Modelos de regressão para variáveis categóricas ordinais com aplicações ao problema de classificação

Roberta Irie Sumi Okura

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

Área de Concentração: Estatística

Orientadora: Profa. Dra. Silvia Nagib Elian

São Paulo, abril de 2008

Modelos de regressão para variáveis categóricas ordinais com aplicações ao problema de classificação

Este exemplar corresponde à redação
final da dissertação devidamente corrigida
e defendida por Roberta Irie Sumi Okura
e aprovada pela Comissão Julgadora.

Banca Examinadora:

- Profa. Dra. Silvia Nagib Elian - IME-USP.
- Profa. Dra. Viviana Giampaoli - IME-USP.
- Profa. Dra. Vera Lucia Damasceno Tomazella - UFSCar.

Para meus queridos pais,
dedico este trabalho
com o mesmo amor que
deles sempre recebi.

Agradecimentos

Agradeço à minha orientadora, professora Silvia Nagib Elian, por toda a ajuda a mim dedicada durante esses anos, pela paciência, pela compreensão, pelo apoio e pela amizade.

Agradeço aos meus pais, que sempre estiveram ao meu lado e que não pouparam esforços e incentivos para eu me aprofundasse em meus estudos e descobrisse que o mundo poderia ser grande, com muito a me oferecer.

Agradeço a Ulisses Duarte Nehmi, por todo carinho e todo apoio que me ofereceu durante essa etapa da minha vida, mesmo nas horas mais difíceis. Sou imensamente grata por ter me compreendido em diversas situações e, principalmente, por ter me presenteado com os momentos mais maravilhosos e mais confortantes dos últimos anos.

Agradeço aos meus amigos de faculdade, que participaram das principais decisões de minha vida relacionadas ao meu desenvolvimento acadêmico e profissional. Agradeço também por estarem comigo me ajudando em tantas ocasiões e por todas as alegrias proporcionadas.

Agradeço a Bruno Fernandes Cerqueira Leite, por ter sido um grande amigo desde que iniciamos a faculdade, pela ajuda incondicional, por seus conselhos e por ter me ensinado tanto ao mesmo tempo que aprendia comigo.

Agradeço, por fim, a Deus, por tudo que tenho e por ter tido a oportunidade de chegar até aqui.

Resumo

Neste trabalho, apresentamos algumas metodologias para analisar dados que possuem variável resposta categórica ordinal. Descrevemos os principais Modelos de Regressão conhecidos atualmente que consideram a ordenação das categorias de resposta, entre eles: Modelos Cumulativos e Modelos Seqüenciais. Discutimos também o problema de discriminação e classificação de elementos em grupos ordinais, comentando sobre os preditores mais comuns para dados desse tipo. Apresentamos ainda a técnica de Análise Discriminante Ótima e sua versão aprimorada, baseada na utilização de métodos *bootstrap*. Por fim, aplicamos algumas das técnicas descritas a dados reais da área financeira, com o intuito de classificar possíveis clientes, no momento da aquisição de um cartão de crédito, como futuros bons, médios ou maus pagadores. Para essa aplicação, discutimos as vantagens e desvantagens dos modelos utilizados em termos de qualidade da classificação.

Abstract

In this work, some methods to analyse data with ordinal categorical response are presented. We describe the most important and widely used Regression Models which consider the ordering of response categories like: Cumulative Models and Sequential Models. We also discuss the problem of how to discriminate and classify elements in ordinal groups, commenting on the most common predictors to this kind of data. Also we present the technique known as optimal discriminant analysis and its improved version, based on the use of bootstrap methods. Finally, we apply some of the described techniques to real financial data, intending to classify possible consumers, on acquistion of a credit card, as high, medium and low risk customers. With this application, we discuss the advantages and disadvantages of the models used in terms of quality of classification.

Sumário

Capítulo 1

Introdução

Modelos classificatórios são geralmente utilizados para categorizar um particular elemento com base em suas características. Como exemplo, vamos supor que se deseja selecionar uma área metropolitana para expansão de vendas de um certo produto. É natural que, antes de promover tal expansão, a área seja classificada em: provável de possuir consumidores do produto ou não provável de possuir consumidores do produto. Faz sentido também que o método de classificação se baseie nas características dos habitantes da área.

De modo geral, $\mathbf{x}$ é um vetor p-dimensional de atributos observados em um objeto ou indivíduo, pertencendo a exatamente uma de k populações mutuamente exclusivas $\Pi_1, \Pi_2, \ldots, \Pi_k$. O problema será então o de discriminar as populações e, a partir daí, decidir a qual população $\mathbf{x}$ pertence. Fisher (1936) sugere o uso de funções lineares discriminantes das componentes de $\mathbf{x}$ para efetuar a classificação. Técnicas baseadas em testes da razão de verossimilhanças, teoria da informação e análise bayesiana são também empregadas.

Modelos de regressão logística são usualmente construídos com o objetivo de relacionar variáveis resposta qualitativas com um conjunto de p variáveis explicativas. Se $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ é o vetor de variáveis explicativas, o modelo de regressão logística é da forma:

$$P(\mathbf{x}) = P(E|\mathbf{x}) = \frac{\exp(\beta_0 + \beta'\mathbf{x})}{1 + \exp(\beta_0 + \beta'\mathbf{x})}$$

onde E é um particular evento de interesse e (β_0, β') é o vetor de parâmetros, estimado a partir do conjunto de dados.

Modelos de regressão logística podem também ser utilizados para classificar um particular elemento em uma de duas populações. Seja E o evento segundo o qual o elemento pertence à primeira população e $\mathbf{x}$ o vetor contendo os valores de seus atributos. Obtido

$$\hat{p}(\mathbf{x}) = \hat{P}(E|\mathbf{x}) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}'\mathbf{x})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}'\mathbf{x})},$$

o elemento é classificado em E se $\hat{P}(E|\mathbf{x}) \geq 0,5$ e em E^c se $\hat{P}(E|\mathbf{x}) < 0,5$.

Esta abordagem é descrita em inúmeros artigos, dentre os quais destacamos Efron (1975) e Press e Wilson (1978), que comparam a eficiência da análise discriminante com a da regressão logística na classificação de um particular elemento em uma de duas populações ($k = 2$).

Na prática, no entanto, podemos ter um número maior de populações. Por exemplo, no caso das vendas do produto, as populações poderiam ser da forma:

- Π_1 = áreas com alta indicência de consumidores;
- Π_2 = áreas de mediana incidência de consumidores;
- Π_3 = áreas de baixa incidência de consumidores do produto.

Esta situação é bastante diferente da original, pois envolve um número maior de populações ($k > 2$) e as populações podem ser ordenadas segundo um particular critério, no caso, incidência de consumidores do produto: baixa, média ou alta.

Existindo apenas a primeira restrição, o problema de classificação poderia ser abordado através de análise discriminante com $k > 2$ grupos (Johnson e Wichern (2007)). Do ponto de vista de regressão, poderiam ser utilizados modelos de regressão com resposta politômica (Fahrmeir e Tutz (1994)).

A presença da ordenação, entretanto, tornaria pouco razoável a aplicação de análise discriminante na sua forma usual. Em particular, nesse caso, as probalidades de erro de classificação merecem uma nova interpretação. Isto porque o erro associado à classificação incorreta de uma cidade de Π_1 em Π_3 será mais grave do que se ela for classificada em Π_2.

Quanto à análise de regressão, o caráter ordinal da variável resposta permitiria a adoção de modelos mais refinados que o modelo de regressão logística politômica.

Os objetivos do presente trabalho são:

1. Descrever possíveis modelos de regressão para variável resposta categórica ordinal;
2. Descrever alguns métodos para classificação em grupos ordinais;
3. Apresentar a técnica de análise discriminante ótima para resposta ordinal;
4. Analisar as técnicas mais comuns através de uma aplicação a um conjunto de dados real.

No Capítulo 2, será feita uma descrição dos modelos de regressão mais conhecidos atualmente para casos em que a variável resposta é categórica ordinal. Estão divididos em três partes: Modelos Cumulativos, Modelos Seqüenciais e Modelos em Dois Estágios. É apresentada também a Inferência Estatística, usada na estimação dos parâmetros por máxima verossimilhança, nos testes de hipóteses e na análise de qualidade do ajuste dos modelos.

O Capítulo 3 descreve o problema de discriminação e classificação de elementos em grupos que podem ser ordenados, apresentando algumas opções de preditores para variável resposta categórica ordinal.

No Capítulo 4, será discutida uma diferente proposta para o problema de classificação em categorias ordenadas, a Análise Discriminante Ótima. Será apresentado também um aperfeiçoamento de sua robustez através do uso de amostras *bootstrap*.

O Capítulo 5, por fim, apresenta a aplicação de algumas técnicas de discriminação para dados reais da área financeira. Foram utilizados modelos que não consideram a ordenação da variável resposta categórica, mas são bastante conhecidos no mercado, como o Modelo Logístico Multinomial e a Análise Discriminante Normal, assim como modelos de regressão para resposta ordinal, descritos no Capítulo 2, e a Análise Discriminante Ótima, apresentada no Capítulo 4. Os resultados foram então comparados, em termos de qualidade da classificação, e vantagens e desvantagens na estimação dos parâmetros.

Capítulo 2

Modelos de regressão para variáveis categóricas ordinais

2.1 Introdução

Variáveis categóricas ordinais são importantes em muitas áreas de estudo, principalmente, em situações onde medidas exatas não são possíveis. Caracterizam-se por apresentar uma ordenação entre seus possíveis valores e estão muito presentes em Ciências Sociais, em particular, para medir atitudes e opiniões sobre vários assuntos, assim como *status* de diversos tipos. Também costumam ocorrer em campos como *marketing* (por exemplo, em escalas de preferência ordinais ou escalas de pesquisa) e em disciplinas médicas e de saúde pública (por exemplo: gravidade de um ferimento, grau de recuperação de uma doença, estágios de uma doença, nível de exposição a uma substância potencialmente danosa). As variáveis ordinais podem originar-se de mecanismos completamente diferentes. Anderson (1984) as classifica em: **variáveis contínuas agrupadas** e **variáveis categóricas naturalmente ordenadas**.

O primeiro tipo representa uma versão categorizada de uma variável contínua. Como exemplo, temos a variável número de anos de estudo, que pode ser medida ordinalmente por meio da categorização 0–8, 9–12, 13–16, 17 ou mais anos.

Já o segundo tipo de variável ordinal ocorre quando é avaliada uma informação não quantificável, associada a níveis de uma escala categórica ordinal. Este tipo, que

consiste em uma coleção de categorias naturalmente ordenadas, origina-se de casos onde uma medida precisa nem sempre é possível. Isso ocorre, por exemplo, em ramos como Psicologia, Sociologia e Pesquisa de Mercado. Como ilustração, temos a classe social, que pode ser classificada como "alta", "média" ou "baixa", e a filosofia política, que pode ser medida como "liberal", "moderada" ou "conservadora".

Uma variável categórica é referida como ordinal ao invés de intervalar quando há uma ordem clara das categorias, mas as distâncias absolutas entre elas são desconhecidas. Por exemplo, a variável educação é ordinal quando medida com as categorias "ensino fundamental", "ensino médio", "ensino superior", "pós-graduação", mas é intervalar quando medida em valores inteiros 0, 1, 2, ..., representando número de anos de estudo. Já a filosofia política é ordinal, porque uma pessoa classificada como moderada é mais liberal que uma pessoa classificada como conservadora, mas não há uma maneira óbvia de quantificar numericamente quão mais liberal essa pessoa é.

Assim, para muitas variáveis categóricas ordinais, é sensato imaginar a existência de uma variável contínua subjacente. Para se aproximar da escala subjacente, é freqüentemente útil associar um conjunto "razoável" de escores às categorias.

Uma variável ordinal ou intervalar é quantitativa, porque cada nível sobre sua escala pode ser comparado a um outro nível em termos de magnitude maior ou menor de certa característica. Esse tipo de variável é de uma natureza bem diferente das variáveis qualitativas, que são medidas em uma escala nominal. Exemplos de variáveis nominais são: raça, religião e situação civil. Os níveis dessas variáveis diferem em qualidade, não em quantidade. Então, a ordem de apresentações das categorias de uma variável nominal não é relevante.

O objetivo deste capítulo é descrever parte dos modelos de regressão para variáveis resposta categóricas ordinais.

2.2 Modelos Cumulativos

Modelos de Regressão ordinal levam em consideração a ordem das categorias da variável resposta. Particulamente, para dados categóricos em que o tamanho amostral é freqüentemente crítico, torna-se necessário fazer uso de toda a informação disponível. Por isso, a ordem das categorias de resposta deve ser levada em consideração.

O modelo mais utilizado em regressão ordinal é baseado nos "limites das categorias" ou na "aproximação limiar", ambas as expressões não tão conhecidas.

Assume-se que a variável observada Y é uma categorização da variável contínua latente U. No caso de variáveis respostas contínuas agrupadas, U pode ser considerada a variável subjacente não observada. Já no caso de variáveis categóricas naturalmente ordenadas, U é uma avaliação sobre uma escala contínua subjacente. Em ambos os casos, a variável latente é utilizada apenas para facilitar a interpretação e a construção do modelo.

Para um dado vetor $\mathbf{x}$ de variáveis explicativas, a abordagem do limite das categorias sugere que a variável observada Y, com $Y \in \{1, \ldots, k\}$, e a variável latente U estão conectadas por:

$$Y = r \quad \Leftrightarrow \quad \theta_{r-1} < U \leq \theta_r, \qquad r = 1, \ldots, k \tag{2.1}$$

onde $-\infty = \theta_0 < \theta_1 \ldots < \theta_k = \infty$. Isso significa que Y é uma versão categorizada de U, determinada pelos pontos $\theta_1, \ldots, \theta_{k-1}$. Além disso, assume-se que a variável latente U é determinada pelas variáveis explicativas de forma linear:

$$U = -\mathbf{x}'\gamma + \epsilon \tag{2.2}$$

onde $\gamma = (\gamma_1, \ldots, \gamma_p)$ é o vetor de parâmetros e ϵ é uma variável aleatória com distribuição F.

Desses fatos, segue que a distribuição de probabilidades da variável observada Y é dada por:

$$P(Y \leq r|\mathbf{x}) = F(\theta_r + \mathbf{x}'\gamma). \tag{2.3}$$

Observa-se que o lado esquerdo da equação (2.3) é a soma $P(Y = 1|\mathbf{x}) + P(Y = 2|\mathbf{x}) + \cdots + P(Y = r|\mathbf{x})$ e, por isso, o modelo é denominado Modelo Cumulativo com função de distribuição F. O modelo (2.3) também é chamado de Modelo Limiar, devido à derivação baseada nos limiares das variáveis latentes.

Para verificarmos (2.3), utilizamos a equação (2.1) e obtemos:

$$P(Y \leq r|\mathbf{x}) = P(Y = 1|\mathbf{x}) + P(Y = 2|\mathbf{x}) + \cdots + P(Y = r|\mathbf{x})$$

$$= P(\theta_0 < U \leq \theta_1) + P(\theta_1 < U \leq \theta_2) + \cdots + P(\theta_{r-1} < U \leq \theta_r)$$

$$= F_U(\theta_r) - F_U(\theta_0).$$

Como $\theta_0 = -\infty$, temos $F_U(\theta_0) = 0$ e, sendo U determinada por (2.2), $F_U(\theta_r) = P(-\mathbf{x}'\gamma + \epsilon \leq \theta_r) = P(\epsilon \leq \theta_r + \mathbf{x}'\gamma) = F(\theta_r + \mathbf{x}'\gamma)$, onde F é a função de distribuição da variável aleatória ϵ.

2.2.1 Principais Modelos Cumulativos

A seguir, apresentaremos alguns dos modelos cumulativos mais utilizados em casos onde a variável resposta é ordinal.

Modelo Logístico Cumulativo ou Modelo de Chances Proporcionais

Os modelos cumulativos são definidos por meio da escolha da função de distribuição F. Uma escolha comum de F é a função de distribuição logística:

$$F(z) = \frac{1}{1 + \exp(-z)} = \frac{\exp(z)}{1 + \exp(z)}, \tag{2.4}$$

que leva ao modelo logístico cumulativo da forma:

$$P(Y \leq r|\mathbf{x}) = \frac{\exp(\theta_r + \mathbf{x}'\gamma)}{1 + \exp(\theta_r + \mathbf{x}'\gamma)}, \tag{2.5}$$

para $r = 1, \ldots, q = k - 1$. É simples chegar em formas equivalentes ao se trabalhar a expressão (2.5):

$$P(Y \leq r|\mathbf{x}) + P(Y \leq r|\mathbf{x}) \exp(\theta_r + \mathbf{x}'\gamma) = \exp(\theta_r + \mathbf{x}'\gamma)$$

$$P(Y \leq r|\mathbf{x}) = [1 - P(Y \leq r|\mathbf{x})] \exp(\theta_r + \mathbf{x}'\gamma)$$

$$\frac{P(Y \leq r|\mathbf{x})}{P(Y > r|\mathbf{x})} = \exp(\theta_r + \mathbf{x}'\gamma)$$

$$\log\left\{\frac{P(Y \leq r|\mathbf{x})}{P(Y > r|\mathbf{x})}\right\} = \theta_r + \mathbf{x}'\gamma, \tag{2.6}$$

e então podemos perceber que o logaritmo da chance acumulada, $\log\left\{\frac{P(Y \leq r|\mathbf{x})}{P(Y > r|\mathbf{x})}\right\}$, é determinado por uma forma linear das variáveis explicativas, ou então

$$\frac{P(Y \leq r|\mathbf{x})}{P(Y > r|\mathbf{x})} = \exp(\theta_r + \mathbf{x}'\gamma) = \frac{P(Y \leq r|\mathbf{x})}{1 - P(Y \leq r|\mathbf{x})},$$

onde temos, do lado esquerdo, a chance de ocorrer o evento $\{Y \leq r|\mathbf{x}\}$.

O Modelo Logístico Cumulativo tem sido também chamado de Modelo de Chances Proporcionais, devido a uma propriedade especial: se duas populações caracterizadas por variáveis explicativas x_1 e x_2 são consideradas, a razão das chances acumuladas para as duas populações é dada por

$$\frac{P(Y \leq r|x_1)/P(Y > r|x_1)}{P(Y \leq r|x_2)/P(Y > r|x_2)} = \exp\{(x_1 - x_2)'\gamma\}$$

e, assim, não depende de r.

Modelo de Cox Agrupado ou Modelo de Riscos Proporcionais

O Modelo de Cox Agrupado é obtido ao se adotar como função de distribuição F a distribuição do mínimo valor extremo, $F(z) = 1 - \exp(-\exp(z))$ que, ao ser substituída em (2.3), leva a:

$$P(Y \leq r|\mathbf{x}) = 1 - \exp\{-\exp(\theta_r + \mathbf{x}'\gamma)\},\ r = 1, \ldots, q \tag{2.7}$$

ou, equivalentemente, utilizando-se a ligação complementar log-log:

$$\log[-\log P(Y > r|\mathbf{x})] = \theta_r + \mathbf{x}'\gamma,$$

que mostra que o termo $\log[-\log P(Y > r|\mathbf{x})]$ é determinado de forma linear pelas variáveis explicativas.

Para chegar nesta última expressão, basta trabalharmos (2.7) como a seguir:

$$P(Y \leq r|\mathbf{x}) = 1 - \exp\{-\exp(\theta_r + \mathbf{x}'\gamma)$$

$$P(Y \leq r|\mathbf{x}) - 1 = -\exp\{-\exp(\theta_r + \mathbf{x}'\gamma)\}$$

$$1 - P(Y \leq r|\mathbf{x}) = \exp\{-\exp(\theta_r + \mathbf{x}'\gamma)\}$$

$$\log P(Y > r|\mathbf{x}) = -\exp(\theta_r + \mathbf{x}'\gamma)$$

$$\log[-\log P(Y > r|\mathbf{x})] = \theta_r + \mathbf{x}'\gamma.$$

Este modelo é determinado Modelo de Cox Agrupado, porque pode ser visto como uma versão agrupada do Modelo de Cox ou Modelos de Riscos Proporcionais, bem conhecidos em Análise de Sobrevivência.

Modelo com distribuição máximo valor extremo

Ao invés da distribuição do mínimo valor extremo, é possível utilizar a distribuição do máximo valor extremo, $F(z) = \exp(-\exp(-z))$. O modelo baseado nessa distribuição é dado por:

$$P(Y \leq r|\mathbf{x}) = \exp[-\exp\{-(\theta_r + \mathbf{x}'\gamma)\}] \tag{2.8}$$

ou, equivalentemente, usando a ligação log-log:

$$\log[-\log P(Y \leq r|\mathbf{x})] = -(\theta_r + \mathbf{x}'\gamma),$$

que é obtida a partir da expressão (2.8):

$$P(Y \leq r|\mathbf{x}) = \exp[-\exp\{-(\theta_r + \mathbf{x}'\gamma)\}]$$

$$\log P(Y \leq r|\mathbf{x}) = -\exp(-(\theta_r + \mathbf{x}'\gamma))$$

$$\log[-\log P(Y \leq r|\mathbf{x})] = -(\theta_r + \mathbf{x}'\gamma).$$

Apesar de o modelo (2.7) não ser equivalente ao modelo (2.8), seus parâmetros podem ser estimados utilizando apenas o último. Para isso, é preciso inicialmente construir a variável $\widetilde{Y} = k + 1 - Y$, com possíveis valores $1, \ldots, k$, mas com ordem inversa das categorias, e então estimar os parâmetros do modelo (2.8) usando $\widetilde{Y}$ ao invés de Y. Em um segundo passo, as estimativas obtidas devem ser multiplicadas por -1 para produzir as estimativas dos parâmetros do modelo (2.7) e as ordens dos parâmetros θ_r precisam ser revertidas. Mais formalmente, deixa-se as ordens das categorias serem revertidas por meio de $\widetilde{r} = k + 1 - Y$. Então, o modelo do mínimo valor extremo para Y com parâmetros θ_r e β é equivalente ao modelo do máximo valor extremo para $\widetilde{Y}$ com parâmetros $\widetilde{\theta}_r = -\theta_{k-\widetilde{r}}$ e $\widetilde{\beta} = -\beta$.

