

UNIVERSIDADE FEDERAL DO PARANÁ

FERNANDO GOMES MORO

ESTRUTURAS DE COVARIÂNCIA DEFINIDAS POR COVARIÁVEIS

CURITIBA

2019

FERNANDO GOMES MORO

ESTRUTURAS DE COVARIÂNCIA DEFINIDAS POR COVARIÁVEIS

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Métodos Estatísticos Aplicados à Engenharia no Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Métodos Estatísticos Aplicados à Engenharia.*

Orientador: Prof. *PhD* Paulo Justiniano Ribeiro Jr..

CURITIBA

2019

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

M867e Moro, Fernando Gomes
Estruturas de covariância definidas por covariáveis [recurso eletrônico] / Fernando Gomes Moro – Curitiba, 2019.

Dissertação - Universidade Federal do Paraná, Setor de Ciências Exatas, Programa de Pós-Graduação em Métodos Numéricos em Engenharia.

Orientador: Prof. PhD Paulo Justiniano Ribeiro Jr.

1. Análise de covariância. 2. Modelos lineares (Estatística). I. Universidade Federal do Paraná. II. Ribeiro Jr., Paulo Justiniano. III. Título.

CDD: 519.536

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em MÉTODOS NUMÉRICOS EM ENGENHARIA da Universidade Federal do Paraná foram convocados para realizar a arguição da Dissertação de Mestrado de **FERNANDO GOMES MORO** intitulada: **Estruturas de covariância definidas por covariáveis**, sob orientação do Prof. Dr. PAULO JUSTINIANO RIBEIRO JÚNIOR, que após terem inquirido o aluno e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestre está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 04 de Setembro de 2019.



PAULO JUSTINIANO RIBEIRO JÚNIOR

Presidente da Banca Examinadora (UNIVERSIDADE FEDERAL DO PARANÁ)



RENATO MARTINS ASSUNÇÃO

Avaliador Externo (UNIVERSIDADE FEDERAL DE MINAS GERAIS)



WAGNER HUGO BONAT

Avaliador Externo (UNIVERSIDADE FEDERAL DO PARANÁ)

À todos aqueles da academia científica que buscam aprimorar seu conhecimento técnico

AGRADECIMENTOS

Ao meu orientador Professor Paulo Justiniano Ribeiro Jr. por toda dedicação e paciência.
Ao professor Wagner Hugo Bonat por todo auxílio prestado, e aos demais membros da banca.
À minha família nuclear, aos meus avós Maria Shirley Soares Gomes e Waldemar Bittencourt Gomes (*In Memoriam*), e aos meus tios e padrinhos Leila Soares Gomes, Marlene Gomes Terra e Benonio Gomes Terra.

RESUMO

Modelos de regressão usualmente modelam a relação, da variável resposta com as covariáveis, através da especificação explícita de uma função para média. A estrutura de covariância é, geralmente, especificada sob independência das observações ou expressa a dependência entre as mesmas causada por fontes de variabilidade desconhecidas porém decorrentes de alguma indexação dos dados, como exemplo as estruturas de dependência longitudinais, espaciais e temporais. No presente trabalho explora-se a modelagem da estrutura de covariância marginal através das especificações baseadas no espaço gerado por um conjunto de covariáveis. Predições condicionais permitem obter trajetórias não lineares que descrevem o comportamento da variável resposta. Diferentes especificações da estrutura de covariância implicam em diferentes características da trajetória produzida pela predição condicional, como suavidade e continuidade. Tal abordagem permite explorar padrões dos dados não capturados pela especificação de modelos para média e substituir ou complementar estruturas de dependência temporais ou espaciais. Foram consideradas aqui três tipos distintos de funções de distâncias que definem a estrutura de covariância: funções contínuas, funções discretas e funções autorregressivas. Os modelos lineares generalizados de covariância linear (MCGLM) foram considerados como ferramenta inferencial por, dentre outros aspectos, possibilitar a modelagem da estrutura de covariância através de uma combinação linear de matrizes conhecidas. Com intuito de ilustrar os conceitos apresentados, foram analisados diferentes conjuntos de dados. Para os mesmos foram ajustados modelos com as covariáveis no componente de média e/ou na estrutura de covariância, comparando-os segundo critérios de qualidade de ajuste global e qualidade preditiva. Os resultados obtidos mostraram que a modelagem da estrutura de covariância em função das covariáveis abre possibilidades para flexibilizar e ampliar opções para ajustes de modelos que melhor explorem a informação das covariáveis.

Palavras-chave: Covariância. Covariáveis. Modelos marginais..

ABSTRACT

Regression models usually model the relationship, between the response variable and the covariates, by an explicit mean function. The covariance structure is, generally, specified under independence of observations or express the dependence among observations induced by unknown variability sources but arising from data indexing, as example longitudinal, spatial and temporal dependence structures. The present work explores the modelling marginal covariance structure through specifications based upon the space spanned by covariates. Conditional predictions allow to obtain non linear curves that describe the behavior of the variable response. Different specifications of the covariance structure implies different features of the curve produced by conditional predictions, as smoothness and continuity. Such approach allows to explore data patterns not taken in account by mean function model specification and replace or complement temporal or spatial dependence structures. It was considered in three distinct types of distance functions that define the covariance structures: continuous functions, discrete functions and autoregressive functions. Multivariate covariance generalized linear models (MCGLM) was considered as an inferential tool for, among several features, allows the modelling of the covariance structure through a linear combination of matrices. In order to illustrate the presented concepts, different datasets were analyzed. For the same, were fitted models with the covariates in the mean structure and/or in the covariance structure, comparing them according to goodness of fit and prediction quality criteria. The results show that modelling the covariance structure as a function of covariates open up a set of possibilities to flexibilize and to expand options for fitting models which better explore information from covariates.

Keywords: Covariance. Covariates. Marginal models..

LISTA DE FIGURAS

2.1	Relação entre variável resposta Y e a covariável X	17
2.2	Ajuste de Y por uma relação linear com X	18
2.3	Ajustes da relação de Y com X por polinômios.	19
2.4	Ajustes via <i>kernel</i> função <i>Matérn</i> com $\kappa = 1,5$ e três valores distintos de ϕ	22
2.5	Ajuste da relação entre Y e X via <i>kernel</i> contínuo e <i>kernel</i> discreto.	23
2.6	Alocação das observações em dois <i>clusters</i> via algoritmo k-médias.	24
2.7	Ajustes com <i>kernels</i> via K-médias.	25
2.8	Ajuste via combinação <i>kernel</i> gaussiano e <i>kernel</i> por k-médias.	26
2.9	Divisão político administrativa dos 399 municípios do Paraná.	27
2.10	<i>Kernel</i> autorregressivo.	28
2.11	Dispersão entre as variáveis X_1 e X_2 em (a), e diagrama de Voronoi aplicado nas variáveis X_1 e X_2 em (b).	28
2.12	Ajustes produzidos pelo <i>Kernel</i> autorregressivo com vizinhança de primeira ordem em (a), e ajustes produzidos pelos <i>kernels</i> autorregressivos conforme a ordem de vizinhança em (b).	29
2.13	Dados longitudinais com 9 indivíduos representados no eixo Y avaliados em 20 observações indexadas em X em (a), e séries individuais para os 9 indivíduos com as 20 observações indexadas em X em (b)	30
2.14	<i>Kernel</i> exponencial para as observações e sem X no preditor linear de média em (a), e <i>kernel</i> exponencial para as observações e com X no preditor linear de média em (b).	31
2.15	<i>Kernel</i> discreto via k-médias para as observações e sem X no preditor linear de média em (a), e <i>Kernel</i> discreto via k-médias para as observações e com X no preditor linear de média em (b).	32
2.16	<i>Kernel</i> gaussiano para os indivíduos e sem X no preditor linear de média em (a), e <i>kernel</i> gaussiano para indivíduos e com X no preditor linear de média em (b).	33
2.17	<i>Kernel</i> discreto via k-médias para os indivíduos e sem x no preditor linear de média em (a), e <i>Kernel</i> discreto via k-médias para indivíduos e com X no preditor linear de média em (b).	33
3.1	Distribuição da variável resposta (a), e dispersão com a covariável (b).	40
3.2	Modelos ajustados por especificação de função de média: Ajuste com modelo linear normal em (a), com modelo linear normal com expansão polinomial de ordem 5 na covariável em (b), com GAM com apenas a função suave em (c), com GAM com a combinação das funções linear e suave em (d).	41

3.3	Modelos ajustados por especificação de função de covariância: Ajuste com <i>kernel</i> gaussiano em (a), com <i>kernel</i> exponencial em (b), com <i>kernel</i> via k-médias em (c), com <i>kernel</i> autorregressivo de ordem 1 em (d).	42
3.4	Ajustes para função linear da covariável combinada com estrutura para covariância: Ajuste com <i>kernel</i> gaussiano em (a), com <i>kernel</i> exponencial em (b), com <i>kernel</i> via k-médias em (c), com <i>kernel</i> autorregressivo de ordem 1 em (d).. . . .	43
3.5	<i>Kernel</i> gaussiano com $\phi = 2,6$ + <i>kernel</i> via k-médias com $k = 11$	44
3.6	Distribuição da variável resposta.	46
3.7	Dispersão da variável resposta com as quatro primeiras covariáveis.	46
3.8	Dispersão da variável resposta com as três últimas covariáveis.	47
3.9	Distribuição da variável resposta na dispersão dos dois primeiros componentes.	49
3.10	Classificações por modelo logístico e análise de discriminante: Modelo de regressão logística em (a), análise discriminante linear de Fisher em (b); análise discriminante quadrático de Fisher em (c).	50
3.11	Classificações por SVM, GAM e árvore de classificação: SVM com base radial em (a), GAM em (b), árvore de classificação em (c)..	51
3.12	Classificações por modelos logísticos com diferentes estruturas de dependência e com covariáveis no componente de média: Ajuste com <i>kernel</i> exponencial em (a), ajuste com <i>kernel</i> gaussiano em (b), ajuste com <i>kernel</i> k-médias em (c).	52
3.13	Classificações por modelos logísticos com diferentes estruturas de dependência e com média constante: Ajuste com <i>kernel</i> exponencial em (a), ajuste com <i>kernel</i> gaussiano em (b), ajuste com <i>kernel</i> k-médias em (c).	53
3.14	Gastos em (log)dólares com policiamento por habitante no estado do Mississipi.	57
3.15	Dispersão da variável resposta com as cinco primeiras covariáveis.	57
3.16	Dispersão da variável resposta com as cinco últimas covariáveis.	58
3.17	Mapa com legenda de cores para os valores ajustados pelos melhores modelos, com as covariáveis no componente de média e estrutura de dependência, segundo o pAIC: Valores observados em (a), modelo com <i>kernel</i> autorregressiva no espaço geográfico em (b), modelo com <i>kernel</i> autorregressiva no espaço das covariáveis em (c), modelo com <i>kernel</i> autorregressiva no espaço geográfico+ <i>kernel</i> autorregressivo no espaço das covariáveis em (d).	59
3.18	Taxas transformadas de prevalência de leucemia no estado de Nova York no período de 1978-1982.	61
3.19	Mapa com legenda de cores para os valores ajustados pelos melhores modelos, com as covariáveis no componente de média e estrutura de dependência, segundo o pAIC: Valores observados em (a), modelo com <i>kernel</i> autorregressiva no espaço geográfico em (b), modelo com <i>kernel</i> autorregressiva no espaço das covariáveis em (c), modelo com <i>kernel</i> autorregressiva no espaço geográfico+ <i>kernel</i> autorregressiva no espaço das covariáveis em (d).	63
3.20	Taxas de mortalidade infantil no período de 2008-2012 nos municípios do estado do Paraná.	64

3.21	Mapa com legenda de cores para os valores ajustados pelos melhores modelos, com as covariáveis no componente de média e estrutura de dependência, segundo o pAIC: Valores observados em (a), modelo com <i>kernel</i> exponencial das distâncias geográficas em (b), modelo com <i>kernel</i> K-médias no espaço das covariáveis em (c), modelo com <i>kernel</i> exponencial das distâncias geográficas+ <i>kernel</i> K-médias no espaço das covariáveis em (d)..	66
------	---	----

LISTA DE TABELAS

2.1	Disposição da variável y_{ij} conforme o indivíduo ID e a observação OBS	30
3.1	Tempo de processamento (TP) e soma dos erros quadrados (EQ) conforme os modelos ajustados.	45
3.2	Medidas de qualidade de ajuste dos modelos ajustados para a classe dos MCGLMs..	45
3.3	Tempo de processamento (TP) em segundos e erro quadrático médio (EQM) das predições conforme os modelos ajustados.	48
3.4	Relação dos dados observados Y com os preditos Y^*	51
3.5	Proporção total de acertos nos dados de teste (PTA), área sob a curva ROC (AUC) e Tempo de processamento (TP) em segundos..	54
3.6	Proporção total de acertos nos dados de teste (PTA), área sob a curva ROC (AUC) e tempo de processamento (TP) para cada um dos modelos ajustados.	54
3.7	Modelos ajustados ordenados pelo pAIC.	58
3.8	Estimativas dos parâmetros τ_s para os modelos com função de ligação $h()$ identidade..	59
3.9	Modelos ajustados ordenados pelo pAIC.	61
3.10	Estimativas dos parâmetros τ_s para os modelos com função de ligação $h()$ identidade..	62
3.11	Modelos ajustados ordenados pelo pAIC.	65
3.12	Estimativas dos parâmetros τ_s para os modelos com função de ligação $h()$ identidade..	65

LISTA DE ACRÔNIMOS

CPU	Da sigla em inglês, unidade de processamento central
GAM	Da sigla em inglês, modelo aditivo generalizado
GLM	Da sigla em inglês, modelo linear generalizado
LDA	Da sigla em inglês, análise discriminante linear
MCGLM	Da sigla em inglês, modelo linear generalizado de covariância linear
NLM	Da sigla em inglês, modelo linear normal
QDA	Da sigla em inglês, análise discriminante quadrático
SVM	Da sigla em inglês, máquina de vetores de suporte

SUMÁRIO

1	INTRODUÇÃO	13
2	MODELAGEM	16
2.1	MODELOS LINEARES NORMAIS	16
2.2	PROCESSOS GAUSSIANOS	18
2.2.1	Predições	20
2.3	ESTRUTURAS DE DEPENDÊNCIA	21
2.3.1	<i>Kernel</i> contínuo	21
2.3.2	<i>Kernel</i> discreto	22
2.3.3	<i>Kernel</i> autorregressivo	26
2.4	DADOS LONGITUDINAIS	29
2.5	MODELOS LINEARES GENERALIZADOS	33
2.6	MCGLM	34
2.6.1	Especificação de $\Omega(\boldsymbol{\tau}, \mathbf{Z})$	34
2.6.2	Especificação de $v(\boldsymbol{\mu}_Y; p)$	35
2.6.3	Inferência	35
2.6.4	Recurso computacional	36
2.6.5	Critério de Informação Escore	36
2.6.6	Medidas de qualidade de ajuste	37
2.6.7	Especificação do parâmetro do <i>kernel</i>	38
3	ANÁLISE DE DADOS	39
3.1	QUALIDADE DO AR DE NOVA YORK	39
3.2	PERFORMANCE RELATIVA DE CPUS	45
3.3	RECONHECIMENTO DE CÉDULAS FALSAS	48
3.4	TRATAMENTO DE REABILITAÇÃO DA FALA	52
3.5	GASTOS POLICIAIS NO ESTADO DO MISSISSIPI	55
3.6	CASOS DE LEUCEMIA NO ESTADO DE NOVA YORK	56
3.7	TAXAS DE MORTALIDADE INFANTIL NOS MUNICÍPIOS DO ESTADO DO PARANÁ	60
4	CONSIDERAÇÕES FINAIS	67
	REFERÊNCIAS	68

1 INTRODUÇÃO

Modelos de regressão expressam a variabilidade de uma variável aleatória, dita variável resposta, relacionando a variação de seus valores aos medidos nas covariáveis. Para variáveis respostas unidimensionais, as duas classes de modelos há mais tempo estabelecidas e largamente utilizadas são, os modelos lineares normais (NLM, da sigla em inglês), conhecido também por modelo de regressão linear, e a dos modelos lineares generalizados (GLM, da sigla em inglês). Os NLM modelam a relação entre a variável resposta e as covariáveis, por meio de uma função linear paramétrica. Além disso, assumem que a variável resposta segue uma distribuição normal com variância constante.

Modelos para variáveis com respostas não normais foram desenvolvidos posteriormente e vários deles foram unificados quando Nelder e Wedderburn (1972) propuseram a classe dos GLM, que estendem os NLM permitindo modelar a variável resposta com distribuições pertencentes à família exponencial de dispersão.

A partir das estruturas do NLM e do GLM, várias extensões foram propostas com objetivo de obter ajustes mais flexíveis para os dados. Dentre as quais, pode-se mencionar os modelos não-lineares, regressões suavizadas por *splines*, modelos aditivos generalizados (GAM, da sigla em inglês) (Hastie e Tibshirani, 1986), árvores de regressão e classificação (Quinlan, 1986) e *support vector machine* (SVM) (Evgeniou e Pontil, 2001).

Os modelos não lineares modelam a média da variável resposta através de funções paramétricas não lineares. Porém, essas funções devem ser especificadas pelo usuário conforme características das relações entre variável resposta e covariáveis levando em conta a interpretação desejada. No caso de múltiplas covariáveis, a especificação dessas funções pode ser complexa.

Os *splines* e GAM são modelos flexíveis que modelam a esperança da variável resposta através de um componente paramétrico e outro não paramétrico. O componente não paramétrico é formado por funções suaves, trazendo maior flexibilidade ao ajuste. Porém, estas funções podem não apresentar boa performance para o caso de múltiplas covariáveis.

O SVM trabalha com a expansão das covariáveis em uma série de funções, conhecidas como funções de base. É semelhante aos *splines* e aos GAM por buscar modelar a esperança da variável resposta através de funções suaves, porém, diferentemente destes, o SVM apresenta bom desempenho no caso de regressão com múltiplas covariáveis. Todavia, por tratar-se de um método matemático, na qual maximiza uma função objetivo conforme um conjunto de restrições, não há interpretabilidade dos parâmetros do modelo e também não há informações acerca da decomposição da covariância nas estruturas de média e de covariância.

Os modelos citados anteriormente buscam flexibilizar o ajuste da esperança da variável resposta em função das covariáveis, seja através da especificação de uma função paramétrica não linear nos parâmetros ou incorporando funções suaves. Porém, todos estes modelos possuem em comum o fato de especificar de forma explícita uma função para a média da variável resposta.

Uma outra forma de generalizar os modelos é pela inclusão de estruturas de dependência entre as observações. Tais estruturas podem implicar em trajetórias suavizadas para a variável resposta, condicionada aos valores observados, ao longo de alguma dimensão que indexe a estrutura de dependência. Um exemplo são os modelos espaciais onde a indexação se dá em um

conjunto de coordenadas geográficas. Em Lindgren et al. (2015), Wall (2004) e Rodrigues e Assunção (2011) encontram-se propostas e discussões de diferentes estruturas de dependência voltadas a dados espaciais. Séries temporais ou longitudinais também podem ser modeladas de tal forma. Em Pan e Pan (2017), Bates et al. (2015) e Halekoh et al. (2006) podem ser encontrados detalhamentos técnicos acerca das estruturas de covariância no contexto de dados longitudinais. De forma mais geral qualquer variável ordenada permite tal indexação e desta maneira pode-se construir a modelagem da dependência indexada a uma ou um conjunto de covariáveis.

Tome-se por exemplo dados espaciais de variação espacial contínua. Neste contexto o GAM e os modelos espaciais geoestatísticos ilustram bem a dualidade na estratégia para modelagem. Enquanto a primeira classe de modelos descreve trajetórias suaves (por *splines* ou análogos) da média como função das coordenadas, na segunda adota-se uma estrutura de covariância como função da distância geográfica levando a funções suavizadas de média na distribuição da variável resposta condicionada às observações. Em certos casos é possível estabelecer relações explícitas entre funções de covariância e *splines* (Dubrule, 1983, 1984).

Os modelos de efeitos aleatórios envolvem uma extensa de classes de modelos, mas de forma geral adicionam no componente de média um termo conhecido por variável latente ou efeito aleatório, sendo o mesmo geralmente descrito por um processo gaussiano. Um processo gaussiano define uma variável aleatória que segue uma distribuição normal com uma matriz de covariância de estrutura conhecida. A especificação dessa última é crucial na caracterização do processo gaussiano, uma vez que determina o comportamento da variável aleatória. Um exemplo clássico é quando a variável aleatória descreve um fenômeno espacial, quando então a covariância será especificada conforme a proximidade de observações no espaço geográfico. Esse mesmo conceito se aplica para descrever fenômenos temporais, efeitos de grupos, dentre outras possibilidades. Ou seja, é possível especificar de forma implícita, através da matriz de covariância, uma função de média condicionada às observações com certas características.

Rasmussen e Williams (2006) apresentam os processos gaussianos na perspectiva de uma classe de modelos de aprendizado de máquina, em que a matriz de covariância é especificada genericamente em função de distâncias no espaço de um conjunto de n covariáveis reais quaisquer. Dependendo de como estas covariáveis são incorporadas na covariância do processo gaussiano, é possível encontrar equivalências com alguns outros modelos como o SVM e os *splines*, como mostram os autores no Capítulo 6.

No presente trabalho explora-se a ideia de especificar modelos através de estruturas de covariância, que tanto podem reproduzir particular escolhas de estrutura de médias quanto incluir outras formas de relacionamento da resposta com as covariáveis. As estruturas de covariâncias se baseiam nas utilizadas em modelos espaciais, o que garante a validade dos modelos. Porém são definidas em espaços gerados pelas covariáveis, em substituição ao espaço geográfico, sobre os quais se definem modelos espaciais. Desta forma, busca-se modelar a covariância por alguma forma de proximidade no espaço das covariáveis. O objetivo é ajustar a relação da variável resposta com as covariáveis por uma superfície que pode ser suave e especificada implicitamente pela covariância. Isso será aplicado, primeiramente, sob suposição de normalidade da variável resposta e, posteriormente, estendido para variáveis respostas com outras distribuições de probabilidades. A proximidade no espaço gerado pelas covariáveis é especificada por uma função das distâncias entre os pares de pontos destas covariáveis, sendo que esta função pode ser especificada de diversas formas. No presente trabalho considera-se funções: (i) contínuas; (ii) discretas; (iii) de janelas móveis. Todas estas serão abordadas em detalhes na seção 2.3.

O presente trabalho segue a linha de modelar estruturas de covariância a partir de covariáveis como em Gattiker et al. (2015), Choudhuri et al. (2007) e Kuss e Rasmussen

(2006), os quais modelam a relação da variável resposta com as covariáveis através de uma variável latente. Os modelos de efeitos aleatórios são largamente utilizados na literatura e para variáveis aleatórias não gaussianas são estimados, convenientemente, via inferência bayesiana. Poucas implementações que estimam estes modelos via verossimilhança, envolvem técnicas de simulações Monte Carlo. O modelagem da covariância da variável latente está limitada a poucas estruturas.

Entretanto, é possível especificar modelos marginais que definem a estrutura de covariância diretamente para a variável resposta, sem a introdução de variáveis latentes. Desta forma, optou-se no presente trabalho pelos modelos marginais por apresentarem um processo de estimação mais simples, utilizando algum tipo de quasi-verossimilhança, pressupondo apenas os dois primeiros momentos, o que compactua com o objetivo de uma modelagem mais flexível.

Especificamente, adotou-se a classe dos modelos lineares generalizados de covariância linear (MCGLM). Os mesmos são uma classe de modelos marginais proposta por Bonat e Jørgensen (2016), e estendem os GLM permitindo a modelagem direta da estrutura de covariância da variável resposta por meio de uma combinação linear de matrizes, a qual os autores chamam de preditor linear matricial. Através do preditor linear matricial serão feitas combinações entre as matrizes especificadas pelas diferentes funções de distâncias no espaço das covariáveis. Se o conjunto de dados possuir indexação espacial ou temporal, pode-se ainda combinar as matrizes de proximidades espaciais ou temporais com as matrizes de proximidades no espaço das covariáveis.

O trabalho está estruturado em quatro capítulos: 1 Introdução, 2 Modelagem 3 Análise de dados e 4 Considerações finais. O capítulo 2 começa revisando os modelos de regressão linear, na seção 2.1, processos gaussianos são brevemente revisados e na seção seguinte, estendendo os modelos de regressão linear. Da mesma forma, na seção 2.5 é feita uma breve revisão dos GLM e a extensão destes pelos MCLGM.

A especificação da covariância marginal e das proximidades no espaço de covariáveis serão abordadas em detalhes na seção 2.3 com uma série de exemplos ilustrativos. O cálculo das predições condicionais serão apresentados na seção 2.2.1.

No capítulo 3 são analisados sete conjuntos de dados de características diferentes, com o objetivo de ilustrar e exemplificar a análise dos mesmos com as diferentes especificações de modelos. O capítulo 4 oferece um resumo e discussão das principais contribuições e resultados do trabalho desenvolvido.

2 MODELAGEM

Seja um vetor $y = (y_1, y_2, \dots, y_n)$ de n observações de uma variável aleatória resposta Y descrita por alguma distribuição de probabilidades com vetor de médias μ_Y e matriz de covariância Σ_Y expressos por:

$$E(Y) = \mu_Y \quad (2.1)$$

$$V(Y) = \Sigma_Y \quad (2.2)$$

Considere também a matriz \mathbf{X} com colunas (X_1, X_2, \dots, X_r) com valores observados de covariáveis $X_1 = (x_{11}, x_{12}, \dots, x_{1n}), X_2 = (x_{21}, x_{22}, \dots, x_{2n}), \dots, X_r = (x_{r1}, x_{r2}, \dots, x_{rn})$. O objetivo é modelar os dois primeiros momentos 2.1 e 2.2 de Y por especificações de \mathbf{X} .

Em particular, a especificação da matriz de variâncias e covariâncias é o foco da contribuição deste trabalho. Inicia-se pela especificação de $E(Y)$ por classes de modelos mais difundidos na literatura. Na sequência discute-se a especificação da matriz de variâncias e covariâncias.

2.1 MODELOS LINEARES NORMAIS

Uma primeira forma de modelar a relação de Y com \mathbf{X} é através dos modelos lineares normais (NLMs).

Os NLMs pressupõem distribuição normal multivariada para Y :

$$Y \sim NM(\mu_Y, \Sigma_Y), \quad (2.3)$$

em que a média μ_Y da variável resposta Y é modelada através de um preditor linear que envolve uma matriz de planejamento \mathbf{X}^* com n linhas e p colunas e um vetor de parâmetros $\beta = (\beta_0, \beta_1, \dots, \beta_p)$, conforme segue:

$$\mu_Y = \mathbf{X}^* \beta, \quad (2.4)$$

enquanto, em sua forma mais simples, a matriz de variâncias e covariâncias Σ_Y é especificada sob independência das observações:

$$\Sigma_Y = \sigma^2 \mathbf{I}, \quad (2.5)$$

em que σ^2 é o parâmetro de variância e \mathbf{I} é a matriz identidade de ordem n .

Os NLMs também podem ser especificados na forma condicional:

$$Y|\mathbf{X} = \mu_Y + \epsilon \quad (2.6)$$

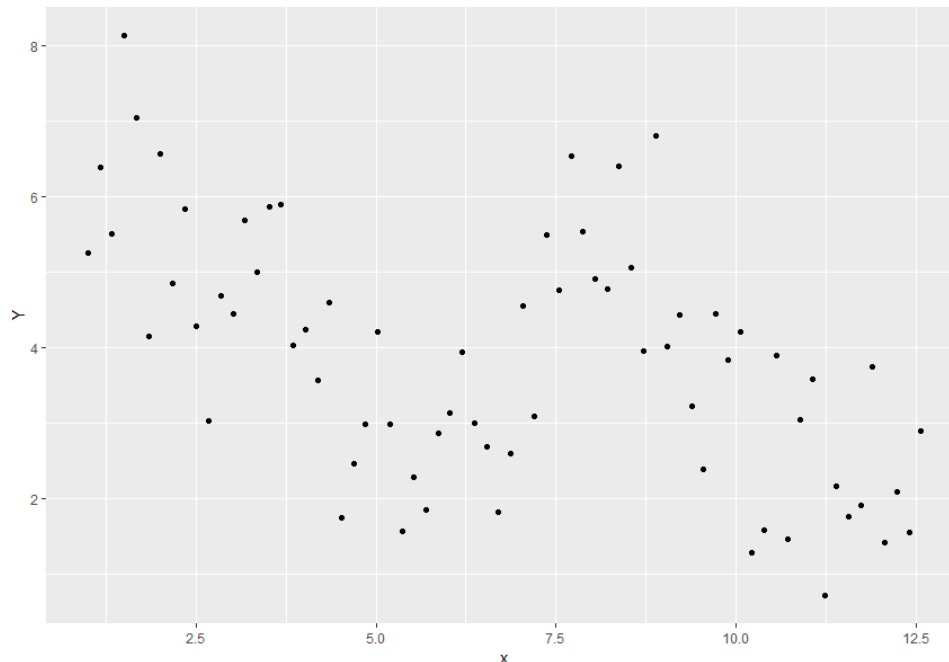
$$\epsilon \sim NM(\mathbf{0}, \Sigma_Y), \quad (2.7)$$

em que μ_Y e Σ_Y são especificados em 2.4 e em 2.5 respectivamente.

