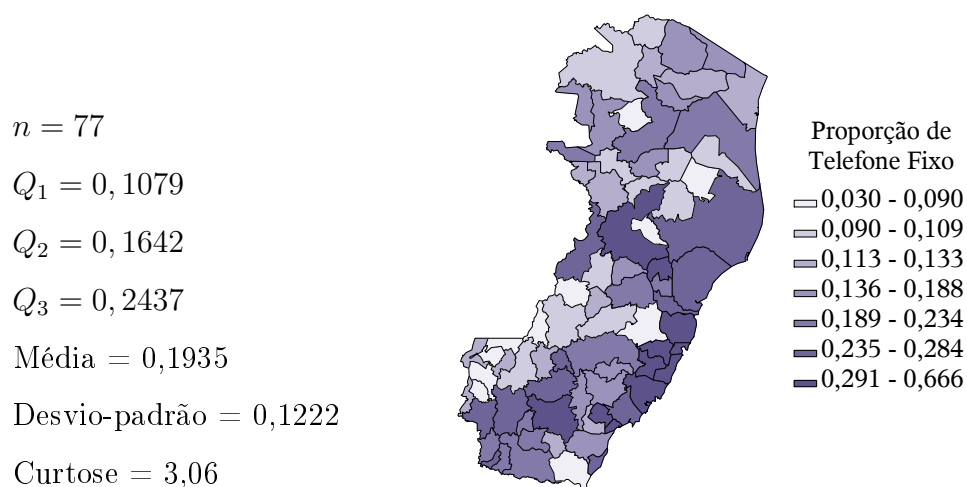


Figura 4.1: Distribuição espacial da variável proporção de telefone fixo

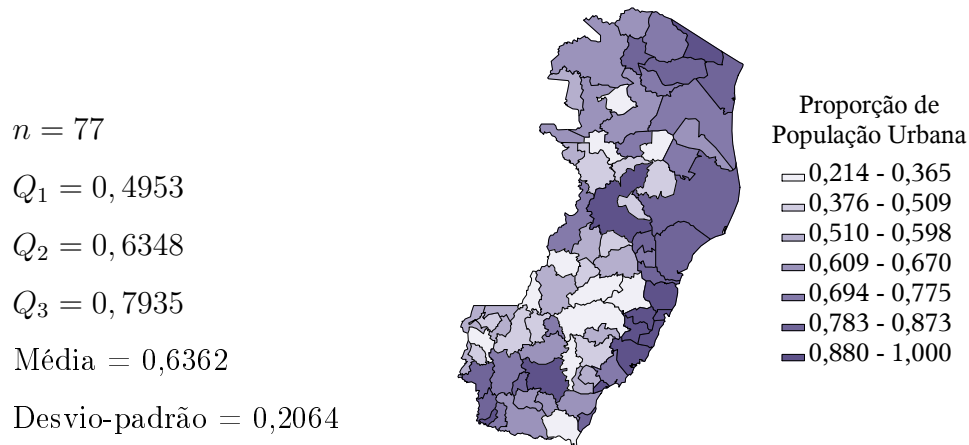


Figura 4.2: Distribuição espacial da variável proporção de população urbana

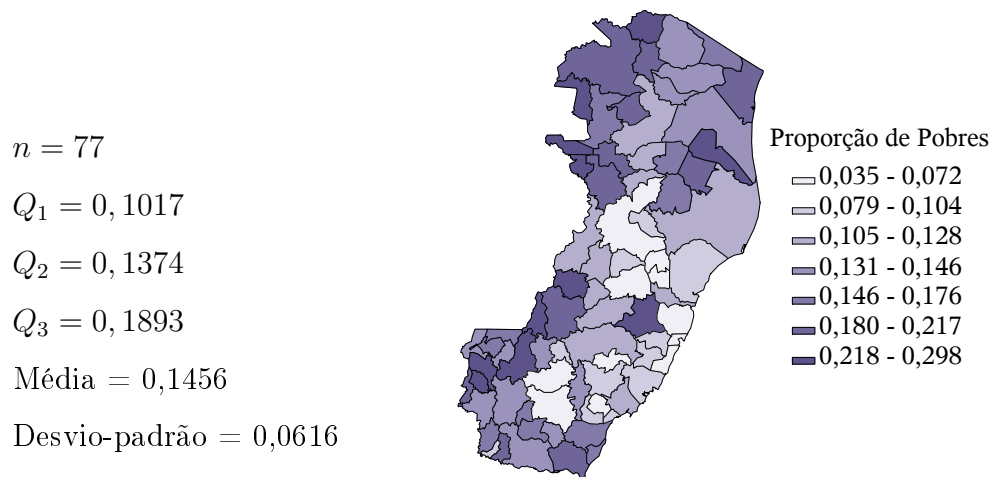


Figura 4.3: Distribuição espacial da variável proporção de pobres

A disposição das variáveis no espaço nas referidas Figuras sugerem a existência de associação espacial das variáveis, porém, conforme mencionado no Capítulo 2, apenas a visualização dos dados no mapa não é suficiente para garantir que a associação exista, sendo aconselhável realizar análises complementares para conclusões mais concretas. Nesse sentido, apresentam-se os diagramas de Moran bem como os respectivos índices globais nas Figuras 4.4 a 4.6, os quais foram calculados a partir da matriz de vizinhança binária padronizada.

Observe que mesmo com escalas diferentes nos gráficos, é possível perceber que em todos os diagramas os valores se concentram basicamente no primeiro e terceiro

quadrantes, à exceção da Figura 4.5 (população urbana), em que os pontos estão um pouco mais dispersos. Esse comportamento revela uma associação espacial moderadamente positiva para a variável resposta, com Índice de Moran Global de 0,5, e também para a variável explicativa proporção de pobres, com índice de 0,42. Já para a variável proporção de população urbana verifica-se uma associação mais baixa, com índice de 0,3.

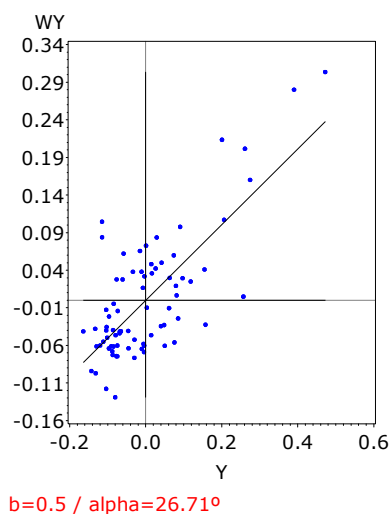


Figura 4.4: Diagrama de Moran da variável proporção de telefone fixo

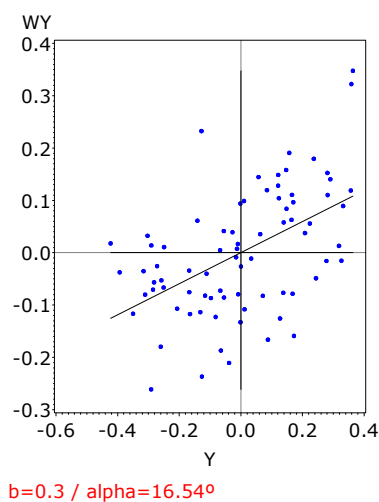


Figura 4.5: Diagrama de Moran da variável proporção de população urbana

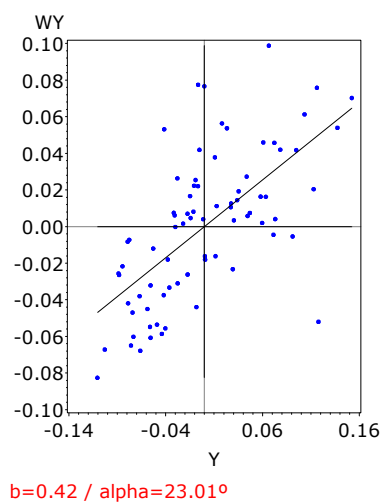


Figura 4.6: Diagrama de Moran da variável proporção de pobres

Além disso, apresentam-se ainda o mapa de espalhamento de Moran (Figura 4.7) e o mapa de Moran (Figura 4.8) para a variável resposta. O primeiro mostra a relação

entre os índices locais, exibindo novamente a relação de valores mais altos no litoral e mais baixos no interior. O segundo mostra aqueles locais que são significativos ao nível de significância de 5% utilizando a aproximação pela distribuição normal. A existência dessas áreas provém indicação de que os dados não são estacionários no espaço.

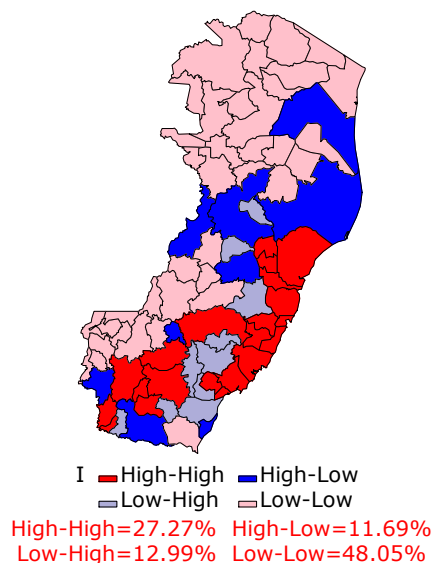


Figura 4.7: Mapa de espalhamento de Moran para a proporção de telefone fixo

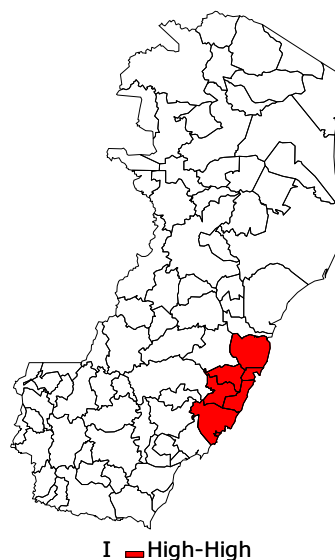


Figura 4.8: Mapa de Moran (95%) para a proporção de telefone fixo

4.3 Ajuste dos modelos globais

A partir dessas análises exploratórias, fica mais evidente a existência de associação espacial das variáveis e da não estacionariedade da variável resposta no espaço, sinalizando que modelos globais talvez não sejam os mais adequados para descrever a possível relação entre a variável resposta e as variáveis explicativas. Ainda, de acordo com a Figura 4.9, que mostra a distribuição da variável resposta, e com as medidas descritivas a essa variável relacionadas na Figura 4.1, observa-se que os dados estão muito concentrados entre 0 e 0,3 e apresentam assimetria à direita, apontando que a distribuição normal também pode não ser a mais adequada.