Várias famílias de funções de ligação têm sido propostas para modelos de resposta binária. A princípio, essas famílias podem também ser usadas para variáveis resposta ordinais, produzindo diferentes modelos cumulativos.

2.2.2 Modelos Cumulativos Generalizados

O Modelo Cumulativo Simples (2.3) baseia-se na suposição de que as variáveis explicativas provocam uma mudança na escala latente, mas não alteram os pontos

$\theta_1, \ldots, \theta_q$. Em um modelo mais geral, esses pontos podem ser dependentes das variáveis explicativas $w = (w_1, \ldots, w_m)$, de forma linear, ou seja:

$$\theta_r = \beta_{r0} + w'\beta_r$$

onde $\beta_r = (\beta_{r0}, \ldots, \beta_{rm})$ é um vetor de parâmetros chamados de categoria específica. O modelo estendido segue diretamente do mecanismo (2.1) e da parametrização da variável latente (2.2), sendo dado por:

$$P(Y \leq r|\mathbf{x}, w) = F(\beta_{r0} + w'\beta_r + \mathbf{x}'\gamma). \tag{2.9}$$

Este modelo também assume o mecanismo de limite das categorias, mas agora somente para as variáveis explicativas x. A variável w determina apenas os pontos que estão sobre a escala latente. Como será visto a seguir, a suposição $w_i = x_i$, $i = 1, \ldots, m$, $m = p$, torna γ um parâmetro não identificável. Assim, é preciso distinguir estritamente entre variáveis limiares w_i e variáveis de deslocamento x_i. As primeiras são sempre ponderadas por um vetor de parâmetros β_r de categorias específicas enquanto as variáveis de deslocamento são ponderadas por um vetor de parâmetros global γ.

2.2.3 Funções de Ligação e Matrizes de Planejamento

Para utilizar a abordagem de modelos lineares generalizados no ajuste dos modelos descritos anteriormente, é preciso especificar uma função de ligação (ou uma função resposta) e uma matriz de planejamento para os modelos cumulativos. A função de ligação $g = (g_1, \ldots, g_q)$ é imediatamente determinada por (2.3) ou (2.9), ou seja,

$$g_r(\pi_1, \ldots, \pi_q) = F^{-1}(\pi_1 + \cdots + \pi_r),$$

$r = 1, \ldots, q$.

Denotando $\pi_r = P(Y = r|\mathbf{x})$ e considerando, por exemplo, o modelo (2.3), temos que:

$$\begin{aligned} P(Y \leq r|\mathbf{x}) &= F(\theta_r + \mathbf{x}'\gamma) \\ \pi_1 + \pi_2 + \cdots + \pi_r &= F(\theta_r + \mathbf{x}'\gamma) \\ F^{-1}(\pi_1 + \cdots + \pi_r) &= \theta_r + \mathbf{x}'\gamma \end{aligned} \tag{2.10}$$

e, assim, $g_r(\pi_1, \ldots, \pi_q) = F^{-1}(\pi_1 + \cdots + \pi_r)$ é a função de ligação.

Considerando o Modelo de Cox agrupado, temos a seguinte função de ligação:

$$g_r(\pi_1, \ldots, \pi_q) = \log\{-\log(1 - \pi_1 - \cdots - \pi_r)\}, \quad r = 1, \ldots, q.$$

A matriz de planejamento deve fazer a distinção entre o modelo simples (2.3) e o modelo geral (2.9). No caso do modelo simples, o termo linear $\eta_i = Z_i\beta$ é determinado por:

$$Z_i = \begin{bmatrix} 1 & 0 & & 0 & \mathbf{x}_i' \\ 0 & 1 & & 0 & \mathbf{x}_i' \\ & & \ddots & & \vdots \\ 0 & 0 & & 1 & \mathbf{x}_i' \end{bmatrix}$$

e pelo vetor de parâmetros $\beta' = (\theta_1, \ldots, \theta_q, \gamma)$. O modelo geral, com variáveis explicativas w_i e x_i, tem matriz de planejamento da forma:

$$Z_i = \begin{bmatrix} 1 & w_i' & 0 & 0 & & 0 & 0 & \mathbf{x}_i' \\ 0 & 0 & 1 & w_i' & & 0 & 0 & \mathbf{x}_i' \\ & & & & \ddots & & & \vdots \\ 0 & 0 & 0 & 0 & & 1 & w_i' & \mathbf{x}_i' \end{bmatrix}$$

O vetor de parâmetros β é dado por $\beta' = (\beta_{10}, \beta_1, \ldots, \beta_{q0}, \beta_q, \gamma)$.

Algumas vezes, como em modelos de coeficientes aleatórios ou em modelagem dinâmica, é útil considerar uma forma alternativa para a função de ligação e para a matriz de planejamento. Em casos de modelos cumulativos, os parâmetros podem não variar livremente. Os parâmetros para os modelos simples (2.3) estão restritos por $\theta_1 < \ldots < \theta_q$. Já para o modelo generalizado, a restrição é determinada por $\beta_{10} + \beta_1' w < \ldots < \beta_{q0} + \beta_q' w$ para todos os valores possíveis da variável w. Verifica-se que para as funções de ligação e matriz de planejamento dadas anteriormente, essas restrições não são levadas em consideração. Se a restrição deve valer para todo w, sua severidade vai depender do intervalo das covariáveis w. Se a limitação não é explicitamente utilizada na estimação, o procedimento de estimação iterativa pode falhar fornecendo estimativas inadmissíveis. No modelo simples, não haverá problema se os limiares estiverem bem separados. No entanto, podem ocorrer problemas numéricos no procedimento de estimação se alguns limiares forem muito similares. Esses problemas podem ser evitados de forma simples utilizando-se uma formulação alternativa,

reparametrizando o modelo da seguinte forma:

$$\alpha_1 = \theta_1, \quad \alpha_r = \log(\theta_r - \theta_{r-1}), \quad r = 2, \ldots, q,$$

ou, respectivamente,

$$\theta_1 = \alpha_1, \quad \theta_r = \theta_1 + \sum_{i=2}^{r} \exp(\alpha_i), \quad r = 2, \ldots, q.$$

Neste caso, os parâmetros $\alpha_1, \ldots, \alpha_q$ não sofrem restrições, pois $(\alpha_1, \ldots, \alpha_q) \in \mathbb{R}^q$ e a estrutura linear do modelo torna-se clara na forma:

$$F^{-1}(P(Y = 1|\mathbf{x})) = \alpha_1 + \mathbf{x}'\gamma = F^{-1}(F(\alpha_1 + \mathbf{x}'\gamma))$$

$$\log\{F^{-1}[P(Y \leq r|\mathbf{x})] - F^{-1}[P(Y \leq r-1|\mathbf{x})]\} = \alpha_r, \ r = 2, \ldots, q.$$

Com isso, a função de ligação é determinada. Verificamos as expressões anteriores utilizando (2.10), obtendo:

$$F^{-1}(P(Y = 1|\mathbf{x})) = F^{-1}(\pi_1) = \theta_1 + \mathbf{x}'\gamma = \alpha_1 + \mathbf{x}'\gamma,$$

para $r = 1$ e, para $r = 2, \ldots, q$,

$$\begin{aligned}
&\log[F^{-1}\{P(Y \leq r|\mathbf{x})\} - F^{-1}\{P(Y \leq r-1|\mathbf{x})\}] \\
&= \log[F^{-1}\{F(\theta_r + \mathbf{x}'\gamma)\} - F^{-1}\{F(\theta_{r-1} + \mathbf{x}'\gamma)\}] \\
&= \log[\theta_r + \mathbf{x}'\gamma - \theta_{r-1} - \mathbf{x}'\gamma] = \log[\theta_r - \theta_{r-1}] \\
&= \log\left[\theta_1 + \sum_{i=2}^{r} \exp(\alpha_i) - \theta_1 - \sum_{i=2}^{r-1} \exp(\alpha_i)\right] \\
&= \log(\exp(\alpha_r)) = \alpha_r.
\end{aligned}$$

No caso especial do modelo logístico cumulativo, obtemos a função de ligação do modelo reparametrizado a partir da função de ligação dada em (2.6):

$$g_1(\pi_1, \ldots, \pi_q) = F^{-1}(P(Y = 1|\mathbf{x})) = F^{-1}(\pi_1) = \log\left(\frac{P(Y \leq 1|\mathbf{x})}{P(Y > 1|\mathbf{x})}\right) = \log\left(\frac{\pi_1}{1 - \pi_1}\right)$$

e

$$
\begin{aligned}
g_r(\pi_1, \ldots, \pi_q) &= \log\{F^{-1}[P(Y \leq r|\mathbf{x})] - F^{-1}[P(Y \leq r-1|\mathbf{x})]\} \\
&= \log[F^{-1}(\pi_1 + \cdots + \pi_r) - F^{-1}(\pi_1 + \cdots + \pi_{r-1})] \\
&= \log\left\{\log\left[\frac{\pi_1 + \cdots + \pi_r}{1 - \pi_1 - \cdots - \pi_r}\right] - \log\left[\frac{\pi_1 + \cdots + \pi_{r-1}}{1 - \pi_1 - \cdots - \pi_{r-1}}\right]\right\},
\end{aligned}
$$

$r = 2, \ldots, q$.

Usando esta função de ligação alternativa para o modelo logito torna-se necessária a adaptação da matriz de planejamento. A matriz de planejamento para a observação (y_i, x_i) terá agora a forma:

$$
\mathbf{Z_i} = \begin{bmatrix}
1 & & & & \mathbf{x}_i' \\
& 1 & & & 0 \\
& & \ddots & & \vdots \\
& & & 1 & 0
\end{bmatrix}
$$

e o vetor de parâmetros é dado por $\beta = (\alpha_1, \ldots, \alpha_q, \gamma)$.

2.2.4 Exemplos

Fahrmeir e Tutz (1994) apresenta dois exemplos de aplicação dos Modelos Cumulativos, que serão descritos a segur.

Exemplo 1: Resultados de exames respiratórios

Forthofer e Lehnen (1981) investigaram os efeitos da idade e do fumo nos resultados de exames respiratórios realizados em trabalhadores de indústrias no Texas. Os resultados dos testes foram classificados em três categorias, denominadas "normal", "limite" e "anormal". Dessa forma, a variável "resultados respiratórios" pode ser considerada uma variável categórica ordinal. Os dados encontram-se na Tabela 2.1.

A análise baseia-se em três variantes do modelo cumulativo: modelo cumulativo logístico, modelo de riscos proporcionais e modelo do máximo valor extremo. Utilizando o *deviance* como uma medida de distância entre os dados e os valores ajustados, o modelo de riscos proporcionais apresentou o melhor ajuste, enquanto o modelo do

Tabela 2.1: Resultados dos testes respiratórios.

		Resultados dos testes		
Idade	Fumo	Normal	Limite	Anormal
< 40	Nunca	577	27	7
	Casual	192	20	3
	Sempre	682	46	11
40-59	Nunca	164	4	0
	Casual	145	15	7
	Sempre	245	47	27

máximo valor extremo apresentou o pior ajuste. A medida *deviance* será definida na Seção 2.6.

A Tabela 2.2 apresenta as estimativas dos parâmetros para os três modelos. Todas as variáveis foram codificadas com -1 para a última categoria. Assim, sinais positivos das estimativas dos coeficientes geram altas estimativas das probabilidades para as categorias "normal" e "limite" (categorias 1 e 2). Conforme esperado, as categorias que representam idades menores e baixo fumo produzem melhores resultados nos exames respiratórios.

A análise inferencial foi realizada através do ajuste de Modelo Linear Gereralizado. Tal metodologia será descrita brevemente na Seção 2.6.

As estimativas dos parâmetros devem ser diferentes para os modelos por causa das variações entre as distribuições logística e valor-extremo. A função de distribuição do mínimo valor extremo, subjacente ao modelos de riscos proporcionais, é mais acentuada (abrupta) que a função de distribuição logística. Assim, para atingir a mesma quantidade de deslocamento sobre a escala latente, um efeito maior (medido pelo parâmetro) é necessário para a última função de distribuição. Conseqüentemente, os parâmetros do modelo logístico são maiores para todas as variáveis. No entanto, a tendência dos parâmetros é praticamente a mesma para todos os modelos.

Neste conjunto de dados, é interessante analisar o efeito de interação entre idade e fumo, pois é mais forte que o efeito principal de idade. A estimativa do efeito de interação IDADE[1] * FUMO[1] $= -0,211$ (modelo de riscos proporcionais) mostra

Tabela 2.2: Estimativas dos parâmetros e p-valores dos modelos cumulativos para os dados de testes respiratórios.

	Cumulativo logístico		Riscos proporcionais		Máximo valor extremo	
Limiar 1	2,370	(0,00)	0,872	(0,00)	2,429	(0,00)
Limiar 2	3,844	(0,00)	1,377	(0,00)	3,843	(0,00)
IDADE[1]	0,114	(0,29)	0,068	(0,04)	0,095	(0,37)
FUMO[1]	0,905	(0,00)	0,318	(0,00)	0,866	(0,19)
FUMO[2]	-0,364	(0,01)	-0,110	(0,02)	-0,359	(0,14)
IDADE[1] * FUMO[1]	-0,557	(0,00)	-0,211	(0,00)	-0,529	(0,19)
IDADE[1] * FUMO[2]	0,015	(0,91)	0,004	(0,92)	0,021	(0,14)
Deviance	8,146		3,127		9,514	

que a tendência positiva dada pela forte influência de FUMO[1] $= 0,318$ não é tão intensa quando a pessoa ainda é jovem. Com a codificação dos efeitos das variáveis utilizada, os efeitos de interação não exibidos na Tabela 2.2 são facilmente calculados por meio da restrição de que eles devem somar zero, o que leva à tabela de estimativa das interações a seguir.

Tabela 2.3: Tabela de estimativas dos efeitos de interação.

	FUMO[1]	FUMO[2]	FUMO[3]
IDADE[1]	-0,211	0,004	0,207
IDADE[2]	0,211	-0,004	-0,207

Pela tabela de interações é fácil observar que o histórico de fumo começa a realmente influenciar a partir de idades mais elevadas. Nota-se que a estimativa da interação IDADE[2] * FUMO[3] $= -0,207$ precisa ser adicionada aos efeitos negativos de IDADE[2] $= -0,114$ e FUMO[3] $= -0,541$. As mesmas conclusões são vistas para o modelo logístico, apesar do ajuste ser inferior em comparação ao modelo de riscos

proporcionais.

Exemplo 2: Expectativa de trabalho

Em um estudo sobre as expectativas de estudantes da Universidade de Regentsburg, alunos do curso de Psicologia responderam se esperavam encontrar um emprego adequado dentro de um tempo razoável depois de se formarem. As categorias de resposta foram ordenadas de acordo com as expectativas: 1 (não espera um emprego adequado), 2 (não tem certeza) e 3 (espera um emprego adequado imediatamente após se formar). A Tabela 2.4 apresenta os dados para as diferentes idades dos estudantes.

Tabela 2.4: Dados para expectativa de trabalho dos estudantes de Psicologia em Regensburg.

Idade	Categorias de Resposta			n_i
em anos	1	2	3	
19	1	2	0	3
20	5	18	2	25
21	6	19	2	27
22	1	6	3	10
23	2	7	3	12
24	1	7	5	13
25	0	0	3	3
26	0	1	0	1
27	0	2	1	3
29	1	0	0	1
30	0	0	2	2
31	0	1	0	1
34	0	1	0	1

Foi utilizada como variável explicativa o logaritmo da idade. Dois modelos foram ajustados, o modelo logístico simples cumulativo

$$P(Y \leq r|\text{IDADE}) = F(\theta_r + \gamma \log \text{IDADE}), \quad r = 1, 2, 3$$

e a versão generalizada

$$P(Y \leq r|\text{IDADE}) = F(\beta_{r0} + \beta_r \log \text{IDADE}),$$

onde $\log$ IDADE é uma variável limiar.

Tabela 2.5: Estimativas dos parâmetros e p-valores do Modelo Cumulativo para dados de expectativa de trabalho.

	LOG IDADE como variável global de deslocamento		LOG IDADE como variável limiar	
Limiar 1	14,987	(0,010)	9,467	(0,304)
Limiar 2	18,149	(0,002)	20,385	(0,002)
Coeficiente γ	-5,402	(0,004)	—	
Coeficiente especificador de categorias β_1	—		-3,597	(0,230)
Coeficiente especificador de categorias β_2	—		-6,113	(0,004)
χ^2 de Pearson	42,696	(0,007)	33,503	(0,055)
Deviance	26,733	(0,267)	26,063	(0,248)

As estimativas são dadas na Tabela 2.5. O modelo simples, onde $\log$ IDADE é uma variável de deslocamento, mostrou um ajuste um pouco inferior segundo a estatística quiquadrado de Pearson (χ^2, Seção 2.6). O ajuste da versão generalizada foi razoável. A grande diferença entre as estatísticas deviance e χ^2 de Pearson pode sugerir que as suposições assintóticas associadas ao ajuste do modelo podem ter sido violadas. A versão generalizada cria estimativas para os parâmetros e uma mudança óbvia nos valores limiares. Os valores negativos de $\hat{\beta}_r$ e $(\hat{\gamma})$ sinalizam que o crescimento da idade gera baixas probabilidades para baixas categorias, de modo que os estudantes parecem mais otimistas quando estão próximos ao final do curso. No entanto, a razão de chances cumulativas

$$\frac{P(Y \leq r|x_1)/P(Y > r|x_1)}{P(Y \leq r|x_2)/P(Y > r|x_2)} = \exp(\beta_{r0} + \beta_r(x_1 - x_2))$$

depende da categoria r. Considere os grupos de idade $x_1 > x_2$, então, a chance cumulativa para $r = 1$ mede a tendência em relação a forte expectativa negativa (categoria 1:

não espera um emprego adequado); para $r = 2$, a chance cumulativa mede a tendência em relação a forte expectativa negativa ou incerta (categoria 2: não tem certeza) em comparação à afirmação positiva da categoria 3. Como $|\hat{\beta}_2| > |\hat{\beta}_1|$, o efeito da idade na última chance cumulativa é mais forte que na primeira.

2.3 Modelos com Logitos de Categorias Adjacentes

Quando as categorias de resposta possuem uma ordem natural, os modelos precisam considerar essa ordem, que pode ser incorporada diretamente na maneira como construímos os logitos. Esta seção discute o primeiro tipo de logito para categorias de resposta ordenadas.

Denotando $\pi_r = P(Y = r|x)$, consideramos $\{\pi_1, \ldots, \pi_j\}$ a probabilidade de resposta no valor x para um conjunto de variáveis explicativas. Os logitos para categorias adjacentes são definidos por:

$$L_r = \log\left(\frac{\pi_r}{\pi_{r+1}}\right), \qquad r = 1, \ldots, k-1.$$

A interpretação do logito é a seguinte: a chance de classificação em $Y = r + 1$ ao invés de $Y = r$ é multiplicada por e^{β} para cada variação de uma unidade em x.

Esses logitos são um conjunto básico equivalente aos logitos de categorias *baseline*

$$L_r^* = \log\left(\frac{\pi_r}{\pi_k}\right), \qquad r = 1, \ldots, k-1.$$

Ambos os conjuntos determinam logitos para todos os $\binom{k}{2}$ pares de categorias de resposta.

Podemos expressar modelos logitos de categorias adjacentes como modelos logitos de categorias *baseline*, e ajustá-los utilizando métodos e *softwares* computacionais para os últimos. Por exemplo, suponha que queremos ajustar um modelo logito para categorias-adjacentes

$$L_r = \alpha_r + \beta' x, \qquad j = 1, \ldots, k-1.$$

Poderíamos usar a relação $L_r^* = L_r + L_{r+1} + \cdots + L_{k-1}$ para obter o equivalente

modelo logito de categorias *baseline*

$$L_r^* = \sum_{i=r}^{k-1} \alpha_i + \beta'(k-r)x = \alpha_r^* + \beta' u_r, \qquad r = 1, \ldots, k-1$$

com $u_r = (k-r)x$. O modelo logito de categorias adjacentes corresponde ao modelo logito de categorias *baseline* com a matriz do modelo ajustada.

2.4 Modelos Seqüenciais

Em muitas aplicações, a ordem das categorias de resposta provém de um mecanismo seqüencial e as categorias são ordenadas de tal forma que possam ser alcançadas apenas sucessivamente. Um exemplo é a variável *tamanho das amídalas*, medida no estudo de Holmes e Williams (1954), citado em Fahrmeir e Tutz (1994), descrito a seguir.

Crianças foram classificadas de acordo com o tamanho relativo de suas amídalas e também por serem ou não portadoras de *Streptococcus pyogenes* (Holmes e Williams (1954)). Os dados encontram-se na Tabela 2.6. É razoável admitir que o tamanho das amídalas sempre começa no estado normal. Se as amídalas crescem de forma anormal, elas podem tornar-se "aumentadas". Mas se crescem de maneira exageradamente anormal, primeiro passam pelo estágio "aumentadas", não importa qual seja a duração dessa fase, para depois tornarem-se "bastante aumentadas".

Tabela 2.6: Tamanho das amídalas e Presença de Streptococuss pyogenes.

	Normal	Aumentada	Bastante aumentada
Portadores	19	29	24
Não portadores	497	560	269

Para dados desse tipo, modelos baseados em mecanismos seqüenciais geralmente são mais apropriados. Assim como os modelos cumulativos, os modelos seqüencias podem ser motivados por variáveis latentes.

Considere variáveis latentes U_r, $r = 1, \ldots, k-1$, com a forma linear $U_r = -\mathbf{x}'\gamma + \epsilon_r$, onde ϵ_r é uma variável aleatória com função de distribuição F. O mecanismo funciona

de maneira que a resposta inicia-se na categoria 1 e o primeiro passo é determinado por:

$$Y = 1 \quad \Leftrightarrow \quad U_1 \leq \theta_1$$

onde θ_1 é um parâmetro limiar. Se $U_1 \leq \theta_1$, o processo pára. Para o experimento do tamanho das amídalas, U_1 pode representar a tendência latente de crescimento das amídalas no estado inicial. Se U_1 está abaixo do limiar θ_1, o tamanho das amídalas permanece normal ($Y = 1$), caso contrário, pelo menos está "aumentada" ($Y \geq 2$).