No presente trabalho, a primeira coluna da matriz de planejamento $\mathbf{X}^* = (\mathbf{1}_n, X_1, X_2, \dots, X_r)^T$ é um vetor de 1s e as demais colunas são as covariáveis X_1, X_2, \dots, X_r . Porém, uma vez que a linearidade dos NLMs está nos parâmetros $\beta_0, \beta_1, \dots, \beta_p$, as colunas da matriz \mathbf{X}^* , podem ainda incluir funções das covariáveis.

Tomando a Figura 2.1 como exemplo, a relação entre a variável resposta Y e a covariável x pode ser ajustada por um NLM com a matriz de planejamento especificada como $\mathbf{X}^* = (\mathbf{1}_n, X)$, o que resulta em um ajuste linear de Y com X , conforme a Figura 2.2. Mas é possível melhorar o ajuste expandindo a covariável X em uma sequência de funções de transformação (bases) $\phi_1(X), \phi_2(X), \dots, \phi_{p-1}(X)$, como por exemplo, séries polinomiais de ordem $p - 1$ $\mathbf{X}^* = (\mathbf{1}_n, X, X^2, \dots, X^{p-1})$ conforme a Figura 2.3. A ideia é de que a relação das novas covariáveis X, X^2, \dots, X^{p-1} com Y seja bem ajustada por um plano, de forma a obter um ajuste flexível para a relação de interesse entre Y e X .

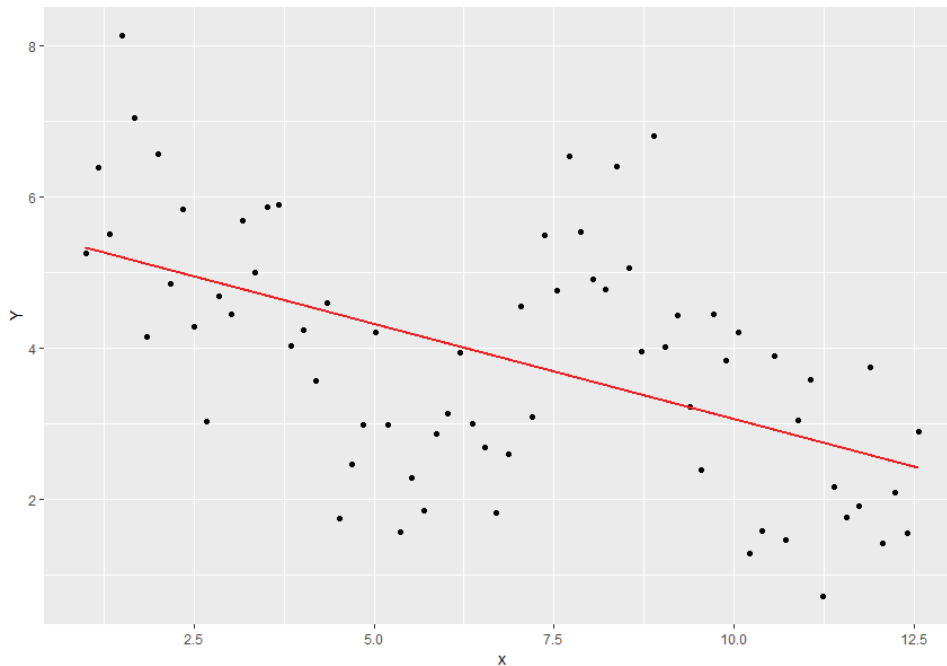
Figura 2.1: Relação entre variável resposta Y e a covariável X



Fonte: Elaborado pelo autor.

Modelar relações não lineares entre as covariáveis e a variável resposta expandindo as covariáveis em um conjunto de funções de transformação é uma estratégia válida, sobretudo quando a expansão é em uma única covariável e a relação não linear é definida por algum modelo relacionado à natureza do fenômeno envolvendo as variáveis. Entretanto, a medida que o número de covariáveis cresce fica mais difícil determinar funções de transformação para cada covariável e ainda ter que considerar possíveis interações entre as mesmas. Além disso, o número de parâmetros a serem estimados aumenta. Se a amostra tiver uma quantidade grande de covariáveis e as mesmas apresentarem relações não lineares com a variável resposta, então essa estratégia acaba sendo não muito atrativa. Por outro lado, os processos gaussianos nos quais a relação é definida através da estrutura de dependência, podem ser uma estratégia válida e flexível para modelar relações não lineares entre a variável resposta e múltiplas covariáveis.

¹ $\mathbf{1}_n = (1, 1, \dots, 1)^T$ é o vetor coluna de 1s de comprimento n .

Figura 2.2: Ajuste de Y por uma relação linear com X .

Fonte: Elaborado pelo autor.

2.2 PROCESSOS GAUSSIANOS

Um processo gaussiano (PG) $\{\gamma(W_i) : W_i \in \mathfrak{R}^k\}$, com $i = 1, \dots, n$, é uma coleção de variáveis aleatórias indexadas em pontos $(w_{11}, w_{12}, \dots, w_{1k}), (w_{21}, w_{22}, \dots, w_{2k}), \dots, (w_{n1}, w_{n2}, \dots, w_{nk})$ que são vetores linha de uma matriz de variáveis $\mathbf{W} = (W_1, W_2, \dots, W_k)$ de dimensão nk , cuja distribuição conjunta de $\gamma(\mathbf{W}) = \{\gamma(W_1), \gamma(W_2), \dots, \gamma(W_n)\}$ é uma normal multivariada com vetor de médias $\boldsymbol{\mu}_\gamma$ e matriz de variâncias e covariâncias $\boldsymbol{\Sigma}_\gamma$:

$$\gamma(\mathbf{W}) \sim NM(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma). \quad (2.8)$$

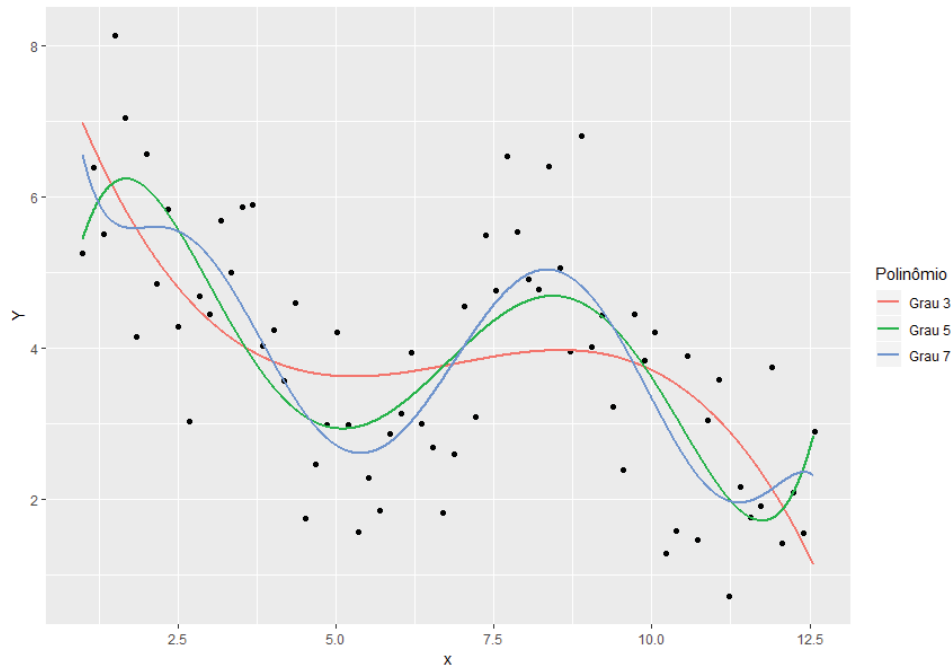
O termo $\gamma(\mathbf{W})$ é também conhecido por variável latente, descreve um fenômeno indexado no vetor linha W_i , com $i = 1, \dots, n$, em geral indexado pelo tempo ($W_i \in \mathfrak{R}$) ou espaço ($W_i \in \mathfrak{R}^2$). Porém, tratando de forma geral, \mathbf{W} pode ser qualquer conjunto de n variáveis reais (Rasmussen e Williams, 2006).

Seja a variável resposta Y , a qual se tem interesse em modelar a relação com um fenômeno através de $\gamma(\mathbf{W})$ indexado em \mathbf{W} . Incluindo $\gamma(\mathbf{W})$ no NLM especificado em 2.6 tem-se:

$$Y|\mathbf{X}, \gamma(\mathbf{W}) = \boldsymbol{\mu}_Y + \gamma(\mathbf{W}) + \boldsymbol{\epsilon} \quad (2.9)$$

$$\gamma(\mathbf{W}) \sim NM(\mathbf{0}, \boldsymbol{\Sigma}_\gamma), \quad (2.10)$$

em que $\boldsymbol{\mu}_Y$ é o preditor linear especificado em 2.4.

Figura 2.3: Ajustes da relação de Y com X por polinômios.

Fonte: Elaborado pelo autor.

Por propriedades da distribuição normal, o modelo expresso em sua forma condicional $Y|\mathbf{X}, \gamma(\mathbf{W})$, conforme 2.9 e 2.10, é equivalente a forma marginal de Y expressa por:

$$Y \sim NM(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y) \quad (2.11)$$

$$\boldsymbol{\mu}_Y = \mathbf{X}^* \boldsymbol{\beta} \quad (2.12)$$

$$\boldsymbol{\Sigma}_Y = \boldsymbol{\Sigma}_\gamma + \sigma^2 \mathbf{I}. \quad (2.13)$$

O termo $\gamma(\mathbf{X})$, seguindo 2.10, é uma função flexível, suave ou não suave, determinada de forma implícita pela especificação de $\boldsymbol{\Sigma}_\gamma$. Ou seja, o processo gaussiano fica caracterizado pela especificação da matriz de covariância $\boldsymbol{\Sigma}_\gamma$. A mesma deve ser especificada com base em algum critério de proximidade entre instantes de tempo, entre pares de coordenadas geográficas (indexação espacial), ou qualquer outro conjunto de variáveis $\mathbf{W} = (W_1, W_2, \dots, W_k)$ sobre as quais o processo está indexado. Se o processo estiver indexado nas covariáveis, isto é, $\mathbf{W} = \mathbf{X}$, então a matriz de covariância $\boldsymbol{\Sigma}_\gamma$ será especificada com base em proximidades entre vetores linha da matriz de covariáveis \mathbf{X} , e $\gamma(\mathbf{X})$ irá descrever a relação de Y com \mathbf{X} .

A matriz de covariância $\boldsymbol{\Sigma}_\gamma$ é o ponto central dos PGs. No presente trabalho, se propões que sua especificação dependa de funções de distâncias paramétricas entre pares de observações (vetores linha) de um conjunto de covariáveis \mathbf{X} , funções estas denominadas de *kernel* por Rasmussen e Williams (2006). Desta forma, adotou-se $k(X_i, X_j)$, com $i, j = 1, \dots, n$, para a expressão algébrica e \mathbf{K} na forma matricial, abordadas na sessão 2.3.

Importante destacar que, por simplificação, \mathbf{X} é usado como notação das variáveis para modelar tanto a média quanto as estruturas de dependência, porém, não utilizando-se necessariamente as mesmas covariáveis.

De forma análoga ao preditor linear de média, a relação em 2.13 pode ser vista como um preditor linear da covariância, em que $\Sigma_Y = \tau^2 \mathbf{K}$. Reescrevendo 2.13 obtém-se:

$$\Sigma_Y = \tau^2 \mathbf{K} + \sigma^2 \mathbf{I}, \quad (2.14)$$

com σ^2 e τ^2 sendo os parâmetros de variância.

As covariâncias podem ser interpretadas como proximidades entre pares de observações. Desta forma, a relação 2.14 pressupõe que as proximidades entre os pares de observações de Y podem ser escritas como combinações lineares das proximidades entre pares de observações de \mathbf{X} .

2.2.1 Predições

O núcleo do presente trabalho é a modelagem de relações não lineares entre a variável resposta Y e as covariáveis \mathbf{X} através da especificação da estrutura de dependência marginal Σ_Y em função das covariáveis X . Na seção 2.2 foram introduzidos os conceitos sobre a especificação da matriz Σ_Y e na presente seção abordadas as predições, responsáveis por fornecer ajustes flexíveis, suaves ou não, por levar em conta a covariância marginal de Y .

Denote por $[\cdot]$ a distribuição da quantidade entre os colchetes. Seja Y_0 o vetor de comprimento p de novos valores, os quais se tem interesse em predizer pela distribuição preditiva, que equivale à condicional $[Y_0|Y]$, e \mathbf{X}_0 a matriz com as r covariáveis (colunas) nos p pontos de predição (linhas). A matriz de planejamento \mathbf{X}_0^* nos pontos de predição é encontrada a partir da matriz \mathbf{X}_0 .

Partindo dos dois primeiros momentos da distribuição conjunta $[Y, Y_0]$ tem-se os mesmos expressos por:

$$E \begin{pmatrix} Y \\ Y_0 \end{pmatrix} = \begin{pmatrix} \mathbf{X}^* \boldsymbol{\beta} \\ \mathbf{X}_0^* \boldsymbol{\beta} \end{pmatrix} \quad (2.15)$$

$$V \begin{pmatrix} Y \\ Y_0 \end{pmatrix} = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YY_0} \\ \Sigma_{Y_0Y} & \Sigma_{Y_0Y_0} \end{pmatrix}. \quad (2.16)$$

Seja a matriz \mathbf{X}_0 de valores fixos das covariáveis \mathbf{X} em p pontos de predição. A partir da especificação conjunta de $(\mathbf{X}, \mathbf{X}_0)$ especifica-se 2.15 e 2.16.

Substituindo $\boldsymbol{\beta}$, em 2.15, por sua estimativa $\hat{\boldsymbol{\beta}}$ tem-se:

$$\hat{E} \begin{pmatrix} Y \\ Y_0 \end{pmatrix} = \begin{pmatrix} \mathbf{X}^* \hat{\boldsymbol{\beta}} \\ \mathbf{X}_0^* \hat{\boldsymbol{\beta}} \end{pmatrix}. \quad (2.17)$$

Da mesma forma é feito para a covariância, substituindo os parâmetros τ^2 , σ^2 e dos parâmetros da função *kernel* $k(\cdot)$ por suas respectivas estimativas:

$$\begin{pmatrix} \hat{\Sigma}_{YY} & \hat{\Sigma}_{YY_0} \\ \hat{\Sigma}_{Y_0Y} & \hat{\Sigma}_{Y_0Y_0} \end{pmatrix} = \hat{\tau}^2 \hat{K}_{\mathbf{X}, \mathbf{X}_0} + \hat{\sigma}^2 \mathbf{I}_{(n+p)}, \quad (2.18)$$

em que $\hat{K}_{\mathbf{X}, \mathbf{X}_0}$ denota a matriz da função *kernel*, com os parâmetros estimados pelo modelo, aplicada na conjunta $(\mathbf{X}, \mathbf{X}_0)$.

Segundo Gotway e Stroup (1997), a partir dos dois primeiros momentos da conjunta $[Y, Y_0]$, expressos em 2.17 e 2.18, é possível encontrar a esperança e a matriz de covariância da distribuição condicional $[Y_0|Y]$:

$$\hat{E}(Y_0|Y) = \mathbf{X}_0^* \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Sigma}}_{Y_0 Y} \hat{\boldsymbol{\Sigma}}_{Y Y}^{-1} (Y - \mathbf{X}^* \hat{\boldsymbol{\beta}}) \quad (2.19)$$

$$\hat{\boldsymbol{\Sigma}}_{Y_0|Y} = \hat{\boldsymbol{\Sigma}}_{Y_0 Y_0} - \boldsymbol{\Sigma}_{Y_0 Y} \hat{\boldsymbol{\Sigma}}_{Y Y}^{-1} \hat{\boldsymbol{\Sigma}}_{Y Y_0}. \quad (2.20)$$

As predições condicionais são um dos pontos chaves dos PGs, em que a covariância marginal de Y pode ser vista como uma correção dos desvios dos dados em relação ao ajuste produzido pelo modelo linear normal, conforme 2.19.

2.3 ESTRUTURAS DE DEPENDÊNCIA

O *kernel* \mathbf{K} é uma matriz ou estrutura de dependência entre as observações determinada através de alguma função de distâncias, em geral paramétrica, entre os vetores linha da matriz \mathbf{X} . A escolha dessa função de distâncias implica diretamente nas características da curva ajustada através da média condicional em 2.19. Neste trabalho optou-se por trabalhar com estruturas de dependência de três tipos diferentes: estruturas contínuas, discretas e autorregressivas. Priorizou-se estruturas de dependência bem conhecidas na literatura, por exemplo de modelos para dados longitudinais, temporais e espaciais, o que garante a validade dos modelos aqui adotados.

2.3.1 *Kernel* contínuo

A função *Matèrn* representa uma classe de funções de correlação biparamétricas originalmente aplicadas no contexto geoestatístico (Diggle e Ribeiro Jr., 2007). Rasmussen e Williams (2006) apresentam as mesmas no contexto de *machine learning* como especificação de *kernel*:

$$k(X_i, X_j) = \{2^{\kappa-1} \Gamma(\kappa)\}^{-1} (\|X_i - X_j\| / \phi)^\kappa B_\kappa(\|X_i - X_j\| / \phi), \quad (2.21)$$

em que B_κ denota a função Bessel modificada de ordem κ (Diggle e Ribeiro Jr., 2007), κ e ϕ são parâmetros e $\|X_i - X_j\|$ denota a distância euclidiana entre os vetores X_i e X_j k -dimensionais.

Na família *Matèrn* (em 2.21) o parâmetro $\phi > 0$ é interpretado como a taxa de decaimento da correlação entre pares de observações em função da distância entre elas. Quanto menor o valor de ϕ menor é a dependência para uma dada distância. A função obtida pelo *kernel* irá capturar padrões locais, conforme representado na Figura 2.4.

O parâmetro κ é o grau de diferenciabilidade da função gerada pelo *kernel* em 2.21, ou seja, o grau de suavidade da função. Pode-se destacar dois casos em particular da família *Matèrn*. Para $\kappa = 0,5$ obtém-se o *kernel* exponencial e para $\kappa \rightarrow \infty$, o *kernel* gaussiano, respecificados como:

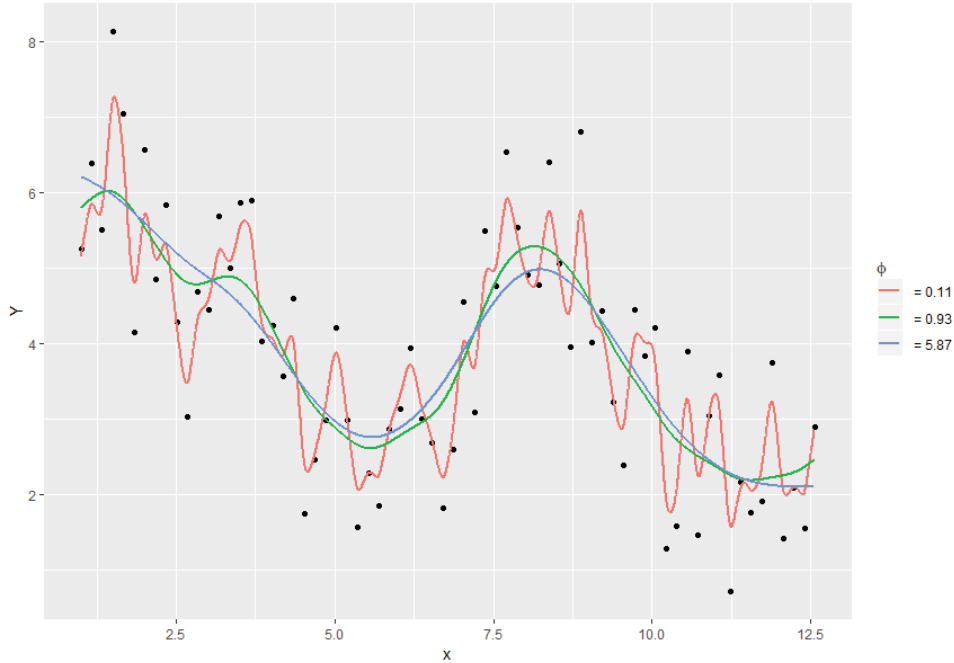
$$k_{exp}(X_i, X_j) = \exp\{-\|X_i - X_j\| / \phi\} \quad (2.22)$$

$$k_{gauss}(X_i, X_j) = \exp\{-(\|X_i - X_j\| / \phi)^2\}. \quad (2.23)$$

O *kernel* exponencial e o *kernel* gaussiano, em 2.22 e 2.23 respectivamente, representam cenários extremos da família *Matèrn*, na qual a primeira é contínua mas não diferenciável e a segunda infinitamente diferenciável, representando o caso mais suave. Na prática, para valores de $\kappa > 2,5$ não há ganhos visualmente perceptíveis em termos de suavidade.

Para fins práticos fixou-se: (i) $\kappa = 0,5$; (ii) $\kappa = 1,5$ e (iii) $\kappa = 2,5$, onde (i) e (iii) representam pouca e muita suavidade respectivamente, e (ii) está no intermediário, conforme a Figura 2.4.

Figura 2.4: Ajustes via *kernel* função *Matérn* com $\kappa = 1,5$ e três valores distintos de ϕ .



Fonte: Elaborado pelo autor.

Importante salientar que ϕ não é ortogonal a κ , sendo a comparação direta, entre valores de ϕ com diferentes valores para κ , não válida. Na seção 2.6.7 foi discutido melhor este aspecto na estimação dos parâmetros da covariância.

2.3.2 Kernel discreto

Também é possível especificar uma estrutura de dependência cuja função gerada (através da predição condicional) tenha característica discreta.

Seja uma variável do tipo fator representada pelo vetor $F = (f_1, f_2, \dots, f_b)$, onde f_1, f_2, \dots, f_b são os fatores. O *kernel* discreto é definido pressupondo independência entre os fatores, ou seja:

$$\text{Corr}(f_1, f_2, \dots, f_b) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \quad (2.24)$$

e uma estrutura do tipo permutável entre as observações de um mesmo fator f_i (para $i = 1, \dots, b$):

$$\text{Corr}(f_i) = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} = \mathbf{1}_t \mathbf{1}_t^T, \quad (2.25)$$

onde $\mathbf{1}_t$ denota um vetor de 1s de comprimento t , e t é a quantidade de observações no fator f_i .

Com 2.24 e 2.25, a estrutura ou *kernel* discreto (\mathbf{K}_{perm}) fica expresso por uma matriz bloco diagonal:

$$\mathbf{K}_{perm} = Corr(f_1, f_2, \dots, f_b) \otimes Corr(f_i) = \begin{pmatrix} \mathbf{1}_t \mathbf{1}_t^T & 0 & \dots & 0 \\ 0 & \mathbf{1}_t \mathbf{1}_t^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{1}_t \mathbf{1}_t^T \end{pmatrix}, \quad (2.26)$$

onde \otimes denota o produto de Kronecker.

A mesma matriz em 2.26 pode ser expressa como:

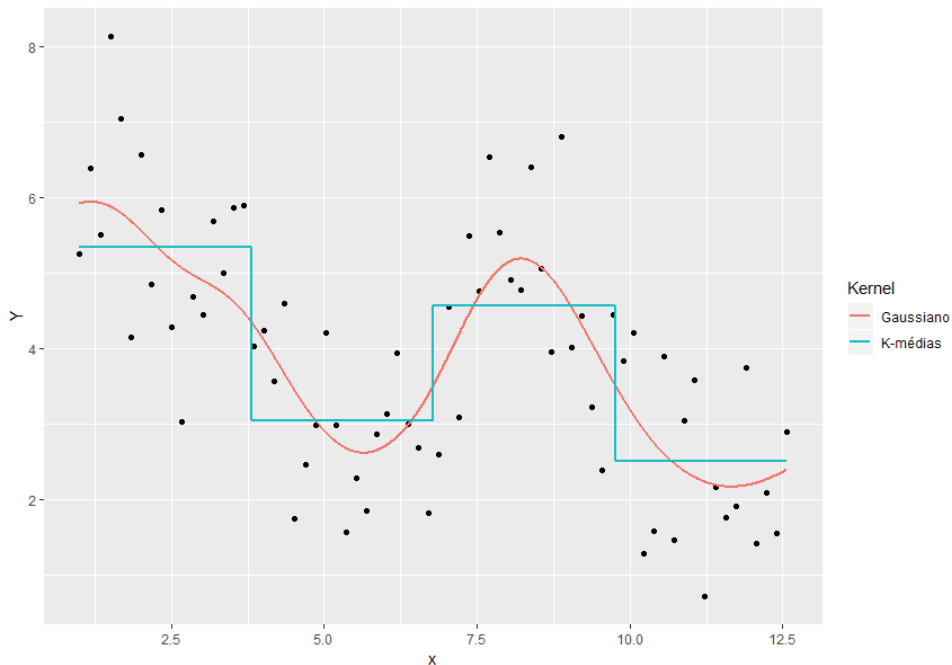
$$k_{perm}(F_i, F_j) = \begin{cases} 1 & \text{se } i \text{ e } j \text{ são observações do mesmo fator} \\ 0 & \text{caso contrário} \end{cases}, \quad (2.27)$$

com $i, j = 1, \dots, n$.

A estrutura discreta, conforme especificada em 2.26 e em 2.27, produz um ajuste em níveis, conforme representado pela função em azul na Figura 2.5.

Se $\mathbf{X} = (X_1, X_2, \dots, X_r)$ é um conjunto de covariáveis X_1, X_2, \dots, X_r contínuas, então é necessário substituir a matriz \mathbf{X} com n linhas e r colunas por um único vetor de comprimento n $F = (f_1, f_2, \dots, f_b)$ com fatores f_1, f_2, \dots, f_b , uma vez que a especificação da estrutura de dependência discreta parte de uma variável do tipo fator.

Figura 2.5: Ajuste da relação entre Y e X via *kernel* contínuo e *kernel* discreto.



Fonte: Elaborado pelo autor.

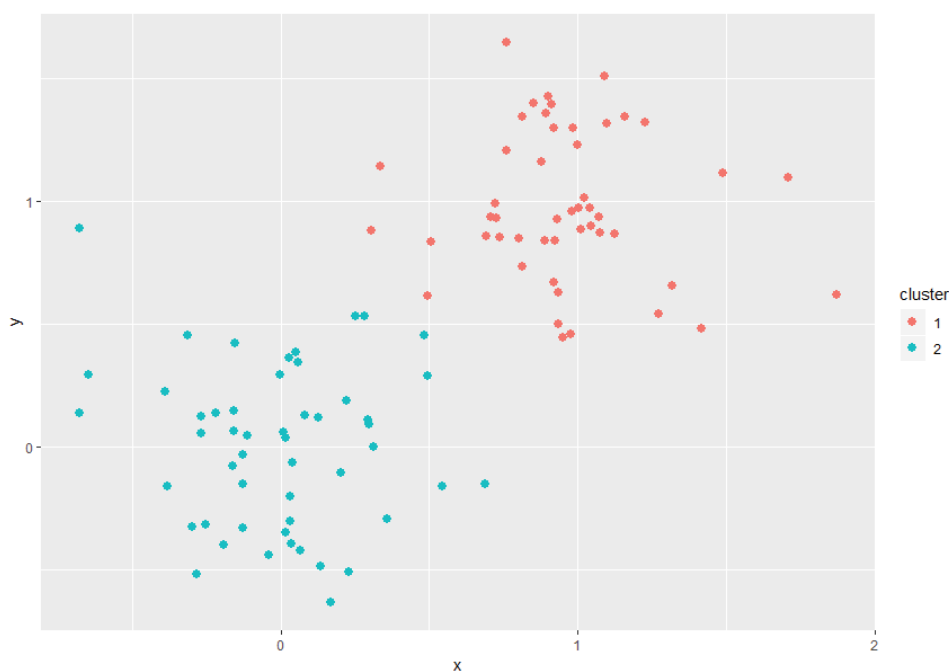
A substituição de um conjunto de covariáveis contínuas \mathbf{X} por uma única variável fatorial F deve ser feita de forma a encontrar F que melhor represente \mathbf{X} . Isso é feito pensando na melhor alocação dos n vetores linhas da matriz \mathbf{X} (observações) nos fatores f_1, f_2, \dots, f_b , em

que o número de fatores b é previamente fixado. Para este propósito, utilizou-se algoritmos de *cluster*.