Não obstante, ajustou-se para esses dados um modelo de regressão clássica e um modelo de regressão beta com função de ligação logito, que é a mais comum e facilita

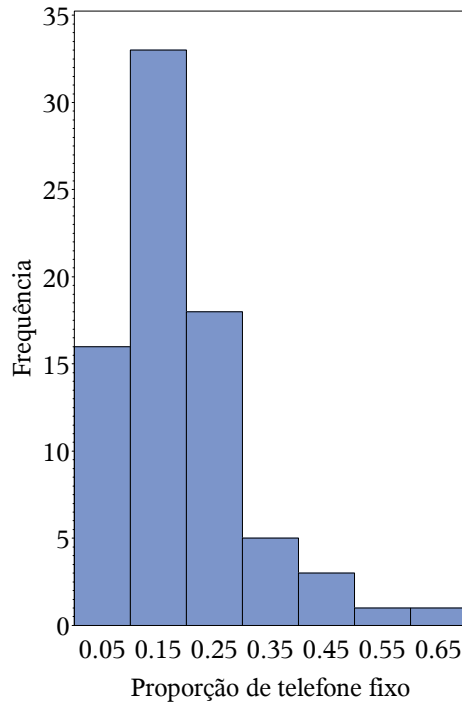


Figura 4.9: Distribuição da variável proporção de telefone fixo

a interpretação. A Tabela 4.1 apresenta os principais resultados para esses modelos. Observe que todos os parâmetros foram significativos ao nível de significância de 5% e que todas as medidas de ajuste apontam para o modelo beta como preferível. Entretanto, cabe aqui destacar que para o modelo de regressão beta utilizou-se uma medida chamada de pseudo R^2 , conforme explanado no Capítulo 2, e que por conta de utilizar metodologias diferentes, deve-se ter cautela ao comparar essas duas medidas.

Tabela 4.1: Resultados dos modelos globais

Variáveis	Regressão Clássica			Regressão Beta		
	Estimativa	t	p -valor	Estimativa	t	p -valor
Intercepto	0,1011	2,28	0,0257	-2,0744	-8,59	< 0,0001
prop_pobres	-0,7891	-5,23	< 0,0001	-5,8998	-6,96	< 0,0001
prop_urbana	0,3258	7,23	< 0,0001	2,1567	8,97	< 0,0001
ϕ	-	-	-	48,5175	6,22	< 0,0001
R^2 ajustado*	0,6604			0,7141		
AICc	-185,0364			-228,3313		
Log-Verossimilhança	95,6526			118,4434		

*: pseudo R^2 para regressão beta

Analisando os valores dos parâmetros, conclui-se pelo sinal negativo que quanto maior for a proporção de população pobre, a tendência é que menor seja a proporção de domicílios que possuem telefone fixo. De forma contrária, quanto maior for a proporção de população urbana, maior será em média a proporção da variável resposta. Apesar de ambos os modelos levarem à mesma conclusão, a interpretação das estimativas não é a mesma para os diferentes modelos. Como os dados são contínuos, existe um impulso em interpretar os parâmetros da regressão beta da mesma forma que a interpretação do modelo clássico, assim, ressalte-se que os parâmetros do modelo beta não devem ser interpretados em seu valor absoluto, como se interpretam os parâmetros do modelo clássico, mas sim em função da razão de chances, já que foi utilizada a função de ligação logito.

Observe a diferença nas interpretações: no modelo clássico, o aumento de 10 pontos percentuais na proporção de pobres mantida a outra variável explicativa constante diminui em média 7,8 pontos percentuais na variável resposta ($-0,7891 \times 0,1$), já o mesmo aumento na proporção de população urbana nas mesmas condições implica aumento médio de 3,2 pontos percentuais ($0,3258 \times 0,1$). Por outro lado, para o modelo de regressão beta, a interpretação via razão de chances indica que um aumento de 10 pontos percentuais na proporção de pobres diminui em média 44% a chance de um domicílio possuir telefone fixo ($e^{-5,8998 \times 0,1} - 1$), com a outra variável explicativa constante, enquanto que o mesmo aumento na variável proporção de população urbana aumenta em média 24% a chance de um domicílio possuir telefone fixo ($e^{2,1567 \times 0,1} - 1$), desde que a proporção de pobres seja mantida constante.

As Figuras 4.10 a 4.13 compõem a análise dos resíduos para os modelos em tela. As duas primeiras Figuras visam verificar a adequação dos resíduos à distribuição normal, uma das premissas do modelo clássico. Pelos gráficos e pelos testes Shapiro-Wilk de normalidade, percebe-se que os resíduos do modelo clássico (Figura 4.10) apresentam maior fuga à normalidade que os resíduos do modelo beta (Figura 4.11), embora para o modelo de regressão beta não existam suposições para a distribuição dos erros. Além disso, a Figura 4.12 apresenta o diagrama de resíduos contra os valores preditos para o modelo clássico, no qual é possível observar que existem valores preditos negativos, os quais não deveriam existir por se tratar de uma pro-

porção, e acabam ratificando os resultados obtidos pela simulação no Capítulo 1 desta dissertação. Ainda, observa-se que existe também uma tendência nesses resíduos, o que geralmente é sinal de má especificação do modelo e é clássico indicativo de associação espacial. Já o modelo de regressão beta mostrou melhor ajuste que o modelo clássico, já que de fato os resíduos possuem menor amplitude e mesmo que ainda seja possível observar um certo padrão em seu gráfico (Figura 4.13), este parece acontecer em menor grau que no modelo clássico.

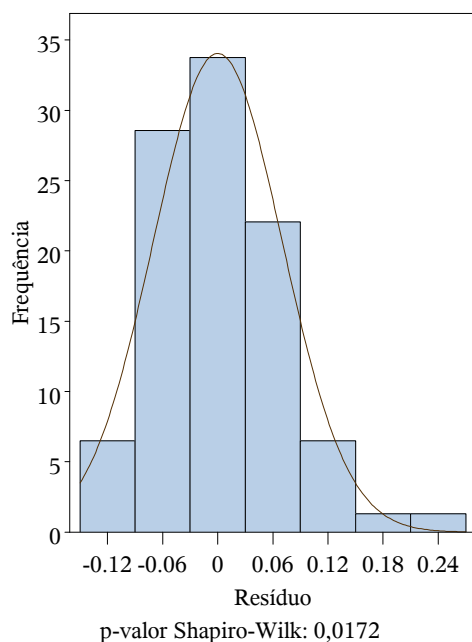


Figura 4.10: Distribuição dos resíduos do modelo clássico

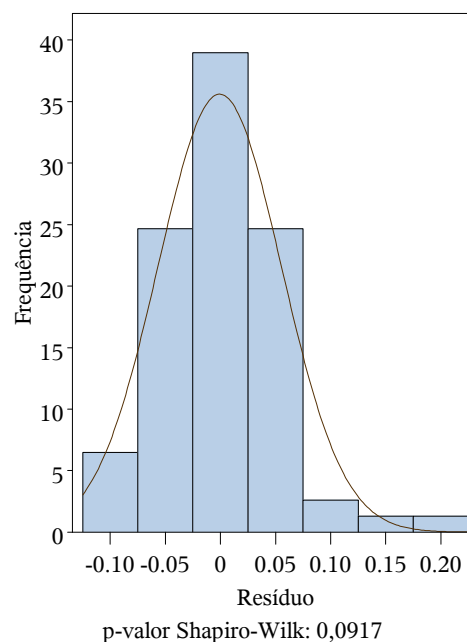


Figura 4.11: Distribuição dos resíduos do modelo beta

Mais ainda, a disposição desses resíduos no mapa do estado de Espírito Santo revela que o ajuste não foi homogêneo no espaço e que os locais com maiores resíduos absolutos são justamente aqueles onde o índice de Moran local apontou significância, ou seja, há indicativo de que para esses locais os parâmetros não devem ser os mesmos que para os demais (Figuras 4.14 e 4.15). Isso pode significar, por exemplo, que a proporção de pobres influencia a proporção de telefone fixo em outra magnitude nessas localidades. Nessas Figuras, apresentam-se ainda os índices de Moran para esses resíduos, que indicam ainda alta correlação (de 0,43 para o modelo clássico e de 0,41 para o modelo beta). Esse cálculo será importante para comparar com os modelos locais e verificar se houve diminuição dessa correlação nos resíduos, caso