Isso significa que, se $U_1 > \theta_1$, o processo continua, de forma que:

$$Y = 2 \quad \text{dado} \quad Y \geq 2 \quad \Leftrightarrow \quad U_2 \leq \theta_2$$

e assim por diante. A variável latente U_2 pode representar a tendência não observável de crescimento quando as amídalas já estão aumentadas. De forma generalizada, o mecanismo seqüencial completo é especificado por:

$$Y = r \quad \text{dado} \quad Y \geq r \quad \Leftrightarrow \quad U_r \leq \theta_r$$

ou, equivalentemente,

$$Y > r \quad \text{dado} \quad Y \geq r \quad \Leftrightarrow \quad U_r > \theta_r,$$

$r = 1, \ldots, k-1$.

O mecanismo seqüencial modela a transição da categoria r para a categoria $r+1$ dado que a categoria r foi alcançada. Uma transição acontece somente se a variável latente determinante da transição está acima de um limiar característico para a categoria sob consideração.

A principal diferença em relação à aproximação dos limites das categorias é a modelagem condicional das transições. O mecanismo seqüencial assume uma decisão binária em cada passo. Dado que a categoria r foi alcançada, é preciso decidir se o processo pára (chegando em r como categoria final) ou se ele continua, resultando numa categoria maior, sendo que somente a categoria resultante final é observável.

O modelo de resposta seqüencial combinado com uma forma linear das variáveis latentes $U_r = -\mathbf{x}'\gamma + \epsilon_r$ leva imediatamente ao modelo seqüencial com distribuição F:

$$P(Y = r | Y \geq r, \mathbf{x}) = F(\theta_r + \mathbf{x}'\gamma) \tag{2.11}$$

Se $r = 1, \ldots, k$, onde $\theta_k = \infty$. As probabilidades definidas neste modelo são dadas por:

$$P(Y = r|\mathbf{x}) = F(\theta_r + \mathbf{x}'\gamma) \prod_{i=1}^{r-1} \{1 - F(\theta_i + \mathbf{x}'\gamma)\},$$

$r = 1, \ldots, k$, onde $\prod_{i=1}^{0} \{\cdot\} = 1$.

Tal fato pode ser demonstrado, pois de (2.11) temos que:

$$P(Y = r|Y \geq r, \mathbf{x}) = F(\theta_r + \mathbf{x}'\gamma)$$
$$\frac{P(Y = r, Y \geq r, \mathbf{x})}{P(Y \geq r, \mathbf{x})} = \frac{P(Y = r, \mathbf{x})}{P(Y \geq r, \mathbf{x})} = \frac{P(Y = r|\mathbf{x})P(\mathbf{x})}{P(Y \geq r|\mathbf{x})P(\mathbf{x})}$$

ou seja,

$$P(Y = r|\mathbf{x}) = F(\theta_r + \mathbf{x}'\gamma)P(Y \geq r|\mathbf{x}).$$

A partir daí, por indução sobre r segue que $r = 1$,

$$P(Y = 1|\mathbf{x}) = F(\theta_1 + \mathbf{x}'\gamma) \overbrace{P(Y \geq 1|\mathbf{x})}^{1}$$
$$= F(\theta_1 + \mathbf{x}'\gamma) \prod_{i=1}^{0} \{1 - \underbrace{F(\theta_0 + \mathbf{x}'\gamma)}_{0}\}$$
$$= F(\theta_1 + \mathbf{x}'\gamma).1 = F(\theta_1 + \mathbf{x}'\gamma),$$

o que mostra que o resultado vale para $r = 1$. Admitindo que vale para $r - 1$,

$$P(Y = r - 1|\mathbf{x}) = F(\theta_{r-1} + \mathbf{x}'\gamma) \underbrace{\prod_{i=1}^{r-2} \{1 - F(\theta_i + \mathbf{x}'\gamma)\}}_{P(Y \geq r-1|\mathbf{x})}$$

onde $P(Y \geq r - 1|\mathbf{x}) = 1 - P(Y = 1|\mathbf{x}) - \cdots - P(Y = r - 2|\mathbf{x})$.

Verificamos, por fim, que a expressão é válida para $Y = r$, pois

$$P(Y = r|\mathbf{x}) = F(\theta_r + \mathbf{x}'\gamma)P(Y \geq r|\mathbf{x})$$

$$= F(\theta_r + \mathbf{x}'\gamma)[\underbrace{1 - P(Y = 1|\mathbf{x}) - P(y = 2|\mathbf{x}) - \cdots - P(Y = r - 2|\mathbf{x})}_{P(Y \geq r-1|\mathbf{x})} - P(Y = r - 1|\mathbf{x})]$$

$$= F(\theta_r + \mathbf{x}'\gamma)\left[\prod_{i=1}^{r-2}\{1 - F(\theta_i + \mathbf{x}'\gamma)\} - F(\theta_{r-1} + \mathbf{x}'\gamma)\prod_{i=1}^{r-2}\{1 - F(\theta_i + \mathbf{x}'\gamma)\}\right]$$

$$= F(\theta_r + \mathbf{x}'\gamma)(1 - F(\theta_{r-1} + \mathbf{x}'\gamma))\prod_{i=1}^{r-2}\{1 - F(\theta_i + \mathbf{x}'\gamma)\}$$

$$= F(\theta_r + \mathbf{x}'\gamma)\prod_{i=1}^{r-1}\{1 - F(\theta_i + \mathbf{x}'\gamma)\}.$$

É importante observar que, neste modelo, nenhuma restrição de ordenação é necessária para os parâmetros $\theta_1, \ldots, \theta_q$, como ocorreu para o modelo cumulativo.

Até o momento, consideramos a forma geral do modelo seqüencial. Assim como no caso dos Modelos Cumulativos, há diversos modelos seqüenciais, dependendo da escolha da função de distribuição F, que serão descritos a seguir.

2.4.1 Principais Modelos Seqüenciais

Modelo Logístico Seqüencial

Se a escolha de F for a distribuição logística (2.4), obtemos o modelo logístico seqüencial:

$$P(Y = r|Y \geq r, \mathbf{x}) = \frac{\exp(\theta_r + \mathbf{x}'\gamma)}{1 + \exp(\theta_r + \mathbf{x}'\gamma)} \tag{2.12}$$

ou, equivalentemente:

$$\log\left\{\frac{P(Y = r|\mathbf{x})}{P(Y > r|\mathbf{x})}\right\} = \theta_r + \mathbf{x}'\gamma.$$

Este fato pode ser verificado, pois a partir da expressão (2.12), temos:

$$P(Y = r|Y \geq r, \mathbf{x}) + P(Y = r|Y \geq r, \mathbf{x})\exp(\theta_r + \mathbf{x}'\gamma) = \exp(\theta_r + \mathbf{x}'\gamma)$$

$$\frac{P(Y = r|Y \geq r, \mathbf{x})}{1 - P(Y = r|Y \geq r, \mathbf{x})} = \exp(\theta_r + \mathbf{x}'\gamma)$$

$$\frac{P(Y = r|Y \geq r, \mathbf{x})}{P(Y \neq r|Y \geq r, \mathbf{x})} = \exp(\theta_r + \mathbf{x}'\gamma)$$

$$\frac{P(Y = r, Y \geq r, \mathbf{x})}{P(Y \neq r, Y \geq r, \mathbf{x})} = \frac{P(Y = r|\mathbf{x})}{P(Y > r|\mathbf{x})} = \exp(\theta_r + \mathbf{x}'\gamma)$$

$$\log\left\{\frac{P(Y = r|\mathbf{x})}{P(Y > r|\mathbf{x})}\right\} = \theta_r + \mathbf{x}'\gamma.$$

Modelo com distribuição do valor extremo

Utilizando a distribuição do valor extremo $F(z) = 1 - \exp(-\exp(z))$, temos:

$$P(Y = r|Y \geq r, \mathbf{x}) = 1 - \exp\{-\exp(\theta_r + \mathbf{x}'\gamma)\}, \tag{2.13}$$

ou, equivalentemente, com a ligação complementar log-log:

$$\log\left[-\log\left\{\frac{P(Y > r|\mathbf{x})}{P(Y \geq r|\mathbf{x})}\right\}\right] = \theta_r + \mathbf{x}'\gamma \tag{2.14}$$

De (2.13), temos:

$$1 - P(Y = r|Y \geq r, \mathbf{x}) = \exp\{-\exp(\theta_r + \mathbf{x}'\gamma)\}$$

$$\log\{P(Y \neq r|Y \geq r, \mathbf{x})\} = -\exp(\theta_r + \mathbf{x}'\gamma)$$

$$\log\left[-\log\left\{\frac{P(Y \neq r, Y \geq r, \mathbf{x})}{P(Y \geq r, \mathbf{x})}\right\}\right] = \theta_r + \mathbf{x}'\gamma$$

$$\log\left[-\log\left\{\frac{P(Y > r|\mathbf{x})}{P(Y \geq r|\mathbf{x})}\right\}\right] = \theta_r + \mathbf{x}'\gamma$$

Modelo exponencial seqüencial

O Modelo exponencial seqüencial é baseado na distribuição exponencial $F(z) = 1 - \exp(-z)$. É dado por:

$$P(Y = r|Y \geq r, \mathbf{x}) = 1 - \exp(-(\theta_r + \mathbf{x}'\gamma))$$

ou, equivalentemente,

$$-\log\left(\frac{P(Y > r|\mathbf{x})}{P(Y \geq r|\mathbf{x})}\right) = \theta_r + \mathbf{x}'\gamma,$$

pois

$$P(Y = r|Y \geq r, \mathbf{x}) = 1 - \exp(-(\theta_r + \mathbf{x}'\gamma))$$

$$1 - P(Y = r|Y \geq r, \mathbf{x}) = \exp(-(\theta_r + \mathbf{x}'\gamma))$$

$$\log\{P(Y \neq r|Y \geq r, \mathbf{x})\} = -(\theta_r + \mathbf{x}'\gamma)$$

$$-\log\left\{\frac{P(Y \neq r, Y \geq r, \mathbf{x})}{P(Y \geq r, \mathbf{x})}\right\} = \theta_r + \mathbf{x}'\gamma$$

$$-\log\left(\frac{P(Y > r|\mathbf{x})}{P(Y \geq r|\mathbf{x})}\right) = \theta_r + \mathbf{x}'\gamma.$$

2.4.2 Modelos Seqüenciais Generalizados

Da mesma maneira que nos modelos cumulativos, os limiares θ_r dos modelos seqüenciais podem ser determinados pelas covariáveis $\mathbf{z} = (z_1, \ldots, z_m)$. Com isso,

$$\theta_r = \delta_{r0} + z'\delta_r,$$

onde $\delta_r = (\delta_{r1}, \ldots, \delta_{rm})$ é um vetor de parâmetros específicos das categorias, e obtém-se o modelo seqüencial generalizado:

$$P(Y = r|Y \geq r, \mathbf{x}) = F(\delta_{r0} + z'\delta_r + \mathbf{x}'\gamma). \tag{2.15}$$

Alternativamente, o modelo (2.15) pode ser derivado diretamente do mecanismo seqüencial. Deve-se assumir que as variáveis latentes U_r têm a forma $U_r = -\mathbf{x}'\gamma - z'\delta_r + \epsilon_r$ e que o mecanismo de resposta

$$Y > r \quad \text{dado} \quad Y \geq r \quad \text{se} \quad U_r > \delta_{r0}$$

é dado pelos limiares fixados δ_{r0}.

Implicitamente, assume-se que a influência das variáveis $z_j, j = 1, \ldots, n$ na transição da categoria r para $r + 1$ depende da categoria. O efeito dessas variáveis é não homogêneo sobre as categorias, ao passo que o efeito das variáveis $\mathbf{x}$ é homogêneo. Dado que a categoria r é alcançada, a transição para uma categoria maior é sempre determinada por $\mathbf{x}'\gamma$. Se a mudança do escore subjacente é determinada por $\mathbf{x}'\gamma$ e é constante sobre as categorias, analogamente ao modelo geral, as variáveis $\mathbf{x}$ são denominadas variáveis de mudança. Embora a suposição de limiares determinados linearmente por $\theta_r = \delta_{r0} + z'\delta_r$ seja opcional, as variáveis presentes em $\mathbf{z}$ com peso de categoria específica δ_r são chamadas variáveis limiares. Assim, a distinção entre variáveis de mudança e de limiar (correspondendo a pesos globais ou de categorias específicas) é a mesma dos modelos cumulativos.

2.4.3 Funções de Ligação e Matrizes de Planejamento

As funções de ligação ou funções resposta podem ser derivadas diretamente do modelo (2.11). A função de ligação $g = (g_1, \ldots, g_q)'$ é dada por:

$$g_r(\pi_1, \ldots, \pi_q) = F^{-1}[P(Y = r|Y \geq r, \mathbf{x})] = F^{-1}\left(\frac{\pi_r}{1 - \pi_1 - \cdots - \pi_{r-1}}\right), \quad r = 1, \ldots, q,$$

e a função resposta $h = (h_1, \ldots, h_q)'$ tal que $P(Y = r|Y \geq r, \mathbf{x}) = h_r$ tem a forma:

$$h_r(\eta_1, \ldots, \eta_q) = F(\eta_r) \prod_{i=1}^{r-1} (1 - F(\eta_i)), \quad r = 1, \ldots, q.$$

Para o modelo logito seqüencial, por exemplo, temos:

$$g_r(\pi_1, \ldots, \pi_q) = \log \left(\frac{\pi_r}{1 - \pi_1 - \cdots - \pi_{r-1}} \right), \quad r = 1, \ldots, q,$$

e

$$h_r(\eta_1, \ldots, \eta_q) = \exp(\eta_r) \prod_{i=1}^{r-1} (1 + \exp(\eta_i))^{-1}, \quad r = 1, \ldots, q.$$

Para os outros modelos, essas funções podem ser facilmente derivadas.

Por outro lado, as matrizes de planejamento dependem das especificações das variáveis de mudança e limiar. No que diz respeito às matrizes de planejamento, não há diferença entre modelos seqüencias e cumulativos. Assim, as matrizes da Seção 2.2.3 aplicam-se no presente caso.

2.4.4 Exemplos

Fahrmeir e Tutz (1994) apresenta dois exemplos de aplicação dos Modelos Seqüencias abordados.

Exemplo 1: Tamanho das amídalas

A Tabela 2.7 fornece as estimativas dos parâmetros do modelo logístico cumulativo e do modelo logístico seqüencial para os dados com variável resposta tamanho das amídalas.

As estatísticas de Pearson e *deviance* sugerem um melhor ajuste do modelo seqüencial. As estimativas dos parâmetros devem ser interpretadas de acordo com o tipo de modelo utilizado. A estimativa $\hat{\gamma} = -0,301$ para o modelo cumulativo significa que a chance $P(Y \leq r|\mathbf{x})/P(Y > r|\mathbf{x})$ de ter amídalas de tamanho normal (r=1), assim como a chance de ter tamanho normal ou aumentada (r=2) é igual a $\exp((x_1 - x_2)'\gamma) = \exp(-0,301(-1 - 1)) \approx 1,8$ vezes a chance para não portadores ($x = -1$) que

Tabela 2.7: Ajustes para os dados das amídalas (estimativas dos parâmetros e p-valores.)

	Modelo cumulativo logístico		Modelo seqüencial logístico	
θ_1	-0,809	(0,013)	-0,775	(0,011)
θ_2	1,061	(0,014)	0,468	(0,012)
γ	-0,301	(0,013)	-0,264	(0,010)
Pearson	0,301		0,005	
Deviance	0,302		0,006	
Graus de liberdade	1		1	

para portadores de *Streptococuss pyogenes* ($x = 1$). Dentro dos modelos seqüenciais, o parâmetro γ dá a força com qual a transição da categoria 1 para 2, e da 2 para 3, é determinada. O valor estimado $\hat{\gamma} = -0,264$ significa que a chance $P(Y = r|\mathbf{x})/P(Y > r|\mathbf{x})$ de ter tamanho normal de amídala ($r = 1$) é igual a $\exp((x_1 - x_2)'\gamma) = \exp(-0,264(-1-1)) \approx 1,7$ vezes maior para não portadores do que para portadores de *Streptococcus pyogenes*. A mesma proporção vale para a chance $P(Y = 2|\mathbf{x})/P(Y \leq 2|\mathbf{x})$ de ter as amídalas meramente aumentadas dado que as amídalas não são normais.

Exemplo 2: Resultados de exames respiratórios

O efeito da idade e do histórico de fumo sobre os resultados de exames respiratórios já foram investigados utilizando o modelo cumulativo. No entanto, as categorias, "normal", "aumentadas" e "bastante aumentadas" podem ser vistas como um mecanismo seqüencial começando de "normal". Conseqüentemente, o modelo seqüencial pode ser utilizado para esse tipo de conjunto de dados. A Tabela 2.8 dá as estimativas de dois modelos seqüenciais logísticos.

O primeiro é do tipo simples (2.11) com IDADE, FUMO e a interação IDADE * FUMO. O segundo é do tipo generalizado (2.15), onde IDADE é uma variável de deslocamento (com peso global), e FUMO e a interação IDADE * FUMO são variáveis limiares (com peso das categorias específicas). O *deviance* para o primeiro modelo é

Tabela 2.8: Modelos seqüenciais logísticos para os dados de testes respiratórios (p-valores entre parênteses).

	Modelo Seqüencial Logístico Simples		Modelo Seqüencial Logístico com variáveis limiares	
Limiar 1	2,379	(0,00)	2,379	(0,00)
Limiar 2	1,516	(0,00)	1,510	(0,00)
IDADE	0,094	(0,37)	0,092	(0,39)
FUMO[1]	0,882	(0,00)	0,915	(0,00)
			0,675	(0,11)
FUMO[2]	-0,356	(0,01)	-0,375	(0,01)
			-0,163	(0,61)
IDADE * FUMO[1]	-0,601	(0,00)	-0,561	(0,00)
			-0,894	(0,05)
IDADE * FUMO[2]	0,092	(0,49)	0,015	(0,91)
			0,532	(0,16)

4,310 sobre 5 graus de liberdade, o que mostra que o modelo seqüencial ajusta os dados melhor que o modelo cumulativo. As estimativas do modelo generalizado são dadas para comparação. É visível que para a variável FUMO os efeitos de categoria específica não são tão diferentes para os dois limiares. O primeiro tem efeitos significantes, o segundo limiar tem p-valores um pouco altos, efeito devido ao baixo número de observações na categoria 3. Os limiares para a interação IDADE * FUMO[2] têm efeitos um pouco diferentes mas ambos chegam a ser não significantes.

2.5 Modelos em Dois Estágios

Ambos os tipos de modelos, cumulativo e seqüencial, só fazem uso da ordenação das categorias de resposta $1, \ldots, k$. No entanto, é freqüente as categorias de resposta poderem ser divididas de forma natural em conjuntos de categorias com respostas muito homogêneas, sendo os conjuntos heterogêneos entre si.

Mehta, Patel e Tsiatis (1984) analisaram dados de pacientes com artrite. Um novo medicamento foi comparado com um tratamento controle e cada paciente foi avaliado segundo uma escala de 5 pontos: "muita melhora", "melhora", "nenhuma mudança", "piora" e "muita piora". Os dados encontram-se na Tabela 2.9. Uma análise global nesse exemplo pode ser subdividida em uma resposta grosseira do tipo: "melhora", "nenhuma mudança" e "piora", sendo que as primeiras categorias de melhora foram unidas, assim como as duas últimas categorias de piora. Como resultado, temos três conjuntos de resposta internamente homogêneos.

Para dados desse tipo, é útil modelar primeiramente a resposta grosseira e, em seguida, a resposta que está dentro dos grupos homogêneos. Generalizando, é necessário subdividir as categorias $1, \ldots, k$ em t conjuntos básicos $S_1, \ldots, S_t$ onde $S_j = \{m_{j-1} + 1, \ldots, m_j\}$, $j = 1, \ldots, t$, $m_0 = 0$, $m_t = k$.

Tabela 2.9: Tratamento clínico de um novo medicamento e tratamento controle

	Avaliação global				
Medicamento	Muita melhora	Melhora	Nenhuma mudança	Piora	Muita piora
Novo	24	37	21	19	6
Controle	11	51	22	21	7

No primeiro passo, a resposta em um dos conjuntos é determinada por um modelo cumulativo, baseado numa variável latente subjacente $U_0 = -\mathbf{x}'\gamma_0 + \epsilon$, onde a variável aleatória ϵ tem função de distribuição F. O mecanismo de resposta é dado por:

$$Y \in S_j \quad \Leftrightarrow \quad \theta_{j-1} < U_0 \leq \theta_j.$$

No segundo passo, o mecanismo condicional em S_j é determinado por um modelo cumulativo baseado numa variável latente $U_j = -\mathbf{x}'\gamma_j + \epsilon_j$, com ϵ_j também seguindo função de distribuição F. Assumimos:

$$Y = r|Y \in S_j \quad \Leftrightarrow \quad \theta_{j,r-1} < U_j \leq \theta_{jr}.$$

Supondo que ϵ_j sejam variáveis aleatórias independentes com média 0, o modelo resultante é dado por:

$$P(Y \in T_j|\mathbf{x}) = F(\theta_j + \mathbf{x}'\gamma_0)$$

$$P(Y \leq r|Y \in S_j, \mathbf{x}) = F(\theta_{jr} + \mathbf{x}'\gamma_j) \tag{2.16}$$

onde $T_j = S_1 \cup \ldots \cup S_j,\ \theta_1 < \ldots < \theta_{t-1},\ \theta_t = \infty,$

$$\theta_{j,m_{j-1}+1} < \ldots < \theta_{j,m_j-1},\ \theta_{j,m_j} = \infty$$

$j = 1, \ldots, t.$

O processo subjacente em ambos os casos é baseado em mecanismos cumulativos. Assim, o modelo é chamado de cumulativo em dois estágios.

Uma vantagem do modelo em dois estágios é que parâmetros diferentes estão envolvidos em passos diferentes. A escolha entre os conjuntos básicos é determinada pelo parâmetro γ_0, a escolha sobre o nível mais alto é determinada pelos parâmetros γ_i. Assim, no modelo, a influência das variáveis explicativas sobre as variáveis dependentes pode variar. A escolha entre os conjuntos básicos, por exemplo as alternativas "melhora" ou "não melhora", pode ser influenciada por fatores diferentes ou mesmo por variáveis diferentes da escolha dentro do conjunto "melhora" e do conjunto "não melhora".