Os algoritmos de *cluster* são métodos de aprendizagem não supervisionada que aloca as observações da amostra em um número b , previamente especificado, de *clusters* guiado por medidas de dissimilaridade das variáveis \mathbf{X} fornecidas, obtendo observações mais similares entre si alocadas a um mesmo *cluster* (Johnson e Wichern, 2007).

A Figura 2.6 ilustra o resultado de um algoritmo de *cluster* no caso bivariado. Os pontos com a mesma coloração representam observações em um mesmo *cluster*.

Figura 2.6: Alocação das observações em dois *clusters* via algoritmo k-médias.



Fonte: Elaborado pelo autor.

De forma geral, é possível dividir os algoritmos de *cluster* em dois tipos, algoritmos hierárquicos e algoritmos não hierárquicos (Mingoti, 2013). Por conveniência, optou-se trabalhar com os algoritmos não hierárquicos, por permitirem a alocação de novas observações e, facilitar o cálculo das previsões com o modelo ajustado via *kernel* discreto. Em específico, utilizou-se o algoritmo de k-médias (Johnson e Wichern, 2007).

O algoritmo de k-médias aloca as n observações da amostra em $b < n$ *clusters* através de realocações sucessivas das observações nos mesmos. A partir das distâncias euclidianas entre os centroides e os vetores linha da matriz \mathbf{X} são feitas as decisões das alocações para cada observação nos respectivos *clusters*, em seguida são recalculados os centroides com as novas realocações, e assim o processo é feito até não haver mais possibilidade de realocações.

Ao entrar com a matriz \mathbf{X} no algoritmo de k-médias, o mesmo retorna um vetor de comprimento n com o *cluster* atribuído à cada uma das n observações. Esse vetor é a variável fatorial F onde os fatores f_1, f_2, \dots, f_b são os *clusters*.

O algoritmo de k-médias necessita de duas informações prévias: o número b de *clusters* e os respectivos centroides iniciais. As especificações dos centroides iniciais foram feitas de forma aleatória (Mingoti, 2013), já b foi considerado como parâmetro e sua estimação foi tratada na seção 2.6.7.

A especificação do *kernel* conforme 2.26 assume um mesmo parâmetro de dispersão τ^2 para todos os *clusters*, ou seja:

$$\Sigma_Y = \tau^2 \mathbf{K}_{perm}, \quad (2.28)$$

mas é possível considerar diferentes τ_b^2 para cada *cluster* (Figura 2.7):

$$\Sigma_Y = \tau_1^2 \begin{pmatrix} \mathbf{1}_{t_1} \mathbf{1}_{t_1}^T & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} + \tau_2^2 \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & \mathbf{1}_{t_2} \mathbf{1}_{t_2}^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix} + \dots + \tau_b^2 \begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{1}_{t_b} \mathbf{1}_{t_b}^T \end{pmatrix}, \quad (2.29)$$

em que $\mathbf{1}_{t_1}, \mathbf{1}_{t_2}, \dots, \mathbf{1}_{t_b}$ são vetores de 1s de comprimentos t_1, t_2, \dots, t_b respectivamente.

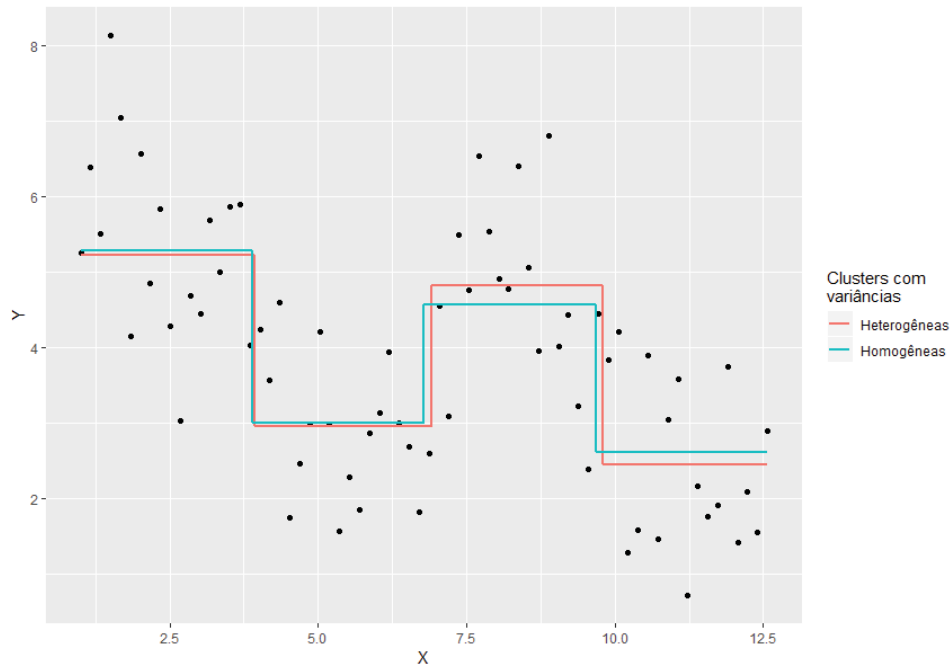
Porém, é importante ressaltar que considerar um parâmetro de variância $\tau_1^2, \tau_2^2, \dots, \tau_b^2$ por bloco pode ser custoso, principalmente quando o número de blocos é elevado, uma vez que consome mais graus de liberdade.

Além disso, é possível combinar um dos *kernels* contínuos \mathbf{K}_{exp} ou \mathbf{K}_{gauss} e outro discreto \mathbf{K}_{perm} , conforme a Figura 2.8, especificando:

$$\Sigma_Y = \tau_1^2 \mathbf{K}_{exp} + \tau_2^2 \mathbf{K}_{perm}. \quad (2.30)$$

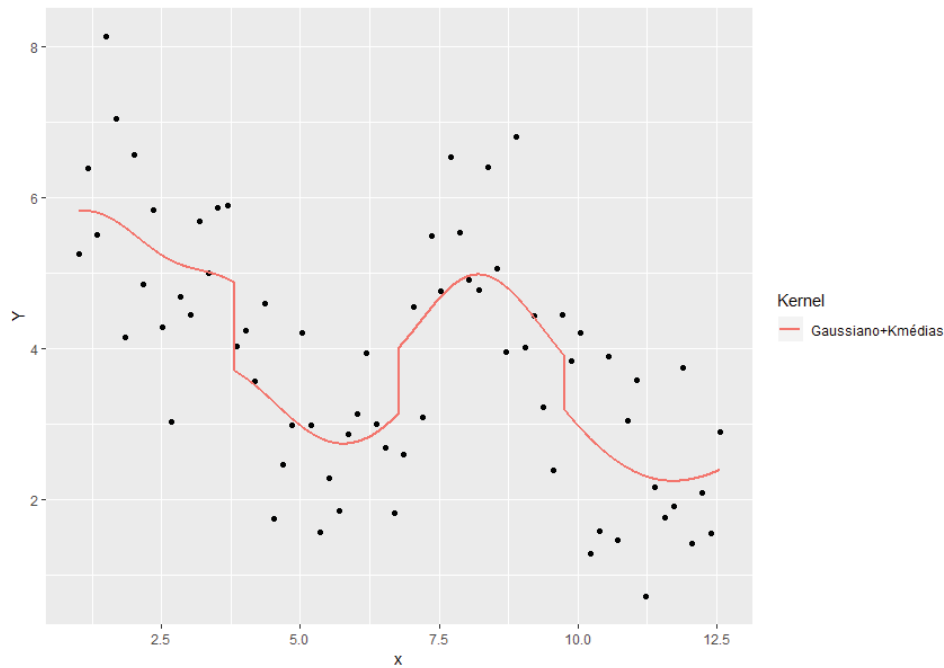
Na especificação da covariância marginal de Y em 2.30 foi feita a combinação de uma estrutura discreta \mathbf{K}_{perm} e uma estrutura contínua exponencial \mathbf{K}_{exp} , mas poderia ser uma estrutura contínua gaussiana também \mathbf{K}_{gauss} .

Figura 2.7: Ajustes com *kernels* via K-médias.



Fonte: Elaborado pelo autor.

Figura 2.8: Ajuste via combinação *kernel* gaussiano e *kernel* por k-médias.



Fonte: Elaborado pelo autor.

2.3.3 *Kernel* autorregressivo

O terceiro tipo de estrutura é a autorregressiva, bastante recorrente em séries temporais e em dados por área.

Dados de área são um processo espacial discreto $\{Y(A_i) : A_i \in D\}$, em que o domínio geográfico $D \subset \mathcal{R}^2$ é fragmentado em A_1, A_2, \dots, A_n áreas disjuntas $\{A_i \cap A_j = \emptyset, \forall i \neq j\}$, como o exemplo da Figura 2.9.

Comumente, dados de área são modelados por um processo autorregressivo condicional (CAR) expresso por:

$$Y(A_i) | Y(A_{-i}) \sim N\left(\mu_i + \sum_{j=1}^n \rho C_{ij} (Y(A_j) - \mu_j), \tau_i^2\right), \quad (2.31)$$

em que $A_{-i} = \{A_j | i \neq j\}$ e $\rho \in (-1, 1)$ é o parâmetro de dependência espacial.

A matriz \mathbf{C} é conhecida como matriz de vizinhança. A mesma estabelece conexões entre as áreas do espaço geográfico (vizinhança) da seguinte forma:

$$C_{ij} = \begin{cases} 1 & \text{se } A_i \sim A_j \\ 0 & \text{caso contrário} \end{cases}, \quad (2.32)$$

em que $A_i \sim A_j$ implica A_i vizinho de A_j .

Considerando como critério de conexão áreas que compartilham da mesma divisão geográfica, conhecida também como vizinhança de áreas adjacentes, 2.31 recai na propriedade markoviana de independência condicional. Ou seja, a distribuição de $Y(A_i)$ condicionado a todas as outras variáveis $Y(A_{-i})$ dependerá apenas das variáveis cujas áreas são geograficamente mais próximas, isto é, as áreas vizinhas especificadas por 2.32.

Figura 2.9: Divisão político administrativa dos 399 municípios do Paraná.



Fonte: Elaborado pelo autor.

Marginalmente, a distribuição de $Y(A)$ fica expressa por:

$$Y(A) \sim NM(\boldsymbol{\mu}_Y, \mathbf{Q}^{-1}), \quad (2.33)$$

em que $\mathbf{Q} = (\tau)^{-1}(\mathbf{I} - \rho\mathbf{C})$ é uma matriz de precisão esparsa, mas que a covariância $\boldsymbol{\Sigma}_Y = \mathbf{Q}^{-1} = \tau(\mathbf{I} - \rho\mathbf{C})^{-1}$ resulta em uma matriz densa. Ou seja, mesmo que A_j e A_i não sejam especificadas como vizinhas, a covariância entre as mesmas é diferente de 0, uma vez que todas as áreas do espaço geográfico estão conectadas indiretamente.

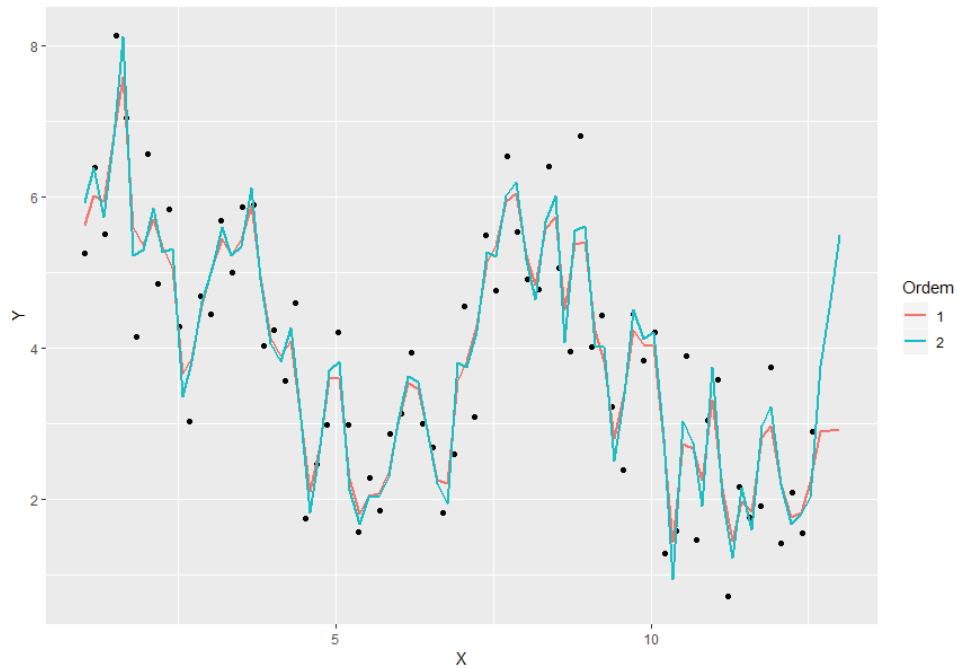
Tratando de uma forma mais geral, o espaço $D \subset \mathfrak{R}^2$ pode ser gerado por duas covariáveis $\mathbf{X} = (X_1, X_2)$ quaisquer (Figura 2.11(a)). A partir da dispersão de X_1 com X_2 , fragmenta-se o espaço gerado pelas mesmas através da técnica conhecida como diagrama de Voronoi (Klein et al., 2009).

Os polígonos, como na Figura 2.11(b), gerados pelo diagrama de Voronoi são disjuntos e formam um domínio $D \subset \mathfrak{R}^2$, em decorrência disso. A teoria para modelos espaciais para dados de área dá suporte a tal estratégia. Desta forma, é possível especificar uma estrutura de vizinhança baseada no espaço das covariáveis fragmentado pelo diagrama de Voronoi. Essa estrutura de vizinhança é o *kernel* autorregressivo denotado por \mathbf{K}_{ar} .

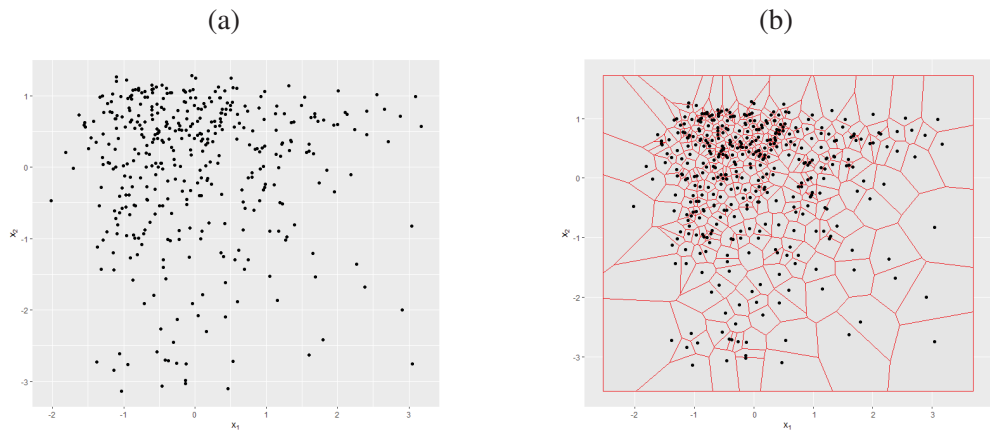
Para $\mathbf{X} = (X_1, X_2, \dots, X_p)$, com $p > 2$, optou-se por aplicar a técnica de componentes principais (Johnson e Wichern, 2007) em \mathbf{X} , e em seguida aplicar o diagrama de Voronoi nas duas primeiras componentes.

Reescrevendo a matriz de precisão como:

$$\mathbf{Q} = \sigma^2\mathbf{I} + \tau^2\mathbf{K}_{ar}, \quad (2.34)$$

Figura 2.10: *Kernel* autorregressivo.

Fonte: Elaborado pelo autor.

Figura 2.11: Dispersão entre as variáveis X_1 e X_2 em (a), e diagrama de Voronoi aplicado nas variáveis X_1 e X_2 em (b).

Fonte: Elaborado pelo autor.

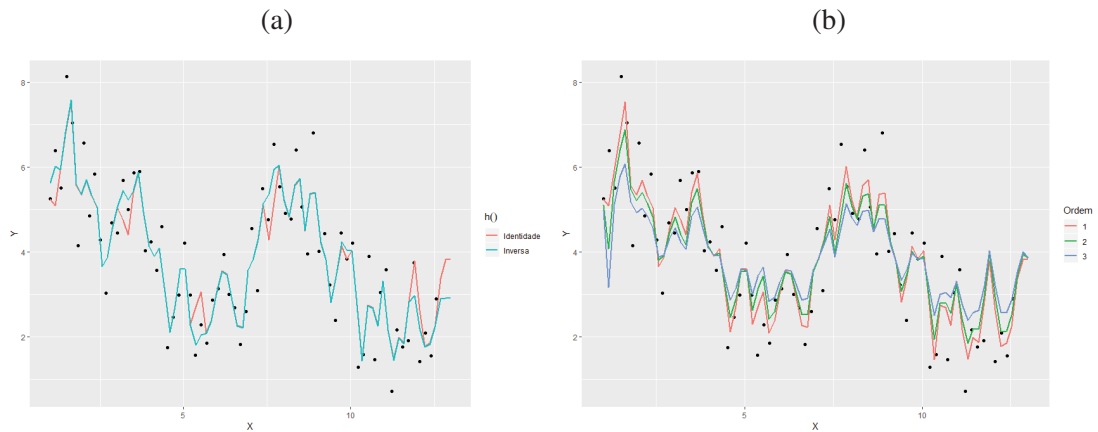
obtem-se uma equação similar a 2.14 com \mathbf{K} o *kernel* autorregressivo \mathbf{K}_{ar} . A matriz de covariância é dada pela inversa de 2.34 conforme segue:

$$\Sigma_Y = (\sigma^2 \mathbf{I} + \tau^2 \mathbf{K}_{ar})^{-1}. \quad (2.35)$$

Também é possível considerar a matriz de precisão \mathbf{Q} (conforme 2.34) como matriz de covariância, porém não é uma prática usual. Na Figura 2.12(a) estão representadas as curvas ajustadas (através da predição condicional) pelo modelo em 2.11 com: (i) matriz de covariância dada pela inversa da matriz de precisão, conforme 2.35 e denotada na legenda

da figura por $h(\cdot)$ inversa; (ii) matriz de covariância dada diretamente pela matriz de precisão, conforme 2.34 e denotada na legenda da figura por $h(\cdot)$ identidade. Pela Figura 2.12(a) observa-se que especificando a covariância como a inversa da matriz de precisão, obtém-se uma curva ajustada mais suave do que àquela obtida pela especificação da matriz de precisão como a própria matriz de covariância.

Figura 2.12: Ajustes produzidos pelo *Kernel* autorregressivo com vizinhança de primeira ordem em (a), e ajustes produzidos pelos *kernels* autorregressivos conforme a ordem de vizinhança em (b).



Fonte: Elaborado pelo autor.

Quando se tem uma única variável X , como no exemplo da Figura 2.1, a vizinhança \mathbf{W} é estabelecida de forma similar a uma série temporal. Pode-se considerar que x_i depende apenas da observação anterior X_{i-1} , conforme um passeio aleatório, ou que X_i depende da observação anterior X_{i-1} e da próxima X_{i+1} , conhecidas como conexão ou vizinhança de ordem 1 e de ordem 2, respectivamente, conforme o exemplo da Figura 2.10.

2.4 DADOS LONGITUDINAIS

Até aqui considerou-se a modelagem da variável resposta como avaliações de um único indivíduo indexadas em algum conjunto de covariáveis $\mathbf{X} = (X_1, X_2, \dots, X_r)$. Porém, as observações podem representar avaliações de l indivíduos em u observações indexadas em \mathbf{X} , sendo o número total de observações $n = lu$.

Em Diggle et al. (2002) podem ser encontrados diversos exemplos de dados longitudinais. De forma análoga ao que foi apresentado na seção 2.2, a indexação das observações de Y foi feita genericamente em r covariáveis \mathbf{X} no lugar de uma variável tempo, como ilustrado nas Figuras 2.13(a) e 2.13(b).

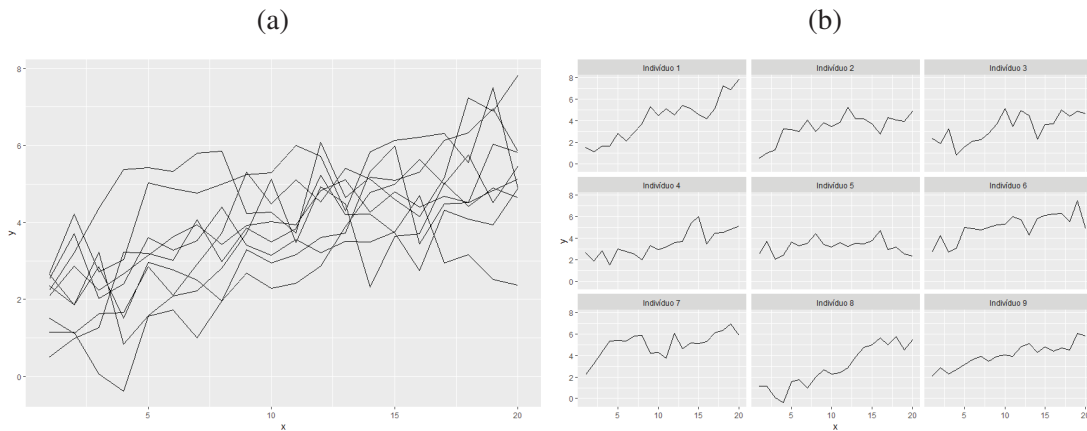
O objetivo aqui é discutir a construção da matriz de covariância marginal Σ_Y no contexto de dados longitudinais. Considera-se a ordenação das observações y_1, y_2, \dots, y_n de Y conforme a Tabela 2.1.

Sob pressuposição de normalidade de Y tem-se:

$$Y \sim NM(\boldsymbol{\mu}_Y, \Sigma_Y) \quad (2.36)$$

$$\Sigma_Y = \tau_1^2 \mathbf{M} + \tau_2^2 \mathbf{I}, \quad (2.37)$$

Figura 2.13: Dado longitudinal com 9 indivíduos representados no eixo Y avaliados em 20 observações indexadas em X em (a), e séries individuais para os 9 indivíduos com as 20 observações indexadas em X em (b)



Fonte: Elaborado pelo autor.

Tabela 2.1: Disposição da variável y_{ij} conforme o indivíduo ID e a observação OBS .

y_{ij}	ID	OBS
y_{11}	1	1
y_{12}	1	2
$y_{1.}$	1	\vdots
y_{1u}	1	n
\vdots	\vdots	\vdots
y_{l1}	l	1
y_{l2}	l	2
$y_{l.}$	l	\vdots
y_{lu}	l	n

Fonte: Elaborado pelo autor.

em que μ_Y é o vetor de médias marginal de Y especificado em 2.4, e \mathbf{M} é uma matriz semi-definida positiva de ordem $n = lu$ dada por:

$$\mathbf{M} = \mathbf{K}_{ind} \otimes \mathbf{K}_{obs}, \quad (2.38)$$

com \otimes denotando o produto de Kronecker, e \mathbf{K}_{ind} e \mathbf{K}_{obs} são matrizes simétricas semi-definidas positivas de ordens l e u , respectivamente.

Em dados longitudinais é possível decompor a estrutura de dependência em dois componentes, a dependência entre os indivíduos representada pela matriz \mathbf{K}_{ind} , e a dependência entre as observações representada pela matriz \mathbf{K}_{obs} , conforme 2.38. Tradicionalmente, pressupõe-

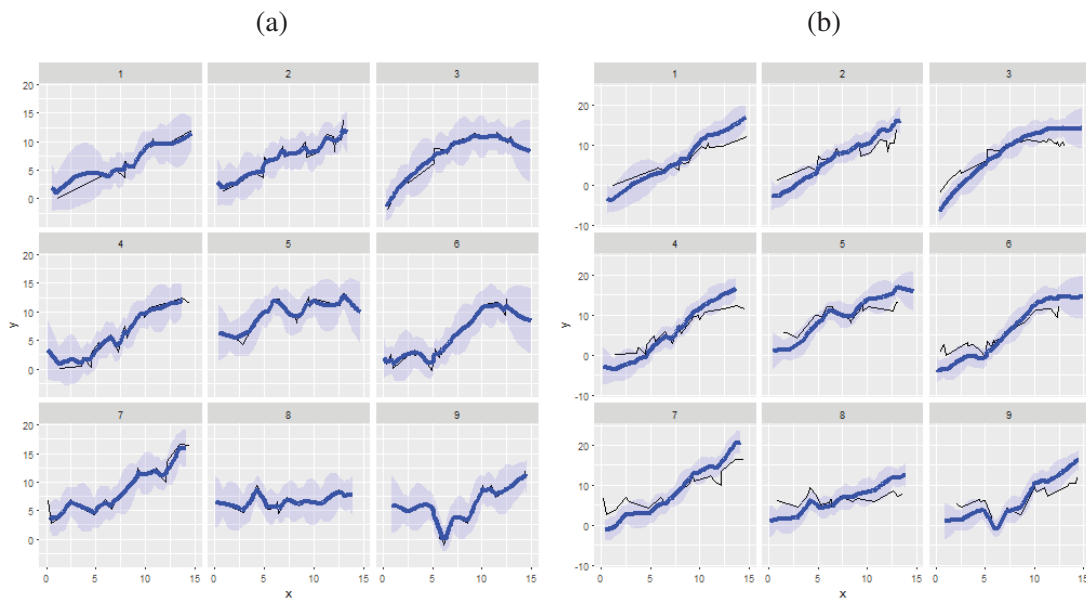
se independência entre os l indivíduos e a dependência entre as u observações é modelada pela matriz \mathbf{K}_{obs} , conforme segue:

$$\mathbf{M} = \mathbf{I}_u \otimes \mathbf{K}_{obs} = \begin{pmatrix} \mathbf{K}_{obs} & 0 & \dots & 0 \\ 0 & \mathbf{K}_{obs} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{K}_{obs} \end{pmatrix}, \quad (2.39)$$

e \mathbf{I}_u denota a matriz identidade de ordem u .

Sob independência das observações, a matriz M fica expressa por uma matriz bloco diagonal com a matriz \mathbf{K}_{obs} na diagonal conforme 2.39. A matriz \mathbf{K}_{obs} é uma estrutura de dependência definida para cada um dos l indivíduos a partir de algum *kernel* visto na seção 2.3. Como exemplo, nas Figuras 2.14 e 2.15 estão representados os ajustes (via predição condicional) dos modelos com o *kernel* exponencial e o *kernel* discreto respectivamente.

Figura 2.14: *Kernel* exponencial para as observações e sem X no preditor linear de média em (a), e *kernel* exponencial para as observações e com X no preditor linear de média em (b).



Fonte: Elaborado pelo autor.