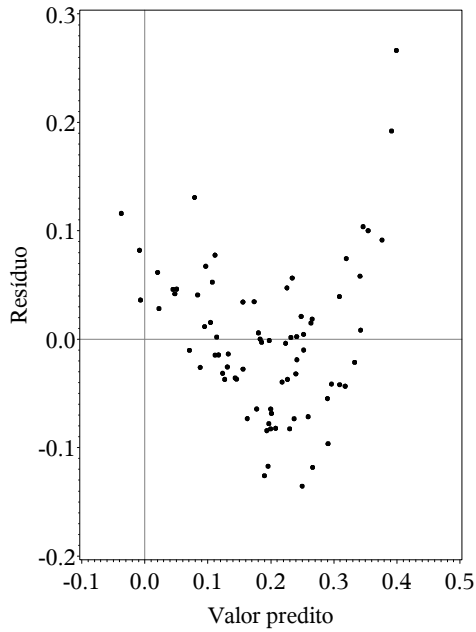


Figura 4.12: Resíduos pelos valores preditos do modelo clássico

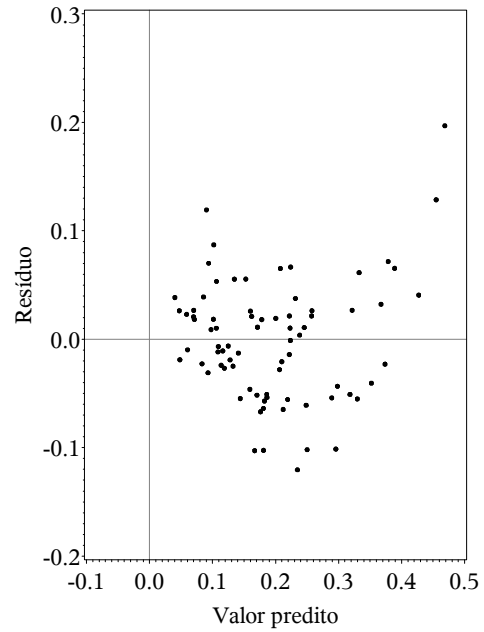
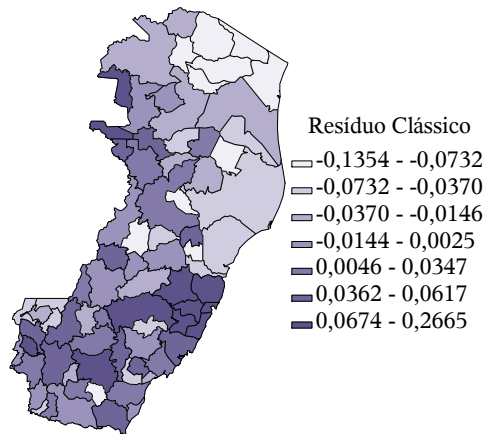
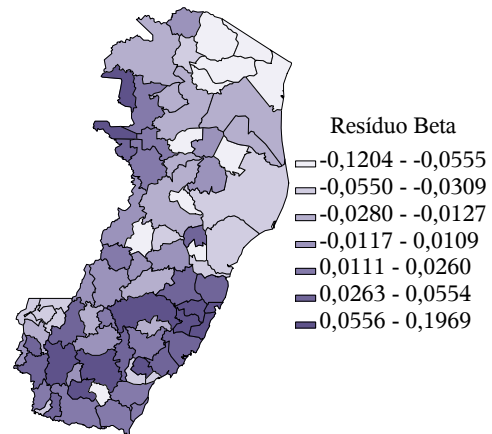


Figura 4.13: Resíduos pelos valores preditos do modelo beta



I=0,43

Figura 4.14: Disposição espacial dos resíduos do modelo clássico



I=0,41

Figura 4.15: Disposição espacial dos resíduos do modelo beta

em que a correlação espacial foi agregada na estimação do modelo.

De forma complementar e seguindo a orientação de Cribari-Neto e Zeileis (2010) sobre a não utilização dos resíduos puros no modelo beta, analisou-se por fim os

resíduos padronizados (1.24) tanto para o modelo beta quanto para o modelo clássico com o auxílio das Figuras 4.16 a 4.19.

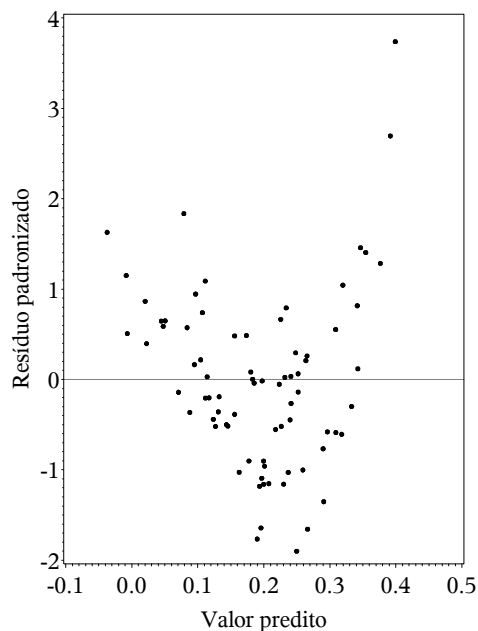


Figura 4.16: Resíduos padronizados pelos valores preditos do modelo clássico

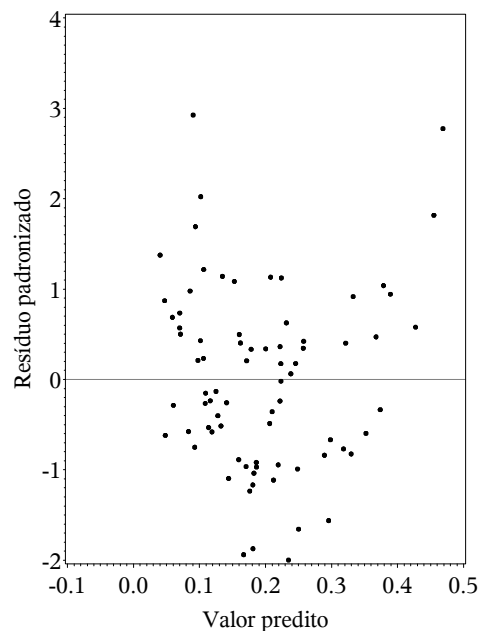
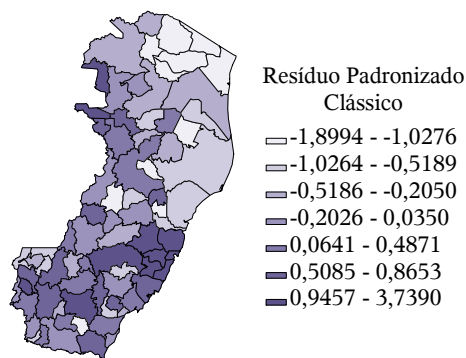


Figura 4.17: Resíduos padronizados pelos valores preditos do modelo beta

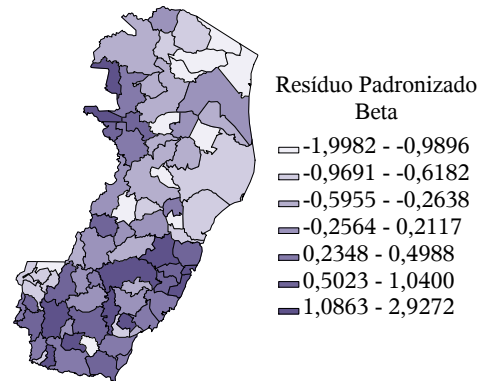
Para o modelo beta, os resíduos padronizados aparentam ser um pouco mais aleatórios que os resíduos puros, com menor índice de Moran. Já para o modelo clássico, não houve alteração no índice de Moran e novamente apresentou maior amplitude que os resíduos do modelo beta.

Considerando suficientes para comparação as análises dos modelos globais, a próxima Seção apresenta os resultados para os modelos locais.



I=0,43

Figura 4.18: Disposição espacial dos resíduos padronizados do modelo clássico



I=0,35

Figura 4.19: Disposição espacial dos resíduos padronizados do modelo beta

4.4 Ajuste dos modelos locais

As constatações da Seção anterior apontam para a utilização de um modelo local no lugar de um modelo global. Sendo assim, ajustou-se o modelo de Regressão Geograficamente Ponderada (RGP) e o modelo de Regressão Beta Geograficamente Ponderada, tanto com ϕ local (RBGP) como com ϕ global (RBGPg), todos considerando a matriz de ponderação exponencial fixa com otimização do *bandwidth* por meio da Seção Áurea com validação cruzada, conforme apontam as Figuras 4.20 a 4.22. Adicionalmente, os modelos que levam em consideração a distribuição beta utilizaram função de ligação logito.

Corroborando as conclusões dos índices locais, a Tabela 4.2 apresenta um resumo dos resultados para os modelos locais, que foram consideravelmente melhores que os resultados dos modelos globais, não apenas pelas medidas aferidas, mas também pela adequação à natureza espacial dos dados. Lembrando ainda que aqui também foi utilizada a medida pseudo R^2 para os modelos RBGP e RBGPg, logo, a comparação dessa medida com a R^2 do modelo RGP deve ser realizada com precaução, utilizando-a sempre que possível em conjunto com outras medidas de ajuste.

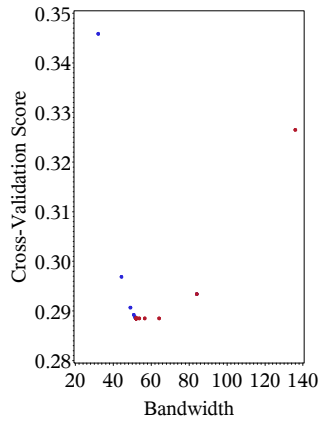


Figura 4.20: Otimização do *bandwidth* para o modelo RGP

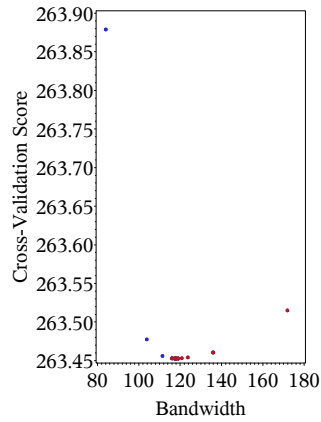


Figura 4.21: Otimização do *bandwidth* para o modelo RBGP

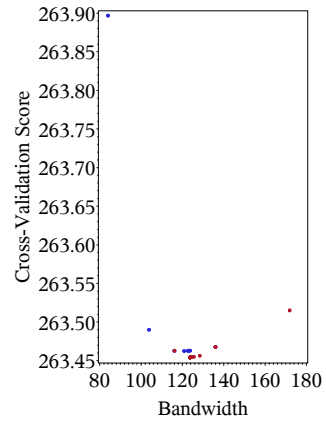


Figura 4.22: Otimização do *bandwidth* para o modelo RBGPg

Tabela 4.2: Resultados dos modelos locais

Variáveis	RGP			RBGP			RBGPg		
	Mín.	Média	Máx.	Mín.	Média	Máx.	Mín.	Média	Máx.
Intercepto	-0,0218	0,0589	0,1912	-2,4589	-2,2256	-1,8918	-2,4273	-2,2222	-1,9020
prop_pobres	-1,0788	-0,5058	0,0530	-6,4385	-5,1115	-3,3263	-6,4076	-5,1577	-3,6052
prop_urbana	0,0849	0,3238	0,4699	1,1951	2,2149	2,4916	1,2481	2,2227	2,4870
ϕ	-	-	-	43,8898	57,7713	81,5703	48,5175	48,5175	48,5175
R^2 ajustado**		0,8101			< 0,7731*			0,7701	
AICc		-203,6208			> -248,2436*			-242,0002	
Log-verossimilhança		165,9224			133,1329			129,6081	
# de parâmetros		18,9697			> 7,9622*			7,6419	
<i>bandwidth</i>		51,9476			117,8854			123,7369	

*: como não se sabe exatamente o número de parâmetros estimados pelo modelo, não é possível precisar estes valores, os quais contabilizam apenas 1 parâmetro para a precisão ϕ para fins comparativos.