O modelo (2.16) é apenas uma representação dos modelos em dois estágios. É possível obter modelos alternativos, por exemplo, o modelo cumulativo-seqüencial, se o primeiro passo é baseado num modelo cumulativo e o segundo passo é baseado num modelo seqüencial. Um modelo desse tipo é apropriado se tivermos três conjuntos básicos, onde S_2 é um tipo de ponto de partida e S_1, S_3 são estágios de mudança em diferentes direções. O exemplo da artrite é desse tipo. Entretanto, neste caso, S_1 e S_3 possuem apenas duas categorias e por isso não há diferença entre um modelo cumulativo ou seqüencial no segundo passo. Para modelos alternativos, ver Tutz (1989) e Morawitz e Tutz (1990).

2.5.1 Funções de Ligação e Matriz de Planejamento

Os modelos em dois estágios podem ser escritos na forma de modelos lineares generalizados. Para o modelo cumulativo, a função de ligação é dada por:

$$g_j(\pi_1, \ldots, \pi_q) \ = \ F^{-1}(\pi_1 + \pi_2 + \cdots + \pi_{m_j}),\ j = 1, \ldots, t,$$

$$
\begin{aligned}
g_{jr}(\pi_1, \ldots, \pi_q) &= F^{-1}\left\{\frac{\pi_{m_{j-1}+1} + \cdots + \pi_{m_{j-1}+1+r}}{\pi_{m_{j-1}+1} + \cdots + \pi_{m_j}}\right\} \\
j = 1, \ldots, t, & \qquad r = 1, \ldots, m_j - 1.
\end{aligned}
$$

A matriz de planejamento é um pouco mais complicada que a do modelo cumulativo simples e a do modelo seqüencial. É da forma:

$$
\begin{bmatrix}
1 & & & \mathbf{x}_i' & & & & & & & & & & & & \\
& \ddots & & \vdots & & & & & & & & & & & & \\
& & 1 & \mathbf{x}_i' & & & & & & & & & & & & \\
& & & & 1 & & & \mathbf{x}_i' & & & & & & & & \\
& & & & & \ddots & & \vdots & & & & & & & & \\
& & & & & & 1 & \mathbf{x}_i' & & & & & & & & \\
& & & & & & & & 1 & & & \mathbf{x}_i' & & & & \\
& & & & & & & & & \ddots & & \vdots & & & & \\
& & & & & & & & & & 1 & \mathbf{x}_i' & & & & \\
& & & & & & & & & & & & 1 & & & \mathbf{x}_i' \\
& & & & & & & & & & & & & \ddots & & \vdots \\
& & & & & & & & & & & & & & 1 & \mathbf{x}_i'
\end{bmatrix}
$$

com vetor de parâmetros

$$
\beta' = (\theta_1, \ldots, \theta_t, \gamma_0', \theta_{11}, \ldots, \theta_{1,m_1-1}, \gamma_1', \ldots, \theta_{t1}, \ldots, \theta_{t,m_t-1}, \gamma_t').
$$

2.6 Inferência Estatística

A Análise de Regressão em modelos lineares generalizados é baseada na estimação dos parâmetros pelo método da máxima verossimilhança. Esta seção contém as ferramentas inferenciais básicas para estimação dos parâmetros, testes de hipóteses e testes de qualidade do ajuste. Assume-se, portanto, que o modelo está completamente e corretamente especificado dentro da classe de modelos lineares generalizados. Em muitas aplicações, essa suposição pode ser idealista demais. Neste caso, modelos de quasi-verossimlhança, onde apenas a média e a função de variância precisam ser especificadas, devem então ser considerados.

2.6.1 Estimação por Máxima Verossimilhança

Dada a amostra de vetores $y_1, \ldots, y_i, \ldots$, juntamente com as covariáveis $x_1, \ldots, x_i, \ldots$, e vetores de planejamento $z_1, \ldots, z_i$, a abordagem do Modelos Lineares Generalizados utiliza um estimador de máxima verossimilhança do vetor de parâmetros desconhecidos β no modelo $E(y_i|x_i) = \mu_i = h(x_i'\beta)$.

A seguir, será descrito o método de estimação na forma generalizada de famílias exponenciais multivariadas. Com base na função densidade da família exponencial

$$f(y_i|\theta_i, \phi, \omega_i) = \exp\left\{\frac{y_i'\theta_i - b(\theta_i)}{\phi}\ \omega_i + c(y_i, \phi, \omega_i)\right\}$$

para os vetores observados $y_1, \ldots, y_n$, a estimação por máxima verossimilhança pode ser derivada de forma análoga ao caso unidimensional. O logaritmo da função de verossimilhança para a amostra tem a forma

$$l(\beta) = \sum_{i=1}^{n} l_i(\mu_i),$$

com

$$l_i(\mu_i) = \frac{y_i'\theta_i - b(\theta_i)}{\phi}\omega_i, \qquad \theta_i = \theta(\mu_i).$$

Para tratar casos de dados individuais ($i = 1, \ldots, n$) e de dados agrupados ($i = 1, \ldots, g$), simultaneamente, omitimos na notação n ou g como limite superior nos somatórios. Logo, a soma pode ir de 1 a n ou de 1 a g, e os pesos w_i devem ser iguais a 1 para dados indiviuais e n_i para dados agrupados.

Primeiramente, assumimos que o parâmetro escalar ϕ é conhecido. Como ϕ aparece como um fator na verossimilhança, consideramos $\phi = 1$, neste caso, sem perda de generalidade se estamos apenas interessados em uma estimativa pontual de β. Note, no entanto, que ϕ (ou uma estimativa consistente) é necessário para calcular variâncias do estimador de máxima verossimilhança. O parâmetro ϕ pode ser considerado um parâmetro de superdispesão, e pode ser tratado formalmente da mesma maneira que um parâmetro escalar.

Para evitar complexidades adicionais em torno do parâmetro de identificablidade, assume-se, neste momento, que a matriz de planejamento

$$Z = (z_1, \ldots, z_i, \ldots)'$$

tem posto completo p, ou, equivalentemente,

$$\sum_i z_i z_i' = Z'Z$$

tem posto p.

Note que, para o caso agrupado, n será igual ao número de grupos, enquanto para o caso não agrupado, n é o número de observações. Usando a ligação $\mu_i = h(Z_i\beta)$, a função escore $s(\beta) = \partial l/\partial\beta = \sum_{i=1}^{n} s_i(\beta)$ tem componentes

$$s_i = Z_i' \; D_i(\beta) \; \Sigma_i^{-1}(\beta) \; [y_i - \mu_i(\beta)]$$

onde

$$D_i(\beta) = \frac{\partial h(\eta_i)}{\partial \eta}$$

é a derivada de $h(\eta)$ calculada em $\eta_i = Z_i\beta$ e

$$\Sigma_i(\beta) = cov(y_i)$$

denota a matriz de covariância do vetor de observações y_i dado o vetor de parâmetros β. A forma alternativa

$$s_i(\beta) = Z_i' \; W_i(\beta) \; \frac{\partial g(\mu_i)}{\partial \mu'} \; [y_i - \mu_i(\beta)]$$

faz uso da matriz ponderada

$$W_i(\beta) = D_i(\beta) \; \Sigma_i^{-1}(\beta) \; D_i'(\beta) = \left\{ \frac{\partial g(\mu_i)}{\partial \mu'} \; \Sigma_i(\beta) \; \frac{\partial g(\mu_i)}{\partial \mu} \right\}^{-1},$$

que pode ser considerada uma aproximação do inverso da matriz de covariância da observação "transformada" $g(y_i)$ em casos onde $g(y_i)$ existe. A informação de Fisher esperada é dada por

$$F(\beta) = cov(s(\beta)) = \sum_{i=1}^{n} Z_i' \; W_i(\beta) \; Z_i.$$

Na notação matricial, a função escore e a matriz de Fisher têm a mesma forma que no caso univariado:

$$s(\beta) = Z' \; D(\beta) \; \Sigma^{-1}(\beta) \; [y - \mu(\beta)], \qquad F(\beta) = Z' \; W(\beta) \; Z$$

onde y e $\mu(\beta)$ são dados por

$$y' = (y_1', \ldots, y_n'), \qquad \mu(\beta)' = (\mu_1(\beta)', \ldots, \mu_n(\beta)').$$

As matrizes são da forma bloco diagonal

$$\Sigma(\beta) = \text{diag}(\Sigma_i(\beta)), \qquad W(\beta) = \text{diag}(W_i(\beta)), \qquad D(\beta) = \text{diag}(D_i(\beta))$$

e a matriz de planejamento total é dada por

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{bmatrix}.$$

Essas fórmulas são dadas para observações individuais $y_1, \ldots, y_n$. Para observações agrupadas, que são mais convenientes computacionalmente, as fórmulas são as mesmas. A única diferença é que a soma é sobre as observações agrupadas $y_1, \ldots, y_g$ onde y_i é a média sobre n_i observações, e $\Sigma_i(\beta)$ é substituído por $\Sigma_i(\beta)/n_i$. Sob as suposições de regularidade, obtemos a normalidade assintótica do estimador de máxima verossimilhança de β,

$$\hat{\beta} \overset{a}{\sim} N(\beta, F^{-1}(\hat{\beta})).$$

Dessa forma, assintoticamente, $\hat{\beta}$ tem distribuição aproximadamente normal com matriz de covariância $\text{cov}(\hat{\beta}) = F^{-1}(\hat{\beta})$.

A estimativa de máxima verossimilhança de β pode ser obtida pelo método de Escore de Fisher como

$$\hat{\beta}^{k+1} = \hat{\beta}^k + (Z'W(\hat{\beta}^k)Z)^{-1}s(\hat{\beta}^k)$$

sendo $\hat{\beta}^k$ a estimativa de β no passo k.

2.6.2 Testes de hipóteses e Análise da Qualidade do Ajuste

Teste de hipóteses lineares

A maior parte dos problemas testados para β são da forma $H_0 : C\beta = \xi$ contra $H_1 : C\beta \neq \xi$, no qual a matriz C possui posto completo $S \leq p$. Um caso especial e

importante é considerar $H_0 : \beta_r = 0$ contra $H_1 : \beta_r \neq 0$, onde β_r é um subvetor de β. Isso corresponde a testar o submodelo definido por $\beta_r = 0$ contra o modelo completo. A seguir, assumimos que parâmetros escalares ou de overdispersão desconhecidos ϕ são substituídos por estimativas consistentes.

A estatística de razão de verossimilhanças

$$\lambda = -2\{l(\tilde{\beta}) - l(\hat{\beta})\}$$

compara o máximo irrestrito $l(\hat{\beta})$ da (log-)verossimilhança com o máximo $l(\tilde{\beta})$ obtido pelo estimador de máxima verossimilhança restrito $\tilde{\beta}$, calculado sob a restrição $C\beta = \xi$ de H_0. Se o máximo irrestrito $l(\hat{\beta})$ é significantemente maior que $\tilde{\beta}$, implicando que λ é grande, H_0 será rejeitada em favor de H_1. Testar um submodelo definido por $\beta_r = 0$ requer novas interações de escore para o submodelo. No caso de parâmetros escalares desconhecidos ϕ, todos os resultados permanecem válidos se esse parâmetro for substituído por um estimador consistente $\hat{\phi}$.

Existe ainda a possibilidade de se utilizar o teste de Wald e o teste de escore.

O teste de hipóteses lineares da forma $H_0 : C\beta = \xi$ contra $H_1 : C\beta \neq \xi$ no caso multivariado é o mesmo que o considerado para o caso univariado, com a diferença de se substituir as funções de escore e as matrizes de Fisher por suas versões multivariadas.

Estatísticas de qualidade do ajuste

A qualidade do ajuste dos modelos pode ser avaliada pelas estatísticas de Pearson e *deviance*. Como estamos considerando dados multinomiais, o vetor de médias das variáveis aleatórias μ_i é o vetor de probabilidade $\pi_i = (\pi_{i1}, \ldots, \pi_{iq})$, $\pi_{ik} = 1 - \pi_{i1} - \cdots - \pi_{iq}$, onde $\pi_{ir} = P(Y = r|x_i)$. A estimativa de π_i baseada no modelo é denotada por $\hat{\mu}_i = \hat{\pi}_i = (\hat{\pi}_{i1}, \ldots, \hat{\pi}_{iq})$. Para o cálculo das estatísticas de ajuste, os dados devem ser agrupados de modo que o vetor de observações $y_i = (y_{i1}, \ldots, y_{iq})$ seja formado pelas freqüências relativas das categorias dentro do i-ésimo grupo.

A estatística de Pearson geral é dada por

$$\chi^2 = \sum_{i=1}^{g} (y_i - \hat{\mu}_i)' \; \Sigma_i^{-1}(\hat{\beta}) \; (y_i - \hat{\mu}_i).$$

No caso de variável resposta multicategórica com distribuição multinomial $n_i y_i \sim M(n_i, \pi_i)$, χ^2 pode ser escrita de uma forma mais familiar

$$\chi^2 = \sum_{i=1}^{g} \chi_P^2(y_i, \hat{\pi}_i)$$

onde

$$\chi_P^2(y_i, \hat{\pi}_i) = n_i \sum_{j=1}^{k} \frac{(y_{ij} - \hat{\pi}_{ij})^2}{\hat{\pi}_{ij}}$$

é o resíduo de Pearson para a i-ésima observação (agrupada) com $y_{ik} = 1 - y_{i1} - \cdots - y_{iq}$, $\hat{\pi}_{ik} = 1 - \hat{\pi}_{i1} - \ldots - \hat{\pi}_{iq}$. O *deviance* ou estatística de razão de verossimilhanças é dada por

$$D = -2 \sum_{i=1}^{g} \{l_i(\hat{\pi}_i) - l_i(y_i)\}.$$

Para dados multinoniais, a forma mais familiar é dada por

$$D = 2 \sum_{i=1}^{g} \chi_D^2(y_i, \hat{\pi}_i)$$

onde

$$\chi_D^2(y_i, \hat{\pi}_i) = n_i \sum_{j=1}^{k} y_{ij} \log\left(\frac{y_{ij}}{\hat{\pi}_{ij}}\right)$$

é o resíduo deviance. Se $y_{ij} = 0$, o termo $y_{ij} \log(y_{ij}/\hat{\pi}_{ij})$ tende a zero.

Sob "condições de regularidade" incluindo, em particular, aumento do tamanho da amostra $n_i \to \infty, i = 1, \ldots, g$ tal que $n_i/n \to \lambda_i > 0$, obtemos a distribuição aproximadamente quiquadrado das estatísticas de qualidade do ajuste

$$\chi^2, D \overset{a}{\sim} \chi^2(g(k-1) - p)$$

onde g denota o número de grupos, k é o número de categorias da resposta e p é o número de parâmetros estimados. Para dados escassos com n_i pequeno, alternativas assintóticas podem ser consideradas.

2.7 Considerações Finais

Nesse capítulo, procuramos descrever os principais modelos de regressão para variável resposta do tipo ordinal. Maior ênfase foi dada à apresentação dos mode-

los, já que a inferência estatística associada é feita através da metodologia de modelos lineares generalizados.

O próximo capítulo descreverá o problema de discriminação e classificação de elementos em mais detalhes e apresentará algumas opções de preditores da variável resposta categórica ordinal.

Capítulo 3

Discriminação entre grupos ordenados

3.1 Introdução

O objetivo do presente capítulo é descrever o problema da discriminação e classificação de elementos em populações ou grupos que podem ser ordenados segundo algum critério. Para que essa ordenação faça sentido, o número de grupos deve ser maior ou igual a três.

O problema clássico da discriminação é predizer a qual grupo $(\Pi_1, \Pi_2, \ldots, \Pi_k)$ um determinado indivíduo pertence baseado em observações $\mathbf{x}$ feitas sobre ele. De acordo com Anderson e Philips (1981), o resultado bem conhecido de que a probabilidade de correta alocação é maximizada pela regra que aloca um indivíduo em Π_r se $\hat{P}(\Pi_r|\mathbf{x}) \geq \hat{P}(\Pi_t|\mathbf{x})$, $t = 1, \ldots, k$, vale tanto para grupos ordenados como para grupos não ordenados.

Suponha que a ordenação do grupo Π_r é tal que Π_r é classificada abaixo de Π_t se $r < t$. Então, os modelos apresentados no capítulo anterior para $P(Y = r|\mathbf{x})$ podem ser utilizados para estimar $P(\Pi_r)$ e identificar Π_r com o valor da variável categórica $Y = r$, levando a uma solução para o problema da discriminação ordenada descrita. Por consistência, vamos utilizar a notação anterior e nos referir ao r-ésimo grupo ordenado Π_r $(r = 1, \ldots, k)$ como a r-ésima categoria de Y.

Anderson e Philips (1981) considera o problema de predizer Y a partir de $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ usando o modelo logístico (2.4). O modelo é da forma

$$P(Y \leq r|\mathbf{x}) = \frac{\exp(\theta_r - \beta'\mathbf{x})}{1 + \exp(\theta_r - \beta'\mathbf{x})}, \qquad r = 0, 1, \ldots, k, \tag{3.1}$$

onde $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_{k-1}$, $\theta_0 = -\infty$ e $\theta_k = +\infty$. O vetor $\beta = (\beta_1, \ldots, \beta_k)'$ representa os coeficientes desconhecidos da regressão e os parâmetros presentes no vetor $\theta = (\theta_1, \ldots, \theta_{k-1})'$ também são desconhecidos.

Os autores consideram ainda a existência de uma variável latente não observável, z, tal que

$$y = r \quad se \quad \theta_{r-1} \leq z < \theta_r, \quad r = 1, 2, \ldots, k.$$

Além disso, para o modelo (3.1), verificam que $E(z|\mathbf{x}) = \mathbf{x}\beta$.

A partir de um conjunto de dados apropriados, os parâmetros β e θ serão estimados por $\hat{\beta}$ e $\hat{\theta}$, que podem ser substituídos em (3.1) para obtenção das estimativas de $P(Y = r|\mathbf{x}), s = 1, \ldots, k$, das probabilidades $P(\Pi_r|\mathbf{x})$.

Marshall e Olkin (1968), citado em Anderson e Philips (1981), propõe utilizar como previsor de y a categoria $\tilde{y}$ com a maior estimativa desta probabilidade. Sob esse critério de maximização, está sendo considerado que todas as alocações incorretas possuem igual consequência. No caso de categorias ordenadas, no entanto, a alocação incorreta em uma categoria vizinha não é tão séria como uma alocação incorreta em uma categoria mais extrema. Mesmo assim, por acreditarem que não existe uma função de perda generalizada que leve em conta esse fato, os autores utilizam apenas o previsor $\tilde{y}$ que maximiza a estimativa da probabilidade $P(\Pi_r|\mathbf{x})$.

Adicionalmente, na busca de um preditor de y com boas propriedades, Anderson e Philips (1981) sugere uma diferente abordagem. Dado $\mathbf{x}$, um preditor de z é $\hat{z} = \hat{\beta}'\mathbf{x}$ e y pode ser predito por $\hat{y}$ onde $\hat{y} = r$ se $\hat{\theta}_{r-1} \leq \hat{z} < \hat{\theta}_r, r = 1, \ldots, k$. Segundo os autores, tal procedimento, além de muito simples, teria o atrativo adicional da ligação com a idéia de uma escala de medida contínua z.

Os dois preditores $\tilde{y}$ e $\hat{y}$ foram comparados no exemplo da área médica, descrito a seguir.

3.2 Exemplo

No exemplo considerado, foram analisados 101 pacientes que sofriam de dores nas costas e tiveram várias características observadas inicialmente. Tais pacientes foram tratados por um de diversos métodos e, após três semanas, o progresso obtido (variável Y) foi avaliado como: 1 (pior), 2 (igual), 3 (pequena melhora), 4 (melhora moderada), 5 (grande melhora) e 6 (alívio completo). O objetivo do estudo era verificar quão bem o *status* de melhora (identificado por Y) poderia ser predito a partir dos dados de pré-tratamento. Investigações preliminares da relação de Y com cada uma das variáveis preditoras potenciais, uma de cada vez, levaram à seleção de oito possíveis preditoras $\mathbf{X}$, indicadas na Tabela 3.1. Todas as variáveis preditoras são binárias, exceto a variável X_6, que possui três níveis ordenados, aos quais foram associados postos 1, 2 e 3.

O modelo logístico (3.1) para a probabilidade condicional $P(Y = r|\mathbf{x})$ foi adotado devido a facilidades computacionais e a inferência foi baseada em L_c, a função de verossimilhança do modelo.

Na seleção de variáveis preditoras para compor o modelo final, foi utilizado um procedimento de stepwise que incluia uma nova variável no modelo se $2 \log L_c$ excedia $3,8$, que é o quantil de ordem $0,95$ da distribuição qui-quadrado com um grau de liberdade.

Com base nesses critérios, o melhor conjunto de variáveis preditoras obtido era formado pelas variáveis X_3, X_6 e X_7. Tal fato evidencia a hipótese de que há uma relação significante entre a resposta Y e as variáveis preditoras X_3, X_6 e X_7, e indica que há um ganho muito pequeno se forem incluídas as variáveis preditoras restantes.

A estimativa do veetor de parâmetros θ é:

$$\hat{\theta} = (-7,94; -6,37; -5,38; -4,42; -2,63)$$

com estimativas dos erros padrão (1,12; 1,04; 1,01; 0,98; 0,90). As estimativas dos parâmetros do vetor β considerando as variáveis $\mathbf{x}$ selecionadas são os coeficientes da equação:

$$\hat{z} = -1,51x_3 - 0,49x_6 - 0,87x_7$$

Tabela 3.1: Variáveis utilizadas no estudo.

Variável	**Definição da variável**	**Valores possíveis**
X_1	Ataques anteriores	0 (nenhum)
		1 (um ou mais)
X_2	Tipo de ataque atual	1 (repentino)
		2 (gradual)
X_3	Duração do ataque anterior	1 (curta)
		2 (longa)
X_4	Formigamento ou dormência nas pernas	1 (sim)
		2 (não)
X_5	Tosse causa dor	1 (sim)
		2 (não)
X_6	Status de progresso da dor	1 (melhor)
		2 (igual)
		3 (pior)
X_7	Lordose	1 (presente/aumentando)
		2 (ausente/reduzindo)
X_8	Limitação de flexão pela dor	1 (sim)
		2 (não)

e as estimativas dos respectivos erros padrão são (0,39; 0,25; 0,36).