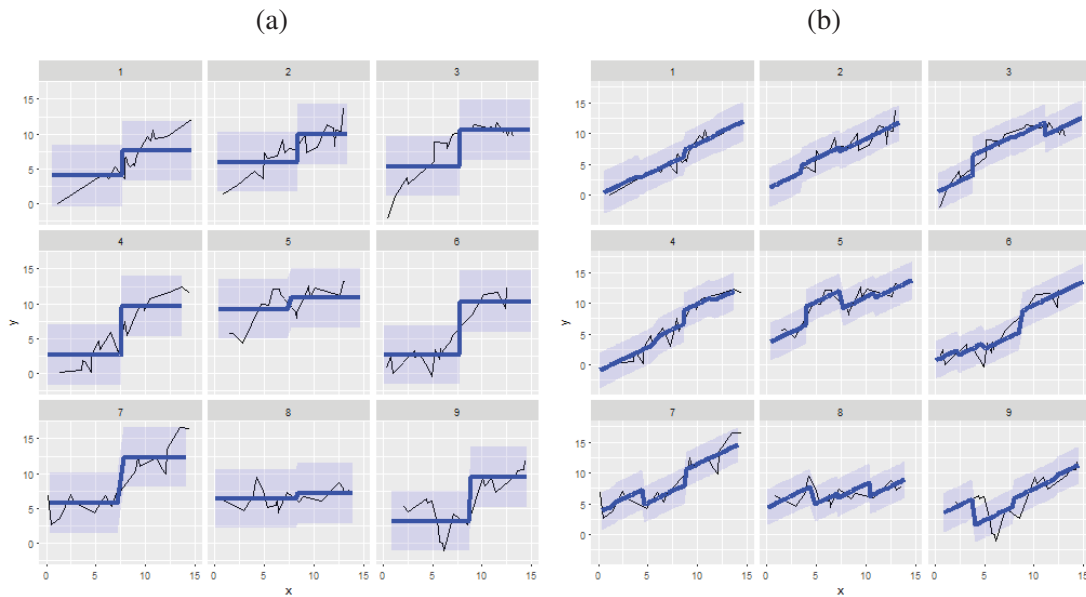
Quando as observações de Y_i ² para o indivíduo i não estão indexadas nos mesmos valores de \mathbf{X} que as observações de Y_j para o indivíduo j , então $\mathbf{K}_{obs_i} \neq \mathbf{K}_{obs_j}$ ($\forall i \neq j$ e $i, j = 1, 2, \dots, l$). Desta forma, cada um dos l indivíduos terá sua respectiva matriz *kernel* $\mathbf{K}_{obs_1}, \mathbf{K}_{obs_2}, \dots, \mathbf{K}_{obs_l}$, e \mathbf{M} fica expresso por:

$$\mathbf{M} = \text{Bdiag}\{\mathbf{K}_{obs_1}, \mathbf{K}_{obs_2}, \dots, \mathbf{K}_{obs_l}\} = \begin{pmatrix} \mathbf{K}_{obs_1} & 0 & \dots & 0 \\ 0 & \mathbf{K}_{obs_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{K}_{obs_l} \end{pmatrix}, \quad (2.40)$$

e Bdiag denota uma matriz bloco diagonal.

² Y_i denota todas as u observações associadas ao indivíduo i .

Figura 2.15: *Kernel* discreto via k-médias para as observações e sem X no preditor linear de média em (a), e *Kernel* discreto via k-médias para as observações e com X no preditor linear de média em (b).



Fonte: Elaborado pelo autor.

Agora, pressupondo que as observações em cada indivíduo são independentes, a modelagem corresponde a *kernel* \mathbf{K}_{ind} . \mathbf{M} ficará expressa por:

$$\mathbf{M} = \mathbf{K}_{ind} \otimes \mathbf{I}_l, \quad (2.41)$$

e \mathbf{I}_l denota a matriz identidade de ordem l .

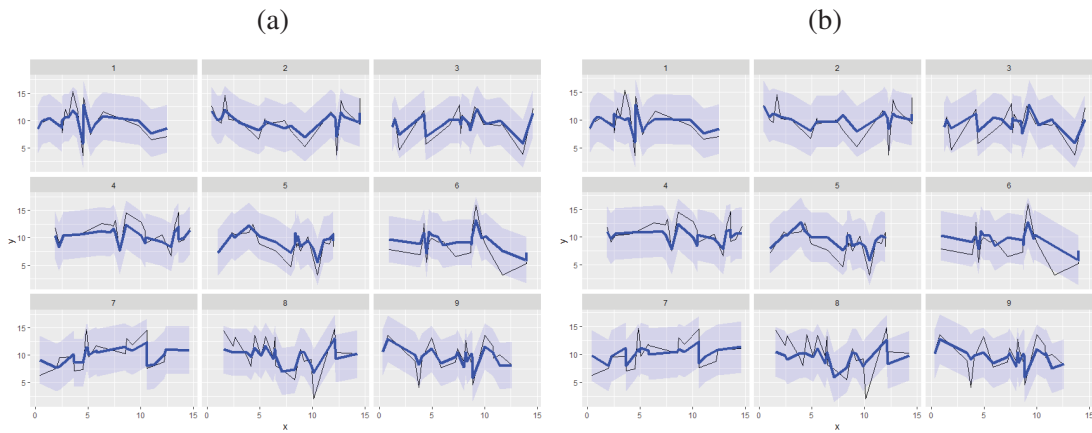
Assume-se aqui que a dependência entre os indivíduos em todas as u observações pode ser modelada por uma única matriz *kernel* \mathbf{K}_{ind} . Porém, vale ressaltar que a dependência entre indivíduos só é possível quando a indexação das observações de Y_i se dá em valores de \mathbf{X} distintos das observações de Y_j .

O *kernel* \mathbf{K}_{ind} pode ainda ser especificado via alguma função que expresse dependência que pode ser contínua, discreta, ou definindo uma estrutura autorregressiva. Como exemplo, nas Figuras 2.16 e 2.17 estão representadas as curvas ajustadas (via predição condicional) com os modelos dados por *kernels* gaussiano e discreto respectivamente, com a média constante ou com a covariável X .

Também pode-se modelar a dependência das observações e dos indivíduos conjuntamente combinando 2.41 e 2.40 na covariância:

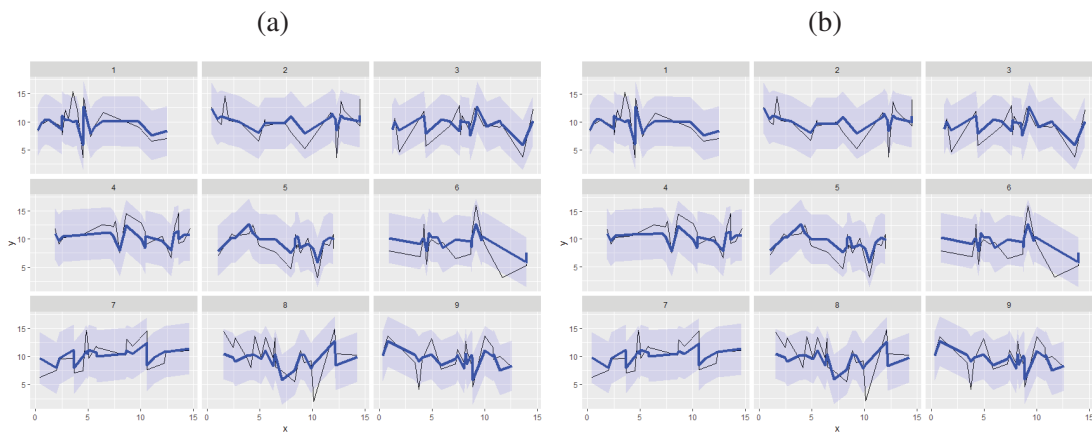
$$\Sigma_Y = \tau_1^2 \mathbf{I}_u \otimes \mathbf{K}_{obs} + \tau_2^2 \mathbf{K}_{ind} \otimes \mathbf{I}_l + \tau_3^2 \mathbf{I}. \quad (2.42)$$

Figura 2.16: *Kernel* gaussiano para os indivíduos e sem X no preditor linear de média em (a), e *kernel* gaussiano para indivíduos e com X no preditor linear de média em (b).



Fonte: Elaborado pelo autor.

Figura 2.17: *Kernel* discreto via k-médias para os indivíduos e sem x no preditor linear de média em (a), e *Kernel* discreto via k-médias para indivíduos e com X no preditor linear de média em (b).



Fonte: Elaborado pelo autor.

2.5 MODELOS LINEARES GENERALIZADOS

A classe dos GLM Nelder e Wedderburn (1972) permite a modelagem de variáveis aleatórias com distribuições de probabilidade pertencentes à família exponencial de dispersão, a qual é expressa em sua forma canônica:

$$f(Y, \theta, \tau) = \exp[\tau^{-1}\{Y\theta - b(\theta)\} + c(Y, \tau)], \quad (2.43)$$

em que τ é o parâmetro de dispersão e θ é o parâmetro canônico.

Os dois primeiros momentos de 2.43 são dados por:

$$E(Y) = b'(\theta) = \mu_Y \quad (2.44)$$

$$V(Y) = b''(\theta)(\tau\mathbf{I}) = \Sigma_Y. \quad (2.45)$$

Os GLMs modelam a média em 2.44 através de dois componentes: o preditor linear $\eta = \mathbf{X}^* \boldsymbol{\beta}$, e uma função de ligação $g(\cdot)$ responsável por garantir que os valores de η estejam no intervalo de variação da média da distribuição associada e ainda, linearizar a relação de Y com η .

Conforme 2.45, a $V(Y)$ pode ser fatorada em dois componentes: (i) O parâmetro de dispersão $\tau > 0$ intrínseco à variável resposta Y ; (ii) a função de variância $v(\boldsymbol{\mu}) = b''(\boldsymbol{\theta})$.

Reescrevendo os dois primeiros momentos em 2.44 e 2.45 respectivamente, tem-se:

$$E(Y) = \boldsymbol{\mu}_Y = g(\mathbf{X}^* \boldsymbol{\beta}) \quad (2.46)$$

$$V(Y) = \boldsymbol{\Sigma}_Y = \vartheta(\boldsymbol{\mu}_Y)^{1/2} (\tau \mathbf{I}) \vartheta(\boldsymbol{\mu}_Y)^{1/2}, \quad (2.47)$$

em que $\vartheta(\boldsymbol{\mu}_Y) = v(\boldsymbol{\mu}_Y)^T \mathbf{I}$.

A função de variância $v(\boldsymbol{\mu}_Y)$ expressa a relação média-variância que caracterizam as distribuições pertencentes à família exponencial. Para cada distribuição da família exponencial de dispersão existe uma única função de variância, e vice versa (Nelder e Wedderburn, 1972).

Algumas das distribuições mais conhecidas da família exponencial e as respectivas funções de variância são: normal ($v(\boldsymbol{\mu}_Y) = \boldsymbol{\mu}_Y^0$), Poisson ($v(\boldsymbol{\mu}_Y) = \boldsymbol{\mu}_Y^1$), Gamma ($v(\boldsymbol{\mu}_Y) = \boldsymbol{\mu}_Y^2$) e Binomial ($v(\boldsymbol{\mu}_Y) = \boldsymbol{\mu}_Y(1 - \boldsymbol{\mu}_Y)$).

2.6 MCGLM

Os Modelos Lineares Generalizados de Covariância Linear (MCGLM, sigla em inglês) são uma classe de modelos marginais proposta por Bonat e Jørgensen (2016), na qual pressupõe os dois primeiros momentos de Y finitos e especificados por:

$$E(Y) = \boldsymbol{\mu}_Y = g(\mathbf{X}^* \boldsymbol{\beta}) \quad (2.48)$$

$$V(Y) = \boldsymbol{\Sigma}_Y = \vartheta(\boldsymbol{\mu}_Y; p)^{1/2} \boldsymbol{\Omega}(\boldsymbol{\tau}, \mathbf{Z}) \vartheta(\boldsymbol{\mu}_Y; p)^{1/2}, \quad (2.49)$$

em que $\vartheta(\boldsymbol{\mu}_Y; p) = v(\boldsymbol{\mu}_Y; p)^T \mathbf{I}$.

O modelo especificado em 2.48 e 2.49 é um caso particular do MCGLM para variável resposta univariada. Em Bonat e Jørgensen (2016) encontra-se a especificação multivariada do modelo.

2.6.1 Especificação de $\boldsymbol{\Omega}(\boldsymbol{\tau}, \mathbf{Z})$

Os MCGLM estendem os GLM permitindo a modelagem da covariância marginal de Y em 2.49, a qual depende de $v(\boldsymbol{\mu}_Y; p)$ e da matriz $\boldsymbol{\Omega}(\boldsymbol{\tau}, \mathbf{Z})$. Bonat e Jørgensen (2016) denotam $\boldsymbol{\Omega}(\boldsymbol{\tau}, \mathbf{Z})$ de matriz de dispersão, a qual é dada por um preditor linear matricial $\tau_0 \mathbf{I} + \tau_1 \mathbf{Z}_1 + \dots + \tau_m \mathbf{Z}_m$ e uma função de ligação $h(\cdot)$:

$$h\{\boldsymbol{\Omega}(\boldsymbol{\tau}, \mathbf{Z})\} = \tau_0 \mathbf{I} + \tau_1 \mathbf{Z}_1 + \dots + \tau_m \mathbf{Z}_m, \quad (2.50)$$

em que $\tau_0, \tau_1, \dots, \tau_m$ são parâmetros de dispersão, \mathbf{I} é a matriz identidade de ordem n e $\mathbf{Z}_1, \dots, \mathbf{Z}_m$ são matrizes simétricas conhecidas também de ordem n .

Bonat (2018) apresenta três funções de ligação para $\boldsymbol{\Omega}$: identidade, inversa, exponencial matricial. A especificação da função $h(\cdot)$ não é intuitiva como ocorre com a função de ligação de média. Normalmente a escolha de $h(\cdot)$ recai sobre alguma forma que garanta a convergência na estimação dos parâmetros.

As matrizes $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m$ em 2.50 refletem a dependência entre as observações de Y . Para cada uma dessas matrizes são atribuídos parâmetros de dispersão $\tau_0, \tau_1, \dots, \tau_m$. Bonat e Jørgensen (2016) argumentam que é difícil definir o espaço paramétrico para $\tau_0, \tau_1, \dots, \tau_m$, pois os mesmos podem assumir valores negativos. Importante destacar que as matrizes $\mathbf{Z}_1, \dots, \mathbf{Z}_m$ são simétricas mas não necessariamente semi-definidas positivas. Porém, a matriz resultante da combinação $\tau_0\mathbf{I} + \tau_1\mathbf{Z}_1 + \dots + \tau_m\mathbf{Z}_m$ deve ser semi-definida positiva.

No contexto do presente trabalho, o MCGLM foi tratado como extensão dos Processos Gaussianos, uma vez que a estrutura de covariância é modelada marginalmente, porém pressupondo apenas os dois primeiros momentos de Y ao invés da distribuição gaussiana.

Substituindo as matrizes $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_m$ por *kernels* $\mathbf{K}_1, \dots, \mathbf{K}_m$ tem-se:

$$h\{\Omega(\boldsymbol{\tau}, \mathbf{K})\} = \tau_0\mathbf{I} + \tau_1\mathbf{K}_1 + \dots + \tau_m\mathbf{K}_m \quad (2.51)$$

Quando a ligação $h\{\}$ é identidade, os parâmetros $\tau_0, \tau_1, \dots, \tau_k$ podem ser diretamente interpretados como grau de importância atribuídos às respectivas matrizes *kernels*. Quando padronizados, isto é $\tau_j^* = \frac{\tau_j}{\sum_{j=1}^k \tau_j}$, oferecem a interpretação em termos da proporção de contribuição de cada *kernel* para a covariância.

É importante destacar que quando o *kernel* é especificado como autorregressivo, 2.3.3, a expressão 2.35 é equivalente a $h\{\Omega(\boldsymbol{\tau}, \mathbf{K})\} = \tau_1\mathbf{I} + \tau_2\mathbf{W}$, sendo \mathbf{W} a matriz de vizinhança entre as observações e $h\{\}$ a função de ligação inversa. Todavia, nem sempre é possível obter convergência do modelo com $h\{\}$ inversa. Uma alternativa é especificar $h\{\}$ como identidade, ou seja, especificar a matriz de vizinhança \mathbf{W} diretamente na covariância, o que resulta em ajustes menos suaves, conforme ilustrado em 2.12(a) e 2.12(b).

2.6.2 Especificação de $v(\mu_Y; p)$

Os MCGLMs não pressupõem e especificação explícita de uma distribuição de probabilidade para Y , mas sim seus dois primeiros momentos marginais. A covariância marginal de Y depende da função de relação média-variância $v(\mu_Y; p)$. Tal relação acaba por caracterizar implicitamente alguma família de distribuição para resposta.

Bonat e Jørgensen (2016) implementam ainda a família Tweedie (Jørgensen, 1987), que é caracterizada pela função de variância $v(\mu_Y; p) = \mu_Y^p$. Esta família tem como casos particulares: a distribuição gaussiana ($p = 0$); Poisson ($p = 1$); gamma ($p = 2$) e normal-inversa ($p = 3$). O parâmetro p pode ser estimado quando na presença de covariáveis significativas no preditor linear da média.

Bonat e Jørgensen (2016) apresentam duas variações da família Tweedie: a Poisson-Tweedie, que abrange distribuições para dados de contagens, caracterizada pela função de dispersão $v(\mu_Y; p) = \mu_Y + \mu_Y^p$; e uma variação voltada à modelagem de dados binários com função de variância $v(\mu_Y; p) = \mu_Y^{p+1}(1 - \mu_Y)^{p+2}$ (Bonat, 2018).

Como um dos objetivos do presente trabalho é manter a estrutura dos modelos tradicionais, ML e GLM, e adicionalmente modelando os desvios pela covariância, então o parâmetro p da função de variância é fixo no valor que caracterize a distribuição de probabilidade conhecida que mais se aproxima para a resposta Y .

2.6.3 Inferência

Seja $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\tau}^T)^T$ o conjunto de parâmetros do modelo, com $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_k)^T$ e $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_m)^T$.

Os parâmetros β da média são estimados utilizando função quase escore:

$$\psi_{\beta}(\beta, \tau) = (\partial \mu / \partial \beta)^T \Sigma_y^{-1} (Y - \mu). \quad (2.52)$$

Já o vetor de parâmetros τ da matriz de dispersão são estimados através da função de estimação de Pearson:

$$\psi_{\tau}(\beta, \tau) = \text{Tr} \{ (-\partial \Sigma_y^{-1} / \partial \tau) [(Y - \mu)^T (Y - \mu) - \Sigma_y] \}, \quad (2.53)$$

em que $\text{Tr} \{ \}$ é o operador traço da matriz.

Seja $\psi_{\theta}(\beta, \tau) = (\psi_{\beta}(\beta, \tau)^T, \psi_{\tau}(\beta, \tau)^T)^T$, a matriz de sensibilidade \mathbf{S}_{θ} fica expressa por:

$$\mathbf{S}_{\theta} = \begin{pmatrix} E(\partial \psi_{\beta} / \partial \beta) & E(\partial \psi_{\beta} / \partial \tau) \\ E(\partial \psi_{\tau} / \partial \beta) & E(\partial \psi_{\tau} / \partial \tau) \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{\beta} & \mathbf{S}_{\beta\tau} \\ \mathbf{S}_{\tau\beta} & \mathbf{S}_{\tau} \end{pmatrix} \quad (2.54)$$

Similarmente, a matriz de variabilidade \mathbf{V}_{θ} é dada por:

$$\mathbf{V}_{\theta} = \begin{pmatrix} \text{var}(\psi_{\beta}) & \text{var}(\psi_{\beta\tau}) \\ \text{var}(\psi_{\tau\beta}) & \text{var}(\psi_{\tau}) \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{\beta} & \mathbf{V}_{\beta\tau} \\ \mathbf{V}_{\tau\beta} & \mathbf{V}_{\tau} \end{pmatrix} \quad (2.55)$$

A estimação dos parâmetros do modelo é feita através da resolução do sistema $\psi_{\theta} = \mathbf{0}$. Porém, como μ não depende dos parâmetros $\tau_0, \tau_1, \dots, \tau_k$ da matriz de dispersão, ou seja, $\mathbf{S}_{\beta\tau} = \mathbf{0}$, então pode-se separar $\psi_{\theta} = \mathbf{0}$ nas duas equações $\psi_{\beta} = \mathbf{0}$ e $\psi_{\tau} = \mathbf{0}$, as quais são resolvidas reciprocamente pelo algoritmo de *Chaser* modificado conforme segue:

$$\beta^{(i+1)} = \beta^{(i)} - \mathbf{S}_{\beta}^{-1} \psi_{\beta}(\beta^{(i)}, \tau^{(i)}) \quad (2.56)$$

$$\tau^{(i+1)} = \tau^{(i)} - \mathbf{S}_{\tau}^{-1} \psi_{\tau}(\beta^{(i+1)}, \tau^{(i)}). \quad (2.57)$$

Bonat e Jørgensen (2016) detalham ainda o algoritmo de verossimilhança recíproca, o qual controla o tamanho do passo das iterações do algoritmo.

Seja o vetor $\hat{\theta} = (\hat{\beta}, \hat{\tau})$ dos parâmetros estimados com distribuição assintótica expressa por:

$$\hat{\theta} \sim NM(\theta, \mathbf{J}_{\theta}^{-1}), \quad (2.58)$$

em que $\mathbf{J}_{\theta}^{-1} = \mathbf{S}_{\theta}^{-1} \mathbf{V}_{\theta} (\mathbf{S}_{\theta}^{-1})^T$ é a inversa da matriz de informação de Godambe. \mathbf{V}_{θ} e \mathbf{S}_{θ}^{-1} são a matriz de variabilidade e a inversa da matriz de suavidade, respectivamente.

2.6.4 Recurso computacional

Os MCGLMs foram implementados no pacote `mcglm` (Bonat, 2018) do software R (R Core Team, 2015), o qual foram feitas as análises dos dados discutidos no capítulo 3.

2.6.5 Critério de Informação Escore

O Critério de Informação Escore (SIC) foi originalmente proposto por Stoklosa et al. (2014) como forma de seleção de covariáveis para a estrutura de média do modelo. É similar ao AIC, mas utiliza como base as equações de estimação em lugar do valor de verossimilhança.

Bonat et al. (2017) estendem o SIC para a seleção de matrizes $\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_k$ no preditor linear da covariância.

Seja um modelo com o respectivo preditor linear matricial:

$$h\{\boldsymbol{\Omega}\} = \tau_0 \mathbf{I} + \tau_1 \mathbf{Z}_1. \quad (2.59)$$

Temos interesse em testar o efeito de \mathbf{Z}_1 conforme a especificação da hipótese nula $H_0 : \tau_1 = 0$ e a alternativa $H_1 : \tau_1 \neq 0$.

Seja o vetor $\hat{\boldsymbol{\tau}} = (\hat{\tau}_0, \hat{\tau}_1) = (\hat{\tau}_0, 0)$ sob H_0 . Aplicando a função de estimação de Pearson em $\hat{\boldsymbol{\tau}}$ temos:

$$\psi_{\boldsymbol{\tau}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\tau}}) = (\psi_{\tau_0}(\hat{\boldsymbol{\beta}}, \hat{\tau}_0), \psi_{\tau_1}(\hat{\boldsymbol{\beta}}, 0)). \quad (2.60)$$

Uma vez que $\hat{\tau}_0$ é raiz de $\psi_{\tau_0} = 0$ então $\psi_{\tau_0}(\hat{\boldsymbol{\beta}}, \hat{\tau}_0) = 0$, logo 2.60 fica expresso por:

$$\psi_{\boldsymbol{\tau}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\tau}}) = (0, \psi_{\tau_1}(\hat{\boldsymbol{\beta}}, 0)). \quad (2.61)$$

Seja o SIC expresso por:

$$SIC(\boldsymbol{\beta}, \boldsymbol{\tau}) = -\mathbf{T}_{\boldsymbol{\tau}_1}(\boldsymbol{\beta}, 0) + \delta |\boldsymbol{\tau}|, \quad (2.62)$$

$\mathbf{T}_{\boldsymbol{\tau}_2}(\boldsymbol{\beta}, \tau_0)$ é conhecido por estatística escore generalizada expressa por:

$$\mathbf{T}_{\boldsymbol{\tau}_2}(\boldsymbol{\beta}, \tau_0) = \psi_{\tau_1}^T(\hat{\boldsymbol{\beta}}, 0) \mathbf{Var}(\psi_{\tau_1}(\hat{\boldsymbol{\beta}}, 0))^{-1} \psi_{\tau_1}(\hat{\boldsymbol{\beta}}, 0). \quad (2.63)$$

A estatística escore generalizada pode ser vista como uma aproximação quadrática do teste da razão de verossimilhanças, a qual sob H_0 , tem distribuição Chi-quadrado com 1 grau de liberdade.

Assim como no caso do teste escore usual, a vantagem do SIC é que há necessidade de ajustar apenas o modelo sob H_0 , desta forma podemos testar o efeito da matriz \mathbf{Z}_1 sem precisar ajustar o modelo com a mesma no preditor linear matricial.

O termo $\delta |\boldsymbol{\tau}|$ é a penalização da complexidade do modelo, para $\delta = 2$ o SIC, em 2.62, fica análogo ao critério de informação de Akaike, e para $\delta = \log n$ fica análogo ao critério de informação Bayesiano.

2.6.6 Medidas de qualidade de ajuste

Nas análises dos dados do capítulo 3 foram utilizadas medidas de qualidade de ajuste como critérios de comparação de modelos. Os MCGLMs fazem pressuposição apenas dos dois primeiros momentos de Y , portanto não há uma função de verossimilhança. Todavia, Bonat (2018) implementam uma aproximação gaussiana da verossimilhança denominada pseudo verossimilhança (plogLik), conforme segue:

$$\text{plogLik} = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |\hat{\boldsymbol{\Sigma}}_Y| - (Y - \hat{\boldsymbol{\mu}}_Y)^T \hat{\boldsymbol{\Sigma}}_Y^{-1} (Y - \hat{\boldsymbol{\mu}}_Y), \quad (2.64)$$

onde N é o número total de observações na amostra, $\hat{\boldsymbol{\Sigma}}_Y$ e $\hat{\boldsymbol{\mu}}_Y$ denotam a matriz de covariância e o vetor de médias de Y com seus respectivos parâmetros estimados.

A partir da plogLik em 2.64 são calculadas as medidas de qualidade de ajuste, pseudo AIC (pAIC) e pseudo BIC (pBIC) similares à equações do AIC e do BIC, porém com o valor de pseudo verossimilhança no lugar do valor de verossimilhança.

2.6.7 Especificação do parâmetro do *kernel*

Seja um modelo com preditor linear matricial dado por:

$$h\{\boldsymbol{\Omega}\}(\boldsymbol{\tau}, \mathbf{Z}) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}(\zeta), \quad (2.65)$$

em que $\mathbf{K}(\zeta)$ é um *kernel* que depende de um parâmetro ζ . Na implementação computacional utilizada, este deve ser estimado externamente ao MCGLM, uma vez que o mesmo requer que as matrizes do preditor linear matricial não dependam de parâmetros. A solução adotada é bastante simples, e consiste na discretização de ζ .

Seja uma sequência de valores para $\zeta : \zeta_1, \zeta_2, \dots, \zeta_g$, os quais avaliados em $\mathbf{K}(\zeta)$ produzem $\mathbf{K}(\zeta_1), \mathbf{K}(\zeta_2), \dots, \mathbf{K}(\zeta_g)$. Logo, teremos g modelos candidatos com os respectivos preditores lineares matriciais:

$$\begin{aligned} \boldsymbol{\Omega}_1(\boldsymbol{\tau}, \mathbf{Z}) &= \tau_0 \mathbf{I} + \tau_1 \mathbf{K}(\zeta_1) \\ \boldsymbol{\Omega}_2(\boldsymbol{\tau}, \mathbf{Z}) &= \tau_0 \mathbf{I} + \tau_1 \mathbf{K}(\zeta_2) \\ &\vdots \\ \boldsymbol{\Omega}_g(\boldsymbol{\tau}, \mathbf{Z}) &= \tau_0 \mathbf{I} + \tau_1 \mathbf{K}(\zeta_g). \end{aligned}$$

Todos são ajustados e o modelo escolhido será aquele que retornar o melhor SIC.

Para um *kernel* discreto $\mathbf{K}(b)$ com b blocos, a especificação da sequência para b é de números naturais $b : 2, 3, \dots, g - 1$. Os valores $b = 1$ e $b = g$ serão desconsiderados, pois são dois casos extremos. O primeiro pressupõe que todas as observações pertencem ao mesmo bloco, ou seja, todas as observações dependem de todas as outras. O segundo recai no caso de independência entre as observações, uma vez que resulta cada observação como um grupo diferente, e estes são independentes entre si.

Para $\mathbf{K}(\phi)$ um *kernel* contínuo que depende do parâmetro ϕ , conforme especificado em 2.3.1, Diggle e Ribeiro Jr. (2007) apresentam uma regra prática para o alcance de ϕ , na família *Matérn*, em função da distância máxima $max_d = \max(\|X_i - X_j\|)$ e mínima $min_d = \min(\|X_i - X_j\|)$ de \mathbf{X} .

Como ϕ e κ não são parâmetros ortogonais, diferentes valores de κ retornam diferentes sequências de valores para ϕ em que:

- Para $\kappa = 0,5$:

$$\phi : min_d/3, min_d/3 + h, min_d/3 + 2h, \dots, (min_d + max_d)/2 \quad (2.66)$$

- Para $\kappa = 1,5$:

$$\phi : min_d/4,75, min_d/4,75 + h, min_d/4,75 + 2h, \dots, (min_d + max_d)/2 \quad (2.67)$$

- Para $\kappa = 2,5$:

$$\phi : min_d/5,92, min_d/5,92 + h, min_d/5,92 + 2h, \dots, (min_d + max_d)/2, \quad (2.68)$$

sendo h o espaçamento entre os valores das sequências 2.66, 2.67 e 2.68. Quanto menor o valor de h mais valores a sequência terá e mais precisa será a estimação de ϕ .