** : pseudo R^2 para regressão beta.

De fato, observa-se que o modelo RGP conseguiu ainda boas medidas de ajuste para esses dados. Ao contrário do modelo global, este modelo local não produziu previsões fora do intervalo, conforme pode ser verificado na Figura 4.26. Entretanto, em comparação com os modelos locais que levam em consideração a distribuição beta, o modelo RGP peca no excesso de parâmetros necessários para o ajuste, quase o dobro em relação ao modelo RBGPg, o que acaba por penalizar o AICc da RGP, medida a qual foi pior que na regressão beta global e deixa a vantagem para os modelos propostos neste trabalho, a saber RBGP e RBGPg. Estes, por sinal, tiveram performances muito semelhantes entre si, como pode ser observado pelas medidas descritivas da Tabela 4.2 e também pela análise dos resíduos com auxílio das Figuras

4.24, 4.25, 4.27, 4.28, 4.30 e 4.31. Isso mostra que para esse estudo de caso em específico, utilizar o parâmetro de precisão estimado pelo modelo global não gera muito prejuízo para a análise.

Sobre os resíduos, as Figuras 4.23 a 4.25 junto com os resultados dos testes de normalidade mostram que os resíduos do modelo RGP são os que trazem maior fuga à normalidade, em comparação com os modelos RBGP e RBGPg, o que também ocorreu para os modelos globais. As Figuras 4.26 a 4.28 apresentam os resíduos contra os valores preditos para cada um dos modelos. De modo geral, observa-se que ainda há um certo padrão semelhante em todos, mas o modelo RBGP parece ser o mais aleatório. Nos mapas de distribuição espacial dos resíduos para os três modelos (Figuras 4.29, 4.30 e 4.31) encontram-se ainda os índices de Moran para os respectivos resíduos. Observe que as referidas Figuras refletem exatamente o que se espera ao se considerar o índice de Moran. Todos os modelos diminuíram o índice calculado para os modelos globais, mas o modelo RGP mostrou melhor resultado, já que se apresenta com maior aleatoriedade no espaço. Isso pode ocorrer porque globalmente os dados são assimétricos, conforme a Figura 4.9, mas localmente os dados podem ser bem comportados e não espelhar a realidade global. Nos resíduos da RBGP e da RBGPg, embora seja possível identificar áreas mais concentradas, o índice reduziu em mais da metade.

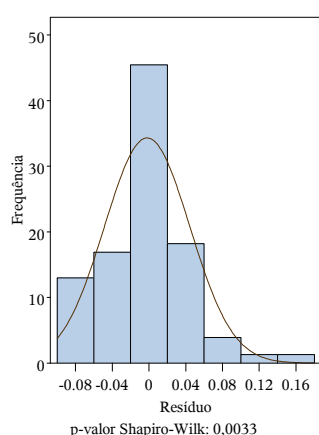


Figura 4.23: Distribuição dos resíduos do modelo RGP

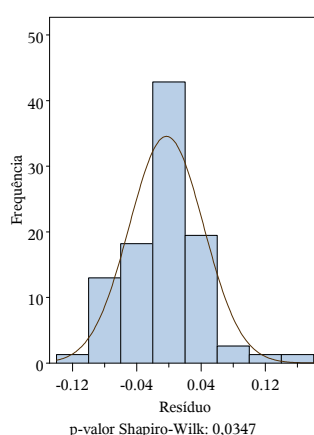


Figura 4.24: Distribuição dos resíduos do modelo RBGP

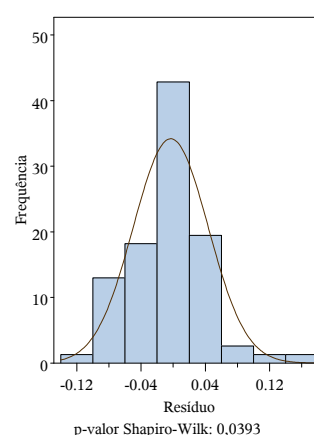


Figura 4.25: Distribuição dos resíduos do modelo RBGPg

Da mesma forma que na Seção anterior e de modo suplementar apresentam-se nas

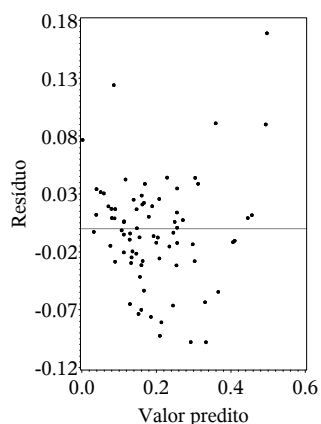


Figura 4.26: Resíduos pelos valores preditos do modelo RGP

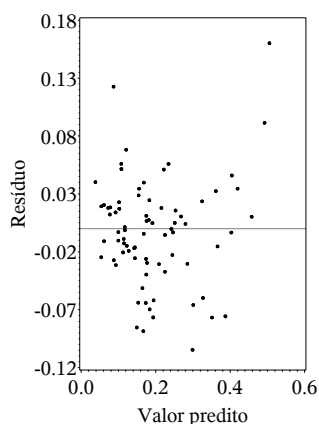


Figura 4.27: Resíduos pelos valores preditos do modelo RBGP

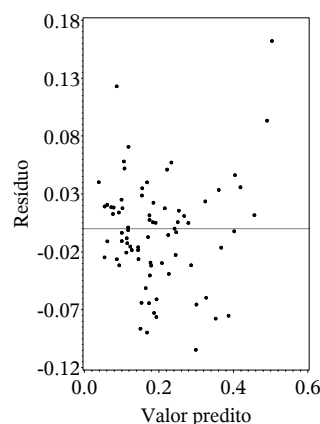
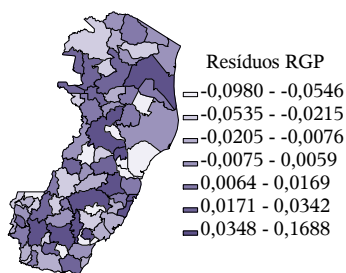
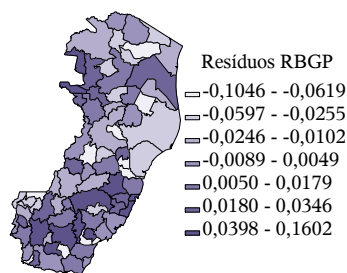


Figura 4.28: Resíduos pelos valores preditos do modelo RBGPg



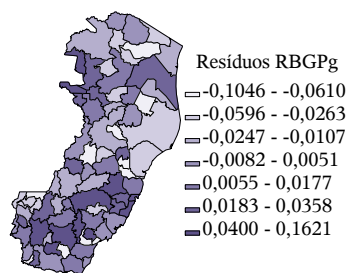
I=0,02

Figura 4.29: Disposição espacial dos resíduos da RGP



I=0,18

Figura 4.30: Disposição espacial dos resíduos da RBGP



I=0,2

Figura 4.31: Disposição espacial dos resíduos da RBGPg

Figuras 4.32 a 4.37 os resíduos padronizados contra os valores preditos e a disposição espacial dos mesmos para os modelos em análise. Note que o padrão observado nessas Figuras ainda segue o padrão obtido a partir dos resíduos puros, mas para o caso dos modelos que levam em consideração a distribuição beta os respectivos índices de Moran diminuíram ainda mais ao serem calculados sob os resíduos padronizados, acompanhando o que foi obtido também no caso dos modelos globais.

Para completar a análise desses modelos, apresentam-se ainda os parâmetros que foram significativos ao nível de significância de 5% nas Figuras 4.38 a 4.46, de acordo com o trabalho de da Silva e Fotheringham (2015), com adaptação do fator

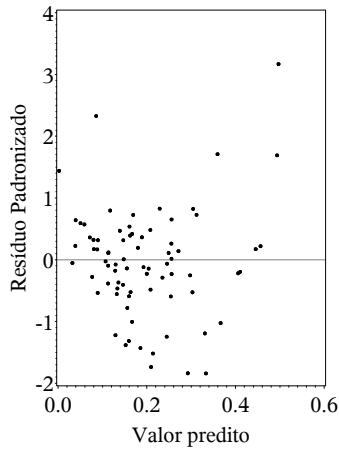


Figura 4.32: Resíduos padronizados pelos valores preditos do modelo RGP

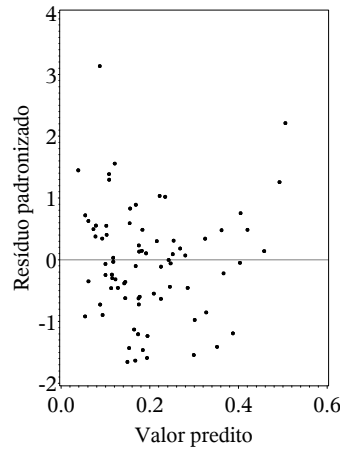


Figura 4.33: Resíduos padronizados pelos valores preditos do modelo RBGP

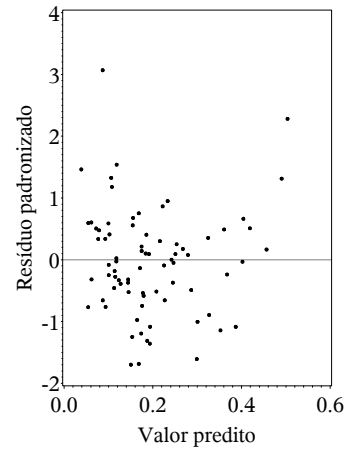
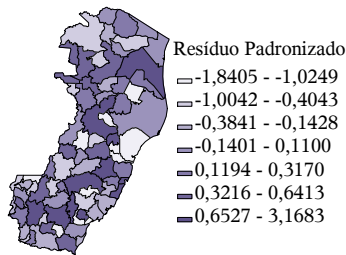
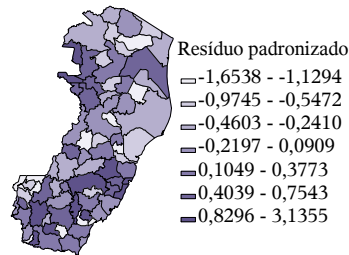


Figura 4.34: Resíduos padronizados pelos valores preditos do modelo RBGPg



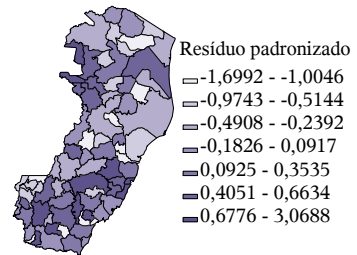
I=0,02

Figura 4.35: Disposição espacial dos resíduos padronizados da RGP



I=0,14

Figura 4.36: Disposição espacial dos resíduos padronizados da RBGP



I=0,17

Figura 4.37: Disposição espacial dos resíduos padronizados da RBGPg

de correção de k/ν_1 para $(k+1)/(\nu_1+1)$ devido à contribuição do parâmetro de precisão, novamente subestimado para o modelo RBGP.