Com o objetivo de obter o previsor $\hat{y}$, os valores de $\hat{z}$ foram calculados para os 101 pacientes no estudo. Há apenas 12 valores distintos para $\hat{z}$, já que X_3 e X_7 são variáveis binárias, enquanto X_6 tem três categorias. O vetor $\hat{\theta}$ foi acrescido dos valores $\hat{\theta}_0 = -\infty$ e $\hat{\theta}_6 = \infty$. Desta forma, o previsor $\hat{y}$ é dado por

$$\hat{y} = \begin{cases} 1, & \text{se} \quad \hat{z} \leq -7,94 \\ 2, & \text{se} \quad -7,94 < \hat{z} \leq -6,37 \\ 3, & \text{se} \quad -6,37 < \hat{z} \leq -5,38 \\ 4, & \text{se} \quad -5,38 < \hat{z} \leq -4,42 \\ 5, & \text{se} \quad -4,42 < \hat{z} \leq -2,63 \\ 6, & \text{se} \quad \hat{z} > -2,63. \end{cases}$$

A Tabela 3.2 apresenta as seis categorias da variável Y, os 12 valores de $\hat{z}$ e a correspondente previsão $\hat{y}$. A principal informação contida na tabela é a classificação conjunta das variáveis y e $\hat{y}$. Há uma concordância absoluta entre $\hat{y}$ e y para 31% dos pacientes enquanto 76% dos casos estão corretamente alocados, ou alocados para um valor de y adjacente ao valor correto. No entanto, embora existam 35 pacientes nas categorais 1, 2 e 6, para nenhum deles $\hat{y} = 1, 2, 6$, implicando que as categorias preditas estão agrupadas mais próximas da categoria central do que as categorias reais. De acordo com os autores, tal fato pode estar relacionado a um viés dos estimadores dos parâmetros β em modelos de regressão logística ajustados com pequenos tamanhos de amostra. Esse viés tende a ocorrer também em regressões logísticas ordinais e pode ser reduzido usando o método proposto por Anderson e Richardson (1979).

Tabela 3.2: Classificação conjunta das variáveis y e $\hat{y}$.

$\hat{y}$	z \ y	1	2	3	4	5	6	Total
3	-6,23	2	-	2	3	-	-	20
	-5,74	1	4	4	3	-	1	
4	-5,36	2	2	1	5	2	-	41
	-5,25	-	1	-	-	3	-	
	-4,87	-	3	4	5	6	2	
	-4,72	-	-	1	1	3	-	
5	-4,38	-	-	3	-	1	2	40
	-4,23	-	1	-	2	-	1	
	-3,85	-	-	-	-	2	2	
	-3,74	-	-	-	1	3	-	
	-3,36	-	2	3	-	6	4	
	-2,87	-	1	-	-	2	4	
Total		5	14	18	20	28	16	101

Os resultados do preditor alternativo $\tilde{y}$ são dados na Tabela 3.3. Observa-se uma concordância absoluta entre $\tilde{y}$ e y para 33% dos pacientes e 70% foram corretamente alocados, ou alocados para valores de y adjacentes ao valor correto. Tais resultados são similares aos obtidos com $\hat{y}$ e, conforme esperado, $\tilde{y}$ fornece mais pacientes corretamente alocados do que $\hat{y}$. Entretanto, embora y=4 seja o segundo valor mais frequentemente

observado da variável categórica Y, não se observa $\tilde{y} = 4$ para nenhum paciente. Isso sugere que nesta aplicação $\tilde{y}$ é menos satisfatório, mas a escolha entre $\hat{y}$ e $\tilde{y}$ como preditor de y depende do contexto.

Tabela 3.3: Classificação conjunta das variáveis y e $\tilde{y}$

y	$\tilde{y}$						Total
	1	2	3	4	5	6	
1	0	3	2	0	0	0	5
2	0	4	2	0	7	1	14
3	0	6	1	0	11	0	18
4	0	6	5	0	9	0	20
5	0	0	2	0	24	2	28
6	0	1	0	0	11	4	16
Total	0	20	12	0	62	7	101

Ambos os procedimentos têm a vantagem de permitir que variáveis preditoras contínuas sejam usadas tão facilmente quanto as binárias.

Finalizando o estudo, os autores refazem a análise combinando as categorias 3, 4 e 5 em uma única, denominada "categoria de melhora", e utilizam o modelo (3.1) com um número menor de parâmetros. A análise obtida se mostrou semelhante à anterior. No entanto, os autores sugerem o uso desse procedimento com cautela. Alertam ainda para o fato de que a união de categorias não pode ser considerada como uma solução para o problema de erros de medida em variáveis categóricas.

3.3 Considerações Finais

A principal contibuição de Anderson e Philips (1981) foi o uso de modelos de regressão para categorias ordenadas no problema de discriminação e classificação.

Esse enfoque tem a vantagem de permitir a análise quando existem $k \geq 3$ categorias ordenáveis, além de contemplar a situação em que as variáveis preditoras são qualitativas e quantitativas.

Devido ao caráter ordinal da variável resposta, recomendamos que, nesta análise, além da porcentagem de classificação correta, seja também avaliada, como no exemplo, a porcentagem de classificação em categorias adjacentes.

O próximo capítulo discutirá uma diferente proposta para o problema de classificação em categorias ordenadas.

Capítulo 4

Análise Discriminante Ótima para Respostas Ordinais

4.1 Introdução

Conforme comentado anteriormente, uma resposta categórica ordinal pode ser relacionada a um grupo de covariáveis independentes (preditoras), que podem ser medidas numa escala nominal, ordinal, intervalar ou razão. Por isso, modelos têm sido criados para a análise de variáveis resposta ordinais, de forma a descrever a relação entre elas e variáveis explicativas, ou, mais freqüentemente, para prever a que categoria da resposta indivíduos ou determinados elementos pertencem (classificação). Nesse último caso, a discriminação entre categorias é normalmente baseada no cálculo de escores ou valores das covariáveis (por exemplo, uma combinação linear desses valores).

Vários artigos apresentam revisões dos modelos ordinais propostos até o momento, com diversas aplicações de um alguns deles a conjuntos de dados da área médica. Apesar de terem como atrativo a parsimônia dos parâmetros, o sistema de escoragem muito complexo e as regras de decisão geralmente pouco práticas têm limitado seu uso, especialmente em prognósticos médicos. Além disso, muitos autores acreditam que, em certas circunstâncias, esses modelos classificam menos acuradamente que processos não ordinais, como a análise discriminante na distribuição normal ou a regressão logística multinominal.

Finalizando, devido à teoria da qual são derivados (geralmente, estimação por máxima verossimilhança), o ajuste de modelos ordinais, bem como os procedimentos clássicos de discriminação não ordinal, não otimizam nenhum critério de classificação explícito quando aplicados a um conjunto de observações.

O objetivo deste capítulo é apresentar um novo método, proposto por Coste, Walter, Wasserman e Venot (1997), para problemas de discriminação com respostas ordinais, que pode oferecer acurácia ótima de classificação, ou seja, otimiza diretamente um critério de classificação explícito baseado na amostra em estudo.

4.2 Método de discriminação ótima para classificação ordinal

A principal utilidade de métodos de discriminação é classificar observações em um de G grupos mutuamente exclusivos, Y_g $(g = 1, \ldots, G)$ com base na informação de um conjunto de k variáveis independentes (preditoras) x_i $(i = 1, \ldots, k)$.

A regra de decisão para associar uma observação a um grupo particular é estabelecida usando um conjunto de observações para o qual já é conhecido a que grupo realmente cada observação pertence. Tal conjunto é denominado amostra de treino. Uma observação é considerada incorretamente classificada se não pertencer ao grupo predito pelo processo de discriminação. Os autores sugerem que a performance do método de discriminação seja avaliada em amostras de validação, mesmo porque a classificação obtida na amostra de treino não é necessariamente um bom indicador de acurácia na classificação para a amostra de validação.

Em discriminação com respostas ordinais, normalmente assume-se que classificações incorretas em diferentes grupos não são igualmente sérias. Assim, classificações incorretas em categorias adjacentes são mais aceitáveis que aquelas em categorias não adjacentes ou mais distantes. A proporção de observações classificadas incorretamente, apesar de sua importância óbvia, claramente não reflete as diferenças na seriedade relativa da classificação incorreta.

Critérios mais adequados, baseados em custos relativos de erros ou perdas relativas, podem ser utilizados, como, por exemplo, a perda média (PM), que é a média sobre n

observações das perdas l_j associadas com a diferença entre a resposta real e predita para a j-ésima observação. Funções de perda podem ser constantes, lineares, quadráticas ou decididas com base em experiências anteriores. Por exemplo, o critério PM_{FPQ}, baseado na função de perda quadrática, adota $l_j = 0$ se a resposta predita é o grupo verdadeiro, $l_j = (1)^2 = 1$ se o grupo predito é adjacente ao grupo verdadeiro, $l_j = (2)^2 = 4$ se o grupo predito é duas categorias distante do grupo verdadeiro, $l_j = (3)^2 = 9$ se o grupo predito é três categorias distante do verdadeiro grupo e assim por diante.

Apresentaremos a seguir os princípios básicos da discriminação ótima para análise de respostas ordinais.

Seja X um vetor linha k-dimensional de variáveis preditoras e Y um inteiro escalar que indica resposta ou resultado. A resposta Y pode assumir G valores Y_g ($g = 1, \ldots, G$) que são supostamente ordenados, com Y_1 sendo o "melhor" resultado e Y_G, o "pior" segundo algum critério.

Uma função f é construída por meio de valores de duas ou mais variáveis preditoras x_i ($i = 1, \ldots, k$) que são combinadas formando um escore simples $\hat{Y}^* = f(X, \hat{\beta})$, em que β é um vetor coluna k-dimensional de parâmetros. Esse escore intermediário $\hat{Y}^*$ é subsequentemente comparado com os pontos de corte $C_l, (l = 1, \ldots, G-1)$ para classificar observações em classes de resultados preditos $\hat{Y}$ ($\hat{Y}$ tomando G valores). Se os pontos de corte estão indexados em ordem crescente e C_0 e C_G são $-\infty$ e $+\infty$, respectivamente, então a regra para determinação da classe predita, $\hat{Y}_j$, para a j-ésima observação, $j = 1, 2, \ldots, n$, é:

$$\text{se} \quad C_{g-1} < \hat{Y}^*(j) \leq C_g \quad \Rightarrow \quad \hat{Y}_j = Y_g, \quad j = 1, 2, \ldots, n. \tag{4.1}$$

A análise discriminante ótima estima os parâmetros do vetor β e determina os pontos de corte C_g que otimizam a acurácia da classificação numa amostra inicial, de acordo com um critério de classificação, como, por exemplo, a perda média. A formulação deste método é geral e permite ao usuário escolher, de acordo com o contexto prático da pesquisa:

1. A relação funcional entre preditores e o escore intermediário (estrutura subjacente do modelo);

   Modelos lineares, log-lineares e logísticos lineares são os modelos mais frequente-

mente utilizados, principalmente na área médica, para descrever a relação entre variáveis preditoras e resposta. Na análise discriminante ótima, o escore intermediário $\hat{Y}^*$ pode então ser calculado por $\hat{Y}^* = X\hat{\beta}$, $\hat{log(Y^*)} = X\hat{\beta}$ ou $\hat{logit(Y^*)} = X\hat{\beta}$, respectivamente. No entanto, como as transformações logaritmo e logito são monotônicas, a posterior classificação é independente do modelo adotado. Por esse motivo, os autores sugerem a utilização do modelo linear.

2. O critério de classificação a ser otimizado;

   Diversos critérios podem ser usados para medir a acurácia da classificação ordinal. No entanto, perdas médias como PM_{FPL} (baseada na função de perda linear) ou PM_{FPQ} (baseada na função de perda quadrática), devem, na maioria dos casos, ser priorizadas à taxa de classificação incorreta, que é incapaz de considerar a seriedade relativa da classificação incorreta. Todos esses critérios são funções não-lineares, geralmente muito complexas, dos parâmetros de escore intermediários (β) e pontos de corte (C).

3. O método de otimização.

Finalizando o estudo, os autores comparam o procedimento proposto com a análise discriminante tradicional e com a análise obtida através do ajuste de modelos de regressão logística com resposta ordinal. Foi calculado o valor da medida PM_{FPQ} para os três casos, tanto na amostra teste quanto na amostra de validação.

Em todas as situações consideradas, o desempenho da análise discriminante com resposta ordinal foi superior ao dos demais métodos.

4.3 Aperfeiçoamento da robustez das estimativas dos parâmetros do Modelo através de métodos *bootstrap*

Na seção anterior, foi discutida a superioridade da análise discriminante ordinal tanto em termos de classificação como em simplicidade de implementação, se comparada a métodos como regressão logística e análise discriminante de Fisher. No entanto, em um posterior trabalho, Le Teuff, Quantin, Venot, Walter e Coste (2005)

constatam que, nessa análise, as estimativas dos parâmetros mostraram sensibilidade excessiva à amostra usada na estimação. De acordo com os autores, essa falta de robustez deve-se principalmente ao viés de super ajuste que ocorre no estágio da validação e é caracterizado por predições ruins ou variabilidade de desempenho.

Neste caso, a proposta é utilizar a técnica de *bootstrap* para construir modelos no caso de análise de dados categóricos ordinais. Muitos autores propuseram apoiar-se nas técnicas *bootstrap* para lidar com o problema de "super ajuste". O princípio da técnica introduzida por Le Teuff, Quantin, Venot, Walter e Coste (2005) consiste em ajustar o modelo em cada amostra *bootstrap* e utilizar a média das predições via *bootstrap* para fornecer a predição generalizada do modelo.

Além disso, a estimativa dos parâmetros aqui é obtida via uma técnica de otimização global, denominada Busca Adaptativa Aleatória. Com isso, temos uma combinação de uma técnica de otimização global, que fornece estimativas potenciais no espaço paramétrico, com *bootstrap* para reduzir a variabilidade amostral. O papel do *bootstrap* é fornecer um critério agregado em replicações *bootstrap* de um critério de classificação, que é otimizado nas estimações dos parâmetros.

Descrição do Método

Esse novo enfoque é baseado no mesmo princípio da Análise discriminante ótima descrita na seção anterior, só difere na estimação dos parâmetros. Ao invés da perda média quadrática, PM_{FPQ}, uma função de perda agregada é minimizada, por exemplo, a média das perdas quadráticas nas réplicas. Além disso, essa minimização não é baseada na amostra teste mas sim em replicações *bootstrap*.

O processo de Análise discriminante ótima com *bootstrap* pode ser dividido em seis etapas:

1. O conjunto de dados original é dividido em dois subconjuntos, denominados amostra de treino e de validação.

2. Amostras *bootstrap* são geradas do subconjunto de treino por reamostragem aleatória uniforme.

3. O critério de classificação por perda média quadática (PM_{FPQ}) associado com

um conjunto de estimativas dos parâmetros do modelo (β, C) é computado para cada amostra *bootstrap*. Os autores sugerem que a estimação seja realizada pelo método de Busca Adaptativa Aleatória, que é um procedimento probabilístico de otimização global, descrito com detalhes no apêndice de Le Teuff, Quantin, Venot, Walter e Coste (2005).

4. Otimização da perda agregada, calculada como uma medida associada aos valores da PM_{FPQ} nas amostras *bootstrap* obtidas. São considerados cinco critérios do cálculo da perda agregada:

   - a perda média (C1);
   - o percentil de ordem 75 da perda (C2);
   - a perda máxima (C3);
   - uma combinação da média e do desvio padrão (C4) (a menor média e, para médias iguais, o menor desvio padrão);
   - a soma da média quadrática e da variância (C5).

   Tais medidas foram propostas após aprofundados estudos sobre a distruição empírica da perda quadrática PM_{FPQ}, realizados por Coste, Walter, Wasserman e Venot (1997).

5. Finalização da otimização global, obtendo-se uma solução da forma $(\hat{\beta}, \hat{C})$.

6. A qualidade da solução é avaliada classificando os elementos da amostra de validação de acordo com o modelo ajustado para as estimativas $(\hat{\beta}, \hat{C})$ obtidas.

4.4 Considerações Finais

De acordo com Le Teuff, Quantin, Venot, Walter e Coste (2005), o uso de métodos *bootstrap* na Análise Discriminante Ótima para respostas ordinais tem a vantagem da robustez e menor dependência da amostra utilizada. Mostra-se ainda vantajoso para o caso de pequenas amostras e em situações em que o processo de discriminação é difícil. Como desvantagens, os autores apontam a dificuldade da escolha de uma medida de perda agregada, no passo 4 do procedimento, ou ainda, a determinação do número de

amostras necessárias. O número de amostras geralmente utilizado, 200, é recomendado por Efron (1979), mas os autores sugerem estudos adicionais nessa linha.

Com relação aos métodos de classificação para categorias ordinais, Coste, Walter, Wasserman e Venot (1997) acredita que tais procedimentos são pouco utilizados na área econômica e que a Análise Discriminante com resposta ordinal é adotada com maior frequência na área médica.

No próximo capítulo, utilizaremos alguns dos modelos e técnicas descritas até então para classificar um conjunto de dados cuja resposta é categórica ordinal. Em seguida, vamos comparar a qualidade de classificação de cada modelo ajustado, discutindo vantagens e desvantagens de cada aplicação.

Capítulo 5

Aplicação a um conjunto de dados reais

5.1 Introdução

Um problema comum e importante na área financeira é determinar se um certo cliente, para o qual será dado algum crédito, vai honrar esse compromisso depois de um tempo. Para essa determinação, usualmente, é utilizada uma previsão baseada em uma função ponderada de variáveis cadastrais e financeiras do solicitante, técnica conhecida no mercado como *credit scoring*.

A grande dificuldade dessa aplicação é que a amostra disponível para ser utilizada, normalmente, é composta de supostos "bons" pagadores, já que são pessoas que foram consideradas "merecedoras de crédito" no passado e por isso se tornaram clientes. A população de "maus" pagadores para criação do modelo será, basicamente, composta dos erros de decisão do passado, ou seja, pessoas que receberam crédito mas não se mostraram boas pagadoras com o passar do tempo. Aqueles que, no passado, não receberam crédito por não terem perfil de bom pagador são desconsiderados da modelagem e, por esse motivo, é recomendável que o modelo de *credit scoring* seja utilizado como auxiliador da decisão de crédito e não como um critério único e inflexível.

O "escore" do cliente, muito utilizado nessa aplicação, refere-se à probabilidade à posteriori (ou a uma função de probabilidade, por exemplo, o valor da função discrim-

inante de Fisher calculada em $\mathbf{x}$) de um cliente ser um bom pagador dado os valores de suas variáveis discriminadoras. Exemplos de variáveis que possuem poder discriminatório em problemas ligados a áreas financeiras são, no caso de pessoa física, número de dependentes, renda, idade etc.

O objetivo do presente capítulo é apresentar uma aplicação a dados reais de concessão de crédito, levando em consideração alguns dos modelos de regressão e algumas das técnicas de discriminação para variáveis respostas ordinais apresentadas nos capítulos anteriores. Em seguida, deseja-se comparar os resultados, avaliando as vantagens e desvantagens da utilização de cada método no conjunto de dados escolhido, assim como suas peculiaridades.

5.2 Descrição do estudo

A aplicação que apresentaremos no presente capítulo foi feita com dados de uma empresa prestadora de serviços de cartões de crédito. A amostra, composta de clientes que adquiriram o cartão de crédito, consistia de 78.020 propostas, classificadas como de clientes BONS, MÉDIOS e MAUS. A definição de "bom", "médio" e "mau" pagador foi feita de acordo com o seguinte critério::

- BOM (Y=1): clientes que permaneceram ativos após um período definido, sem apresentar atrasos;
- MÉDIO (Y=2): clientes que apresentaram algum atraso de pagamento maior que 30 dias ou pagaram com cheque sem fundos, mas que permaneceram ativos após o período definido;
- MAU (Y=3): clientes que tiveram seus cartões cancelados durante o período definido por motivo de falta de pagamento da fatura;

A classificação das variáveis preditoras consideradas na análise é dada a seguir, sendo que o significado de cada uma foi omitido a pedido do agente financeiro que fornececeu os dados.

- X_1: variável qualitativa nominal (binária)

- X_2: variável qualitativa ordinal (4 categorias)
- X_3: variável quantitativa contínua
- X_4: variável quantitativa contínua

Na análise apresentada, em razão da necessidade de sigilo da informação financeira, a distribuição utilizada das categorias de resposta "bom", "médio" e "mau" não reflete a distribuição real dos dados.

Com o intuito de explicar a variável resposta categórica ordinal *tipo de pagador* (Y), com categorias ordinais MAU, MÉDIO (MED) e BOM pagador, foram aplicados e avaliados os seguintes métodos ou modelos:

1. Modelo Logístico Multinomial (sem considerar a ordenação das categorias da variável resposta);
2. Modelo Logístico Cumulativo e Modelo de Riscos Proporcionais;
3. Análise Discriminante Normal adaptada para variáveis explicativas categóricas;
4. Análise Discriminante Ótima para variável resposta ordinal;

Para isso, a base inicial de 78.020 clientes foi dividida aleatoriamente em duas, chamadas de base de treino e base de validação. A primeira delas era composta de 58.515 casos (75% da base total) para serem usados no ajuste dos modelos, enquanto a segunda, utilizada para avaliação da discriminação do modelo, continha 19.505 dados que seriam classificados por meio das estimativas dos parâmetros obtidas da base de treino.

Um estudo descritivo de cada uma das variáveis, considerando a amostra total, é apresentado no Apêndice A.

As parametrizações das variáveis categóricas X_1 e X_2, quando necessárias, foram feitas de maneira que X_1 fosse considerada variável indicadora e X_2 fosse reparametrizada conforme apresentamos na Tabela 5.1.

O pacote computacional utilizado na aplicação descrita a seguir foi o *SAS 9.1.3*.

Tabela 5.1: Reparametrização da variável qualitativa ordinal X_2.