3 ANÁLISE DE DADOS

Este capítulo tem como objetivo principal exemplificar e ilustrar a estratégia de modelagem discutida até então. Isto é feito através da análise de sete conjuntos de dados com três principais características: dados típicos de modelos de regressão, classificação e espaciais.

Os exemplos das seções 3.1 e 3.3 servem para ilustrar as curvas ajustadas aos dados. Dependendo de como o componente de covariância do modelo é especificado, resultam diferentes características da curva ajustada, como continuidade e suavidade, que são visualmente ilustradas. Também são ajustados modelos já difundidos na literatura, como os NLM, GLM, GAM e LDA, com a intenção de comparar os ajustes usuais dos mesmos com os ajustes dos modelos com especificação da dependência entre as observações.

O exemplo da seção 3.2 é um problema de regressão múltipla, em que a finalidade é prever a variável resposta, contínua, segundo as informações das covariáveis. Este exemplo explora uma das vantagens de especificar o componente de covariância nas covariáveis, sobretudo com finalidade de predição, que é o fato de poder acomodar o conjunto de covariáveis multidimensionalmente. Desta forma, possíveis interações entre as covariáveis e relações não lineares podem ser consideradas. Além disso, o número de parâmetros no componente de covariância independe do número de covariáveis, ou seja, através do componente de covariância é possível ajustar modelos em conjuntos de dados onde o número de covariáveis é igual ou superior ao número de observações, como o exemplo de classificação tratado na seção 3.4.

Os exemplos das seções 3.5, 3.6 e 3.7 tratam de conjuntos de dados com indexação espacial, mais especificamente, dados de área. Modelos para dados de área, em geral, tem por finalidade a suavização dos valores observados da variável resposta, ou seja, são voltados em obter o melhor ajuste possível aos dados. Sendo assim, estes três exemplos de dados de área foram utilizados como contexto para comparar a qualidade de ajuste dos modelos com diferentes especificações do componente de média e de covariância.

3.1 QUALIDADE DO AR DE NOVA YORK

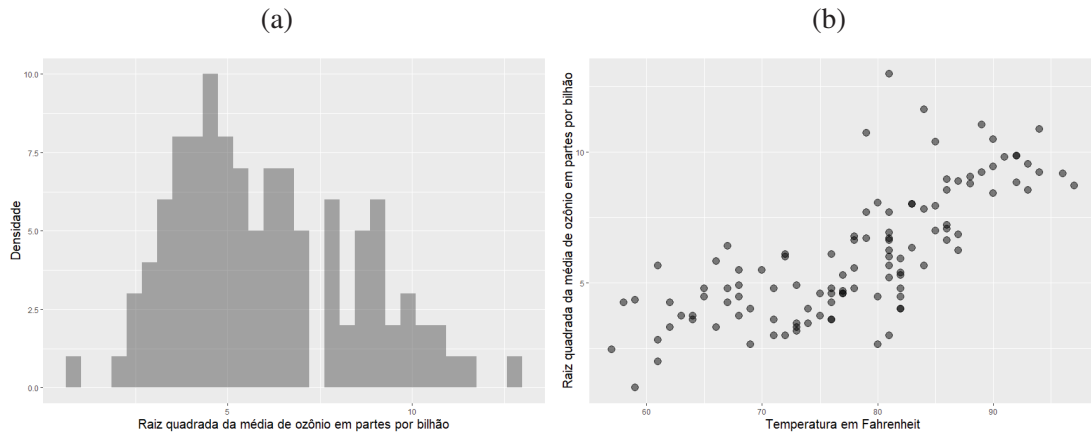
O conjunto de dados (*airquality*), do pacote **MASS** (Venables e Ripley, 2002), contém medidas diárias da qualidade do ar na cidade de Nova York de Maio a Setembro de 1973. São 153 observações de 4 variáveis. Excluindo-se os dados incompletos restam 111 observações.

A variável resposta é a raiz quadrada da média de ozônio em partes por bilhão de 1.300 a 1.500 horas. O objetivo aqui é encontrar um modelo de predição da variável resposta em função das covariáveis. Como exemplo ilustrativo, foi usada apenas a covariável *temperatura*. O principal objetivo neste exemplo é ilustrar de forma a explorar intuitivamente os resultados que são fácil visualização por conter apenas uma covariável. Uma análise mais completa está na página de complementos online da dissertação ¹.

Marginalmente, a variável resposta apresenta assimetria à direita, como ilustrado em 3.1(a). Além disso, a variável resposta apresenta uma relação não linear com a covariável, conforme a Figura 3.1(b).

¹<http://www.leg.ufpr.br/fmorodiss>

Figura 3.1: Distribuição da variável resposta (a), e dispersão com a covariável (b).



Fonte: Elaborado pelo autor.

Conforme argumentado na seção 2.2, existe um conjunto de funções de transformação, para a covariável, que permite tornar a relação vista na Figura 3.1(b) mais próxima da linearidade. Estas funções podem ser especificadas diretamente na média da variável resposta ou implicitamente pela covariância.

O conjunto de dados foi separado em dados de treino e dados de teste. O conjunto de dados de treino é composto por 66 observações, selecionadas por amostragem aleatória simples sem reposição dos dados originais e foi utilizado nos ajustes dos modelos. O conjunto de dados de teste, composto por 45 observações, serviu para quantificar o erro preditivo, este calculado pela soma dos quadrados das diferenças entre os valores observados e preditos.

O modelo base é o modelo linear normal, em que:

$$E(Y) = \beta_0 + \beta_1 X \quad (3.1)$$

$$V(Y) = \tau_0 \mathbf{I}. \quad (3.2)$$

Em 3.1 é pressuposto que a relação da variável resposta com a covariável possa ser aproximada por uma reta, conforme a Figura 3.2(a). Todavia é possível melhorar o ajuste especificando funções suaves para a média, como por exemplo, ajustar um modelo linear normal com expansões da covariável em séries polinomiais:

$$E(Y) = \beta_0 + \sum_{i=1}^d \beta_i X^i \quad (3.3)$$

$$V(Y) = \tau_0 \mathbf{I}. \quad (3.4)$$

Outra forma de obter um ajuste suave é especificando uma função suave $s(X)$ para o componente de média, como em um GAM:

$$E(Y) = \beta_0 + s(X) \quad (3.5)$$

$$V(Y) = \tau_0 \mathbf{I}, \quad (3.6)$$

ou combinando os modelos linear e suave:

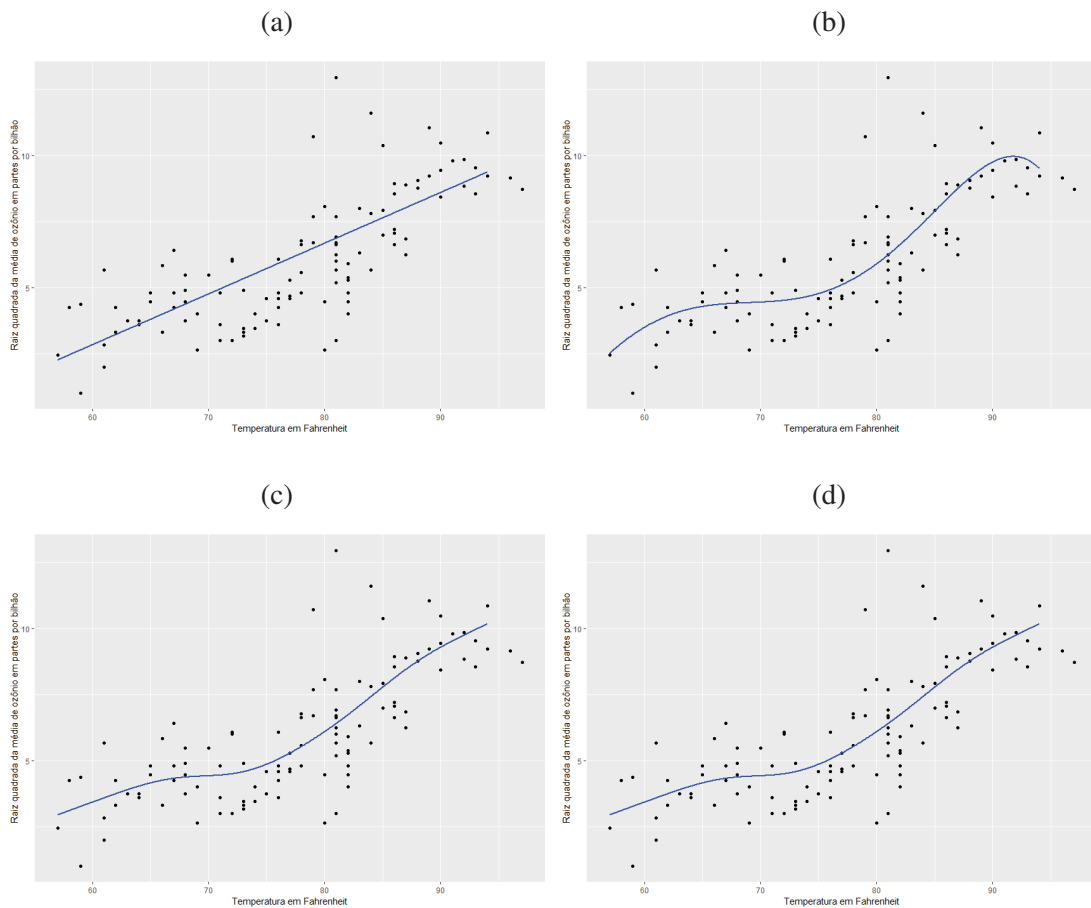
$$E(Y) = \beta_0 + \beta_1 x + s(X) \quad (3.7)$$

$$V(Y) = \tau_0 \mathbf{I}. \quad (3.8)$$

O modelo linear normal com expansão polinomial em X (Figura 3.2(b)) obteve um ajuste melhor aos dados comparado com os ajustes obtidos pelo GAM (Figuras 3.2(c) e 3.2(d)), sobretudo a curvatura nas bordas.

O GAM com média especificada apenas por $s(X)$ (Figura 3.2(c)) produz a mesma curva ajustada que o GAM com média especificada pela combinação da função $s(X)$ e linear (Figura 3.2(d)).

Figura 3.2: Modelos ajustados por especificação de função de média: Ajuste com modelo linear normal em (a), com modelo linear normal com expansão polinomial de ordem 5 na covariável em (b), com GAM com apenas a função suave em (c), com GAM com a combinação das funções linear e suave em (d).



Fonte: Elaborado pelo autor.

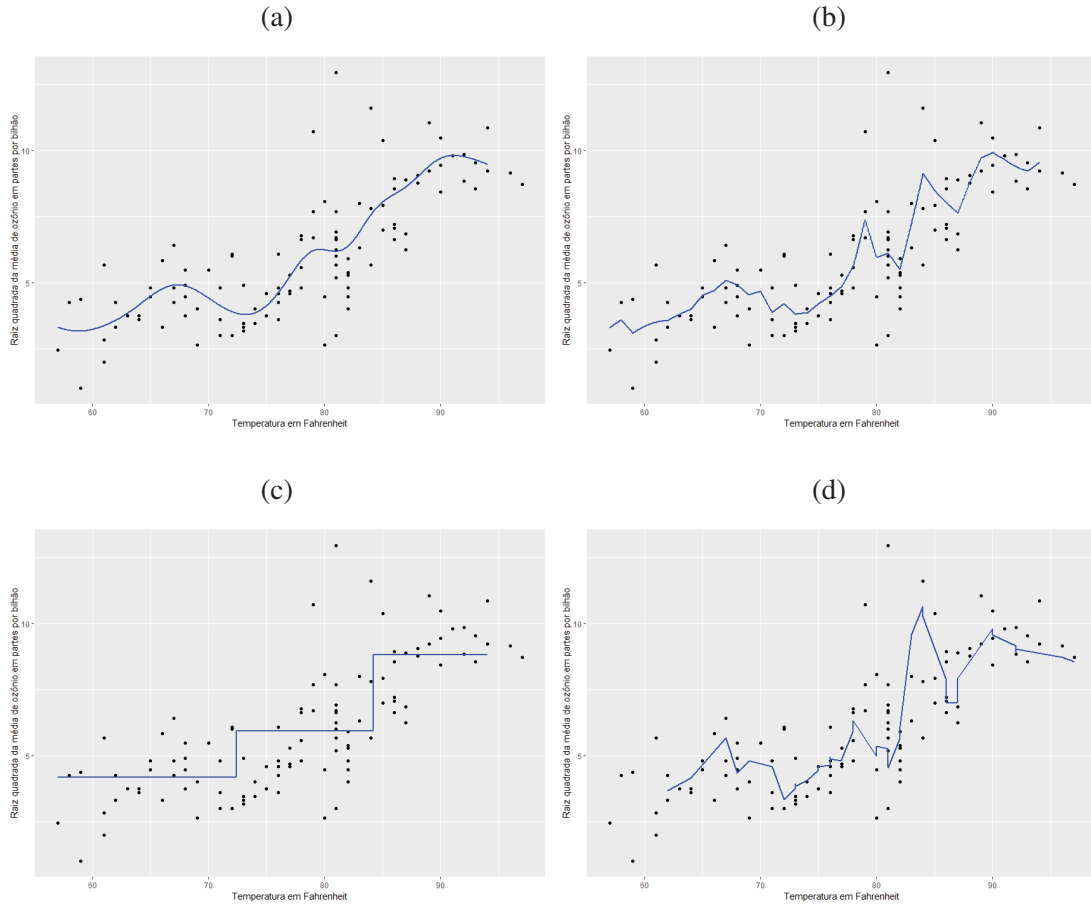
Também é possível manter o componente de média constante e especificar uma matriz de covariância Σ_Y de tal forma que retorne na predição uma função de média que se ajusta aos dados,

$$E(Y) = \beta_0. \quad (3.9)$$

$$V(Y) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}, \quad (3.10)$$

os ajustes de modelos com diferentes especificações para \mathbf{K} são mostrados na Figura 3.3.

Figura 3.3: Modelos ajustados por especificação de função de covariância: Ajuste com *kernel* gaussiano em (a), com *kernel* exponencial em (b), com *kernel* via k-médias em (c), com *kernel* autorregressivo de ordem 1 em (d).



Fonte: Elaborado pelo autor.

O modelo com *kernel* gaussiano (conforme a Figura 3.3(a)) retorna uma curva de ajuste suave, ao contrário dos ajustes com os demais *kernels* (conforme as Figuras 3.3(b), 3.3(c) e 3.3(d)). Se comparado com os ajustes dos modelos de média, o ajuste produzido pelo modelo com *kernel* gaussiano captou mais padrões locais, porém, aumentando o alcance da correlação, parâmetro ϕ , é possível obter ajustes similares aos dos modelos de média.

Os modelos com *kernel* exponencial (Figura 3.3(b)) e com *kernel* autorregressivo de ordem 1 (Figura 3.3(d)) fornecem curvas de ajuste contínuas e não suaves. A diferença é que o ajuste fornecido pelo modelo com o *kernel* autorregressivo de ordem 1 sofreu maior influência de pontos atípicos, como pode ser visto o ajuste da curva entre, aproximadamente, os valores 80 e 85 de X na Figura 3.3(d).

O modelo com *kernel* via k-médias fornece um ajuste do tipo função degrau, em que os *clusters* determinam os degraus. Conforme a Figura 3.3(c), o ajuste não é adequado pois pelo gráfico de dispersão, na Figura 3.1(b), as observações não apresentam ter pontos de quebra ou agrupamentos em torno de um valor de Y .

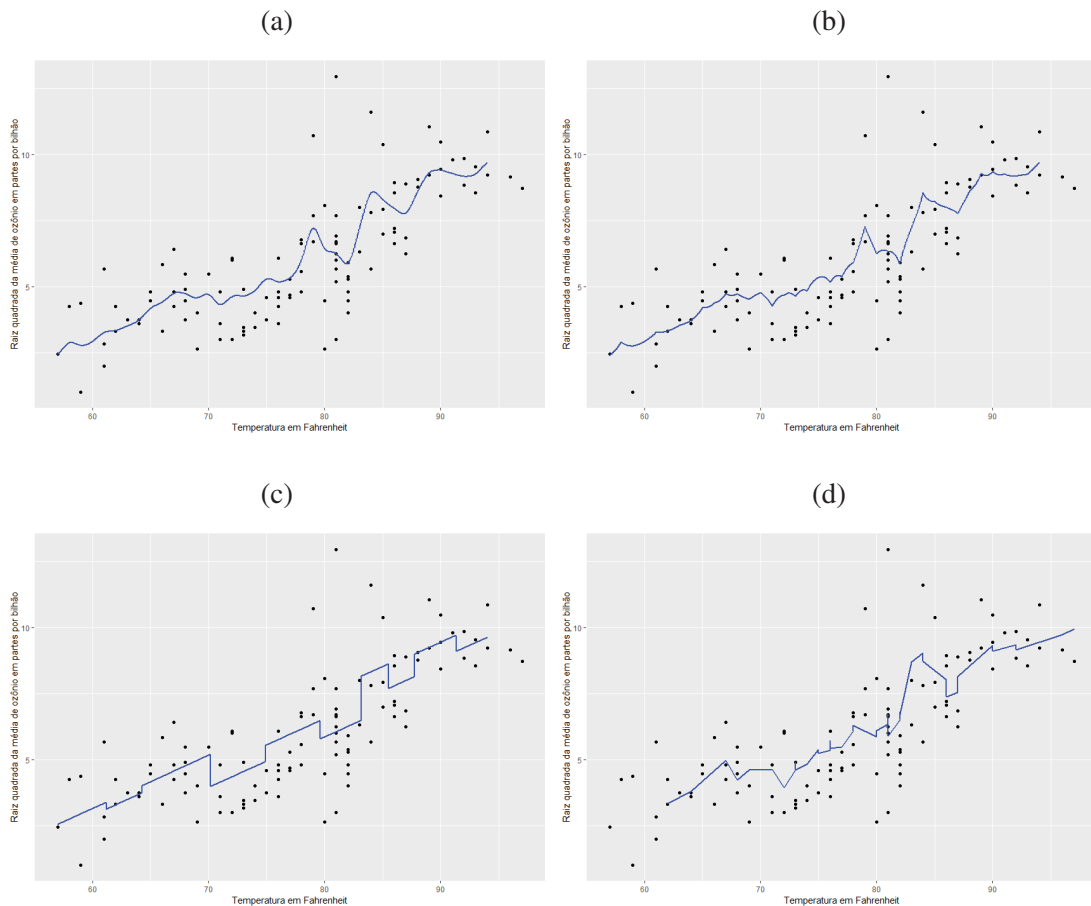
Mantendo o componente linear na média em 3.1 e acrescentando \mathbf{K} no componente de covariância como forma de correção dos desvios tem-se que:

$$E(Y) = \beta_0 + \beta_1 X \quad (3.11)$$

$$V(Y) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}. \quad (3.12)$$

Desta forma, os ajustes deste modelo para mesmas especificações para \mathbf{K} consideradas anteriormente são mostrados na Figura 3.4.

Figura 3.4: Ajustes para função linear da covariável combinada com estrutura para covariância: Ajuste com *kernel* gaussiano em (a), com *kernel* exponencial em (b), com *kernel* via k-médias em (c), com *kernel* autorregressivo de ordem 1 em (d).



Fonte: Elaborado pelo autor.

A inclusão da covariável no componente de média dos modelos com *kernel* fez os ajustes dos mesmos mais próximos da linearidade (Figura 3.4). Além disso, os valores de ϕ dos *kernels* gaussiano e exponencial reduziram, o que resultou em ajustes que melhor capturam padrões locais, conforme as Figuras 3.4(a) e 3.4(b). Se o objetivo for de predição, a captura de padrões locais resulta em aumento do erro de predição pelo fato do modelo ter "decorado" os dados (*overfitting*), como pode ser visto na comparação das somas dos erros quadráticos (EQ) das predições dos modelos com *kernels* gaussiano e exponencial com e sem a covariável na média.

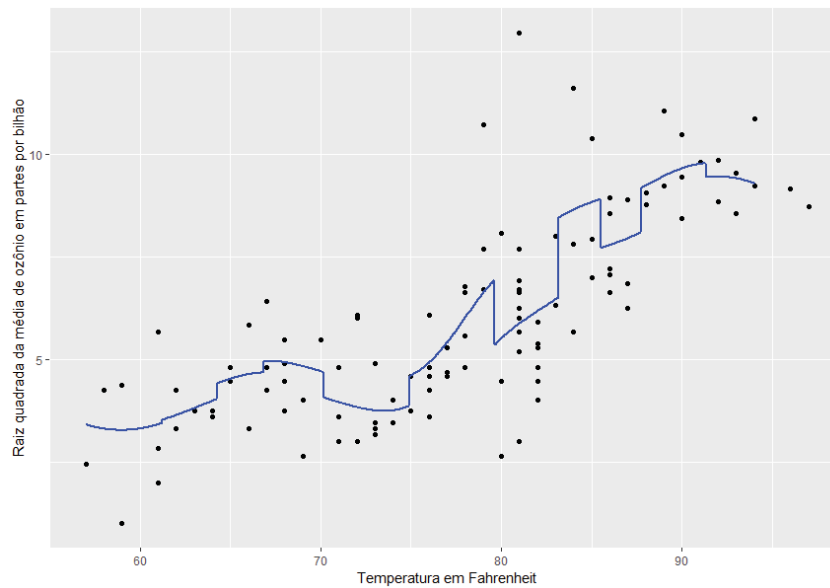
Pode-se tentar melhorar ainda mais o ajuste assumindo uma combinação linear na especificação no componente de covariância. Por exemplo, tomando-se

$$E(Y) = \beta_0 + \beta_1 X \quad (3.13)$$

$$V(Y) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}_1 + \tau_2 \mathbf{K}_2, \quad (3.14)$$

em que \mathbf{K}_1 é definido pelo *kernel* gaussiano com $\phi = 2,6$ e \mathbf{K}_2 pelo *kernel* definido por k-médias com $k = 11$ obtém-se o ajuste da Figura 3.5.

Figura 3.5: *Kernel* gaussiano com $\phi = 2,6$ + *kernel* via k-médias com $k = 11$.



Fonte: Elaborado pelo autor.

A Tabela 3.1 resume os ajustes dos modelos considerados, comparando os mesmos pelo critério de qualidade preditiva dado pelo erro quadrático (EQ). O EQ é dado por $EQ = (Y - Y^*)(Y - Y^*)^T$, em que Y^* é o vetor dos valores preditos pela equação da predição condicional, conforme 2.19. Entre os modelos da classe dos MCGLMs, o melhor modelo foi aquele com o *kernel* exponencial e com componente de média constante. O pior desempenho foi o do modelo com componente de média constante e *kernel* via k-médias. A inclusão do *kernel* via k-médias não ajudou a melhorar a performance preditiva dos modelos com *kernel* gaussiano e com *kernel* exponencial.

Ainda pela Tabela 3.1, destaca-se o desempenho do SVM (*support vector machine*) com base radial. O SVM é um modelo matemático o qual otimiza uma função objetivo conforme um conjunto de restrições. Porém, o SVM é válido se o objetivo for somente prever, uma vez que não especifica um modelo com componentes interpretáveis e portanto não fornece informações acerca da decomposição da dispersão da variável resposta.

Quanto à qualidade de ajuste (Tabela 3.2), o pBIC e pAIC apontaram para os modelos com a covariável X no componente de média, sendo que o modelo com função de média polinomial de grau cinco obteve melhor ajuste pelo pAIC e o modelo com função de média linear na covariável obteve melhor ajuste pelo pBIC. Ainda pela Tabela 3.2, dentre os modelos com *kernels*, tanto o pAIC quanto o pBIC apontaram para os modelos com a covariável na média.

Tabela 3.1: Tempo de processamento (TP) e soma dos erros quadrados (EQ) conforme os modelos ajustados.

Classe	Média	kernel	TP	EQ
MCGLM	$\sim x$	-	1,21	137,01
MCGLM	$\sim x + x^2 + x^3 + x^4 + x^5$	-	1,05	122,07
MCGLM	~ 1	gaussiano	2,26	114,83
MCGLM	$\sim x$	gaussiano	6,55	118,26
MCGLM	~ 1	exponencial	8,57	111,48
MCGLM	$\sim x$	exponencial	11,94	116,85
MCGLM	~ 1	k-médias	7,22	148,17
MCGLM	$\sim x$	k-médias	8,23	131,12
MCGLM	~ 1	Autorregressivo	7,98	158,81
MCGLM	$\sim x$	Autorregressivo	2,75	130,21
MCGLM	~ 1	gaussiano+k-médias	12,74	121,28
MCGLM	~ 1	exponencial+k-médias	10,80	118,12
GAM	$\sim x + s(x)$	-	0,25	121,21
GAM	$\sim s(x)$	-	0,25	121,21
SVM radial	-	-	0,44	111,80

Fonte: Elaborado pelo autor.

Tabela 3.2: Medidas de qualidade de ajuste dos modelos ajustados para a classe dos MCGLMs.

Média	kernel	pLogLik	pAIC	pBIC
~ 1	-	-256.03	516.06	521.48
$\sim x$	-	-210.93	427.86	435.99
$\sim x + x^2 + x^3 + x^4 + x^5$	-	-201.89	417.78	436.74
~ 1	exponencial	-215.78	437.56	445.69
$\sim x$	exponencial	-207.21	422.42	433.26
~ 1	gaussiano	-214.72	435.44	443.57
$\sim x$	gaussiano	-207.58	423.16	434.00
~ 1	k-médias	-218.85	443.70	451.83
$\sim x$	k-médias	-207.03	422.06	432.90
~ 1	exponencial+k-médias	-213.07	434.14	444.98
~ 1	gaussiano+k-médias	-213.60	435.20	446.04

Fonte: Elaborado pelo autor.

3.2 PERFORMANCE RELATIVA DE CPUS

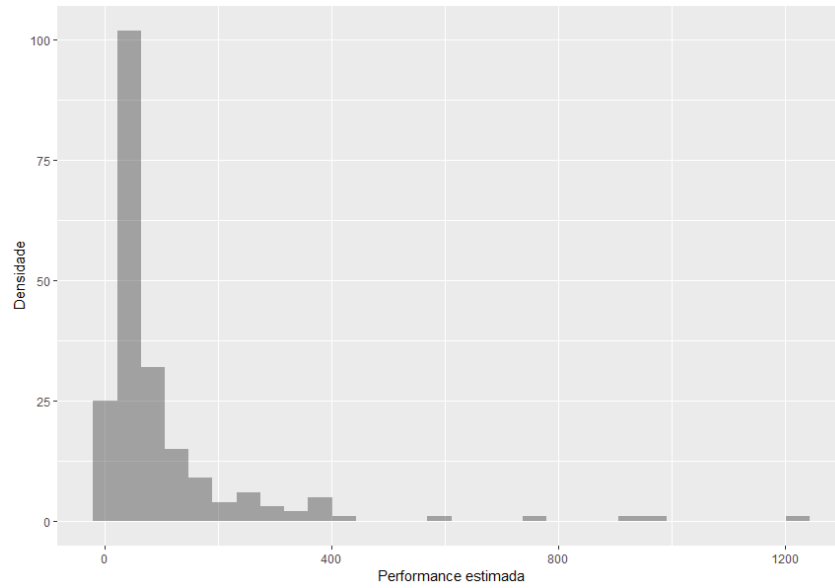
A performance relativa é uma variável adimensional e há interesse em predizê-la segundo uma série de características da CPU avaliada, como a performance publicada pelo vendedor, tempo de ciclo em nanosegundos, tamanho do cache em *kilobytes*, dentre outras variáveis descritas em Ein-Dor e Feldmesser (1987). É neste contexto dos dados do segundo exemplo apresentado aqui.

O conjunto de dados (*cpus*) está disponível no pacote **MASS** (Venables e Ripley, 2002) e conta com 209 observações, das quais foram separadas 125 observações para o ajuste e 84 observações para predição.

O primeiro exemplo teve como objetivo visualizar os ajustes obtidos para cada um dos modelos especificados em uma única covariável. Este segundo exemplo tem como principal objetivo comparar a performance preditiva dos modelos especificados em múltiplas covariáveis.