Em suma, esses parâmetros indicam que de fato pode haver heterogeneidade na relação entre a variável resposta e as variáveis explicativas no espaço, mas idealmente espera-se que mesmo com a variação local, os parâmetros fiquem em torno da média apresentada pelos modelos globais na Tabela 4.1. A partir desse entendimento, a primeira circunstância que chama a atenção é que para o modelo RGP nenhum intercepto foi significativo (Figura 4.38), contrariando os resultados de seu

equivalente modelo global. Fora esse caso, os demais parâmetros desse modelo e dos demais estão em volta do valor médio global.

Além disso, pode-se dizer de modo geral que nas regiões mais escuras das Figura 4.39, 4.42 e 4.45, a proporção de pobres influencia com mais intensidade na proporção de domicílios que possuem telefone fixo em relação às regiões mais claras, mas em todo o caso, influencia sempre diminuindo essa proporção. Já o sinal positivo nas Figuras 4.40, 4.43 e 4.46 indica que a proporção de população urbana contribui para que haja maiores proporções de domicílios que possuem telefone fixo, principalmente nas áreas mais escuras. Observe que essas ilações estão de acordo com aquelas concluídas pela análise dos modelos globais e ainda vão além, indicando em quais regiões a variável resposta é mais ou menos suscetível a mudanças nessas variáveis explicativas.

Novamente, como foi utilizada a função de ligação logito nos modelos que dependem da distribuição beta, tem-se a opção de interpretar esses parâmetros em função da razão de chances, sem deixar de considerar a escala das variáveis respostas. Por exemplo, de acordo com o modelo RBGP, tem-se que em média o aumento em 10 pontos percentuais na proporção de população urbana aumenta em média 24,8% a chance de um domicílio possuir telefone fixo ($e^{0,1 \times 2,2149}$), mantida a outra variável inalterada.

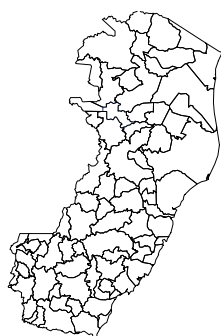


Figura 4.38: Interceptos significativos a 5% na RGP

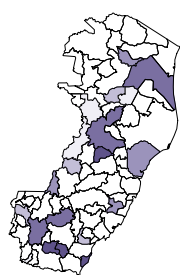


Figura 4.39: Parâmetros prop_pobres significativos a 5% na RGP

Proporção de Pobres
 -1.07880 - -0.97668
 -0.92094 - -0.81902
 -0.78982 - -0.70629
 -0.66652 - -0.64084
 -0.60818 - -0.54510

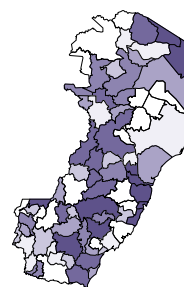


Figura 4.40: Parâmetros prop_urbana significativos a 5% na RGP

Proporção de Urbana
 0.24488 - 0.31712
 0.32700 - 0.34705
 0.34718 - 0.35598
 0.35790 - 0.38272
 0.38507 - 0.43121
 0.43556 - 0.46988

Adicionalmente, para o modelo RBGP faz-se importante analisar ainda a distri-

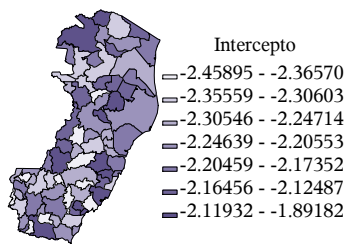


Figura 4.41: Interceptos significativos a 5% na RBGP

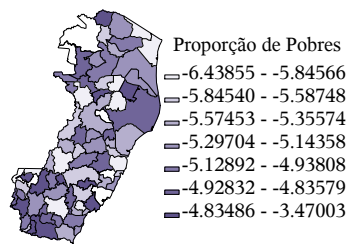


Figura 4.42: Parâmetros prop_pobres significativos a 5% na RBGP

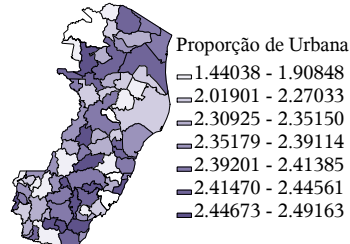


Figura 4.43: Parâmetros prop_urbana significativos a 5% na RBGP

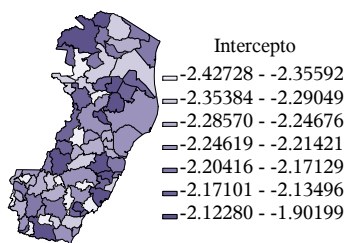


Figura 4.44: Interceptos significativos a 5% na RBGPg

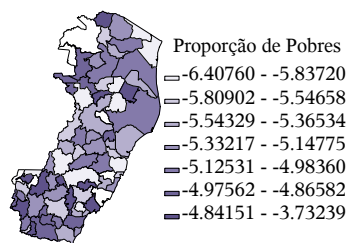


Figura 4.45: Parâmetros prop_pobres significativos a 5% na RBGPg

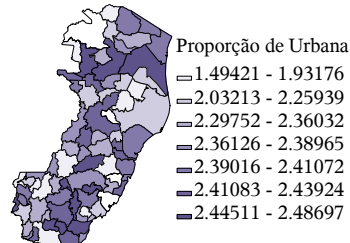


Figura 4.46: Parâmetros prop_urbana significativos a 5% na RBGPg

buição espacial dos parâmetros de precisão ϕ significativos ao nível de significância de 5%. Observe na Figura 4.47 que os valores variam bastante, de 43 a 81 aproximadamente. Os maiores valores encontram-se ao norte do estado, indicando menor variação nesses municípios. Já as áreas mais claras possuem menor precisão e, portanto, maior variância. Observe ainda que esse parâmetro foge mais ao valor médio do modelo global. Vale citar ainda que esse parâmetro foi avaliado somente para o modelo RBGP, pois no modelo RGP esse parâmetro não existe e para o modelo RBGPg esse valor é constante em todos os municípios.

Por fim, a Tabela 4.3 traz os resultados dos testes de pseudo-significância para 999 aleatorizações para os parâmetros dos modelos, a fim de verificar a estacionariedade dos mesmos, ou seja, se a variação deles é tão pequena de modo que poderiam ser descritos por um valor constante em vez de variar localmente. Observe que a

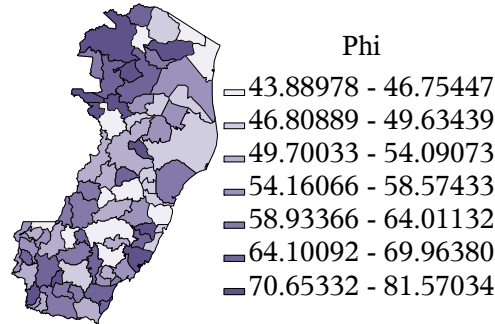


Figura 4.47: Parâmetros ϕ significativos a 5% na RBGP

indicação em todos os testes é de não rejeição da hipótese nula de estacionariedade, e que portanto um modelo global seria uma escolha mais acertada visando ao princípio da parcimônia. Entretanto, note que apesar de não rejeitar a hipótese nula, os modelos RBGP e RBGPg possuem p-valores menores que os do modelo RGP e que para a variável proporção de população urbana a hipótese nula poderia ser rejeitada se fosse utilizado um nível de significância de 15%, por exemplo.

Tabela 4.3: Teste de pseudo-significância para estacionariedade dos parâmetros

Variáveis	RGP	RBGP	RBGPg
	<i>p</i> -valor: Pseudo-Significância	<i>p</i> -valor: Pseudo-Significância	<i>p</i> -valor: Pseudo-Significância
Intercepto	1	0,81	0,792
prop_pobres	0,869	0,522	0,523
prop_urbana	0,735	0,139	0,124
ϕ	-	0,201	1

Findas as análises, observa-se o potencial dos modelos propostos nesta dissertação, principalmente onde o modelo RGP falhou em suas premissas, e cumpre-se o objetivo de prover uma técnica de análise alternativa às já existentes, mas que não se adequam totalmente aos dados analisados, quais sejam provenientes de taxas ou proporções com dependência espacial.

Capítulo 5

Conclusões

De longa data tem-se a consciência de que abordagens globais para modelar interações espaciais podem ocultar muitas informações importantes, mas apenas com a evolução tecnológica nas últimas décadas houve uma evolução nas técnicas para visualização, diagnóstico e modelagem de dados espaciais, as quais agregam muito mais conhecimento sobre os fenômenos que possuem relação com o espaço.

Este trabalho vem corroborar esse pensamento, já que a Regressão Beta Geograficamente Ponderada ora proposta, sendo um modelo que gera estimativas locais, fornece informações mais detalhadas sobre o espaço em conjunto com uma solução intuitiva na interpretação da não estacionariedade, as quais não seriam evidenciadas nas estatísticas globais. E por considerar a natureza dos dados provenientes de taxas ou proporções, o modelo RBGP se adequa melhor a dados que possuem essa condição em comparação com o modelo de Regressão Geograficamente Ponderada Clássico, por exemplo, que apesar de se mostrar com bom ajuste, apresentou fuga à suposição de normalidade, conforme aplicação desenvolvida no Capítulo 4.

5.1 Limitações do trabalho

A maior limitação do trabalho foi com relação à mensuração do número de parâmetros locais de precisão. A alternativa ora proposta, o modelo RBGPg, obteve boa performance na aplicação do Capítulo 4, porém é evidente que ela não é ideal, principalmente por utilizar uma estimativa proveniente de um modelo global, sem

considerar que os outros parâmetros são locais.