X_2	$X_2^{(1)}$	$X_2^{(2)}$	$X_2^{(3)}$
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

5.3 Aplicação

5.3.1 Modelo Logístico Multinomial

O Modelo Logístico Multinomial (Fahrmeir e Tutz (1994)), atualmente, é um dos mais utilizados quando a variável resposta possui mais de duas categorias. No entanto, tal modelo não leva em conta a ordenação que pode existir entre as categorias da variável resposta, considerando essa uma variável qualitativa nominal. Desejamos ajustar esse modelo para depois compará-lo com os modelos que consideram a ordenação da variável *tipo de pagador*.

Para essa aplicação, foi selecionada uma amostra balanceada da base de treino, de forma a termos um conjunto de dados formado pela mesma quantidade de bons, médios e maus pagadores. O total da base de treino foi de 10.908 clientes, sendo cada grupo de resposta composto por 3.636 observações. Para seleção de variáveis, foi utilizado o método *backward*. No entanto, todas as variáveis preditoras utilizadas foram selecionadas.

As estimativas dos parâmetros encontram-se na Tabela 5.2 e o modelo pode ser escrito como:

$$\hat{\log}\frac{P(BOM|\mathbf{x})}{P(MAU|\mathbf{x})} = -1,67 - 0,43X_1 - 0,81X_2^{(1)} - 0,43X_2^{(2)} - 0,09X_2^{(3)} + 0,003X_3 - 81,57X_4$$

$$\hat{\log}\frac{P(MED|\mathbf{x})}{P(MAU|\mathbf{x})} = 0,29 - 0,21X_1 - 0,73X_2^{(1)} - 0,45X_2^{(2)} - 0,16X_2^{(3)} + 0,001X_3 - 68,25X_4$$

Tabela 5.2: Estimativas dos parâmetros e p-valores do Modelo Logístico Multinomial.

Parâmetro	Estimativa	Erro padrão	p-valor
θ_1	-1,671	0,1501	(0,00)
θ_2	0,287	0,1338	(0,03)
X_1[BOM]	-0,428	0,0556	(0,00)
X_1[MED]	-0,209	0,0528	(0,00)
$X_2^{(1)}$[BOM]	-0,812	0,0780	(0,00)
$X_2^{(1)}$[MED]	-0,725	0,0737	(0,00)
$X_2^{(2)}$[BOM]	-0,430	0,0678	(0,00)
$X_2^{(2)}$[MED]	-0,450	0,0659	(0,00)
$X_2^{(3)}$[BOM]	-0,089	0,0625	(0,15)
$X_2^{(3)}$[MED]	-0,164	0,0617	(0,01)
X_3[BOM]	0,003	0,0002	(0,00)
X_3[MED]	0,001	0,0001	(0,00)
X_4[BOM]	-81,574	1,0738	(0,00)
X_4[MED]	-68,245	1,0405	(0,00)

Assim, a presença de X_1 tende a tornar o cliente pior, dado que o risco relativo de se tornar bom é mais reduzido do que o de se tornar médio. O mesmo ocorre para um valor alto de X_4, que diminui mais o risco relativo de um cliente ser bom do que o risco relativo de ele ser médio. Por outro lado, valores altos de X_2 e X_3 parecem aumentar mais o risco relativo de ser um cliente bom do que o risco relativo de ser médio, de tal forma que elevam a chance de um cliente ser bom.

Com o intuito de avaliar o poder de discriminação do modelo multinomial ajustado, os clientes da amostra de treino e de validação foram classificados segundo as estimativas dos parâmetros obtidas pelo modelo, de acordo com o critério descrito no Capítulo 3, que utiliza o preditor $\tilde{y}$, alocando os clientes na categoria com maior probabilidade a posteriori estimada. O outro critério sugerido no mesmo capítulo, que utiliza o preditor $\hat{y}$, foi aplicado, mas não apresentou bom desempenho para a amostra utilizada, já que classificou todos os clientes em uma única categoria. A classificação obtida foi então comparada com a classificação real do cliente. É possível observar o percentual de acertos para a amostra de treino na Tabela 5.8 e para a amostra de validação na

Tabela 5.10.

A amostra de treino apresenta uma concordância absoluta entre o valor predito e o valor real para 43,5% dos clientes, conforme observamos na Tabela 5.9. Se somarmos a esse valor o percentual de clientes alocados em uma categoria adjacente à correta, chegamos a 81,4%, o que indica que apenas 18,6% dos casos foram alocados de maneira extremamente incorreta, ou seja, foram classificados como bons sendo maus e vice-versa.

No entanto, como esses resultados são da amostra utilizada para o ajuste do modelo, essa qualidade de classificação não corresponde exatamente à real. Para isso, é fundamental analisarmos o comportamento da concordância de classificação na amostra de validação, apresentada na Tabela 5.11. Notamos que há uma concordância absoluta entre o valor predito e o valor real para 57,9% dos clientes e, ao mesmo tempo, 75,3% dos casos estão corretamente alocados ou alocados em uma categoria adjacente à correta.

5.3.2 Modelo Logístico Cumulativo e Modelo de Riscos Proporcionais

Foram ajustados aos dados dois dos modelos cumulativos apresentados no Capítulo 2: modelo logístico cumulativo (2.4) e modelo de riscos proporcionais (2.7). Ao contrário do modelo multinomial aplicado aos dados na seção anterior, os modelos cumulativos ajustados consideram a ordenação da variável resposta categórica.

A base de treino utilizada era composta dos mesmos 10.908 clientes considerados na modelagem anterior, ou seja, tratava-se de uma base balanceada, contendo 3.636 observações para cada grupo de resposta.

A Tabela 5.3 apresenta as estimativas dos parâmetros e seus respectivos p-valores. O método de seleção de variáveis preditoras utilizado foi o *backward*. No entanto, assim como no modelo multinomial ajustado anteriormente, todas as variáveis foram selecionadas nos dois modelos cumulativos. Valores negativos das estimativas sinalizam baixas probabilidades para menor categoria (BOM), de forma que altos valores de X_1 e X_4 e menores valores de X_2 e X_3 levam a piores escores de clientes.

Para avaliar o poder de discriminação dos dois modelos ajustados, os clientes da

Tabela 5.3: Estimativas dos parâmetros e p-valores dos Modelos Logístico Cumulativo e de Riscos Proporcionais.

	Modelo Cumulativo logístico			Modelo de Riscos proporcionais		
Parâmetro	Estimativa	Erro padrão	p-valor	Estimativa	Erro padrão	p-valor
θ_1	-1,897	0,1081	(0,00)	-1,728	0,0771	(0,00)
θ_2	-0,415	0,1066	(0,00)	-0,683	0,0757	(0,00)
X_1	-0,307	0,0405	(0,00)	-0,197	0,0277	(0,00)
$X_2^{(1)}$	-0,626	0,0574	(0,00)	-0,468	0,0412	(0,00)
$X_2^{(2)}$	-0,310	0,0500	(0,00)	-0,248	0,0342	(0,00)
$X_2^{(3)}$	-0,052	0,0455	(0,26)	-0,048	0,0298	(0,11)
X_3	0,002	0,00011	(0,00)	0,002	0,00007	(0,00)
X_4	-6,043	0,7885	(0,00)	-4,441	0,5391	(0,00)

amostra de treino e de validação foram classificados segundo o critério descrito no Capítulo 3, $\tilde{y}$, que aloca os clientes na categoria com maior probabilidade a posteriori estimada. A classificação obtida foi então comparada com a classificação real do cliente. É possível observar o percentual de acertos das amostras de treino e de validação nas Tabelas 5.12 a 5.18.

A amostra de treino apresenta uma concordância absoluta entre o valor predito e o valor real para 43,1% dos clientes, quando consideramos o Modelo Logístico Cumulativo (Tabela 5.13). Se considerarmos o Modelo de Riscos Proporcionais (Tabela 5.17), essa concordância absoluta é igual a 43,4%. Ao somarmos a esse valor de concordância absoluta o percentual de clientes alocados em uma categoria adjacente à correta, chegamos a 82,3% para o Modelo Logístico Cumulativo e a 81,7% para o Modelo de Riscos Proporcionais.

Em seguida, ao analisarmos os resultados na amostra de validação, que corresponde a uma melhor avaliação da performance de classificação dos modelos, obtemos os resultados apresentados nas Tabelas 5.15 e 5.19. Ao analisar o ajuste através do Modelo Logístico Cumulativo, observa-se que há uma concordância absoluta entre o valor predito e o valor real para 56,0% dos clientes e, ao mesmo tempo, 76,4% dos casos estão

corretamente alocados ou alocados em um valor adjacente ao valor correto. Quando observamos o Modelo de Riscos Proporcionais, esses percentuais de acerto são iguais a 53,8% e 73,1%, respectivamente. É interessante observar também que os percentuais por linha fornecem uma idéia da eficiência do processo em cada categoria da variável resposta. Nesse caso, é possível notar que essa eficiência de acerto dentro de cada categoria da variável resposta é maior para os grupos de cliente "BOM" e "MAU".

Assim, analisando o comportamento da classificação nas amostras de treino e validação para os dois modelos cumulativos, observamos uma performance superior do Modelo Cumulativo Logístico, que apresentou percentuais maiores de acerto.

Nessa aplicação, percebemos também que o Modelo Multinomial apresentou uma porcentagem de acerto um pouco maior que os Modelos Ordinais quando só avaliamos a porcentagem de acerto de classificação correta. Entretanto, ao considerarmos a classificação em categorias adjacentes, o Modelo Logístico Cumulativo, que leva em conta a ordenação, apresentou um percentual de acerto maior.

5.3.3 Análise Discriminante Normal adaptada para variáveis explicativas categóricas

Os dados foram submetidos a dois métodos de Análise Discriminate Normal adaptada para variáveis explicativas categóricas: o método da função discriminante linear e o método de Krzanowski (Sanda (1990)). Por não considerar a ordenação entre os grupos de resposta, a análise discriminante normal será aplicada para ser comparada, em termo de classificação, com os modelos que consideram a ordem entre as categorias da variável dependente.

Dado que X_1 e X_2 não se tratavam de variáveis contínuas, mas de variáveis binária e qualitativa ordinal, respectivamente, foi necessário fazer uma adaptação na aplicação, de tal forma que essa considerasse as quatro variáveis utilizadas nas análises até então.

A primeira solução para o problema, usada para a aplicação da Análise Discriminante com função linear, foi utilizar X_1 como variável indicadora e reparametrizar X_2 como apresentamos na Tabela 5.1, de acordo com sugestão de Sanda (1990).

As estimativas dos parâmetros das funções discriminante lineares para cada grupo de resposta são dadas na Tabela 5.4 e a regra de decisão foi criada considerando custos

iguais.

Tabela 5.4: Função discriminante linear por grupo de resposta.

Variável	BOM	MED	MAU
Constante	-21,192	-19,646	-19,117
X_1	1,626	1,900	2,167
$X_2^{(1)}$	2,589	2,666	3,584
$X_2^{(2)}$	2,728	2,682	3,120
$X_2^{(3)}$	2,346	2,290	2,454
X_3	0,034	0,032	0,030
X_4	170,750	171,318	177,920

Como outra solução, Sanda (1990) sugere agrupar os dados da amostra em conjuntos consistindo de todas as possíveis combinações de categorias das variáveis explicativas categóricas, procedimento conhecido como método Krzanowski. Seguindo esse critério, como X_1 trata-se de uma variável binária e X_2 de uma variável qualitativa ordinal com quatro categorias, tivemos que dividir nossos dados da amostra de treino em oito caselas tais que:

- Casela 1 (4594 clientes): X_1=0 e X_2=1;
- Casela 2 (7382 clientes): X_1=0 e X_2=2;
- Casela 3 (10347 clientes): X_1=0 e X_2=3;
- Casela 4 (22542 clientes): X_1=0 e X_2=4;
- Casela 5 (1482 clientes): X_1=1 e X_2=1;
- Casela 6 (2445 clientes): X_1=1 e X_2=2;
- Casela 7 (3451 clientes): X_1=1 e X_2=3;
- Casela 8 (6272 clientes): X_1=1 e X_2=4.

Esse método é operacionalmente viável quando as variáveis discriminadoras categóricas formarem poucas caselas ou quando a amostra disponível for grande, pois o número de caselas faz crescer exponencialmente a quantidade mínima necessária de amostra.

Tabela 5.5: Função discriminante linear por casela e por grupo de resposta.

Casela	Variável	BOM	MED	MAU
1	Constante	-18,287	-16,712	-16,739
	X_3	0,031	0,028	0,028
	X_4	158,102	160,719	162,966
2	Constante	-21,235	-19,332	-18,998
	X_3	0,036	0,033	0,032
	X_4	182,211	181,014	188,158
3	Constante	-20,907	-19,301	-18,104
	X_3	0,035	0,033	0,031
	X_4	174,795	173,270	181,096
4	Constante	-20,460	-18,894	-17,414
	X_3	0,034	0,032	0,029
	X_4	171,574	170,729	177,973
5	Constante	-20,186	-18,516	-18,642
	X_3	0,034	0,032	0,031
	X_4	182,405	182,394	189,072
6	Constante	-18,059	-16,779	-16,299
	X_3	0,030	0,029	0,028
	X_4	162,025	163,995	168,480
7	Constante	-18,583	-17,412	-16,966
	X_3	0,031	0,029	0,028
	X_4	160,751	166,489	176,833
8	Constante	-20,361	-18,618	-17,244
	X_3	0,034	0,032	0,029
	X_4	170,205	174,937	179,127

Após essa segmentação da base de treino, foi ajustada a análise discriminante normal dentro de cada casela, baseada nas variáveis contínuas X_3 e X_4. As estimativas dos parâmetros encontram-se na Tabela 5.5. Supondo que as variáveis categóricas sejam não informativas, isto é, a probabilidade a priori de cada casela não varia dependendo dos grupos de resposta, a regra do Método Krzanowski é equivalente à regra da função discriminante linear calculada exclusivamente com os dados pertencentes a esta casela. Essa suposição foi considerada para os dados ajustados.

A fim de avaliar a qualidade da discriminação dos dois métodos de análise discriminante normal ajustados, os clientes da amostra de treino e de validação foram classificados segundo as estimativas dos parâmetros obtidas pelo modelo e a regra de decisão que considera custos iguais. A classificação obtida foi então comparada com a classificação real do cliente. Podemos observar o percentual de acertos para a amostra de treino nas Tabelas 5.20 e 5.24 e para a amostra de validação nas Tabelas 5.22 e 5.26.

No caso da análise discriminante linear, a amostra de treino apresenta uma concordância absoluta entre o valor predito e o valor real para 57,7% dos clientes, conforme observamos na Tabela 5.21. Se somarmos a esse valor o percentual de clientes alocados em uma categoria adjacente à correta, obtemos a concordância de 76,2%. Para o método de Krzanowski, esses percentuais são iguais a 59,6% e 74,8%, respectivamente, conforme observamos na Tabela 5.25.

Como esses resultados são da amostra utilizada para a estimação dos parâmetros, sabemos que essa qualidade de classificação não corresponde exatamente à real. Para isso, é fundamental analisarmos o comportamento da concordância de classificação na amostra de validação, apresentada nas Tabelas 5.23 e 5.27. Nelas, notamos que há uma concordância absoluta entre o valor predito e o valor real, no caso da análise discriminante linear, para 57,9% dos clientes e que, ao mesmo tempo, 75,8% dos casos estão corretamente alocados ou alocados em uma categoria adjacente à correta. No caso do método de Krzanowski, esses acertos são da ordem de 59,6% e 74,1%, respectivamente.

Assim, verificamos um melhor desempenho de classificação para o método Krzanowski, quando consideramos apenas o acerto na categoria correta. Se considerarmos também o acerto nas categorias adjacentes, o primeiro método apresentou um melhor resultado.

5.3.4 Análise Discriminante Ótima para variável resposta ordinal

Nesta etapa do estudo, foi utilizada a metodologia de análise discriminante ótima descrita no Capítulo 4.

A aplicação da análise discriminante ótima necessita, em primeiro lugar, que haja uma transformação da variável categórica ordinal em variável contínua. Nesse caso, a categoria "BOM" recebeu o valor 1, a categoria "MÉDIO" recebeu o valor 2 e a categoria "MAU" recebeu o valor 3. É importante observar que essa transformação leva em conta que as distâncias entre as categorias são iguais, ou seja, a categoria "BOM" está tão distante da categoria "MÉDIO" quanto a categoria "MAU" está. Caso fosse de interesse alterar essas distâncias, os valores da variável resposta transformada poderiam ser diferentes dos utilizados, por exemplo: 1 para "BOM", 2 para "MÉDIO" e 4 para MAU". Nesse caso, ser "MAU" seria mais distante de "MÉDIO".

A partir da nova variável resposta, foi ajustado um Modelo Linear com função de ligação log. As estimativas dos parâmetros desse modelo encontram-se na Tabela 5.6. Observa-se pelos valores estimados que categorias ou números mais altos de X_2 e X_3 levam a uma redução da variável resposta, sinalizando melhores clientes. $X_1 = 1$ ou um valor maior de X_4, por sua vez, levam um aumento da variável resposta, sinalizando piores clientes.

Tabela 5.6: Estimativas dos parâmetros e p-valores do Modelo Log-linear.

Parâmetro	Estimativa	Erro padrão	p-valor
Intercepto	0,5640	0,0117	(0,00)
X_1	0,0853	0,0046	(0,00)
$X_2^{(1)}$	0,1343	0,0063	(0,00)
$X_2^{(2)}$	0,0554	0,0056	(0,00)
$X_2^{(3)}$	0,0157	0,0052	(0,00)
X_3	-0,0006	0,0000	(0,00)
X_4	1,0386	0,0896	(0,00)
Escalar	0,6276	0,0018	

Em seguida, foi calculada a variável resposta predita para cada um dos clientes da amostra de treino. O próximo passo era calcular $\hat{\theta_1}$ e $\hat{\theta_2}$ que são os pontos de corte na variável contínua Y de modo a dividir sua escala contínua em três partes, formando três grupos ordenados. Para isso, foi feito um programa computacional em *SAS* (Apêndice C) que calculava os dois pontes de corte que levam à melhor classificação, sendo que essa última foi avaliada segundo os seguintes critérios:

- Menor perda média linear,

$$PM_{FPL} = \frac{1}{n}\sum_{j=1}^{n}(y(j) - \hat{y}(j)),$$

onde $y(j)$ é o valor real da variável resposta do cliente j e $\hat{y}(j)$ é o valor predito da resposta para o cliente j;

- Menor perda média quadrática,

$$PM_{FPQ} = \frac{1}{n}\sum_{j=1}^{n}(y(j) - \hat{y}(j))^2,$$

onde $y(j)$ é o valor real da variável resposta do cliente j e $\hat{y}(j)$ é o valor predito da resposta para o cliente j;

- Maior estatística de concordância Kappa ponderada,

$$K_w = \frac{P_{o(w)} - P_{e(w)}}{1 - P_{e(w)}},$$

onde $P_{o(w)} = \sum_i \sum_j w_{ij} p_{ij}$, $P_{e(w)} = \sum_i \sum_j w_{ij} p_{i.} p_{.j}$ e os pesos w_{ij} são definidos de tal forma que $0 \leq w_{ij} < 1$ para todo $i \neq j$, $w_{ii} = 1$ para todo i e $w_{ij} = w_{ji}$.

Os pesos do coeficiente Kappa utilizados foram do tipo Cicchetti-Allison, definidos em Cicchetti e Allison (1971):

$$w_{ij} = 1 - \frac{|C_i - C_j|}{C_C - C_1},$$

em que C_i é o escore para coluna i e C é o número de categorias ou colunas.

Conforme comentamos, para obtenção de $\hat{\theta_1}$ e $\hat{\theta_2}$, foi elaborado um programa em linguagem *SAS* que calculava, inicialmente, os valores de $\hat{y}$ com base na amostra de validação. Obtidos $\hat{y}_{min}$ e $\hat{y}_{max}$, respectivamente, o mínimo e o máximo desses valores para a amostra, foram propostos valores de $\hat{\theta_1}$ e $\hat{\theta_2}$ dentro deste intervalo, com $\hat{\theta_1} < \hat{\theta_2}$, totalizando quase 600 possibilidades. Para cada possibilidade, foi calculada a perda média linear, a perda média quadrática e a estatística Kappa ponderada. Finalmente, foram selecionados pares $(\hat{\theta_1}, \hat{\theta_2})$ de modo a minimizar as perdas médias ou maximizar o valor da estatística Kappa.

A base utilizadas nesta aplicação, diferentemente das bases utilizadas no ajuste dos modelos anteriores, não eram balanceadas. Isso foi feito porque o tamanho de cada um dos grupos de categorias da variável resposta influencia fortemente na definição dos pontos de corte. Assim, em termos de distribuição da variável resposta, se a população for muito diferente da amostra utilizada, esse algoritmo deve apresentar uma classificação inferior.

A Tabela 5.7 apresenta os pontos de corte $\hat{\theta_1}$ e $\hat{\theta_2}$ obtidos segundo cada um dos três critérios considerados.

Tabela 5.7: Estimativas de $\hat{\theta_1}$ e $\hat{\theta_2}$ por critério de definição dos pontos de corte.

Critério considerado	Valor obtido	$\hat{\theta_1}$	$\hat{\theta_2}$
PM_{FPL} mínima	0,2791	0,80429	0,87929
PM_{FPQ} mínima	0,4746	0,40429	0,87929
Kappa ponderada máxima	0,1834	0,32929	0,35429

As Tabelas 5.28, 5.32 e 5.36 mostram a classificação dos clientes da base de treino que levaram aos pontos de corte definidos pelos critérios de perda média linear mínima, perda média quadrática mínima e Kappa ponderada máxima, respectivamente. Os valores de concordância absoluta resultantes de cada critério encontram-se nas Tabelas 5.29, 5.33 e 5.37. Nas mesmas, temos também o valor acumulado de concordância adicionando o percentual de acerto em categorias adjacentes.

Com o objetivo de analisar o desempenho de cada um dos critérios considerados para a aplicação da análise discriminante ótima, após o ajuste do modelo log-linear e da estimação dos pontos de corte $\hat{\theta_1}$ e $\hat{\theta_2}$ obtidos por meio da base de treino, fizemos a

classificação dos clientes da base de validação segundo essas estimativas. Para isso, foi utilizado o critério de predição 4.1, descrito no Capítulo 4 e, em seguida, a classificação predita foi comparada com a classificação real do cliente. A classificação dos clientes em cada uma das categorias da variável resposta encontra-se nas Tabelas 5.30, 5.34 e 5.38.