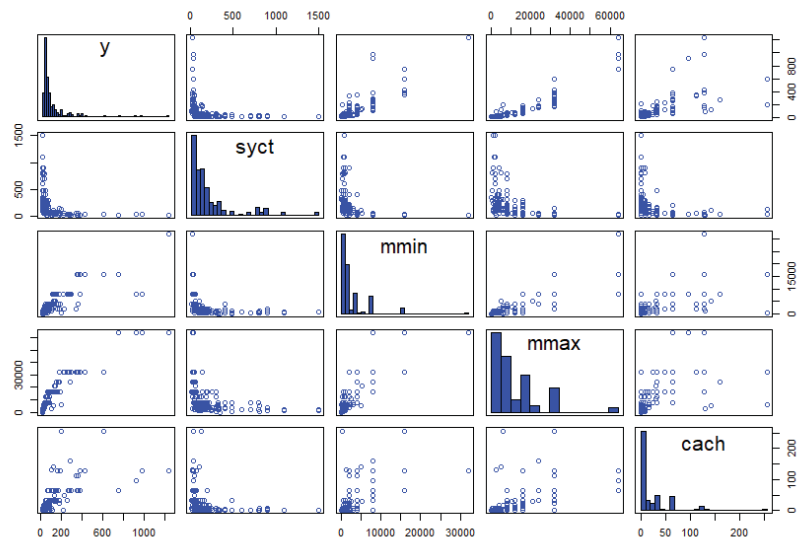
Conforme a Figura 3.6, a performance relativa apresenta uma distribuição bastante assimétrica à direita. Além disso, as Figuras 3.7 e 3.8 mostram uma relação locação-escala da variável resposta, representada por y , com a maior parte das covariáveis. Desta forma, optou-se

Figura 3.6: Distribuição da variável resposta.



Fonte: Elaborado pelo autor.

Figura 3.7: Dispersão da variável resposta com as quatro primeiras covariáveis.



Fonte: Elaborado pelo autor.

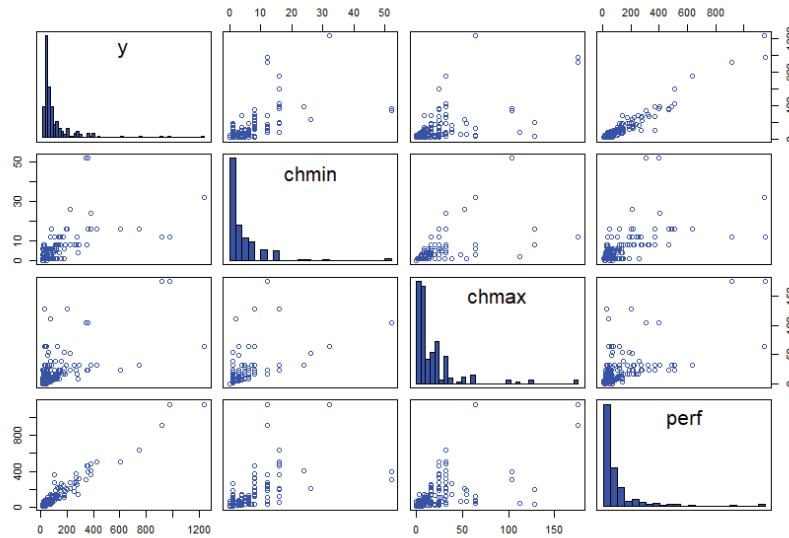
por adotar a função de variância da distribuição gamma $v(\mu_Y) = \mu_Y^2$. O modelo de base é um GLM gamma especificado como:

$$E(Y) = \exp(\beta \mathbf{X}^*) \quad (3.15)$$

$$V(Y) = \text{diag}(\mu_Y) \Omega(\tau, \mathbf{Z}) \text{diag}(\mu_Y) \quad (3.16)$$

$$\Omega(\tau, \mathbf{Z}) = \tau_0 \mathbf{I}. \quad (3.17)$$

Figura 3.8: Dispersão da variável resposta com as três últimas covariáveis.



Fonte: Elaborado pelo autor.

Na equação 3.15 a matriz experimental $\mathbf{X}^* = (\mathbf{1}_n, syct, mmin, mmax, cach, perf)^2$, sendo que *syct*, *mmin*, *mmax*, *cach* e *perf* são cinco entre as sete covariáveis disponíveis, selecionadas via algoritmo *stepwise*.

Foram ajustados modelos com a especificação de matriz de dependência entre as observações. A partir do modelo de base, foram ajustados modelos com a adição de uma matriz *kernel* \mathbf{K} na matriz de dispersão da equação 3.17, com as mesmas covariáveis do componente de média da equação 3.15:

$$E(Y) = \exp(\beta \mathbf{X}^*) \quad (3.18)$$

$$V(Y) = \text{diag}(\mu_Y) \Omega(\tau, \mathbf{Z}) \text{diag}(\mu_Y) \quad (3.19)$$

$$\Omega(\tau, \mathbf{Z}) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K} \quad (3.20)$$

ou com média constante:

$$E(Y) = \exp(\beta_0) \quad (3.21)$$

$$V(Y) = \text{diag}(\mu_Y) \Omega(\tau, \mathbf{Z}) \text{diag}(\mu_Y) \quad (3.22)$$

$$\Omega(\tau, \mathbf{Z}) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}. \quad (3.23)$$

A matriz \mathbf{K} foi especificada no espaço de todas as sete covariáveis de três formas diferentes: como *kernel* gaussiano, exponencial e via k-médias. Também foi ajustado um SVM com base radial e um GAM, este último especificado conforme segue:

$$E(Y) = \exp(\eta) \quad (3.24)$$

$$\eta = s(syct) + s(mmin) + s(mmax) + s(cach) + s(perf) \quad (3.25)$$

$$V(Y) = \text{diag}(\mu_Y) (\tau_0 \mathbf{I}) \text{diag}(\mu_Y). \quad (3.26)$$

² $\mathbf{1}_n$ denota o vetor de 1s de comprimento n .

Na Tabela 3.3 encontram-se informações acerca da qualidade preditiva e tempos computacionais dos ajustes, os quais foram descritos anteriormente. Adotou-se a notação x para quando o componente de média do modelo for especificado pelas covariáveis e 1 para quando o componente de média especificado por uma constante. O erro quadrático médio (EQM) é uma medida de qualidade preditiva dada por $EQM = \frac{1}{n}(Y - Y^*)^T (Y - Y^*)$, sendo Y^* o vetor dos valores preditos pela equação da predição condicional conforme 2.19.

Tabela 3.3: Tempo de processamento (TP) em segundos e erro quadrático médio (EQM) das predições conforme os modelos ajustados.

Classe	Média	kernel	TP	EQM
MCGLM	$\sim x$	–	1,36	163.640,9
MCGLM	~ 1	gaussiano	42,00	6.842,5
MCGLM	$\sim x$	gaussiano	12,53	158.907,0
MCGLM	~ 1	exponencial	35,05	5.125,6
MCGLM	$\sim x$	exponencial	22,43	117.845,8
MCGLM	~ 1	k-médias	20,01	19.742,7
MCGLM	$\sim x$	k-médias	15,68	196.600,1
GAM	$\sim s(x)^3$	–	0,81	531.162,6
SVM radial	–	–	0,13	16.855,9

Fonte: Elaborado pelo autor.

Pela Tabela 3.3 conclui-se que, para os modelos com *kernels* gaussiano, exponencial e k-médias, o componente de média especificado em função das covariáveis resulta em piores performances preditivas, em comparação ao componente de média especificado por uma constante. E comparando os *kernels* entre si, a melhor performance preditiva, segundo o EQM, foi o modelo com *kernel* exponencial. Este mesmo modelo também teve a melhor performance preditiva entre todos os modelos avaliados.

Ainda pela Tabela 3.3, o GAM foi o modelo que forneceu a pior performance preditiva dentre todos os modelos. As funções suaves do GAM não comportam multidimensionalmente um conjunto de covariáveis, o que acaba não levando em conta efeitos de interação das mesmas.

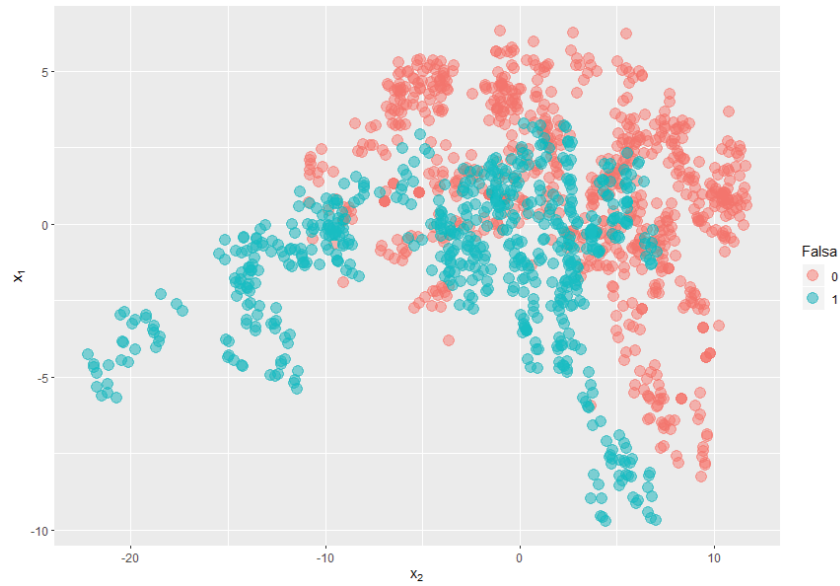
3.3 RECONHECIMENTO DE CÉDULAS FALSAS

O conjunto de dados (*banknote-authentication*) relaciona-se com a identificação de cédulas falsas e genuínas (VANSCHOREN et al., 2013), na qual a variável resposta é binária, assumindo 1 quando a cédula é falsa e 0 caso contrário. Também conta com 4 covariáveis, obtidas a partir da análise *wavelet* de imagens, em escalas de cinza, de 1372 cédulas.

O objetivo em uma análise de tais dados é encontrar o melhor modelo de classificação de cédulas falsas ou genuínas de acordo com valores fornecidos das quatro covariáveis. Aqui, o objetivo principal é ilustrar as diferentes estratégias de ajuste incluindo as propostas de modelagem da estrutura de dependência. Desta forma, optou-se por trabalhar com as duas primeiras componentes de uma análise de componentes principais destas covariáveis, as quais explicam 90,3% da variabilidade das covariáveis originais. Desta forma podemos explorar visualizações das análises com este conjunto de dados, o que é objetivo do exemplo.

Para problemas de classificação supervisionada, sobretudo classificação binária, provavelmente os modelos estatísticos mais conhecidos são o modelo de regressão logística e o análise discriminante linear de Fisher (ADL). Ambos os modelos tentam separar as classes tanto quanto possível através de uma função linear, conforme segue ilustrado em 3.10(a) e 3.10(b), respectivamente.

Figura 3.9: Distribuição da variável resposta na dispersão dos dois primeiros componentes.



Fonte: Elaborado pelo autor.

A análise discriminante linear de Fisher reduz as r covariáveis $\mathbf{X} = (X_1, X_2, \dots, X_r)$ em uma única variável, a partir de combinações lineares $z = C^T \mathbf{X}$. O problema consiste em determinar o vetor C^T de coeficientes de forma a obter a melhor discriminação entre os grupos (Johnson e Wichern, 2007, cap. 11), que no caso atual seria o grupo das cédulas falsas e das cédulas genuínas. Importante destacar que a ADL não requer pressuposição de normalidade dos dados, todavia, pressupõe implicitamente que a covariância das observações para ambos os grupos seja a mesma. Quando a covariância de um dos grupos considerada diferente da covariância do outro grupo, a função de discriminação passa a ser quadrática, conforme segue 3.10(c).

O modelo de regressão logística é um caso particular dos MLGs, em que a variável resposta é modelada pela distribuição binomial, com relação média-variância $v(\mu_Y) = \mu_Y(1 - \mu_Y)$, e a função de ligação logito. Com isso, o modelo de regressão logística pode ser expresso como

$$E(\mathbf{Y}) = \mu_Y = \frac{\exp(\mathbf{X}^* \boldsymbol{\beta})}{\exp(\mathbf{X}^* \boldsymbol{\beta}) + 1} \quad (3.27)$$

$$V(\mathbf{Y}) = v(\mu_Y)^{1/2} \boldsymbol{\Omega} v(\mu_Y)^{1/2} \quad (3.28)$$

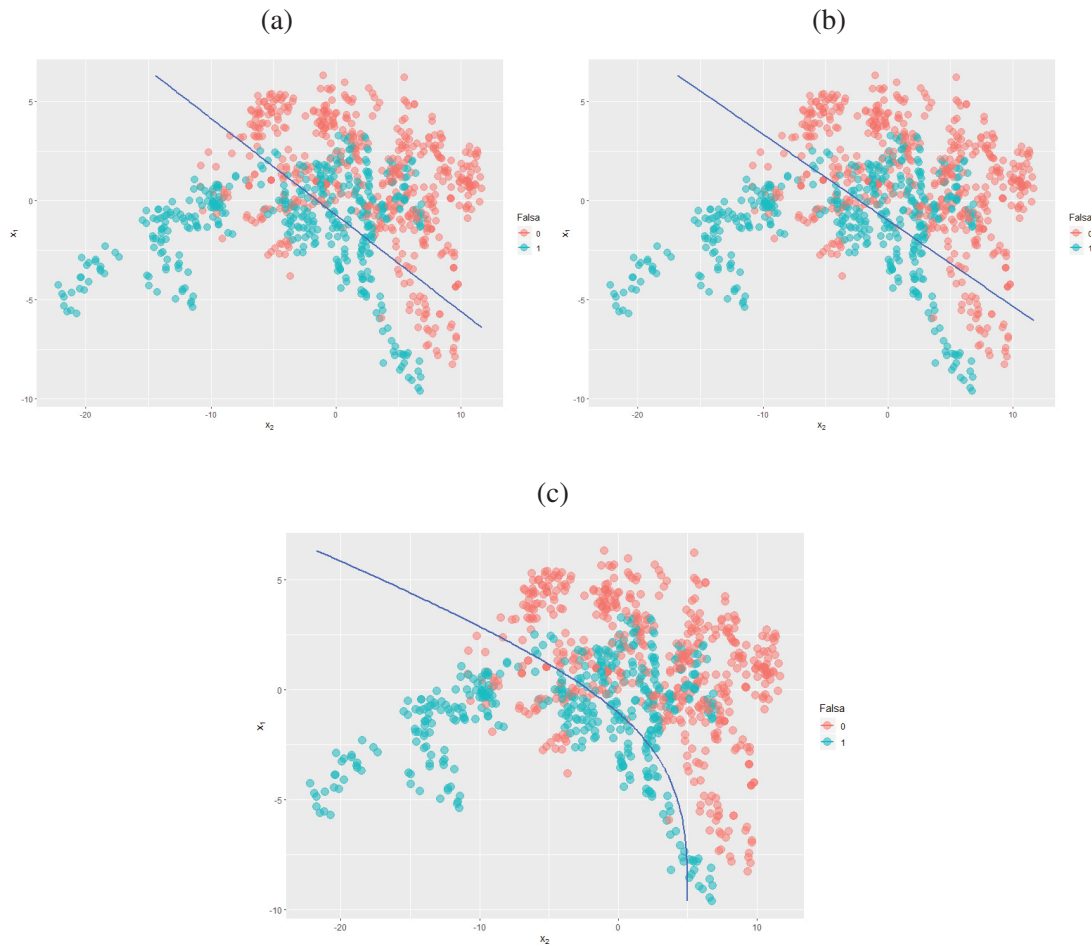
$$\boldsymbol{\Omega} = \tau_0 \mathbf{I}. \quad (3.29)$$

Tais métodos foram inicialmente propostos na literatura e são atrativos pois possuem estruturas matemática e computacionalmente simples. Entretanto não apresentam grande flexibilidade para separação dos grupos e outros modelos para classificação podem ser adotados. Neste exemplo, foram ajustados diversos modelos de classificação binária que foram comparados conforme a qualidade preditiva, mensurada pela proporção total de acertos nos dados de teste.

Além das análises discriminantes e do modelo logístico comum também foram ajustados SVM, GAM, árvore de classificação com resultados que podem ser visualizados na Figura 3.11.

A árvore de classificação retorna uma função de classificação não linear e não suave, conforme ilustrado na Figura 3.11(c), uma vez que este modelo funciona através da decisão

Figura 3.10: Classificações por modelo logístico e análise de discriminante: Modelo de regressão logística em (a), análise discriminante linear de Fisher em (b); análise discriminante quadrático de Fisher em (c).



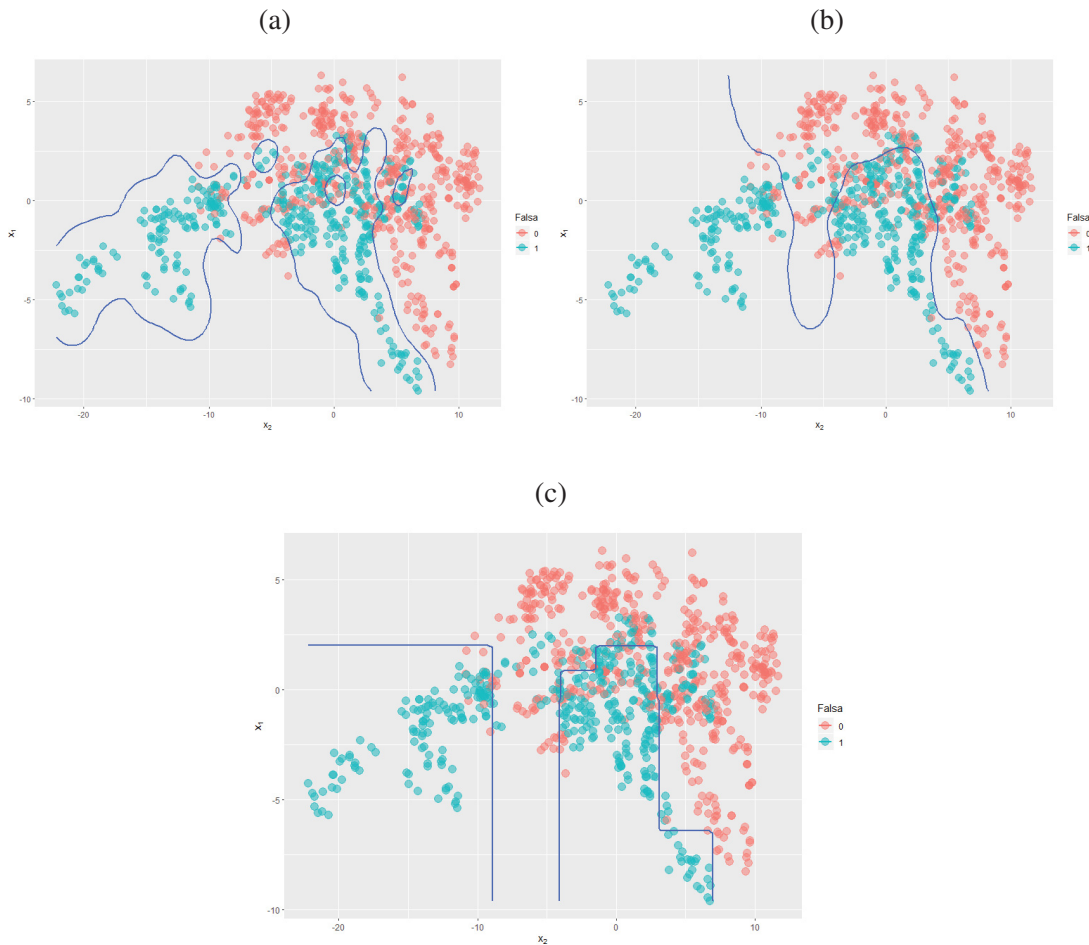
Fonte: Elaborado pelo autor.

sequencial de pontos de quebra das covariáveis. O SVM e o GAM (Figuras 3.11(a) e 3.11(b), respectivamente) também retornam uma função de classificação não linear, sendo que o GAM tende a retornar funções de classificação mais suaves que o SVM.

Finalmente, foram ajustados modelos com a especificação de estrutura na matriz de dependência. Partindo da estrutura do modelo logístico, foram ajustados modelos com especificação da dependência entre as observações a partir de um *kernel* \mathbf{K} na equação 3.29, em que $\mathbf{\Omega} = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}$. Modelos foram ajustados na presença ($\sim x$) ou ausência (~ 1) das covariáveis no componente de média do modelo. Desta forma foi possível obter funções de discriminação não-lineares, sejam funções não suaves, como a obtida pelo modelo com *kernel* via k-médias (Figura 3.13(c)) e com *kernel* exponencial (Figura 3.13(a)), ou funções suaves, como a obtida pelo modelo com *kernel* gaussiano (Figura 3.13(b)). Vale destacar que a adição das covariáveis no componente de média do modelo suaviza a função de discriminação, por exemplo como visto no modelo ajustado com *kernel* exponencial e adição das covariáveis no componente de média (Figura 3.12(a)). A Figura 3.12 ilustra ajustes obtidos com a especificação de algumas possíveis estruturas de dependência.

Na Tabela 3.5 estão as informações sobre o tempo de processamento computacional em segundos (TP) e a proporção total de acertos (PTA), esta última é uma medida de qualidade

Figura 3.11: Classificações por SVM, GAM e árvore de classificação: SVM com base radial em (a), GAM em (b), árvore de classificação em (c).



Fonte: Elaborado pelo autor.

preditiva que consiste em $PTA = (T_{11} + T_{22})/N$, onde N é o total de observações na amostra e T_{11} e T_{22} são contagens da tabela de contingência pela Tabela 3.4. E na Tabela 3.4, o vetor Y representa são os valores observados e o vetor Y^* representa os valores preditos pela equação da predição condicional conforme 2.19.

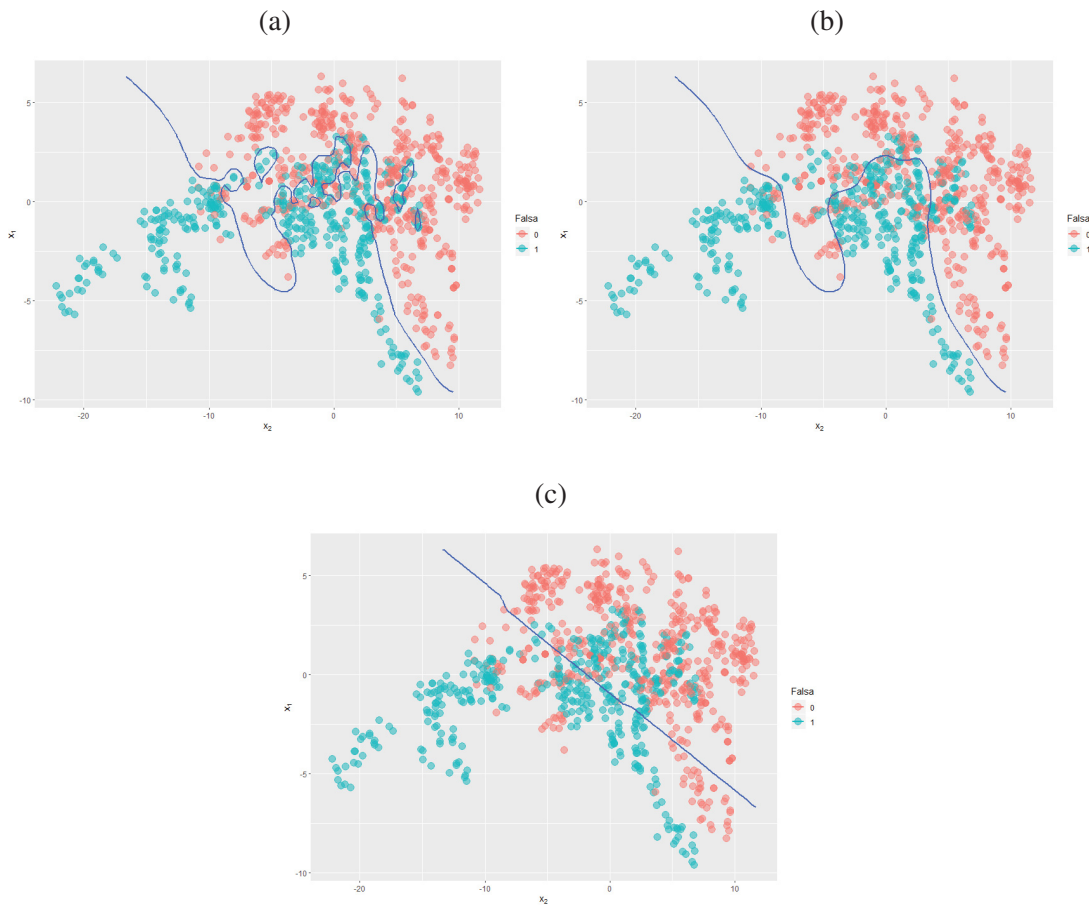
Tabela 3.4: Relação dos dados observados Y com os preditos Y^* .

		Y^*	
		0	1
Y	0	T_{11}	T_{12}
	1	T_{21}	T_{22}

Fonte: Elaborado pelo autor.

Conforme a Tabela 3.5, entre os modelos da classe dos MCGLM, o que obteve melhor performance preditiva considerando a proporção total de acertos, foi o modelo com a covariância modelada pela combinação linear de um *kernel* exponencial mais um *kernel* via K-médias e com as covariáveis inclusas no componente de média. Com exceção deste modelo, a incorporação das

Figura 3.12: Classificações por modelos logísticos com diferentes estruturas de dependência e com covariáveis no componente de média: Ajuste com *kernel* exponencial em (a), ajuste com *kernel* gaussiano em (b), ajuste com *kernel* k-médias em (c).



Fonte: Elaborado pelo autor.

covariáveis no componente de média dos modelos não trouxe nenhum ganho em performance preditiva.

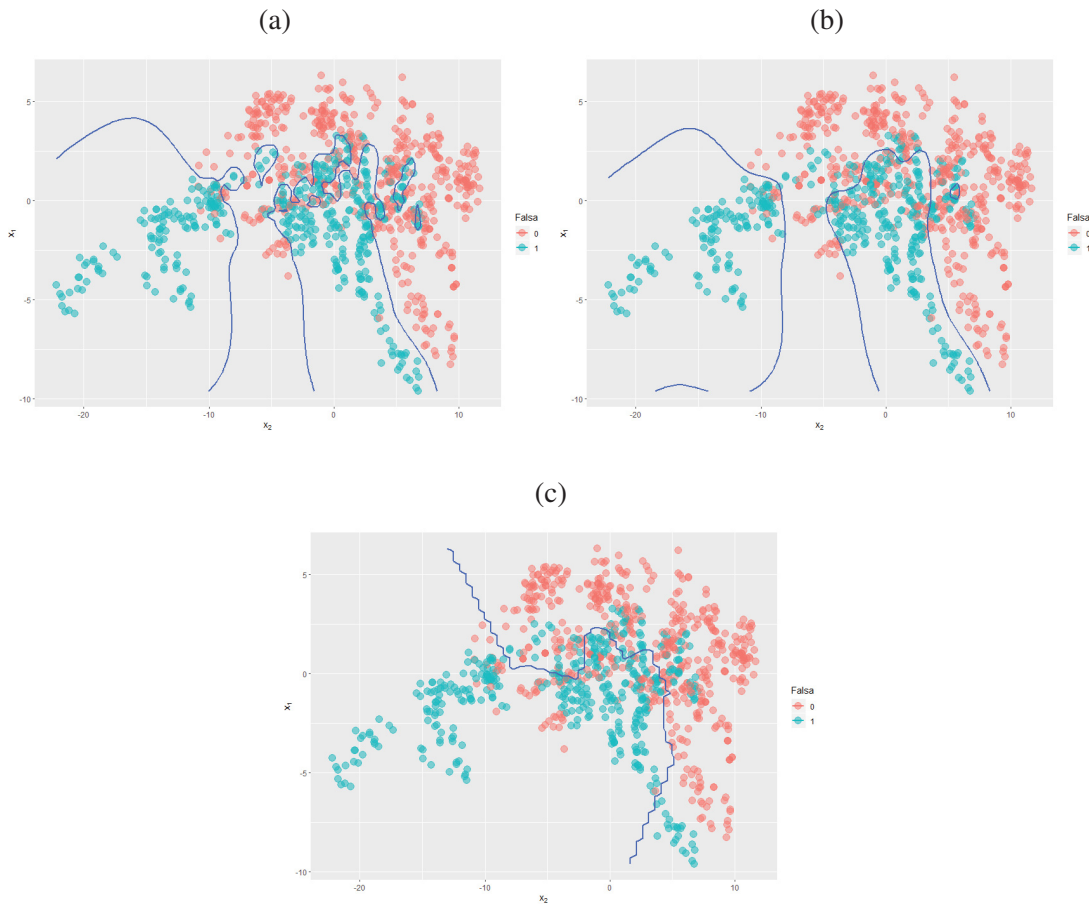
Dos modelos com um único *kernel*, o melhor desempenho preditivo foi o dos modelos com *kernel* exponencial, com e sem as covariáveis no componente de média. Desta forma, indica-se para predição, o modelo sem covariáveis no componente de média, como o de menor número de parâmetros. Entre modelos que não são da classe do MCGLM, o SVM retornou a melhor performance preditiva.