Uma segunda limitação seria a otimização por Seção Áurea, a qual pode levar a otimizações em pontos de mínimos locais em vez de globais, não apenas na RGP, mas também em outras aplicações.

Houve também um intento em adaptar o teste de Leung (Leung *et al.*, 2000) para verificar a estacionariedade dos parâmetros na RBGP, porém constatou-se que a aplicação para a RBGP não é tão direta, visto que as suposições adotadas pelos autores se utilizam da estimação em forma fechada dos parâmetros, $(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}$, o que não ocorre na regressão beta e portanto exige maior foco do pesquisador nesse assunto, o que não é propriamente alvo desta dissertação.

5.2 Trabalhos futuros

Em suma, o modelo RGBP se beneficia de muitos aspectos que já foram abordados no contexto da Regressão Geograficamente Ponderada e da Regressão Beta, mas há ainda outros temas cuja execução exige adaptações mais elaboradas, que podem ser temas de trabalhos futuros, como por exemplo:

- Modelo Semiparamétrico, que envolve estimação tanto de coeficientes locais como de globais no mesmo modelo, diferentemente do modelo RBGPg aqui proposto, que utiliza o parâmetro de precisão estimado pelo modelo global.
- Modelo que permite a existência de valores extremos.
- Modelo com uma abordagem bayesiana.
- Modelo para dados restritos a qualquer outro intervalo da reta. Conforme visto no Capítulo 1, nada obsta que o modelo de regressão beta seja aplicado para dados restritos a qualquer outro intervalo da reta, desde que seja realizada a devida transformação, o que não foi testado para o modelo RBGP nesta dissertação.
- Explorar outras funções de ponderação espacial ou funções de ligação e os efeitos destas sobre o ajuste.

- Estudar a distribuição do erro do modelo.
- Estudar a performance e/ou propor outras técnicas de otimização para o *bandwidth*.
- Adaptar o teste de Leung, se possível de modo a incluir a avaliação dos parâmetros de precisão ϕ .
- Formalizar técnicas para identificar e lidar com observações discrepantes, multicolinearidade e outros problemas que podem surgir na adequabilidade e no ajuste do modelo.
- Implementar e estudar outros resíduos da regressão beta no contexto da RBGP.

Referências Bibliográficas

- Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, 6(3):251–262.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. NATO Asi Series. Series E, Applied Sciences. Springer.
- Anselin, L. (1995). Local indicators of spatial association - lisa. *Geographical Analysis*, 27(2):93–115.
- Anselin, L. (1996). The moran scatterplot as an esda tool to assess local instability in spatial association. *Spatial Analytical Perspectives on GIS*, 111:111–125.
- Anselin, L. (1998). Exploratory spatial data analysis in a geocomputational environment. In: *Geocomputation, a primer*, pages 77–94. Wiley.
- Appleton, D. R., French, J. M., e Vanderpump, M. P. (1996). Ignoring a covariate: An example of simpson’s paradox. *The American Statistician*, 50(4):340–341.
- Atkinson, P. M., German, S. E., Sear, D. A., e Clark, M. J. (2003). Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis*, 35(1):58–82.
- Barclay, M. J., Smith, C. W., e Watts, R. L. (1997). The determinants of corporate leverage and dividend policies. *Journal of Financial Education*, 7(4):1–15.
- Bayer, F. M. e Cribari-Neto, F. (2013). Bartlett corrections in beta regression models. *Journal of Statistical Planning and Inference*, 143(3):531–547.
- Bitter, C., Mulligan, G. F., e Dall’erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, 9(1):7–27.
- Branscum, A. J., Johnson, W. O., e Thurmond, M. C. (2007). Bayesian beta regression: Applications to household expenditure data and genetic distance between

- foot-and-mouth disease viruses. *Australian & New Zealand Journal of Statistics*, 49(3):287–301.
- Brehm, J. e Gates, S. (1993). Donut shops and speed traps: Evaluating models of supervision on police behavior. *American Journal of Political Science*, 37(2):555–581.
- Brunsdon, C., Fotheringham, A. S., e Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4):281–298.
- Brunsdon, C., Fotheringham, A. S., e Charlton, M. E. (1998). Geographically weighted regression - modelling spatial non-stationarity. *The Statistician*, 47(3):431–443.
- Burnham, K. P. e Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.
- Casetti, E. (1972). Generating models by the expansion method: Applications to geographical research. *Geographical Analysis*, 4(1):81–91.
- Castellani, M., Pattitoni, P., e Scorcu, A. E. (2012). Visual artist price heterogeneity. *Economics and Business Letters*, 1(3):16–22.
- Cepeda-Cuervo, E. e Núñez-Antón, V. (2013). Spatial double generalized beta regression models extensions and application to study quality of education in colombia. *Journal of Educational and Behavioral Statistics*, 38(6):604–628.
- Charlton, M., Fotheringham, S., e Brunsdon, C. (2009). Geographically weighted regression. *White paper. National Centre for Geocomputation. National University of Ireland Maynooth*.
- Cliff, A. e Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, 4(3):267–284.
- Comber, A., Brunsdon, C., e Phillips, M. (2012). The varying impact of geographic distance as a predictor of dissatisfaction over facility access. *Applied Spatial Analysis and Policy*, 5(4):333–352.
- Cook, D. O., Kieschnick, R., e McCullough, B. D. (2008). Regression analysis of proportions in finance with self selection. *Journal of Empirical Finance*, 15(5):860–867.

- Cox, C. (1996). Nonlinear quasi-likelihood models: applications to continuous proportions. *Computational Statistics & Data Analysis*, 21(4):449–461.
- Cribari-Neto, F. e Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34(2):1–24.
- da Silva, A. R. e Fotheringham, A. S. (2015). The multiple testing issue in geographically weighted regression. *Geographical Analysis. Forthcoming*.
- de Sousa, A. M. O., Pereira, J. M. O. C., Mota, B. W. C., e dos Santos, J. R. (2013). Análise da distribuição de fogos ativos detectados pelo sensor atsr 2 e sua relação com variáveis demográficas, ambientais e infra-estruturais: Comparabilidade entre métodos de regressão ols e gwr. *Revista Brasileira de Cartografia*, 64(5):703–721.
- Dyke, G. e Patterson, H. (1952). Analysis of factorial arrangements when the data are proportions. *Biometrics*, 8(1):1–12.
- Eskelson, B. N., Madsen, L., Hagar, J. C., e Temesgen, H. (2011). Estimating riparian understory vegetation cover with beta regression and copula models. *Forest Science*, 57(3):212–221.
- Espinheira, P. L., Ferrari, S. L., e Cribari-Neto, F. (2008a). Influence diagnostics in beta regression. *Computational Statistics & Data Analysis*, 52(9):4417 – 4431.
- Espinheira, P. L., Ferrari, S. L., e Cribari-Neto, F. (2008b). On beta regression residuals. *Journal of Applied Statistics*, 35(4):407–419.
- Falls, L. W. (1974). The beta distribution: a statistical model for world cloud cover. *Journal of Geophysical Research*, 79(9):1261–1264.
- Ferrari, S. e Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7):799–815.
- Fleiss, J. L., Levin, B., e Paik, M. C. (2013). *Statistical methods for rates and proportions*. John Wiley & Sons.
- Foster, S. A. e Gorr, W. L. (1986). An adaptive filter for estimating spatially-varying parameters: Application to modeling police hours spent in response to calls for service. *Management Science*, 32(7):878–889.
- Fotheringham, A., Brunson, C., e Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

- Fotheringham, A. S., Charlton, M., e Brunson, C. (1996). The geography of parameter space: an investigation of spatial non-stationarity. *International Journal of Geographical Information Systems*, 10(5):605–627.
- Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5(3):115–146.
- Getis, A. e Ord, K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3):189–206.
- Goldstein, H. (1987). *Multilevel models in education and social research*. Oxford University Press.
- Greig, D. M. (1980). *Optimisation*. Longman Publishing Group.
- Gupta, A. e Nadarajah, S. (2004). *Handbook of Beta Distribution and Its Applications*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis.
- Hope, A. C. (1968). A simplified monte carlo significance test procedure. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(3):582–598.
- Hunger, M., Döring, A., e Holle, R. (2012). Longitudinal beta regression models for analyzing health-related quality of life scores over time. *BMC Medical Research Methodology*, 12(1):144.
- IBGE (2010). Censo demográfico 2010. Disponível em: <http://www.censo2010.ibge.gov.br>. Acesso em: 02 set. 2015.
- Jorgensen, B. (1997). *The theory of dispersion models*. CRC Press.
- Kieschnick, R. e McCullough, B. D. (2003). Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. *Statistical modelling*, 3(3):193–213.
- Krige, D. (1966). Two-dimensional weighted moving average trend surfaces for ore-evaluation. *Journal of the South African Institute of Mining and Metallurgy*, 66:13–38.
- Lawson, F., Schuurman, N., Amram, O., e Nathens, A. B. (2015). A geospatial analysis of the relationship between neighbourhood socioeconomic status and adult severe injury in greater vancouver. *Injury Prevention*, 21(4):260–265.
- LeSage, J. P. (2004). A family of geographically weighted regression models. In: *Advances in spatial econometrics*, pages 241–264. Springer.

- Leung, Y., Chang-Lin, M., e Wen-Xiu, Z. (2000). Statistical tests for spatial non-stationarity based on the geographically weighted regression model. *Environment and Planning A*, 32(1):9–32.
- McCullagh, P., Nelder, J. A., e McCullagh, P. (1989). *Generalized linear models*, volume 2. Chapman and Hall London.
- McDonald, J. B. e Xu, Y. J. (1995). A generalization of the beta distribution with applications. *Journal of Econometrics*, 66(1):133–152.
- Meaney, C. e Moineddin, R. (2014). A monte carlo simulation study comparing linear regression, beta regression, variable-dispersion beta regression and fractional logit regression at recovering average difference measures in a two sample design. *BMC medical research methodology*, 14(1):14.
- Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23.
- Mulatu, D. W., Van der Veen, A., Becht, R., Van Oel, P. R., e Bekalo, D. J. (2013). Accounting for spatial non-stationarity to estimate population distribution using land use/cover. case study: The lake Naivasha basin, Kenya. *Journal of Settlements and Spatial Planning*, 4(1):33–44.
- Nakaya, T., Fotheringham, A., Brunson, C., e Charlton, M. (2005). Geographically weighted poisson regression for disease association mapping. *Statistics in Medicine*, 24(17):2695–2717.
- Nelder, J. A. e Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–384.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., e Wasserman, W. (1996). *Applied linear statistical models*, volume 4. Irwin Chicago.
- Ord, J. K. e Getis, A. (1995). Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis*, 27(4):286–306.
- Ord, J. K. e Getis, A. (2001). Testing for local spatial autocorrelation in the presence of global autocorrelation. *Journal of Regional Science*, 41(3):411–432.
- Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349):120–126.