Se comparados com os ajustes anteriores, valores altos de concordância total da base de validação foram obtidos com as classificações provenientes da análise discriminante ótima pelos critérios de perda média, tanto linear quanto quadrática, conforme vemos nas Tabelas 5.31 e 5.35. Considerando apenas a concordância absoluta de classificação, ou seja, a alocação em uma categoria predita exatamente igual à categoria real, verificamos um percentual de 83,7% de acerto quando utilizado o critério de perda média linear e 78,3% de acerto para o caso da perda média quadrática, ambos os percentuais maiores que os encontrados nas aplicações anteriores, já adicionados os percentuais por acerto em categorias adjacentes. No entanto, ao analisarmos melhor as Tabelas 5.30 e 5.34, percebemos que esses critérios de predição levaram a uma classificação de clientes "questionável", pois resultaram na alocação de quase todos os clientes na primeira categoria (BOM), uma vez que essa representa a grande maioria da população. Assim, em razão da grande diferença entre as probabilidades a priori, esses modelos de classificação concedem crédito a quase todos os clientes.

Em termos práticos, classificar todos os solicitantes de cartão de crédito na categoria "BOM" não parece razoável, pois levaria à aprovação dos mesmos, no momento do requerimento, com base na suposição de que esses terão boa performance no primeiro ano, sendo que uma parte, mesmo que relativamente menor, terá o cartão de crédito cancelado por falta de pagamento. Se levantarmos o custo que cada um desses cancelamentos por falta de pagamento pode gerar para a empresa, é bem provável que cheguemos à conclusão de que esse tipo de decisão pode ser muito perigosa financeiramente. Por fim, também podemos dizer que classificar todos os solicitantes em uma mesma categoria, no caso a categoria "BOM", seria o mesmo que não utilizar nenhum modelo de *credit scoring* na aprovação.

A classificação de clientes pelo critério que utilizava a estatística de concordância Kappa ponderada, por sua vez, levou à alocação de clientes da base de validação apresentada na Tabela 5.39. Diferentemente dos dois critérios de análise discriminante ótima cujos resultados já discutimos, o ajuste obtido pelo critério da Kappa ponder-

ada parece ter levado à resultados de classificação melhores, com uma distribuição de alocação nas categorias preditas mais similar à distribuição das categorias reais. O percentual de acertos absolutos foi igual a 72,6% e o total de acertos considerando a classificação em categorias adjacentes foi igual a 82,2%, mostrando uma performance de classificação melhor que os modelos e técnicas ajustados anteriormente.

Tabela 5.8: Classificação Real e Predita para a Amostra de Treino - Modelo Logístico Multinomial.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	2204	494	938	3636
	MED	1689	603	1344	3636
	MAU	1093	605	1938	3636
	Total	4986	1702	4220	10908
% por linha	BOM	60,6	13,6	25,8	0
	MED	46,5	16,6	37	0
	MAU	30,1	16,6	53,3	0
	Total	45,7	15,6	38,7	100
% pelo total	BOM	20,2	4,5	8,6	33,3
	MED	15,5	5,5	12,3	33,3
	MAU	10	5,6	17,8	33,3
	Total	45,7	15,6	38,7	100

Tabela 5.9: Alocação nas categorias da variável resposta para a Amostra de Treino - Modelo Logístico Multinomial.

Classificação	% Absoluto	% Acumulado
Categoria correta	43,5	43,5
Categorias adjacentes	37,9	81,4
Categorias distantes	18,6	100

Tabela 5.10: Classificação Real e Predita para a Amostra de Validação - Modelo Logístico Multinomial.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	9993	2154	4189	16336
	MED	498	205	421	1124
	MAU	622	330	1093	2045
	Total	11113	2689	5703	19505
% por linha	BOM	61,2	13,2	25,6	0
	MED	44,3	18,2	37,5	0
	MAU	30,4	16,1	53,5	0
	Total	57	13,8	29,2	100
% pelo total	BOM	51,2	11	21,5	83,8
	MED	2,6	1,1	2,2	5,8
	MAU	3,2	1,7	5,6	10,5
	Total	57	13,8	29,2	100

Tabela 5.11: Alocação nas categorias da variável resposta para a Amostra de Validação - Modelo Logístico Multinomial.

Classificação	% Absoluto	% Acumulado
Categoria correta	57,9	57,9
Categorias adjacentes	17,4	75,3
Categorias distantes	24,7	100

Tabela 5.12: Classificação Real e Predita para a Amostra de Treino - Modelo Logístico Cumulativo.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	2144	579	913	3636
	MED	1674	625	1337	3636
	MAU	1025	682	1929	3636
	Total	4843	1886	4179	10908
% por linha	BOM	59	15,9	25,1	0
	MED	46	17,2	36,8	0
	MAU	28,2	18,8	53,1	0
	Total	44,4	17,3	38,3	100
% pelo total	BOM	19,7	5,3	8,4	33,3
	MED	15,4	5,7	12,3	33,3
	MAU	9,4	6,3	17,7	33,3
	Total	44,4	17,3	38,3	100

Tabela 5.13: Alocação nas categorias da variável resposta para a Amostra de Treino - Modelo Logístico Cumulativo.

Classificação	% Absoluto	% Acumulado
Categoria correta	43,1	43,1
Categorias adjacentes	39,2	82,3
Categorias distantes	17,7	100

Tabela 5.14: Classificação Real e Predita para a Amostra de Validação - Modelo Logístico Cumulativo.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	9659	2659	4018	16336
	MED	495	199	430	1124
	MAU	584	391	1070	2045
	Total	10738	3249	5518	19505
% por linha	BOM	59,1	16,3	24,6	0
	MED	44	17,7	38,3	0
	MAU	28,6	19,1	52,3	0
	Total	55,1	16,7	28,3	100
% pelo total	BOM	49,5	13,6	20,6	83,8
	MED	2,5	1	2,2	5,8
	MAU	3	2	5,5	10,5
	Total	55,1	16,7	28,3	100

Tabela 5.15: Alocação nas categorias da variável resposta para a Amostra de Validação - Modelo Logístico Cumulativo.

Classificação	% Absoluto	% Acumulado
Categoria correta	56	56
Categorias adjacentes	20,4	76,4
Categorias distantes	23,6	100

Tabela 5.16: Classificação Real e Predita para a Amostra de Treino - Modelo de Riscos Proporcionais.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	2035	536	1065	3636
	MED	1578	554	1504	3636
	MAU	933	561	2142	3636
	Total	4546	1651	4711	10908
% por linha	BOM	56	14,7	29,3	0
	MED	43,4	15,2	41,4	0
	MAU	25,7	15,4	58,9	0
	Total	41,7	15,1	43,2	100
% pelo total	BOM	18,7	4,9	9,8	33,3
	MED	14,5	5,1	13,8	33,3
	MAU	8,6	5,1	19,6	33,3
	Total	41,7	15,1	43,2	100

Tabela 5.17: Alocação nas categorias da variável resposta para a Amostra de Treino - Modelo de Riscos Proporcionais.

Classificação	% Absoluto	% Acumulado
Categoria correta	43,4	43,4
Categorias adjacentes	38,3	81,7
Categorias distantes	18,3	100

Tabela 5.18: Classificação Real e Predita para a Amostra de Validação - Modelo de Riscos Proporcionais.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	9120	2498	4718	16336
	MED	466	183	475	1124
	MAU	528	320	1197	2045
	Total	10114	3001	6390	19505
% por linha	BOM	55,8	15,3	28,9	100
	MED	41,5	16,3	42,3	100
	MAU	25,8	15,7	58,5	100
	Total	51,9	15,4	32,8	100
% pelo total	BOM	46,8	12,8	24,2	83,8
	MED	2,4	0,9	2,4	5,8
	MAU	2,7	1,6	6,1	10,5
	Total	51,9	15,4	32,8	100

Tabela 5.19: Alocação nas categorias da variável resposta para a Amostra de Validação - Modelo de Riscos Proporcionais.

Classificação	% Absoluto	% Acumulado
Categoria correta	53,8	53,8
Categorias adjacentes	19,3	73,1
Categorias distantes	26,9	100

Tabela 5.20: Classificação Real e Predita para a Amostra de Treino - Análise Discriminante Normal Linear.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	29719	6819	11993	48531
	MED	1688	626	1322	3636
	MAU	1938	1007	3403	6348
	Total	33345	8452	16718	58515
% por linha	BOM	61,2	14,1	24,7	100
	MED	46,4	17,2	36,4	100
	MAU	30,5	15,9	53,6	100
	Total	57	14,4	28,6	100
% pelo total	BOM	50,8	11,7	20,5	82,9
	MED	2,9	1,1	2,3	6,2
	MAU	3,3	1,7	5,8	10,8
	Total	57	14,4	28,6	100

Tabela 5.21: Alocação nas categorias da variável resposta para a Amostra de Treino - Análise Discriminante Normal Linear.

Classificação	% Absoluto	% Acumulado
Categoria correta	57,7	57,7
Categorias adjacentes	18,5	76,2
Categorias distantes	23,8	100

Tabela 5.22: Classificação Real e Predita para a Amostra de Validação - Análise Discriminante Normal Linear.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	10004	2236	4096	16336
	MED	491	209	424	1124
	MAU	632	342	1071	2045
	Total	11127	2787	5591	19505
% por linha	BOM	61,2	13,7	25,1	100
	MED	43,7	18,6	37,7	100
	MAU	30,9	16,7	52,4	100
	Total	57	14,3	28,7	100
% pelo total	BOM	51,3	11,5	21	83,8
	MED	2,5	1,1	2,2	5,8
	MAU	3,2	1,8	5,5	10,5
	Total	57	14,3	28,7	100

Tabela 5.23: Alocação nas categorias da variável resposta para a Amostra de Validação - Análise Discriminante Normal Linear.

Classificação	% Absoluto	% Acumulado
Categoria correta	57,9	57,9
Categorias adjacentes	17,9	75,8
Categorias distantes	24,2	100

Tabela 5.24: Classificação Real e Predita para a Amostra de Treino - Análise Discriminante Normal pelo Método Krzanowski.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	31256	4920	12355	48531
	MED	1847	448	1341	3636
	MAU	2415	787	3146	6348
	Total	35518	6155	16842	58515
% por linha	BOM	64,4	10,1	25,5	100
	MED	50,8	12,3	36,9	100
	MAU	38	12,4	49,6	100
	Total	60,7	10,5	28,8	100
% pelo total	BOM	53,4	8,4	21,1	773,8
	MED	3,2	0,8	2,3	6,2
	MAU	4,1	1,3	5,4	10,8
	Total	60,7	10,5	28,8	100

Tabela 5.25: Alocação nas categorias da variável resposta para a Amostra de Treino - Análise Discriminante Normal pelo Método Krzanowski.

Classificação	% Absoluto	% Acumulado
Categoria correta	59,6	59,6
Categorias adjacentes	15,2	74,8
Categorias distantes	25,2	100

Tabela 5.26: Classificação Real e Predita para a Amostra de Validação - Análise Discriminante Normal pelo Método Krzanowski.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	10497	1601	4238	16336
	MED	524	149	451	1124
	MAU	801	259	985	2045
	Total	11822	2009	5674	19505
% por linha	BOM	64,3	9,8	25,9	100
	MED	46,6	13,3	40,1	100
	MAU	39,2	12,7	48,2	100
	Total	60,6	10,3	29,1	100
% pelo total	BOM	53,8	8,2	21,7	798,8
	MED	2,7	0,8	2,3	5,8
	MAU	4,1	1,3	5	10,5
	Total	60,6	10,3	29,1	100

Tabela 5.27: Alocação nas categorias da variável resposta para a Amostra de Validação - Análise Discriminante Normal pelo Método Krzanowski.

Classificação	% Absoluto	% Acumulado
Categoria correta	59,6	59,6
Categorias adjacentes	14,5	74,1
Categorias distantes	25,8	100

Tabela 5.28: Classificação Real e Predita para a Amostra de Treino - Discriminação Ótima por Perda média linear (PM_{FPL}).

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	48529	1	1	48531
	MED	3635	1	0	3636
	MAU	6347	1	0	6348
	Total	58511	3	1	58515
% por linha	BOM	100	0	0	100
	MED	100	0	0	100
	MAU	100	0	0	100
	Total	100	0	0	100
% pelo total	BOM	82,9	0	0	82,9
	MED	6,2	0	0	6,2
	MAU	10,8	0	0	10,8
	Total	100	0	0	100

Tabela 5.29: Alocação nas categorias da variável resposta para a Amostra de Treino - Discriminação Ótima por Perda média linear (PM_{FPL}).

Classificação	% Absoluto	% Acumulado
Categoria correta	82,9	82,9
Categorias adjacentes	6,2	89,1
Categorias distantes	10,8	100

Tabela 5.30: Classificação Real e Predita para a Amostra de Validação - Discriminação Ótima por Perda média linear (PM_{FPL}).

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	16333	2	1	16336
	MED	1124	0	0	1124
	MAU	2045	0	0	2045
	Total	19502	2	1	19505
% por linha	BOM	100	0	0	100
	MED	100	0	0	100
	MAU	100	0	0	100
	Total	100	0	0	100
% pelo total	BOM	83,7	0	0	83,8
	MED	5,8	0	0	5,8
	MAU	10,5	0	0	10,5
	Total	100	0	0	100

Tabela 5.31: Alocação nas categorias da variável resposta para a Amostra de Validação - Discriminação Ótima por Perda média linear (PM_{FPL}).

Classificação	% Absoluto	% Acumulado
Categoria correta	83,7	83,7
Categorias adjacentes	5,8	89,5
Categorias distantes	10,5	100

Tabela 5.32: Classificação Real e Predita para a Amostra de Treino - Discriminação Ótima por Perda média quadrática (PM_{FPQ}).

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	44949	3581	1	48531
	MED	3135	501	0	3636
	MAU	4902	1446	0	6348
	Total	52986	5528	1	58515
% por linha	BOM	92,6	7,4	0	100
	MED	86,2	13,8	0	100
	MAU	77,2	22,8	0	100
	Total	90,6	9,4	0	100
% pelo total	BOM	76,8	6,1	0	82,9
	MED	5,4	0,9	0	6,2
	MAU	8,4	2,5	0	10,8
	Total	90,6	9,4	0	100

Tabela 5.33: Alocação nas categorias da variável resposta para a Amostra de Treino - Discriminação Ótima por Perda média quadrática (PM_{FPQ}).

Classificação	% Absoluto	% Acumulado
Categoria correta	77,7	77,7
Categorias adjacentes	13,9	91,6
Categorias distantes	8,4	100

Tabela 5.34: Classificação Real e Predita para a Amostra de Validação - Discriminação Ótima por Perda média quadrática (PM_{FPQ}).

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	15108	1227	1	16336
	MED	969	155	0	1124
	MAU	1624	421	0	2045
	Total	17701	1803	1	19505
% por linha	BOM	92,5	7,5	0	100
	MED	86,2	13,8	0	100
	MAU	79,4	20,6	0	100
	Total	90,8	9,2	0	100
% pelo total	BOM	77,5	6,3	0	83,8
	MED	5	0,8	0	5,8
	MAU	8,3	2,2	0	10,5
	Total	90,8	9,2	0	100

Tabela 5.35: Alocação nas categorias da variável resposta para a Amostra de Validação - Discriminação Ótima por Perda média quadrática (PM_{FPQ}).

Classificação	% Absoluto	% Acumulado
Categoria correta	78,3	78,3
Categorias adjacentes	13,4	91,7
Categorias distantes	8,3	100

Tabela 5.36: Classificação Real e Predita para a Amostra de Treino - Discriminação Ótima por Kappa.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	39884	2008	6639	48531
	MED	2583	182	871	3636
	MAU	3562	470	2316	6348
	Total	46029	2660	9826	58515
% por linha	BOM	82,2	4,1	13,7	100
	MED	71	5	24	100
	MAU	56,1	7,4	36,5	100
	Total	78,7	4,5	16,8	100
% pelo total	BOM	68,2	3,4	11,3	82,9
	MED	4,4	0,3	1,5	6,2
	MAU	6,1	0,8	4	10,8
	Total	78,7	4,5	16,8	100

Tabela 5.37: Alocação nas categorias da variável resposta para a Amostra de Treino - Discriminação Ótima por Kappa.

Classificação	% Absoluto	% Acumulado
Categoria correta	72,4	72,4
Categorias adjacentes	10,1	82,5
Categorias distantes	17,4	100

Tabela 5.38: Classificação Real e Predita para a Amostra de Validação - Discriminação Ótima por Kappa.

	Classificação	Classificação predita			Total
	Real	BOM	MED	MAU	
Contagem	BOM	13400	661	2275	16336
	MED	766	75	283	1124
	MAU	1185	168	692	2045
	Total	15351	904	3250	19505
% por linha	BOM	82	4	13,9	100
	MED	68,1	6,7	25,2	100
	MAU	57,9	8,2	33,8	100
	Total	78,7	4,6	16,7	100
% pelo total	BOM	68,7	3,4	11,7	83,8
	MED	3,9	0,4	1,5	5,8
	MAU	6,1	0,9	3,5	10,5
	Total	78,7	4,6	16,7	100

Tabela 5.39: Alocação nas categorias da variável resposta para a Amostra de Validação - Discriminação Ótima por Kappa.

Classificação	% Absoluto	% Acumulado
Categoria correta	72,6	72,6
Categorias adjacentes	9,6	82,2
Categorias distantes	17,7	100

5.4 Considerações Finais

O objetivo deste capítulo era aplicar as técnicas mais comuns para casos em que a resposta é categórica ordinal. Dois dos métodos considerados, o Modelo Logístico Multinomial e a Análise Discriminante Normal, não consideram a ordenação da resposta, mas foram utilizados por serem bastante conhecidos atualmente e para que seus resultados fossem comparados com as técnicas que levam em conta a ordenação da variável resposta em termos de classificação.

A Tabela 5.40 apresenta um resumo do percentual de acerto na categoria exata e nas categorias adjacentes para cada um dos modelos ajustados nessa aplicação. É interessante observar a importância de considerar não apenas o acerto de classificação na categoria exata da resposta, mas também o acerto nas categorias adjacentes, já que o custo de errar em uma categoria mais distante é maior do que em uma categoria adjacente quando trabalhamos com resposta ordinal.

Ao compararmos os modelos de regressão, verificamos que o modelo logístico cumulativo, apesar de ter errado um pouco mais que o multinomial na classificação dentro da categoria exata, obteve um melhor desempenho nas categorias adjacentes. Isso não aconteceu para o modelo de riscos proporcionais que, apesar de considerar a ordenação, teve uma performance inferior. Os modelos cumulativos, atualmente, são fáceis de implementar, dado que boa parte dos pacotes computacionais já estão preparados para ajustá-los. Além disso, assim como os modelos logísticos, possuem a vantagem de exigirem menos em termos de suposição do modelo.

Por outro lado, verificamos que os modelos de análise discriminante que não consideravam a ordenação apresentaram desempenho reduzido em termos de classificação se comparados com os resultados da análise discriminante ótima. Além do mais, esses modelos exigem a normalidade das variáveis preditoras, o que nem sempre é comum de se verificar, principalmente quando se tem variáveis categóricas no conjunto, como aconteceu na nossa aplicação. Caso não existisse esse problema, essa técnica teria a vantagem de ser muito fácil de ser aplicada, dado que é muito conhecida e utilizada atualmente.

É importante observar que os ajustes por análise discriminante ótima que utilizavam os critérios de perda média linear e perda média quadrática não apresentaram uma classificação aceitável, dado que os modelos classificaram todos as pessoas como "BOAS".

Tabela 5.40: Resumo do percentual de alocação em categorias corretas ou adjacentes para os modelos aplicados (Base de validação).

Técnica utilizada	% Classificação na categoria correta	% Classificação em categorias adjacentes	% Total
Modelo Logístico Multinomial	57,9	17,4	75,3
Modelo Logístico Cumulativo	56,0	20,4	76,4
Modelo de Riscos Proporcionais	53,8	19,3	73,1
Análise Discriminante Normal (Função Linear)	57,9	17,9	75,8
Análise Discriminante Normal (Método Krzanowski)	59,6	14,5	74,1
Análise Discriminante Ótima por PM_{FPL}	83,7	5,8	89,5
Análise Discriminante Ótima por PM_{FPQ}	78,3	13,4	91,7
Análise Discriminante Ótima por Kappa ponderada	72,6	9,6	82,2

Isso aconteceu porque a probabilidade a priori dos grupos era bastante diferente. E, por serem modelos extremamente sensíveis ao conjunto de dados do qual se estimam os parâmetros, não era possível ajustar a base fazendo um balanceamento, por exemplo.

A análise discriminante que utilizava o critério de Kappa ponderado, por fim, foi a que, aparentemente, apresentou a melhor classificação da base de validação, trazendo o maior acerto dentro da categoria exata e mantendo a proporção esperada de pessoas por grupo na sua predição. Adicionado ao fato de ser fácil sua aplicação, ela parece ser a melhor opção para classificar os dados que utilizamos nessa aplicação. Contudo, é relevante lembrar que essa técnica continua sendo fortemente sensível a mudanças na população a ser classificada, podendo trazer grandes erros de classificação no momento em que isso ocorrer. Mais que as técnicas anteriores, torna-se necessário revisar os

parâmetros desse modelo com uma freqüência maior.

Uma análise mais aprofundada poderia ser feita utilizando-se a análise discriminante ótima com uso de *bootstrap* e considerando-se custos de erros de classificação diferentes, por exemplo. Além disso, um estudo mais detalhado em termos de diagnóstico poderia ser feito para comparar os modelos ordinais com os não ordinais.

Campbell, Donner e Webster (1991) aplicam alguns dos modelos aqui utilizados a dados simulados e discutem sobre a utilidade dos modelos ordinais em termos de classificação, observando que podem classificar menos acuradamente em uma série de circunstâncias. No entanto, verificamos que o uso de técnicas que consideram a ordenação, em alguns casos, mostrou-se vantajosa para classificação dos dados financeiros utilizados nessa aplicação, levando a um percentual maior de acerto.

Apêndice A

Análise descritiva das variáveis consideradas na aplicação

Tabela A.1: Distribuição Conjunta da Variável Binária X_1 e Y.