3.4 TRATAMENTO DE REABILITAÇÃO DA FALA

O conjunto de dados (`lsvt`) foi disponibilizado por VANSCHOREN et al. (2013), e contém dados de avaliações do tratamento de reabilitação da fala em pessoas diagnosticadas com doença de Parkinson. Os indivíduos foram submetidos a testes de fonemas e classificados como "aceitável" ou "inaceitável".

Uma particularidade deste conjunto de dados é a de que são 126 indivíduos avaliados em 308 covariáveis, além da variável resposta. Uma das vantagens de inserir as covariáveis no componente de covariância é o fato de estimar um menor número de parâmetros, podendo até

Figura 3.13: Classificações por modelos logísticos com diferentes estruturas de dependência e com média constante: Ajuste com *kernel* exponencial em (a), ajuste com *kernel* gaussiano em (b), ajuste com *kernel* k-médias em (c).



Fonte: Elaborado pelo autor.

ser um único associado à matriz de dependência, independentemente da dimensão do conjunto de covariáveis. Neste caso específico, o número de covariáveis é maior que o número de observações, o que torna inviável incluir as covariáveis no componente de média de modelo paramétrico. Sendo assim, este conjunto de dados tem como principal objetivo exemplificar a modelagem de dados em que o número de covariáveis é superior ao número de observações, comparando as performances preditivas dos modelos considerados.

Os modelos ajustados pela classe dos MCGLM seguem a estrutura:

$$E(\mathbf{Y}) = \boldsymbol{\mu}_Y = \frac{\exp(\boldsymbol{\beta}_0)}{\exp(\boldsymbol{\beta}_0) + 1} \quad (3.30)$$

$$V(\mathbf{Y}) = \nu(\boldsymbol{\mu}_Y)^{1/2} \boldsymbol{\Omega} \nu(\boldsymbol{\mu}_Y)^{1/2} \quad (3.31)$$

$$\boldsymbol{\Omega} = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}, \quad (3.32)$$

com $\nu(\boldsymbol{\mu}) = \text{diag}(\boldsymbol{\mu}(1 - \boldsymbol{\mu}))$.

Da equação 3.31, a matriz \mathbf{K} é o *kernel* especificado como gaussiano, exponencial e de k-médias.

Tabela 3.5: Proporção total de acertos nos dados de teste (PTA), área sob a curva ROC (AUC) e Tempo de processamento (TP) em segundos.

Classe	Média	<i>kernel</i>	TP	AUC	PTA
MCGLM	$\sim x$	–	7,81	0,844	0,771
MCGLM	~ 1	gaussiano	349,14	0,946	0,866
MCGLM	$\sim x$	gaussiano	423,12	0,943	0,859
MCGLM	~ 1	exponencial	301,89	0,961	0,886
MCGLM	$\sim x$	exponencial	366,44	0,961	0,886
MCGLM	~ 1	k-médias	37,04	0,871	0,779
MCGLM	$\sim x$	k-médias	70,07	0,847	0,771
MCGLM	~ 1	gaussiano+k-médias	425,31	0,943	0,859
MCGLM	~ 1	exponencial+k-médias	744,77	0,961	0,886
MCGLM	$\sim x$	gaussiano+k-médias	425,86	0,946	0,852
MCGLM	$\sim x$	exponencial+k-médias	971,66	0,961	0,895
GAM	$\sim x + s(x)$	–	1,64	0,919	0,847
SVM	–	–	3,08	0,933	0,886
ADL	–	–	0,44	0,844	0,767
ADQ	–	–	0,20	0,858	0,791
Árvore	–	–	0,60	0,888	0,788

Fonte: Elaborado pelo autor.

A matriz de dispersão também foi especificada como combinação de um *kernel* discreto \mathbf{K}_1 e um *kernel* contínuo \mathbf{K}_2 , $\mathbf{\Omega} = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}_1 + \tau_2 \mathbf{K}_2$, mantendo as especificações conforme as equações 3.30 e 3.31.

Também foram ajustados o SVM, modelo Lasso e modelo Ridge pelo pacote `glmnet` (Simon et al., 2011), os quais também podem ser escritos como um processo gaussiano (Rasmussen e Williams, 2006, Cap. 6).

Tabela 3.6: Proporção total de acertos nos dados de teste (PTA), área sob a curva ROC (AUC) e tempo de processamento (TP) para cada um dos modelos ajustados.

Classe	<i>kernel</i>	TP	AUC	PTA
MCGLM	gaussiano	6,78	0,854	0,873
MCGLM	exponencial	2,81	0,861	0,857
MCGLM	k-médias	60,24	0,748	0,761
MCGLM	gaussiano+k-médias	60,67	0,851	0,841
MCGLM	exponencial+k-médias	43,64	0,839	0,888
SVM	–	0,45	0,893	0,841
Lasso	–	0,71	0,883	0,857
Ridge	–	0,28	0,832	0,841

Fonte: Elaborado pelo autor.

O conjunto de dados foi dividido em 50% de dados de treino e 50% dados de teste. Foram computadas a área sob a curva (AUC) e a proporção total de acertos (PTA) como medidas de qualidade preditivas e disponibilizadas na Tabela 3.6.

Os resultados apresentados pela Tabela 3.6 mostram que as performances preditivas, segundo a proporção total de acertos, foram bastante próximas. O MCGLM com combinação do *kernel* via k-médias e do *kernel* exponencial resultou na melhor proporção de acertos.

3.5 GASTOS POLICIAIS NO ESTADO DO MISSISSIPI

O primeiro, entre os três exemplos de dados de área utilizados aqui, tem como variável resposta os gastos *per capita* em policiamento em 82 localidades do estado de Mississippi, além de dez covariáveis. Os dados foram analisados por Kelijian e Robinson (1992) e estão disponíveis em Justen (2018).

Pelo mapa da Figura 3.14, os gastos com policiamento per capita aparentam ter regiões de concentração de valores mais elevados na borda oeste do estado, mas não um padrão espacial muito nítido. Possivelmente este padrão pode ser descrito ao adicionar as covariáveis no componente de média do modelo. Por outro lado, essas mesmas covariáveis também podem ser inseridas no componente de covariância do modelo, conforme alguma das especificações de \mathbf{K} discutidas na seção 2.3. O objetivo aqui é explorar a relação da variável resposta com as covariáveis, inserindo as mesmas no componente de média e/ou no componente de covariância do modelo, através da especificação de um *kernel*.

Seja um MCGLM especificando:

$$E(Y) = \boldsymbol{\mu}_Y \quad (3.33)$$

$$V(Y) = \boldsymbol{\Omega}(\boldsymbol{\tau}, \mathbf{Z}). \quad (3.34)$$

O componente de média, equação 3.33, foi especificado como constante, $\boldsymbol{\mu}_Y = \beta_0$, ou como combinação linear das covariáveis, $\boldsymbol{\mu}_Y = \mathbf{X}^* \boldsymbol{\beta}$. Da mesma forma, o componente de covariância, equação 3.34, foi especificado sob independência das observações:

$$\boldsymbol{\Omega}(\boldsymbol{\tau}, \mathbf{Z}) = \tau_0 \mathbf{I}. \quad (3.35)$$

O componente de covariância também foi especificado inserindo uma matriz de dependência \mathbf{K} , *kernel*, que pode ser no espaço geográfico ou no espaço das covariáveis:

$$\boldsymbol{\Omega}(\boldsymbol{\tau}, \mathbf{Z}) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}. \quad (3.36)$$

Também foram consideradas combinações do *kernel* definido no espaço das covariáveis \mathbf{K}_x e outro no espaço geográfico \mathbf{K}_{sp} :

$$\boldsymbol{\Omega}(\boldsymbol{\tau}, \mathbf{Z}) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x + \tau_2 \mathbf{K}_{sp}. \quad (3.37)$$

Nas Figuras 3.15 e 3.16 podem ser encontradas as distribuições de cada uma das variáveis e a relação entre as mesmas, com a resposta representada pela variável `POLICE` no gráfico. Pelos gráficos, com exceção das covariáveis *INC*, *OWN* e *WHITE*, as variáveis apresentam forte assimetria à direita. Além disso, existe um ponto atípico na metade superior nos gráficos de dispersão, que pode vir a ser um ponto de alavancagem. Desta forma, foi realizada transformação *log* nas variáveis com forte assimetria à direita, incluindo a variável resposta.

Primeiramente, foi feita uma seleção das covariáveis para o componente de média dos modelos pelo critério de informação *escore*. Sendo selecionadas as covariáveis *LogPOP*, *OWN*, *LogTAX* e a longitude *X2* para permanecerem na média dos modelos. Para estrutura de dependência, os *kernels* foram definidos no espaço de todas as covariáveis.

As medidas de qualidade de ajuste, *pAIC* e *pBIC*, disponibilizadas na Tabela 3.7, apontaram para seleção dos modelos com as covariáveis incluídas no componente de média.

Ainda segundo as medidas apresentadas na Tabela 3.7, dentre os modelos com covariância estritamente espacial, o melhor ajuste, tanto em *pBIC* quanto em *pAIC*, foi aquele com *kernel*

autorregressivo de primeira ordem. Dentre os modelos com covariância estritamente especificada pelas covariáveis, o melhor também foi aquele com *kernel* autorregressivo de primeira ordem. Sendo que por pAIC, o melhor modelo geral foi justamente a combinação de ambos os *kernels*, evidenciando de que o *kernel* autorregressivo no espaço das covariáveis capta um padrão não detectado pelo componente espacial.

Na Figura 3.17 encontram-se mapas com legenda de cores de valores ajustados por modelos. As classes da legenda foram obtidas dividindo os valores observados da variável resposta nos quantis de 20%, 40%, 60%, 80% e 100%, Figura 3.17(a), sendo que a mesma legenda para os dados observados foi utilizada nos demais mapas com finalidade de facilitar a comparação. Além disso, os valores entre parenteses nas legendas são o número de observações na respectiva classe.

Os ajustes obtidos pelo modelo com dependência espacial, Figura 3.17(b), apresentaram concentração maior de observações na classe central, de 5,9 a 6,3, se comparados com os ajustes obtidos pelos modelos com dependência no espaço das covariáveis, Figuras 3.17(c) e 3.17(d). Ou seja, o modelo com dependência espacial resultou em ajustes mais suavizados que os ajustes dos modelos com dependência no espaço de covariáveis.

A Tabela 3.8 mostra as estimativas dos parâmetros associados a cada matriz de dependência dos modelos. Comparando os modelos sob independência das observações, o modelo com as covariáveis na média tem $\hat{\tau}_0 = 0,872$, enquanto o modelo com média constante teve $\hat{\tau}_0 = 0,026$. As covariáveis na média expressam aproximadamente 97% da dispersão. Dos modelos com *kernel* exponencial nas coordenadas geográficas, o modelo com média constante teve uma proporção de aproximadamente 39% da dispersão para o termo independente e 61% para a dependência espacial, enquanto o modelo com as covariáveis na média obtiveram aproximadamente 42% da dispersão no termo independente e 58% da dispersão explicada pela dependência espacial. Do modelo com *kernel* exponencial no espaço das covariáveis, o modelo com média constante resultou em 13,2% da dispersão no termo independente e 86,8% da dispersão explicada pela dependência no espaço das covariáveis. Por outro lado, os modelos com *kernel* autorregressivo apresentaram resultados bem mais modestos, com 7% da dispersão explicada para o caso da dependência espacial e 8,8% da dispersão explicada para o caso da dependência no espaço das covariáveis.

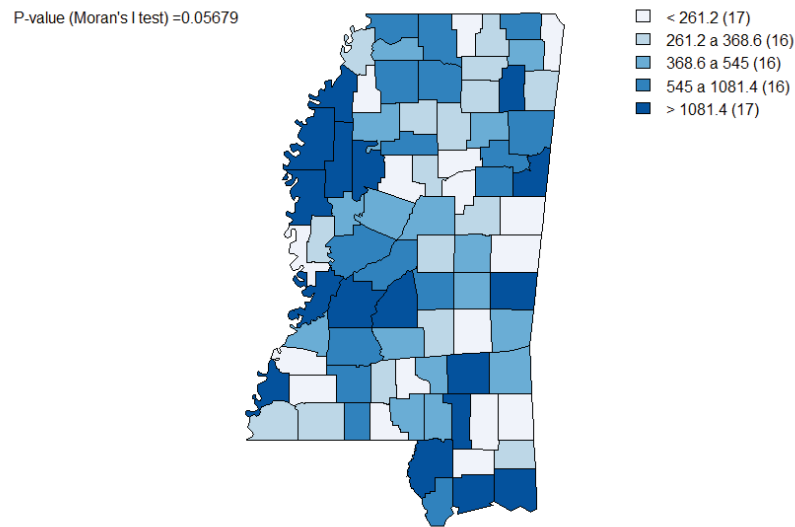
Ainda na Tabela 3.8, o modelo com média constante e combinação de uma dependência espacial com outra no espaço das covariáveis, ambas as dependências com *kernel* exponencial, obteve uma proporção de aproximadamente 28,1% da dispersão para o termo independente, matriz não estruturada, 34,2% para a dependência espacial e 37,5% para a dependência no espaço das covariáveis. Já o modelo com média constante e combinação de uma dependência espacial com outra no espaço das covariáveis, ambas as dependências com *kernel* autorregressiva, obteve uma proporção de aproximadamente 91,1% da dispersão para o termo independente, matriz não estruturada, 0,02% para a dependência espacial e 8,8% para a dependência no espaço das covariáveis.

3.6 CASOS DE LEUCEMIA NO ESTADO DE NOVA YORK

O `nydata` é um conjunto de dados com estrutura espacial de áreas do pacote `spData` (Bivand et al., 2019), e tem como resposta taxas de prevalência de leucemia nas 281 divisões do Estado de Nova York no período 1978-1982, com a população do censo de 1980.

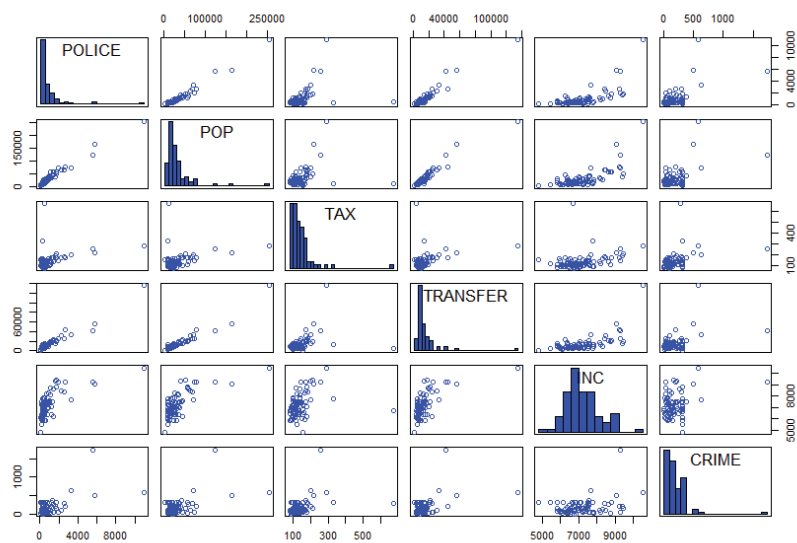
As observações do vetor de resposta Y são contagens dos casos de leucemia, porém são uma estimativa grosseira a partir da população do censo de 1980 e não as contagens originais. O

Figura 3.14: Gastos em (log)dólares com policiamento por habitante no estado do Mississippi.



Fonte: Elaborado pelo autor.

Figura 3.15: Dispersão da variável resposta com as cinco primeiras covariáveis.



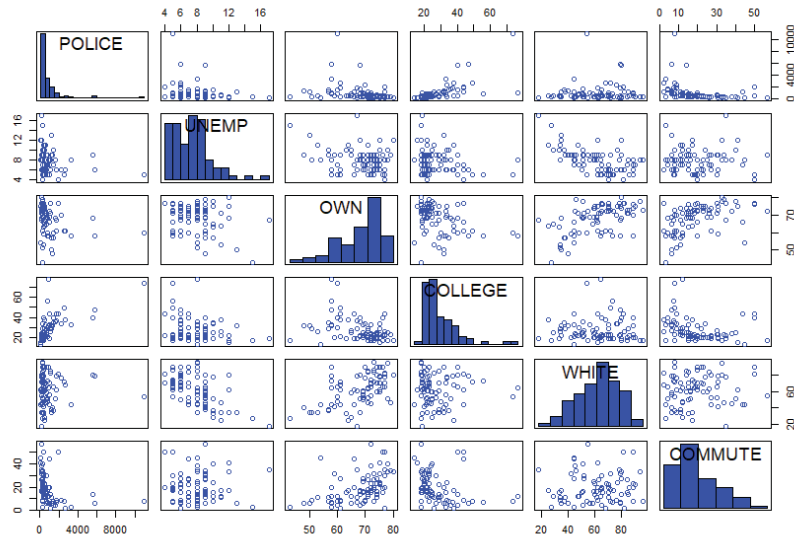
Fonte: Elaborado pelo autor.

conjunto de dados também fornece a variável Z que são as proporções transformadas, a qual é nossa variável resposta modelada por uma distribuição gaussiana.

Assim como no exemplo anterior, o principal objetivo aqui é encontrar o melhor modelo, conforme os critérios de qualidade de ajuste, inserindo as covariáveis no componente de média e/ou de covariância, e combinar as dependências no espaço das covariáveis e espacial.

Pelo mapa da Figura 3.18, aparentemente a variável resposta apresenta algum padrão espacial, porém, por haver uma grande variação entre os tamanhos geográficos das áreas, torna difícil uma conclusão apenas por visualização. O teste I de Moran por permutação (Bivand

Figura 3.16: Dispersão da variável resposta com as cinco últimas covariáveis.



Fonte: Elaborado pelo autor.

Tabela 3.7: Modelos ajustados ordenados pelo pAIC.

<i>kernel</i>	$\Omega(\tau, \mathbf{Z})$	Média	pLogLik	pAIC	pBIC
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p} + \tau_2 \mathbf{K}_x$	$\sim x$	36.21	-56.42	-37.17
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	34.14	-54.28	-37.43
exponencial+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p} + \tau_2 \mathbf{K}_x$	$\sim x$	34.97	-53.94	-34.69
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	32.37	-50.74	-33.89
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p}$	$\sim x$	31.81	-49.62	-32.77
-	$\tau_0 \mathbf{I}$	$\sim x$	30.18	-48.36	-33.92
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p}$	$\sim x$	30.80	-47.60	-30.75
exponencial+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p} + \tau_2 \mathbf{K}_x$	~ 1	-53.33	114.66	124.29
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p} + \tau_2 \mathbf{K}_x$	~ 1	-54.29	116.58	126.21
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	-59.66	125.32	132.54
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	-84.25	174.50	181.72
exponencial+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p} + \tau_2 \mathbf{K}_x$	~ 1	-84.26	176.52	186.15
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p} + \tau_2 \mathbf{K}_x$	~ 1	-84.27	176.54	186.17
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p}$	~ 1	-109.92	225.84	233.06
-	$\tau_0 \mathbf{I}$	~ 1	-111.28	226.56	231.37
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{s_p}$	~ 1	-110.98	227.96	235.18

Fonte: Elaborado pelo autor.

e Piras, 2015; Schabenberger e Gotway, 2009), no canto superior à direita, resultou em um p -valor < 0.05 , adotando um nível de significância de 5%. Desta forma conclui-se que há uma dependência espacial que pode ser absorvida tanto pelo componente de média quanto pelo componente de covariância.

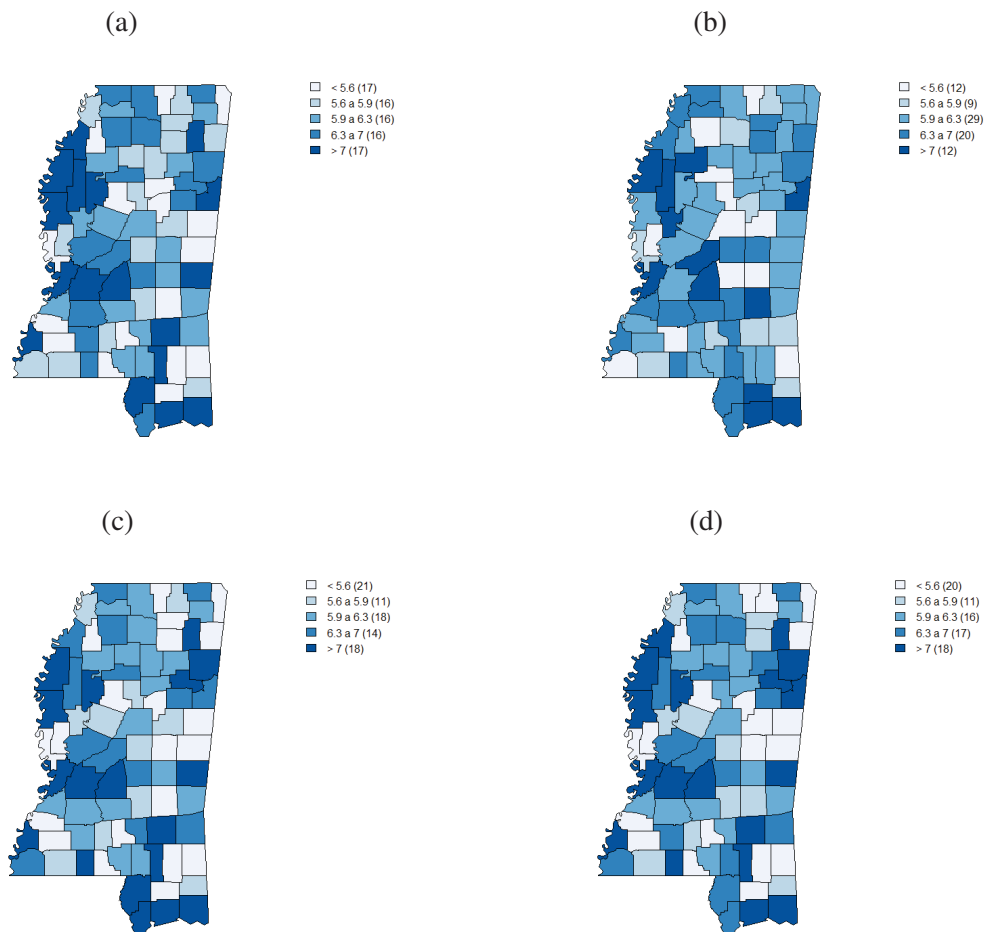
Foi realizada seleção de covariáveis para o componente de média dos modelos pelo critério de informação score, das quais permaneceram a *PCTAGE65P* e *PCTOWNHOME*.

Os resultados da Tabela 3.9, mostram que as covariáveis na média dos modelos trazem ganhos na qualidade de ajuste, conforme os critérios pAIC e pBIC. Ainda na Tabela 3.7, dentre os modelos com covariância estritamente espacial, o melhor ajuste, tanto em pBIC quanto em pAIC, foi aquele com *kernel* exponencial de primeira ordem. Dentre os modelos com covariância

Tabela 3.8: Estimativas dos parâmetros τ s para os modelos com função de ligação $h(\cdot)$ identidade.

<i>kernel</i>	$\Omega(\tau, \mathbf{Z})$	Média	τ_0	τ_1	τ_2
-	$\tau_0 \mathbf{I}$	~ 1	0,872	-	-
-	$\tau_0 \mathbf{I}$	$\sim x$	0,026	-	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	~ 1	-1,598	2,469	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	$\sim x$	-0,069	0,095	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	~ 1	1,486	-0,113	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	0,193	1,265	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	1,285	-0,125	-
exponencial+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	-0,945	1,151	1,26
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	0,362	-0,026	1,348
exponencial+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	-0,216	1,522	-0,126
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	1,287	$-3,66 \times 10^{-4}$	-0,125

Fonte: Elaborado pelo autor.

Figura 3.17: Mapa com legenda de cores para os valores ajustados pelos melhores modelos, com as covariáveis no componente de média e estrutura de dependência, segundo o pAIC: Valores observados em (a), modelo com *kernel* autorregressiva no espaço geográfico em (b), modelo com *kernel* autorregressiva no espaço das covariáveis em (c), modelo com *kernel* autorregressiva no espaço geográfico+*kernel* autorregressivo no espaço das covariáveis em (d).

Fonte: Elaborado pelo autor.

estritamente especificada pelas covariáveis, o melhor também foi aquele com *kernel* exponencial. Sendo que por pAIC, o melhor modelo geral foi a combinação de ambos os *kernels*. Já o melhor modelo por pBIC, foi com as covariáveis no componente de média e especificação do componente de covariância sob independência das observações, ou seja, o modelo linear normal. De forma geral, não houveram muitas diferenças entre as medidas de qualidades de ajustes dos modelos com diferentes especificações do componente de covariância.

Como os modelos foram especificados sob normalidade, isso é, função de variância $v(\mu_Y) = \mu_Y^0$, então a pLogLik coincide com a verossimilhança. Além disso, todos os modelos possuem verossimilhanças diretamente comparáveis pois foram ajustados com mesma função de ligação $h()$ da covariância. Desta forma é possível realizar testes da razão de verossimilhança com estatísticas de teste do tipo $2(pLogLike_1 - pLogLike_0)$ comparando modelos encaixados. Comparando os modelos com as covariáveis na média, na tabela 3.9, percebe-se que a adição de uma matriz dependência espacial, seja através de uma *kernel* do tipo autorregressiva ou contínua, exponencial, não apresenta nenhum ganho significativo em relação ao modelo linear normal com observações independentes. Já dentre os modelos com dependência especificada no espaço das covariáveis, os dois modelos com covariâncias contínuas, *kernel* exponencial, com e sem covariáveis no componente de média, foram melhores em relação ao NLM com observações independentes, tanto em razão de verossimilhanças quanto em pAIC.

A Tabela 3.10 mostra as estimativas dos parâmetros associados a cada matriz de dependência dos modelos. Comparando os modelos sob independência das observações, o modelo com as covariáveis na média tem $\hat{\tau}_0 = 0,424$, o que dá 0,1 a menos comparado ao modelo com média constante. As covariáveis na média exprimem aproximadamente 19% da dispersão. Dos modelos com uma matriz \mathbf{K} no componente de covariância, o *kernel* exponencial expressou mais da dispersão do que o *kernel* autorregressivo. Como exemplo, dos modelos com média constante e dependência espacial, o *kernel* exponencial explicou aproximadamente 34% da dispersão, enquanto o *kernel* autorregressivo explicou aproximadamente 3,5%⁴ da dispersão.

Ainda na Tabela 3.10, o modelo com média constante e combinação de uma dependência espacial com outra no espaço das covariáveis, ambas as dependências com *kernel* exponencial, obteve uma proporção de aproximadamente 44,3% da dispersão para o termo independente, 24,2% para a dependência espacial e 31,3% para a dependência no espaço das covariáveis. Já o modelo com média constante e combinação de uma dependência espacial com outra no espaço das covariáveis, ambas as dependências com *kernel* autorregressiva, obteve uma proporção de aproximadamente 93,3% da dispersão para o termo independente, matriz não estruturada, 1,4% para a dependência espacial e 5,2% para a dependência no espaço das covariáveis.