- Ospina, R., Cribari-Neto, F., e Vasconcellos, K. L. (2006). Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis*, 51(2):960–981.
- Peplonska, B., Bukowska, A., Sobala, W., Reszka, E., Gromadzinska, J., Wasowicz, W., Lie, J. A., Kjuus, H., e Ursin, G. (2012). Rotating night shift work and mammographic density. *Cancer Epidemiology Biomarkers & Prevention*, 21(7):1028–1037.
- Pereira, T. L., Souza, T. C., e Cribari-Neto, F. (2014). Uma avaliação da eficiência do gasto público nas regiões do brasil. *Ciência e Natura*, 36(Ed. Especial):23–36.
- PNUD, IPEA, e FJP (2010). Atlas do desenvolvimento humano no brasil. Disponível em: <http://http://www.atlasbrasil.org.br>. Acesso em: 02 set. 2015.
- Rocha, A. V. e Cribari-Neto, F. (2009). Beta autoregressive moving average models. *Test*, 18(3):529–545.
- Rogerson, P. A. (1999). The detection of clusters using a spatial version of the chi-square goodness-of-fit statistic. *Geographical Analysis*, 31(2):130–147.
- Rydlewski, J. P. (2007). Beta-regression model for periodic data with a trend. *Univ. Iagel. Acta Math*, 45:211–222.
- Schmid, M., Wickler, F., Maloney, K. O., Mitchell, R., Fenske, N., e Mayr, A. (2013). Boosted beta regression. *PloS one*, 8(4):e61623.
- Silva, A. R. e Rodrigues, T. C. V. (2014). Geographically weighted negative binomial regression - incorporating overdispersion. *Statistics and Computing*, 24(5):769–783.
- Simas, A. B., Barreto-Souza, W., e Rocha, A. V. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis*, 54(2):348–366.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13(2):238–241.
- Sulaiman, M. Y., Oo, W. H., Wahab, M. A., e Zakaria, A. (1999). Application of beta distribution model to malaysian sunshine data. *Renewable Energy*, 18(4):573–579.
- Swearingen, C. J., Castro, M. S. M., e Bursac, Z. (2011). Modeling percentage outcomes: the% beta_regression macro. In: *Proceedings of the SAS Global Forum, Las Vegas, NV*, pages 1–12. Citeseer.

- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica: Journal of the Econometric Society*, 26(1):24–36.
- Vasconcellos, K. L. e Cribari-Neto, F. (2005). Improved maximum likelihood estimation in a new class of beta regression models. *Brazilian Journal of probability and Statistics*, 19(1):13–31.
- Zhang, L. e Shi, H. (2004). Local modeling of tree growth by geographically weighted regression. *Forest Science*, 50(2):225–244.
- Zimprich, D. (2010). Modeling change in skewed variables using mixed beta regression models. *Research in Human Development*, 7(1):9–26.

Apêndice A

Estimação, inferência e diagnóstico da Regressão Beta com função de ligação logito

n = número de linhas de \mathbf{y}

$$\mathbf{y}_{transformado} = \check{\mathbf{y}} = g(\mathbf{y}) = \log(\mathbf{y}/(1 - \mathbf{y}))$$

$$\boldsymbol{\beta}_{inicial} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \check{\mathbf{y}}$$

$$\check{\mathbf{e}} = \check{\mathbf{y}} - \mathbf{X} \boldsymbol{\beta}_{inicial}$$

$$\check{\boldsymbol{\mu}} = g^{-1}(\mathbf{X} \boldsymbol{\beta}_{inicial}) = \exp(\mathbf{X} \boldsymbol{\beta}_{inicial}) / (1 + \exp(\mathbf{X} \boldsymbol{\beta}_{inicial}))$$

$$g'(\check{\boldsymbol{\mu}}) = (1/\check{\boldsymbol{\mu}}) + (1/(1 - \check{\boldsymbol{\mu}}))$$

$$\check{\sigma}^2 = \frac{\check{\mathbf{e}}^T \check{\mathbf{e}}}{(n-k)g'(\check{\boldsymbol{\mu}})\#g'(\check{\boldsymbol{\mu}})}$$

$$\phi_{inicial} = 0$$

Para $i = 1$ até n faça{

$$\phi_{inicial} = \phi_{inicial} + \check{\boldsymbol{\mu}}[i](1 - \check{\boldsymbol{\mu}}[i]) / (n\check{\sigma}^2[i])$$

}

$$\boldsymbol{\beta} = \boldsymbol{\beta}_{inicial}$$

$$\phi = \phi_{inicial} - 1$$

Função de Máxima Verossimilhança:

$$\max_like(\boldsymbol{\beta}, \phi) = \{$$

$$\logver1 = \lgamma(\phi \boldsymbol{\mu})$$

$$\logver2 = \lgamma((1 - \boldsymbol{\mu})\phi)$$

$$\logver3 = (\phi \boldsymbol{\mu} - 1)\#\log(\mathbf{y})$$

$$\logver4 = ((1 - \boldsymbol{\mu})\phi - 1)\#\log(1 - \mathbf{y})$$

$$\logver = 0$$

Para $i = 1$ até n faça {
 $max_like = logver + lgamma(\phi) - logver1[i] - logver2[i] + logver3[i] + logver4[i]$
}
}
it=0
dif=1
Enquanto ($|dif| > 10^{-8}$ e $it < maxit$) faça {
 $\boldsymbol{\mu} = g^{-1}(X\boldsymbol{\beta}) = exp(\mathbf{X}\boldsymbol{\beta}) / (1 + exp(\mathbf{X}\boldsymbol{\beta}))$
 $g'(\boldsymbol{\mu}) = (1/\boldsymbol{\mu}) + (1/(1 - \boldsymbol{\mu}))$
 $\mathbf{y}^* = log(\mathbf{y}/(1 - \mathbf{y}))$
 $\boldsymbol{\mu}^* = \psi'(\boldsymbol{\mu}\phi) - \psi'((1 - \boldsymbol{\mu})\phi)$
 $\mathbf{T} = diag(1/g'(\boldsymbol{\mu}))$
 $\mathbf{c} = \phi(\psi''(\boldsymbol{\mu}\phi)\#\boldsymbol{\mu} - \psi''((1 - \boldsymbol{\mu})\phi)\#(1 - \boldsymbol{\mu}))$
 $\mathbf{Z} = diag((\phi\psi''(\boldsymbol{\mu}\phi) + \phi\psi''((1 - \boldsymbol{\mu})\phi)) / (g'(\boldsymbol{\mu})\#g'(\boldsymbol{\mu})))$
 $\mathbf{D} = diag(\psi''(g'(\boldsymbol{\mu})\phi)\#g'(\boldsymbol{\mu})\#g'(\boldsymbol{\mu}) + \psi''((1 - g'(\boldsymbol{\mu}))\phi)\#(1 - g'(\boldsymbol{\mu}))\#(1 - g'(\boldsymbol{\mu})) - \psi''(\phi))$
 $\mathbf{k}_{\beta\beta} = \phi\mathbf{X}^T\mathbf{Z}\mathbf{X}$
 $\mathbf{k}_{\beta\phi} = \mathbf{X}^T\mathbf{T}\mathbf{c}$
 $\mathbf{k}_{\phi\phi} = tr(\mathbf{D})$
 $\mathbf{km} = (\mathbf{k}_{\beta\beta} \|\mathbf{k}_{\beta\phi}\|) / (\mathbf{k}_{\beta\phi}^T \|\mathbf{k}_{\phi\phi}\|)$
 $\mathbf{k}^{-1} = \mathbf{km}^{-1}$
 $U_{\beta}(\boldsymbol{\beta}, \phi) = \phi(\mathbf{X}^T\mathbf{T}(\mathbf{y}^* - \boldsymbol{\mu}^*))$
 $U_{\phi}(\boldsymbol{\beta}, \phi) = 0$
Para $i = 1$ até n faça {
 $U_{\phi}(\boldsymbol{\beta}, \phi) = U_{\phi}(\boldsymbol{\beta}, \phi) + \mu[i](y^*[i] - \mu^*[i]) + log(1 - y[i]) - \psi'((1 - \mu[i])\phi) + \psi'(\phi)$
}
 $\mathbf{U} = U_{\beta}(\boldsymbol{\beta}, \phi) \|\mathbf{U}_{\phi}(\boldsymbol{\beta}, \phi)$
 $param = \boldsymbol{\beta} \|\phi$
 $param^{n+1} = param + \mathbf{k}^{-1}\mathbf{U}$
 $dif = max_like(param) - max_like(param^{n+1})$
 $\boldsymbol{\beta} = \boldsymbol{\beta}^{n+1}$
 $\phi = \phi^{n+1}$
 $it = it + 1$
}
}
Inferência:
 $\boldsymbol{\sigma} = \sqrt{vecdiag(\mathbf{k}^{-1})}$
 $tstat = (\boldsymbol{\beta}/\phi)/\boldsymbol{\sigma}$
 $probt = 2(1 - probt(|tstat|, n - k))$

Resíduos e diagnóstico:

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$$

$$pseudoR^2 = corr(\boldsymbol{\eta}, \hat{\mathbf{y}})^2$$

$$oddsratio = exp(\boldsymbol{\beta}[2 : nrow(\boldsymbol{\beta})])$$

$$\hat{\mathbf{y}} = exp(\boldsymbol{\eta}) / (1 + exp(\boldsymbol{\eta}))$$

$$\mathbf{res} = \mathbf{y} - \hat{\mathbf{y}}$$

Deviance residual:

$$lgamma1t = lgamma(\phi\mathbf{y})$$

$$lgamma2t = lgamma((1 - \mathbf{y})\phi)$$

$$lgamma3t = (\phi\mathbf{y} - 1)\#log(\mathbf{y})$$

$$lgamma4t = ((1 - \mathbf{y})\phi - 1)\#log(1 - \mathbf{y})$$

$$lgamma1 = lgamma(\phi\boldsymbol{\mu})$$

$$lgamma2 = lgamma((1 - \boldsymbol{\mu})\phi)$$

$$lgamma3 = (\phi\boldsymbol{\mu} - 1)\#log(\mathbf{y})$$

$$lgamma4 = ((1 - \boldsymbol{\mu})\phi - 1)\#log(1 - \mathbf{y})$$

$$\ell_t(\tilde{\boldsymbol{\mu}}_t, \hat{\phi}) = lgamma(\phi) - lgamma1t - lgamma2t + lgamma3t + lgamma4t$$

$$\ell_t(\hat{\boldsymbol{\mu}}_t, \hat{\phi}) = lgamma(\phi) - lgamma1 - lgamma2 + lgamma3 + lgamma4$$

$$\mathbf{r}_t^d = sgn(\mathbf{y} - \boldsymbol{\mu})\#sqrt(2|\ell_t(\tilde{\boldsymbol{\mu}}_t) - \ell_t(\hat{\boldsymbol{\mu}}_t, \hat{\phi})|)$$

Resíduo padronizado:

$$var(\mathbf{y}) = \boldsymbol{\mu}\#(1 - \boldsymbol{\mu}) / (1 + \phi)$$

$$\mathbf{r}_t = (\mathbf{y} - \boldsymbol{\mu}) / sqrt(var(\mathbf{y}))$$

Generalized leverage:

$$\mathbf{M} = diag(1 / (\mathbf{y}\#(1 - \mathbf{y})))$$

$$g''(\boldsymbol{\mu}) = 1 / ((1 - \boldsymbol{\mu})\#(1 - \boldsymbol{\mu})) - (1 / (\boldsymbol{\mu}\#\boldsymbol{\mu}))$$

$$\mathbf{Q} = diag((\phi(\psi'(\boldsymbol{\mu}\phi) + \psi'((1 - \boldsymbol{\mu})\phi)) + (\mathbf{y}^* - \boldsymbol{\mu}^*)\#g''(\boldsymbol{\mu}) / g'(\boldsymbol{\mu})) / g'(\boldsymbol{\mu})\#g'(\boldsymbol{\mu}))$$

$$\mathbf{f} = \mathbf{c} - (\mathbf{y}^* - \boldsymbol{\mu}^*)$$

$$\gamma = tr(\mathbf{D}) - (1/\phi)\mathbf{f}^T\mathbf{T}^T\mathbf{X}(\mathbf{X}^T\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{T}\mathbf{f}$$

$$\mathbf{b} = -(\mathbf{y} - \boldsymbol{\mu}) / (\mathbf{y}\#(1 - \mathbf{y}))$$

$$GL(\boldsymbol{\beta}) = \mathbf{TX}(\mathbf{X}^T\mathbf{ZX})^{-1}\mathbf{X}^T\mathbf{TM}$$

$$GL(\boldsymbol{\beta}, \phi) = vecdiag(GL(\boldsymbol{\beta}) + \frac{1}{\gamma\phi}\mathbf{TX}(\mathbf{X}^T\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{T}\mathbf{f}(\mathbf{f}^T\mathbf{TX}(\mathbf{X}^T\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{TM} - \mathbf{b}^T))$$

Cook:

$$\mathbf{H} = vecdiag(sqrt(\mathbf{Z})\mathbf{X}(\mathbf{X}^T\mathbf{ZX})^{-1}\mathbf{X}^T sqrt(\mathbf{Z}))$$

$$cookD = \mathbf{H}\#(\mathbf{r}_t\#\mathbf{r}_t) / (k(1 - \mathbf{H})\#(1 - \mathbf{H}))$$

$lgamma(.) = log(\Gamma(.))$

$\#$: multiplicação elemento por elemento

$diag(.)$: matriz diagonal com os elementos do vetor

$\psi'(\cdot)$: função digamma
 $\psi''(\cdot)$: função trigamma
||: concatenar horizontal
//: concatenar vertical
vecdiag(.): vetor com a diagonal da matriz
probt(.): probabilidade da distribuição t de *student*
corr(.): correlação

Apêndice B

Algoritmo de estimação da RBGP e RBGPg com função de ligação logito

n = número de linhas de \mathbf{y}

$$\mathbf{y}_{transformado} = \tilde{\mathbf{y}} = g(\mathbf{y}) = \log(\mathbf{y}/(1 - \mathbf{y}))$$

Para $i = 1 \dots, n$ faça {

Para $j = 1 \dots, n$ faça {

$$w_i[j, j] = k(d_{ij})$$

}

$$\boldsymbol{\beta}_{inicial}[i] = (\mathbf{X}^T \mathbf{W}_i \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}_i \tilde{\mathbf{y}}$$

$$\hat{\mathbf{y}}[i] = \mathbf{X}[i,] \boldsymbol{\beta}_{inicial}[i]$$

}

$$\tilde{\mathbf{e}} = \tilde{\mathbf{y}} - \hat{\mathbf{y}}$$

$$\boldsymbol{\eta}_{inicial} = \mathbf{X} \boldsymbol{\beta}_{inicial}$$

$$\tilde{\boldsymbol{\mu}} = g^{-1}(\boldsymbol{\eta}) = \exp(\boldsymbol{\eta}) / (1 + \exp(\boldsymbol{\eta}))$$

$$g'(\tilde{\boldsymbol{\mu}}) = (1/\tilde{\boldsymbol{\mu}}) + (1/(1 - \tilde{\boldsymbol{\mu}}))$$

$$\tilde{\boldsymbol{\sigma}}^2 = \frac{\tilde{\mathbf{e}}^T \tilde{\mathbf{e}}}{(n - \nu_1) g'(\tilde{\boldsymbol{\mu}}) \# g'(\tilde{\boldsymbol{\mu}})}$$

Se for o modelo RBGP, então faça {

$$\boldsymbol{\phi}_{inicial} = \mathbf{0}$$

Para $i = 1$ até n faça{

Para $j = 1$ até n faça{

$$\boldsymbol{\phi}_{inicial}[i] = \boldsymbol{\phi}_{inicial}[i] + \tilde{\boldsymbol{\mu}}[j, i](1 - \tilde{\boldsymbol{\mu}}[j, i]) / (n \tilde{\boldsymbol{\sigma}}^2[j, i])$$

}

}

$$\boldsymbol{\phi} = \boldsymbol{\phi}_{inicial} - 1$$

}

Se for o modelo RBGPg, então faça {

$\phi = \text{repita } \phi_g \text{ n vezes}$

}

$\beta = \beta_{inicial}$

Função de Máxima Verossimilhança:

$\max_like(\beta, \phi) = \{$

$\logver1 = lgamma(\phi\mu)$

$\logver2 = lgamma((1 - \mu)\phi)$

$\logver3 = (\phi\mu - 1)\#log(\mathbf{y})$

$\logver4 = ((1 - \mu)\phi - 1)\#log(1 - \mathbf{y})$

$\logver = 0$

Para $i = 1$ até n faça {

$\max_like = \logver +$

$+(lgamma(\phi) - \logver1[i] - \logver2[i] + \logver3[i] + \logver4[i])w[i, i]$

}

}

Para $i = 1 \dots, n$ faça {

Para $j = 1 \dots, n$ faça {

$w_i[j, j] = k(d_{ij})$

}

it=0

dif=1

Enquanto ($|dif| > 10^{-8}$ e $it < \text{maxit}$) faça {

$\mu = g^{-1}(X\beta[, i]) = \exp(\mathbf{X}\beta[, i]) / (1 + \exp(\mathbf{X}\beta[, i]))$

$g'(\mu) = (1/\mu) + (1/(1 - \mu))$

$\mathbf{y}^* = \log(\mathbf{y}/(1 - \mathbf{y}))$

$\mu^* = \psi'(\mu\phi[i]) - \psi'((1 - \mu)\phi[i])$

$\mathbf{T} = \text{diag}(1/g'(\mu))$

$\mathbf{c} = \phi[i](\psi''(\mu\phi[i])\#\mu - \psi''((1 - \mu)\phi[i])\#(1 - \mu))$

$\mathbf{Z} = \text{diag}((\phi[i]\psi''(\mu\phi[i]) + \phi[i]\psi''((1 - \mu)\phi[i]))/(g'(\mu)\#g'(\mu)))$

$d = \psi''(g'(\mu)\phi[i])\#g'(\mu)\#g'(\mu) + \psi''((1 - g'(\mu))\phi[i])\#(1 - g'(\mu))\#(1 - g'(\mu)) - \psi''(\phi[i])$

$\mathbf{D} = \mathbf{W}_i \text{diag}(d)$

$\mathbf{k}_{\beta\beta} = \phi[i]\mathbf{X}^T\mathbf{W}\mathbf{Z}\mathbf{X}$

$\mathbf{k}_{\beta\phi} = \mathbf{X}^T\mathbf{W}\mathbf{T}\mathbf{c}$

$\mathbf{k}_{\phi\phi} = \text{tr}(\mathbf{D})$

```

km = (kββ||kβφ)/(kβφT||kφφ)
k-1 = km-1
Uβ(β, φ) = φ[i](XWTT(y* - μ*))
Uφ(β, φ) = 0
Para l = 1 até n faça {
  Uφ(β, φ) = Uφ(β, φ) + μ[l](y*[l] - μ*[l]) + log(1 - y[l]) - ψ'((1 - μ[l])φ[i]) + ψ'(φ[i])
}
U = Uβ(β, φ)||Uφ(β, φ)
param = β||φ
paramn+1 = param + k-1U
dif = max_like(param) - max_like(paramn+1)
β[,i] = paramn+1[1 : k]
Se for o modelo RBGP, então faça {
  φ = paramn+1[k + 1]
}
Se for o modelo RBGPg, então faça {
  φ = φg
}
it = it + 1
}
}

```

[*i*, *j*]: indica o elemento da linha *i* e coluna *j* de uma matriz

k(.): função de ponderação

*d*_{*ij*}: medida de distância entre o ponto *i* e o ponto *j*

#: multiplicação elemento por elemento

lgamma(.) = log(Γ(.))