X_1	BOM	MED	MAU	Total
0	37.837 (0,78)	2.663 (0,73)	4.365 (0,69)	44.865 (0,77)
1	10.694 (0,22)	973 (0,27)	1.983 (0,31)	13.650 (0,23)
Total	48.531 (1,00)	3.636 (1,00)	6.348 (1,00)	58.515 (1,00)

Tabela A.2: Distribuição Conjunta da Variável Qualitativa Ordinal X_2 e Y.

X_2	BOM	MED	MAU	Total
1	4.497 (0,00)	391 (0,11)	1.188 (0,19)	6.076 (0,10)
2	7.888 (0,16)	599 (0,16)	1.340 (0,21)	9.827 (0,17)
3	11.536 (0,24)	839 (0,23)	1.423 (0,22)	13.798 (0,24)
4	24.610 (0,51)	1.807 (0,50)	2.397 (0,38)	28.814 (0,49)
Total	48.531 (1,00)	3.636 (1,00)	6.348 (1,00)	58.515 (1,00)

Tabela A.3: Medidas Resumo da Variável Contínua X_3 por categoria de resposta Y.

X_3	BOM	MED	MAU	Total
mínimo	0,0	0,0	0,0	0,0
média	818,1	764,7	711.3	803,2
máximo	999,0	999,0	998,0	999,0
desvio padrão	156,1	179,2	179,4	163,9

Figura A.1: Boxplot da Variável X_3 por categoria de resposta.

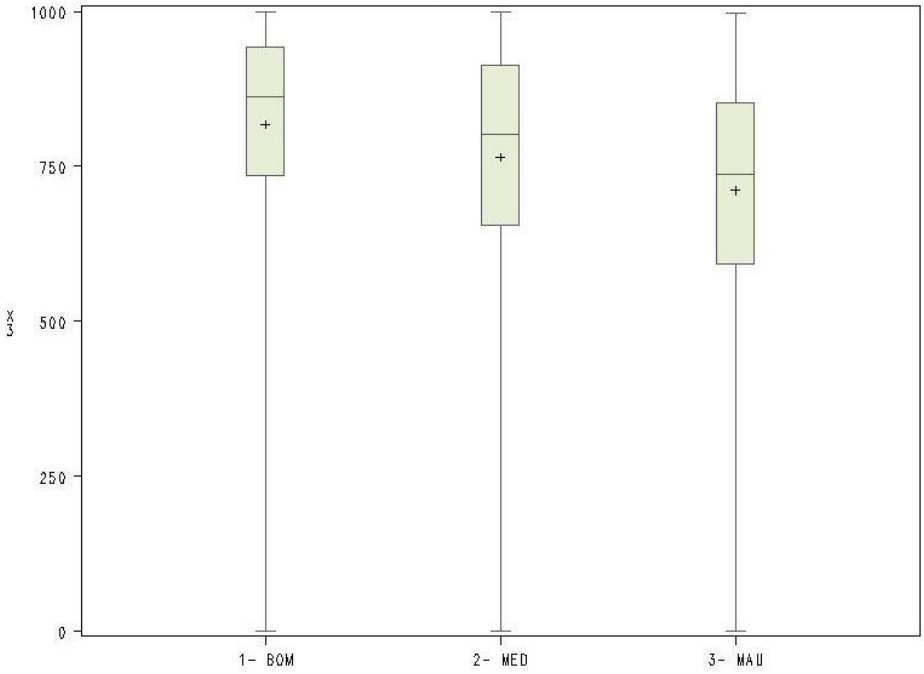
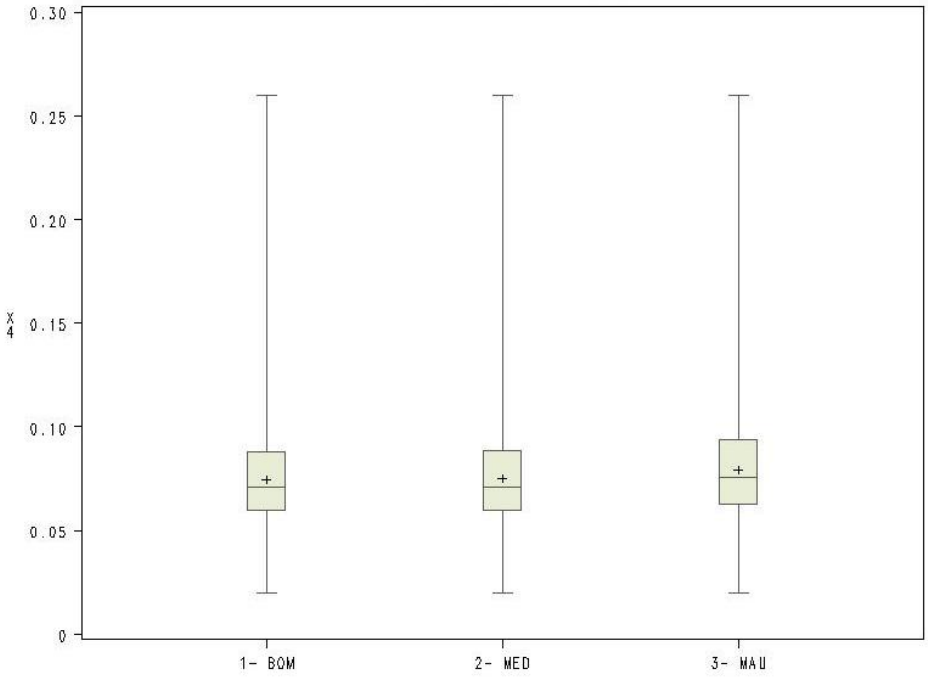


Tabela A.4: Medidas Resumo da Variável Contínua X_4 por categoria de resposta Y.

X_4	BOM	MED	MAU	Total
mínimo	0,020	0,020	0,020	0,020
média	0,075	0,075	0,079	0,075
máximo	0,131	0,260	0,260	0,260
desvio padrão	0,022	0,023	0,023	0,022

Figura A.2: Boxplot da Varíável X_4 por categoria de resposta.



Apêndice B

Comandos SAS para ajustar os modelos da aplicação

```
/************************************/
/* M1 - Modelo Logístico Multinomial */
/************************************/

/* Estimação dos parâmetros */
proc logistic data=mest.base_ord_t_bal outest=betas
outmodel=mest.infomodel1
OUTDESIGN=matriz_delineamento;
class X2 /order=formatted param=ref;
model status_pagador2 =  X2 X1 X3 X4
/ link = glogit selection=backward slstay=0.1;
output PRED = Pred
PREDPROBS = Individual XBETA = Xbeta;
run;


/*******************************************************/
/* M2.A - Modelo Cumulativo com função de ligação log  */
/*******************************************************/
```

```
/* Estimação dos parâmetros */
proc logistic data=mest.base_ord_t_bal outest=betas
outmodel=mest.infomodel2a
OUTDESIGN=matriz_delineamento rorder=data;
class X2 /order=formatted param=ref;
model status_pagador2 = X1 X2 X3 X4
/ link = CUMlogit selection=backward slstay=0.1;
output
PRED = Preda
PREDPROBS = Individual
XBETA = Xbeta;
run;

/*************************************************************************/
/* M2.B - Modelo Cumulativo com função de ligação complementar log-log */
/*************************************************************************/

/* Estimação dos parâmetros */
proc logistic data=mest.base_ord_t_bal outest=betas
outmodel=mest.infomodel2b
OUTDESIGN=matriz_delineamento rorder=data;
class X2
  /order=formatted param=ref;
model status_pagador2 = X1 X2 X3 X4
/ link = CUMcloglog
selection=backward slstay=0.1;
output
PRED = Pred
PREDPROBS = Individual
XBETA = Xbeta;
run;
```

```
/*************************************************************************/
/* M3 - Análise Discriminante com variáveis explicativas categóricas */
/*************************************************************************/

/* MÉTODO 1 */
/* Criar indicadoras para as variáveis que são categóricas (X2) */

data base_ord_t;
set mest.base_ord_t;
X2_1 = 0;
X2_2 = 0;
X2_3 = 0;
if X2 = "1" then X2_1 = 1;
if X2 = "2" then X2_2 = 1;
if X2 = "3" then X2_3 = 1;
run;

data base_ord_v;
set mest.base_ord_v;
X2_1 = 0;
X2_2 = 0;
X2_3 = 0;
if X2 = "1" then X2_1 = 1;
if X2 = "2" then X2_2 = 1;
if X2 = "3" then X2_3 = 1;
run;

proc discrim data=base_ord_t testdata= base_ord_v
out=plot testout=plotp testoutd=plotd
method=normal pool=yes;
class status_pagador2;
var X1 X2_1 X2_2 X2_3 X3 X4;
run;
```

```
/* MÉTODO 2 */
/* Criar as combinações de grupos para as variáveis que são
 categóricas (X2) */

data  baseT_X1_0_X2_1 /*4.594*/
 baseT_X1_0_X2_2 /*7.382*/
 baseT_X1_0_X2_3 /*10.347*/
 baseT_X1_0_X2_4 /*22.542*/
 baseT_X1_1_X2_1 /*1.482*/
 baseT_X1_1_X2_2 /*2.445*/
 baseT_X1_1_X2_3 /*3.451*/
 baseT_X1_1_X2_4 /*6.272*/
;
set  mest.base_ord_t;
if X1  = 0 and X2 = "1" then output  baseT_X1_0_X2_1;
else if X1  = 0 and X2 = "2" then output  baseT_X1_0_X2_2;
else if X1  = 0 and X2 = "3" then output  baseT_X1_0_X2_3;
else if X1  = 0 and X2 = "4" then output  baseT_X1_0_X2_4;
else if X1  = 1 and X2 = "1" then output  baseT_X1_1_X2_1;
else if X1  = 1 and X2 = "2" then output  baseT_X1_1_X2_2;
else if X1  = 1 and X2 = "3" then output  baseT_X1_1_X2_3;
else if X1  = 1 and X2 = "4" then output  baseT_X1_1_X2_4;
keep X1 X2 X3 X4 status_pagador2;
run;

data  baseV_X1_0_X2_1 /*1.537*/
 baseV_X1_0_X2_2 /*2.530*/
 baseV_X1_0_X2_3 /*3.415*/
 baseV_X1_0_X2_4 /*7.648*/
 baseV_X1_1_X2_1 /*444*/
 baseV_X1_1_X2_2 /*810*/
 baseV_X1_1_X2_3 /*1.070*/
 baseV_X1_1_X2_4 /*2.045*/
;
```

```
set  mest.base_ord_V;
if X1  = 0 and X2 = "1" then output  baseV_X1_0_X2_1;
else if X1  = 0 and X2 = "2" then output  baseV_X1_0_X2_2;
else if X1  = 0 and X2 = "3" then output  baseV_X1_0_X2_3;
else if X1  = 0 and X2 = "4" then output  baseV_X1_0_X2_4;
else if X1  = 1 and X2 = "1" then output  baseV_X1_1_X2_1;
else if X1  = 1 and X2 = "2" then output  baseV_X1_1_X2_2;
else if X1  = 1 and X2 = "3" then output  baseV_X1_1_X2_3;
else if X1  = 1 and X2 = "4" then output  baseV_X1_1_X2_4;
keep X1 X2 X3 X4 status_pagador2;
run;

proc discrim data= baseT_X1_0_X2_1 testdata=  baseV_X1_0_X2_1
out=plot testout=plotp testoutd=plotd
method=normal pool=yes;
      class status_pagador2;
      var X3 X4;
  title1 'Discriminant Analysis';
      title2 'Using Normal Density Estimates with Equal Variance';
   run;

proc discrim data= baseT_X1_0_X2_2 testdata=  baseV_X1_0_X2_2
out=plot testout=plotp testoutd=plotd
method=normal pool=yes;
      class status_pagador2;
      var X3 X4;
  title1 'Discriminant Analysis';
      title2 'Using Normal Density Estimates with Equal Variance';
   run;
proc discrim data= baseT_X1_0_X2_3 testdata=  baseV_X1_0_X2_3
out=plot testout=plotp testoutd=plotd
method=normal pool=yes;
      class status_pagador2;
      var X3 X4;
```

```
  title1 'Discriminant Analysis';
      title2 'Using Normal Density Estimates with Equal Variance';
   run;
proc discrim data= baseT_X1_0_X2_4 testdata=  baseV_X1_0_X2_4
out=plot testout=plotp testoutd=plotd
method=normal pool=yes;
      class status_pagador2;
      var X3 X4;
  title1 'Discriminant Analysis';
      title2 'Using Normal Density Estimates with Equal Variance';
   run;
proc discrim data= baseT_X1_1_X2_1 testdata=  baseV_X1_1_X2_1
out=plot testout=plotp testoutd=plotd
method=normal pool=yes;
      class status_pagador2;
      var X3 X4;
  title1 'Discriminant Analysis';
      title2 'Using Normal Density Estimates with Equal Variance';
   run;
proc discrim data= baseT_X1_1_X2_2 testdata=  baseV_X1_1_X2_2
out=plot testout=plotp testoutd=plotd
method=normal pool=yes;
      class status_pagador2;
      var X3 X4;
  title1 'Discriminant Analysis';
      title2 'Using Normal Density Estimates with Equal Variance';
   run;
proc discrim data= baseT_X1_1_X2_3 testdata=  baseV_X1_1_X2_3
out=plot testout=plotp testoutd=plotd
method=normal pool=yes;
      class status_pagador2;
      var X3 X4;
  title1 'Discriminant Analysis';
      title2 'Using Normal Density Estimates with Equal Variance';
```

```
   run;
proc discrim data= baseT_X1_1_X2_4 testdata=  baseV_X1_1_X2_4
out=plot testout=plotp testoutd=plotd
method=normal pool=yes;
      class status_pagador2;
      var X3 X4;
  title1 'Discriminant Analysis';
      title2 'Using Normal Density Estimates with Equal Variance';
   run;
```

Apêndice C

Programa utilizado na Análise Discriminante Ótima

```
/* Estimação dos parâmetros */
/* Relação linear entre resposta e preditoras */
/* Como essa proc não escora depois a base de validação separadamente,
tenho que colocar de input todos os casos (T e V), com Y = . para os
casos de V, pois aí eles não serão considerados na estimação dos
parâmetros, mas serão escorados juntos no mesmo passo */
proc genmod data = mest.base_ord_tot2;
class  X2 /order=formatted;
model status_pagador3 =  X1 X2 X3 X4  /* Considera a variável numérica
que assume valores 1, 2 e 3*/
/link = log;
output out=saida_m4  xbeta=xb;
run;

data ScoredT4 ScoredV4;
set saida_m4;
if status_pagador3 eq . then do;
if status_pagador2='1- BOM'  then status_pagador3 = 1;
else if status_pagador2='2- MED'  then status_pagador3 = 2;
```

```
else             status_pagador3 = 3;
end;
if selected = 0 then output ScoredT4;
else output ScoredV4;
run;

/* A base de treino foi utilizada para definir os parâmetros da regressão
e agora será utilizada para definir os pontos de corte na escala de Y
que vão definir os três grupos da variável resposta. A base de validação
só será utilizada, como em todos os casos, para verificação da
classificação*/

%macro def_pto_corte (arq_in= , arq_out=);

/*options compress=yes symbolgen mprint mlogic NOXWAIT XSYNC;*/

proc datasets library = work nolist;
delete &arq_out;
quit;

/* Registrar o valor mínimo e o valor máximo estimado para variável
resposta pelo modelo linear ajustado*/

PROC UNIVARIATE DATA=&arq_in noprint;
VAR xb;
OUTPUT OUT=aux_pto_corte min=minimo max=maximo;
RUN;

/* Multiplico esses valores por 1000000 para depois fazer o loop by 25000,
criando várias opções de possíveis valores de y chapéu. Se deixasse na
escala original, não seria possível fazer isso, pois não é um valor
inteiro e os contadores k e l não poderiam ser utilizados */

data _null_;
```

```
set aux_pto_corte;
call symput ("min", round(minimo*1000000));
call symput ("max", round(maximo*1000000));
/*call symput ("m", 1);*/
run;

%do k = &min %to &max %by 25000 ;
%do l = &k+25000 %to &max %by 25000;
/* Classifica os casos nas três categorias da variável resposta para cada
par de pontos de corte */
data pontos_corte;
set &arq_in;
%let A=&k/1000000;
%let B=&l/1000000;
if xb < &A then valorpredito = 1;
else if &A <= xb < &B then valorpredito = 2;
else valorpredito = 3;
run;

/* Calcular a estatística Kappa ponderada */
proc freq data=pontos_corte noprint;
table status_pagador3 * valorpredito / missing out=saida_freq agree;
output out=valor_kappa agree;
/*weight valorpredito / zeros;*/
run;

data temporaria;
set saida_freq;
controle= compress(status_pagador3)||compress(valorpredito);
run;
proc transpose data=temporaria prefix=cel out=result_pontos;
var count;
id controle;
run;
```

```
data result_pontos;
set result_pontos;
ponto1 = &A;
ponto2 = &B;
if cel11 = . then cel11=0;
if cel22 = . then cel22=0;
if cel33 = . then cel33=0;
if cel12 = . then cel12=0;
if cel13 = . then cel13=0;
if cel21 = . then cel21=0;
if cel23 = . then cel23=0;
if cel31 = . then cel31=0;
if cel32 = . then cel32=0;

/*acerto = cel11/(cel11 + cel12 + cel13) + cel22/(cel21 + cel22 + cel23)
 + cel33/(cel31 + cel32 + cel33);*/

/* Cálculo da Perda Média Linear */
PML = (cel12 + 2*cel13 + cel21 + cel23 + 2*cel31 + cel32)/
(cel11 + cel22 + cel33 + cel12 + cel13 + cel21 + cel23 + cel31 + cel32);

/* Cálculo da Perda Média Quadrática */
PMQ = (cel12 + 4*cel13 + cel21 + cel23 + 4*cel31 + cel32)/
(cel11 + cel22 + cel33 + cel12 + cel13 + cel21 + cel23 + cel31 + cel32);

run;

data resumo;
merge result_pontos valor_kappa;
keep cel11 cel12 cel13 cel21 cel22 cel23 cel31 cel32 cel33
ponto1 ponto2 /*acerto*/ PML PMQ _WTKAP_;
run;

data &arq_out;
```

```
set &arq_out resumo;
run;

proc datasets library = work nolist;
delete valor_kappa resumo;
quit;

%end;

%end;

PROC UNIVARIATE DATA=&arq_out noprint; VAR PML; OUTPUT OUT=aux_PML
min=min_PML; RUN;
PROC UNIVARIATE DATA=&arq_out noprint; VAR PMQ; OUTPUT OUT=aux_PMQ
min=min_PMQ; RUN;
PROC UNIVARIATE DATA=&arq_out noprint; VAR _WTKAP_; OUTPUT OUT=aux_WTKAP
max=max_WTKAP; RUN;

data _null_; set aux_PML; call symput ("min_PML", min_PML); run;
data _null_; set aux_PMQ; call symput ("min_PMQ", min_PMQ); run;
data _null_; set aux_WTKAP; call symput ("max_WTKAP", max_WTKAP); run;

%put &min_PML
&min_PMQ
&max_WTKAP;

proc print data = &arq_out; run;

%MEND;

%def_pto_corte (arq_in= ScoredT4, arq_out=base_treino_total);
```

Referências Bibliográficas

Anderson, J. A. e Philips, P. R. (1981), 'Regression, discrimination and measurement models for ordered categorial variables', *Applied Statistics* **30**, 22–31.

Anderson, J. A. e Richardson, S. C. (1979), 'Logistic discrimination and bias correction in maximum likelihood estimation', *Technometrics* **31**, 71–78.

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis*, New York: John Wiley and Sons.

Campbell, M. K., Donner, A. e Webster, K. M. (1991), 'Are ordinal models useful for classification?', *Statistics in Medicine* **10**, 383–394.

Cicchetti, D. V. e Allison, T. (1971), 'A new procedure for assessing reliability of scoring eeg sleep recordings', *American Journal of EEG Technology* **11**, 101–109.

Coste, J., Walter, E., Wasserman, D. e Venot, A. (1997), 'Optimal discriminant analysis for ordinal responses', *Statistics in Medicine* **16**, 561–569.

Efron, B. (1975), 'The efficiency of logistic regression compared to normal discriminant analysis', *Journal of the American Statistical Association* **70**, 892–898.

Efron, B. (1979), 'Bootstrap methods: Another look at the jackknife', *The Annals of Statistics* **7**, 1–26.

Fahrmeir, L. e Tutz, G. (1994), *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer Verlag.

Fisher, R. A. (1936), 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics* **7**, 179–188.

Forthofer, R. N. e Lehnen, R. G. (1981), *Public Program Analysis. A new categorical data approach*, Belmont, Calif: Lifetime Learning Publications.

Holmes, M. C. e Williams, R. (1954), 'The distribution of carriers of streptococcus pyogenes among 2413 healthy children', *J. Hyg. Camb.* **52**, 165–179.

Johnson, R. A. e Wichern, D. W. (2007), *Applied Multivariate Statistical Analysis*, Prentice Hall.

Le Teuff, G., Quantin, C., Venot, A., Walter, E. e Coste, J. (2005), 'Improving model robustness with bootstrapping - application to optimal discriminant analysis for ordinal response (odao)', *Methods of Information in Medicine* **44**, 704–711.

Marshall, A. W. e Olkin, I. (1968), 'A general approach to some screening and classification problems', *Journal of the Royal Statistical Society, Ser.* **B 30**, 407–443.

Mehta, C. R., Patel, N. R. e Tsiatis, A. A. (1984), 'Exact significance testing to establish treatment equivalence with ordered categorical data', *Biometrics* **40**, 819–825.

Morawitz, B. e Tutz, G. (1990), 'Alternative parameterizations in business tendency surveys', *Methods and Models of Operations Research* **34**, 143–156.

Press, S. J. e Wilson, S. (1978), 'Choosing between logistic regression and discriminant analysis', *Journal of the American Statistical Association* **73**, 669–705.

Sanda, R. (1990), *Análise Discriminante com mistura de variáveis categóricas e contínuas*, Dissertação de Mestrado, Instituto de Matemática e Estatística da Universidade de São Paulo.

Tutz, G. (1989), 'Compound regression models for categorical ordinal data', *Biometrical Journal* **31**, 259–272.