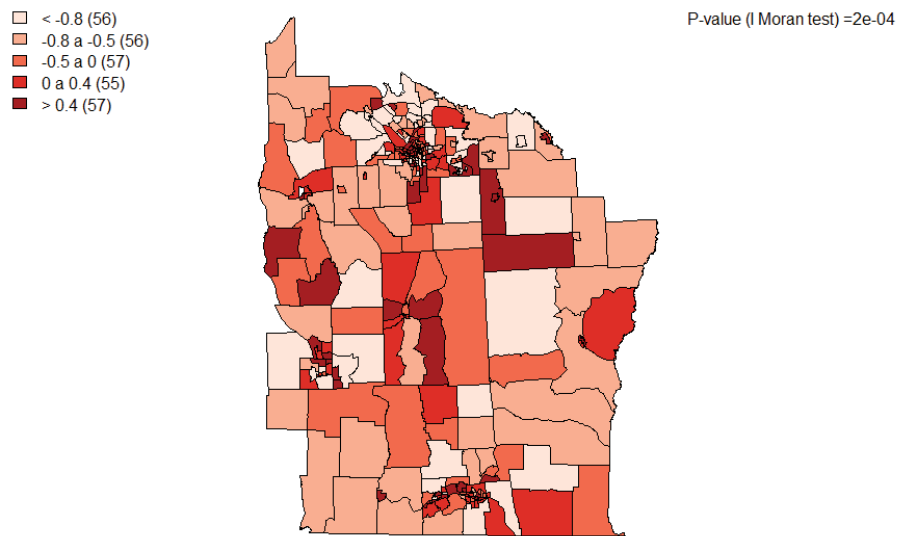
Na Figura 3.19 estão os mapas com legendas de cores dos valores ajustados de três modelos, e mostram que a suavização obtida, em relação aos valores observados da Figura 3.19(a), pelo modelo com dependência espacial. A Figura 3.19(b) é similar com à dos modelos com dependência no espaço das covariáveis, Figura 3.19(c), e combinação entre a dependência espacial e no espaço das covariáveis, Figura 3.19(d).

3.7 TAXAS DE MORTALIDADE INFANTIL NOS MUNICÍPIOS DO ESTADO DO PARANÁ

Este terceiro conjunto de dados espaciais tem como resposta contagens de óbitos infantis nos municípios do estado do Paraná, no período de 2008-2012, e como *offset* o número de nascidos vivos (Figura 3.20) no mesmo período, além de um conjunto de covariáveis socioeconômicas referentes ao ano de 2010.

⁴O modelo estimou $\hat{\tau}_0 = 0,661$ e $\hat{\tau}_1 = -0,024$, ou seja, um contraste da dispersão da matriz não estruturada com a dependência espacial.

Figura 3.18: Taxas transformadas de prevalência de leucemia no estado de Nova York no período de 1978-1982.



Fonte: Elaborado pelo autor.

Tabela 3.9: Modelos ajustados ordenados pelo pAIC.

<i>kernel</i>	$\Omega(\tau, \mathbf{Z})$	Média	pLogLik	pAIC	pBIC
exponencial+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP} + \tau_2 \mathbf{K}_x$	$\sim x$	-275.83	563.66	585.49
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	-276.92	563.84	582.03
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP} + \tau_2 \mathbf{K}_x$	$\sim x$	-276.92	565.84	587.67
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP}$	$\sim x$	-278.30	566.60	584.79
-	$\tau_0 \mathbf{I}$	$\sim x$	-279.72	567.44	581.99
exponencial+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP} + \tau_2 \mathbf{K}_x$	$\sim x$	-277.99	567.98	589.81
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	-279.25	568.50	586.69
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP}$	$\sim x$	-279.73	569.46	587.65
autorregressiva+Voronoi	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP} + \tau_2 \mathbf{K}_x$	$\sim x$	-279.25	570.50	592.33
exponencial+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP} + \tau_2 \mathbf{K}_x$	~ 1	-294.32	596.64	611.19
exponencial+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP} + \tau_2 \mathbf{K}_x$	~ 1	-294.41	596.82	611.37
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	-298.23	602.46	613.38
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP}$	~ 1	-298.26	602.52	613.44
K-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	-298.84	603.68	614.60
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP} + \tau_2 \mathbf{K}_x$	~ 1	-298.07	604.14	618.69
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	-300.38	606.76	617.68
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP} + \tau_2 \mathbf{K}_x$	~ 1	-299.86	607.72	622.27
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{SP}$	~ 1	-306.47	618.94	629.86
-	$\tau_0 \mathbf{I}$	~ 1	308.89	621.78	629.06

Fonte: Elaborado pelo autor.

A variável resposta, por ser de contagem, é assumida seguir uma distribuição de Poisson. Além disso, pelo mapa da Figura 3.20, a resposta aparenta ter algum padrão espacial, com uma concentração de valores mais elevados ao centro do mapa.

Tabela 3.10: Estimativas dos parâmetros τ s para os modelos com função de ligação $h()$ identidade.

<i>kernel</i>	$\Omega(\tau, \mathbf{Z})$	Média	τ_0	τ_1	τ_2
-	$\tau_0 \mathbf{I}$	~ 1	0,525	-	-
-	$\tau_0 \mathbf{I}$	$\sim x$	0,424	-	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	~ 1	0,331	0,172	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	$\sim x$	0,353	0,068	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	~ 1	0,661	-0,024	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	$\sim x$	0,415	0,001	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	0,308	0,240	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	-0,211	0,639	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	0,897	-0,062	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	0,495	-0,012	-
exponencial+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	0,232	0,127	0,164
exponencial+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-0,267	0,059	0,636
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	0,340	-0,003	0,226
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-0,219	0,002	0,635
exponencial+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	0,592	0,143	-0,039
exponencial+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	0,414	0,066	-0,009
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	0,894	-0,014	-0,050
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	0,484	0,002	-0,012

Fonte: Elaborado pelo autor.

Da mesma forma que os exemplos anteriores, foram ajustados modelos com covariáveis na média:

$$E(Y) = \boldsymbol{\mu}_Y = \exp(\mathbf{X}^* \boldsymbol{\beta}) + \text{offset} \quad (3.38)$$

$$V(Y) = \text{diag}(\boldsymbol{\mu}_Y)^{1/2} \boldsymbol{\Omega}(\tau, \mathbf{Z}) \text{diag}(\boldsymbol{\mu}_Y)^{1/2}, \quad (3.39)$$

ou com média constante:

$$E(Y) = \boldsymbol{\mu}_Y = \exp(\beta_0) + \text{offset} \quad (3.40)$$

$$V(Y) = \text{diag}(\boldsymbol{\mu}_Y)^{1/2} \boldsymbol{\Omega}(\tau, \mathbf{Z}) \text{diag}(\boldsymbol{\mu}_Y)^{1/2}. \quad (3.41)$$

Pelo modelo especificado nas equações 3.38 e 3.39 foi feita seleção de covariáveis para o componente de média, pelo critério de informação score, na qual a covariável *trabinf* permaneceu.

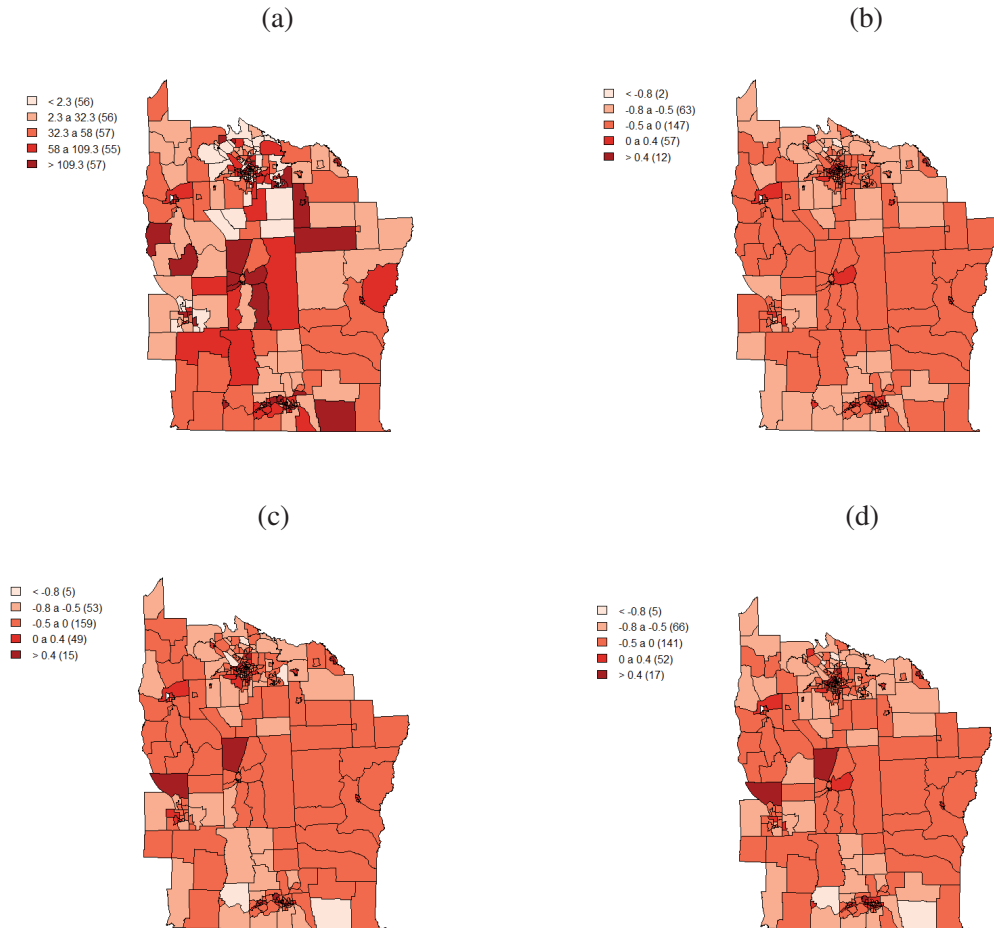
A matriz de dispersão $\boldsymbol{\Omega}(\tau, \mathbf{Z})$ foi especificada de três formas: sob independência das observações, $\boldsymbol{\Omega}(\tau, \mathbf{Z}) = \tau_0 \mathbf{I}$, por uma matriz de dependência espacial $\boldsymbol{\Omega}(\tau, \mathbf{Z}) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$, por uma matriz de dependência no espaço das covariáveis $\boldsymbol{\Omega}(\tau, \mathbf{Z}) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$, ou a combinação das duas últimas, $\boldsymbol{\Omega}(\tau, \mathbf{Z}) = \tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x + \tau_2 \mathbf{K}_{sp}$. Tanto \mathbf{K}_x quanto \mathbf{K}_{sp} foram especificados pelos *kernel* exponencial, gaussiano, via k-médias e autorregressivo.

Na Tabela 3.11 encontram-se os resultados dos modelos ajustados. Pelo pAIC, o modelo com *kernel* k-médias nas covariáveis e a covariável *trabinf* no componente de média resultou no melhor ajuste. E pelo pBIC, o modelo com *kernel* k-médias nas covariáveis e componente de média especificado como constante resultou no melhor ajuste.

Ainda pela Tabela 3.11, dentre as diferentes especificações da dependência espacial no componente de covariância, o *kernel* exponencial apresentou o melhor resultado em pAIC e em pBIC. Quanto às especificações da dependência no espaço gerado pelo conjunto de covariáveis, o *kernel* via k-médias apresentou melhores resultados de qualidade de ajuste, por pAIC e pBIC.

A Tabela 3.12 mostra as estimativas dos parâmetros associados a cada matriz de dependência dos modelos. Comparando os modelos sob independência das observações, o modelo

Figura 3.19: Mapa com legenda de cores para os valores ajustados pelos melhores modelos, com as covariáveis no componente de média e estrutura de dependência, segundo o pAIC: Valores observados em (a), modelo com *kernel* autorregressiva no espaço geográfico em (b), modelo com *kernel* autorregressiva no espaço das covariáveis em (c), modelo com *kernel* autorregressiva no espaço geográfico+*kernel* autorregressiva no espaço das covariáveis em (d).



Fonte: Elaborado pelo autor.

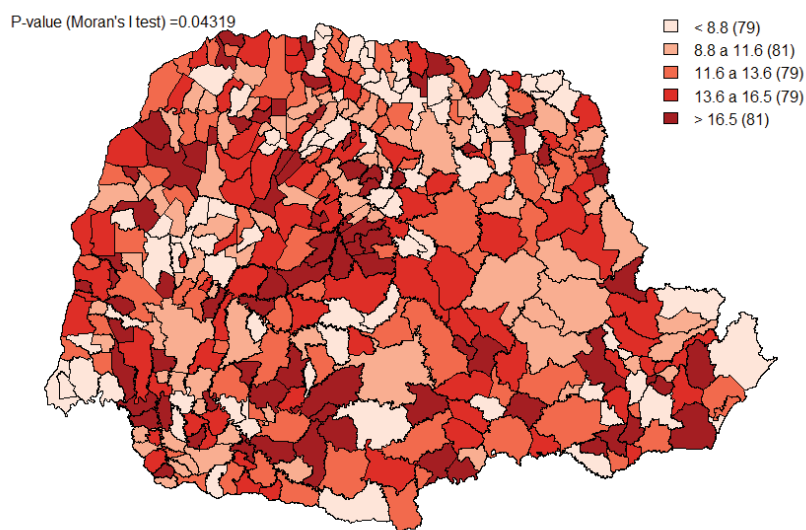
com as covariáveis na média tem $\hat{\tau}_0 = 1,931$, enquanto o modelo com média constante tem $\hat{\tau}_0 = 1,768$. As covariáveis na média exprimem aproximadamente 8,4% da dispersão. Dos modelos com média constante e dependência espacial, o *kernel* exponencial explicou aproximadamente 8,9% da dispersão, enquanto o *kernel* autorregressivo explicou aproximadamente 2,8% da dispersão. Dos modelos com média constante e dependência no espaço das covariáveis, o *kernel* exponencial explicou aproximadamente 15,5% da dispersão, enquanto o *kernel* autorregressivo explicou aproximadamente 1,9% da dispersão.

Ainda na Tabela 3.12, o modelo com média constante e combinação de uma dependência espacial com outra no espaço das covariáveis, ambas as dependências com *kernel* exponencial, obteve uma proporção de aproximadamente 32,5% da dispersão para o termo independente, 1,2% para a dependência espacial e 66,1% para a dependência no espaço das covariáveis. Já modelo com média constante e combinação de uma dependência espacial com outra no espaço das covariáveis, ambas as dependências com *kernel* autorregressiva, obteve uma proporção de aproximadamente 96,2% da dispersão para o termo independente, matriz não estruturada, 2,3% para a dependência espacial e 1,3% para a dependência no espaço das covariáveis. Também, o

modelo com média constante e combinação de uma dependência espacial com *kernel* autorregressivo com outra no espaço das covariáveis com *kernel* via k-médias, obteve uma proporção de aproximadamente 81,5% da dispersão para o termo independente, matriz não estruturada, 1,8% para a dependência espacial e 16,6% para a dependência no espaço das covariáveis.

Na Figura 3.21 encontram-se os mapas dos ajustes dos melhores modelos com, dependência espacial, dependência no espaço das covariáveis e combinação de ambas. Observa-se que os valores ajustados pelo modelo com dependência espacial, Figura 3.21(b), estão concentrados mais na classe central na legenda, de 11,6 a 13,6, em comparação com os valores ajustados pelos modelos com dependência no espaço das covariáveis e com a combinação da dependência espacial e no espaço das covariáveis, na qual resultaram em maior número de observações nas classes adjacentes.

Figura 3.20: Taxas de mortalidade infantil no período de 2008-2012 nos municípios do estado do Paraná.



Fonte: Elaborado pelo autor.

Tabela 3.11: Modelos ajustados ordenados pelo pAIC.

<i>kernel</i>	$\Omega(\tau, \mathbf{Z})$	Média	pLogLik	pAIC	pBIC
k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	-1116.36	2240.72	2256.68
exponencial+k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-1115.37	2240.74	2260.68
k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	-1118.00	2242.00	2253.97
exponencial+k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	-1117.13	2242.26	2258.22
autorregressiva+k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-1116.20	2242.40	2262.34
autorregressiva+k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	-1117.71	2243.42	2259.38
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	$\sim x$	-1119.93	2247.86	2263.82
-	$\tau_0 \mathbf{I}$	$\sim x$	-1121.12	2248.24	2260.21
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	~ 1	-1121.20	2248.40	2260.37
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	-1121.47	2248.94	2260.91
exponencial+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-1119.63	2249.26	2269.20
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	-1120.87	2249.74	2265.70
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	$\sim x$	-1120.90	2249.80	2265.76
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	-1120.91	2249.82	2265.78
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	-1121.08	2250.16	2266.12
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-1120.64	2251.28	2271.22
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-1120.67	2251.34	2271.28
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	~ 1	-1127.43	2260.86	2272.83
-	$\tau_0 \mathbf{I}$	~ 1	-1128.51	2261.02	2269.00
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	-1127.91	2261.82	2273.79
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	-1127.01	2262.02	2277.98

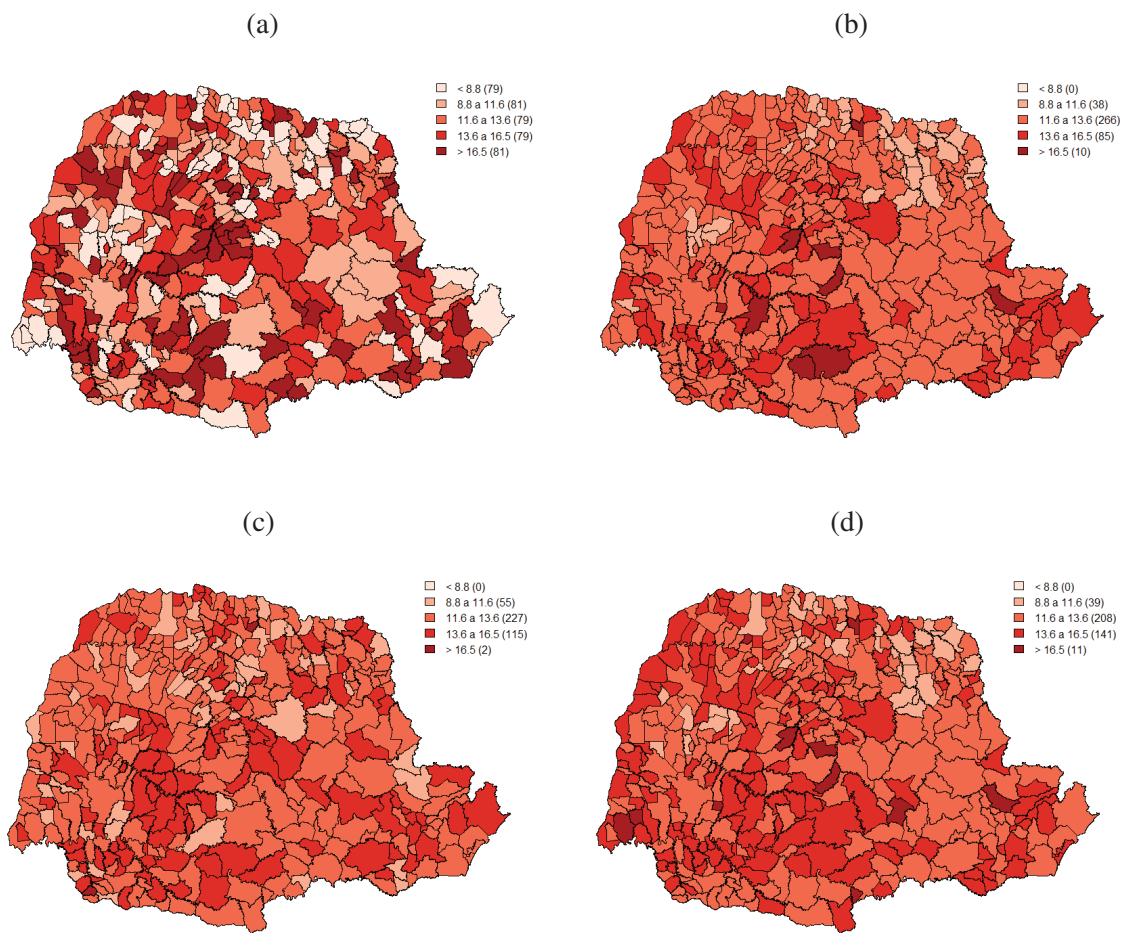
Fonte: Elaborado pelo autor.

Tabela 3.12: Estimativas dos parâmetros τ s para os modelos com função de ligação $h(\cdot)$ identidade.

<i>kernel</i>	$\Omega(\tau, \mathbf{Z})$	Média	τ_0	τ_1	τ_2
-	$\tau_0 \mathbf{I}$	~ 1	1,931	-	-
-	$\tau_0 \mathbf{I}$	$\sim x$	1,768	-	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	~ 1	1,897	0,186	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	$\sim x$	1,722	0,061	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	~ 1	2,326	-0,068	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp}$	$\sim x$	1,894	-0,022	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	1,861	0,342	-
exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	-1,017	2,788	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	2,195	-0,0436	-
autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	1,904	-0,022	-
k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	$\sim x$	1,929	0,0037	-
k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_x$	~ 1	1,800	0,0019	-
exponencial+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-1,670	0,066	3,391
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	2,110	-0,044	0,339
autorregressiva+exponencial	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-1,182	-0,026	3,102
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	2,513	-0,062	-0,036
autorregressiva+autorregressiva	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	2,021	-0,021	-0,022
autorregressiva+k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	~ 1	2,089	-0,047	0,427
autorregressiva+k-médias	$\tau_0 \mathbf{I} + \tau_1 \mathbf{K}_{sp} + \tau_2 \mathbf{K}_x$	$\sim x$	-2,538	-0,023	4,439

Fonte: Elaborado pelo autor.

Figura 3.21: Mapa com legenda de cores para os valores ajustados pelos melhores modelos, com as covariáveis no componente de média e estrutura de dependência, segundo o pAIC: Valores observados em (a), modelo com *kernel* exponencial das distâncias geográficas em (b), modelo com *kernel* K-médias no espaço das covariáveis em (c), modelo com *kernel* exponencial das distâncias geográficas+*kernel* K-médias no espaço das covariáveis em (d).



Fonte: Elaborado pelo autor.

4 CONSIDERAÇÕES FINAIS

A modelagem de estruturas de dependência no espaço das covariáveis se mostrou uma alternativa adicional para flexibilidade e aprimoramento de modelos na busca de melhores ajustes, conforme ilustrado nas análises dos dados apresentados. Desta forma é possível explorar mais amplamente a relação da variável resposta com as covariáveis. Além disso, tal estratégia possui aplicabilidade em dados de diferentes contextos como, dados espacialmente referenciados, dados longitudinais, dados indexados no tempo, por grupos e problemas de classificação.

Não se buscou aqui uma recomendação de melhor estrutura específica de dependência, mas sim prover alternativas que podem proporcionar ajustes que melhorem ainda mais os proporcionados por modelos usuais. A recomendação de formas específicas de estrutura de dependência baseada nas covariáveis pode ser diferentes em diferentes dados e contextos. De maneira geral, nos exemplos relatados aqui e algumas análises adicionais não reportadas, os modelos marginais com *kernels* contínuos e sem covariáveis no componente de média apresentaram as melhores performances preditivas. Por outro lado, os modelos com covariáveis no componente de média e *kernels* via K-médias e autorregressiva apresentaram melhores performances em qualidade de ajuste global. Os parâmetros $\tau_0, \tau_1, \dots, \tau_m$ da matriz de dispersão podem ser interpretados como contribuição relativa dos respectivos *kernels*.

No presente trabalho, os *kernels* foram computados considerando conjuntamente todo o conjunto de covariáveis. Porém, um ponto que pode ser explorado em contextos de grupos de covariáveis de diferentes naturezas, é considerar *kernels* para subconjuntos de covariáveis afins. Neste caso pode-se construir uma estrutura de covariância para cada grupo de covariáveis se aproveitando da possibilidade de combinação linear de estruturas dentro do contexto dos MCGLMs. Desta forma, seria ainda possível avaliar a contribuição associada a cada subconjunto de covariáveis, o que pode permitir maior interpretabilidade dos resultados.

É ainda possível modelar a variância entre as observações, substituindo a matriz identidade por uma matriz diagonal com os valores de alguma variável ou matrizes diagonais com os valores de um conjunto de variáveis.

Os MCGLMs possibilitam a modelagem de variáveis respostas multivariadas, cada variável com diferentes distribuições com respectivos preditores de médias e preditores matriciais associados o que leva à possibilidade de estender as estruturas aqui consideradas para o caso multivariado.

Um possível ponto negativo destes modelos é que, a modelagem de componentes de dependência, e estimação dos respectivos parâmetros, implica em uma complexidade computacional maior, sobretudo no caso de *kernels* contínuos que dão origem a matrizes densas.

REFERÊNCIAS

- Bates, D., Mächler, M., Bolker, B. e Walker, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, 67:1–48.
- Bivand, R., Nowosad, J. e Lovelace, R. (2019). *spData: Datasets for Spatial Analysis*. R package version 0.3.0.
- Bivand, R. e Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, 63:1–36.
- Bonat, W. H. (2018). Multiple response variables regression models in r: The mcglm package. *Journal of Statistical Software*, 84.
- Bonat, W. H. e Jørgensen, B. (2016). Multivariate covariance generalized linear models. *Royal Statistical Society*, 65:649–675.
- Bonat, W. H., Olivero, J., Grande-Vega, M., Farfán, M. A. e Fa, J. A. (2017). Modelling the covariance structure in marginal multivariate count models: Hunting in bioko island. *International Biometric Society*, 22(4):446–464.
- Choudhuri, N., Ghosal, S. e Roy, A. (2007). Nonparametric binary regression using a gaussian process prior. *Statistical Methodology*, 4:227–243.
- Diggle, P. J., Heagerty, P. J., Liang, K. e Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Oxford University Press.
- Diggle, P. J. e Ribeiro Jr., P. J. (2007). *Model-based Geostatistics*. Springer.
- Dubrulle, O. (1983). Two methods with different objectives: Splines and kriging. *Mathematical Geology*, 15:245–257.
- Dubrulle, O. (1984). Comparing splines and kriging. *Computers & Geosciences*, 10:327–338.
- Ein-Dor, P. e Feldmesser, J. (1987). Attributes of the performance of central processing units: A relative performance prediction model. *Commun. ACM*, 30:308–317.
- Evgeniou, T. e Pontil, M. (2001). *Support Vector Machines: Theory and Applications*, volume 2049. Springer.
- Gattiker, J. R., Hamada, M. S., Higdon, D. M., Schonlau, M. e Welch, W. (2015). Using a gaussian process as a nonparametric regression model. *Quality and Reliability Engineering International*, 32.
- Gotway, C. A. e Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(2):157–178.
- Halekoh, V., Højsgaard, S. e Yan, J. (2006). The r package **geepack** for generalized estimation equations. *Journal of Statistical Software*, 15:1–11.

- Hastie, T. e Tibshirani, R. (1986). Generalized additive models. *Biometrics*, 1(3):297–310.
- Johnson, R. A. e Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 6 edition.
- Jørgensen, B. (1987). Exponential dispersion models. *Royal Statistical Society*, 49:127–162.
- Justen, W. (2018). GeoDa data and lab.
- Kelijian, H. e Robinson, D. (1992). Spatial autocorrelation: a new computationally simple test with application to per capita county police expenditures. *Regional Science and Urban Economics*, 22(3):317–331.
- Klein, R., Langetepe, E. e Nilforoushan, Z. (2009). Abstract voronoi diagrams revisited. *Computational Geometry: Theory and Applications*, 42:885–902.
- Kuss, M. e Rasmussen, C. E. (2006). Assessing approximations for gaussian process classification. *Advances in Neural Information Processing Systems*, 18:699–706.
- Lindgren, F., Rue, H. e Lindstrom (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63:1–25.
- Mingoti, S. A. (2013). *Análise de dados através de métodos de estatística multivariada: Uma abordagem aplicada*. UFMG.
- Nelder, J. A. e Wedderburn, W. (1972). Generalized linear models. *Royal Statistical Society*, 135(3):370–384.
- Pan, J. e Pan, Y. (2017). jmcm: An R package for joint mean-covariance modeling of longitudinal data. *Computational Statistics and Data Analysis*, 82.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1:81–106.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rasmussen, C. E. e Williams, C. K. I. (2006). *Gaussian Process for Machine Learning*. Massachusetts Institute of Technology.
- Rodrigues, E. e Assunção, R. (2011). Bayesian spatial models with mixture neighbourhood structure. *Revista da Estatística da UFOP*, 1:92–108.
- Schabenberger, O. e Gotway, C. A. (2009). *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC.
- Simon, N., Friedman, J., Hastie, T. e Tibshirani, R. (2011). Regularization paths for cox’s proportional hazards model via coordinate descent. *Journal of Statistical Software*, 39(5):1–13.
- Stoklosa, J., Gibb, H. e Warton, D. I. (2014). Fast forward selection for generalized estimating equations with a large number of predictor variables. *Biometrics*, 70:110–120.
- VANSCHOREN, J., VAN RIJN, J. N., BISCHL, B. e TORGO, L. (2013). Openml: Networked science in machine learning. *SIGKDD Explorations*, 15(2):49–60.

Venables, W. N. e Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

Wall, M. M. (2004). A close look at the spatial structure implied by the car and sar models. *Journal of Statistical Planning and Inference*, (121):311–324.