

Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

MODELO DE CALIBRAÇÃO BETA

Mileno Tavares Cavalcante

Trabalho submetido ao Programa de Pós-Graduação em
Estatística do Departamento de Estatística da Universidade
Federal de Pernambuco como requisito parcial para obtenção
do grau de Mestre em Estatística.

Área de Concentração: **Estatística Aplicada**
Orientadora: **Profa. Dra. Betsabé Grimalda Blas Achic**

Recife, fevereiro de 2013

Catálogo na fonte
Bibliotecária Jane Souto Maior, CRB4-571

Cavalcante, Mileno Tavares
Modelo de calibração beta. / Mileno Tavares Cavalcante.
- Recife: O Autor, 2013.
xii, 58 folhas: il., fig., tab.

Orientadora: Betsabé Grimalda Blas Achic.
Dissertação (mestrado) - Universidade Federal de
Pernambuco. CCEN, Estatística, 2013.

Inclui bibliografia e apêndice.

1. Estatística aplicada - Modelagem. 2. Modelos de calibração.
I. Achic, Betsabé Grimalda Blas (orientadora). II. Título.

519.536

CDD (23. ed.)

MEI2013 – 038

Universidade Federal de Pernambuco
Pós-Graduação em Estatística

Recife, 19 de fevereiro de 2013.

Nós recomendamos que a dissertação de mestrado de autoria de

Mileno Tavares Cavalcante

Intitulada

“Modelo de Calibração Beta”

Seja aceita como cumprimento parcial dos requerimentos para o grau de Mestre em Estatística.

Coordenador da Pós-Graduação em Estatística

Banca Examinadora:

Betsabé Grimalda Blas Achic

Orientadora / UFPE

Francisco Cribari Neto

UFPE

Mônica Carneiro Sandoval

USP

Este documento será anexado à versão final da tese.

Dedicatória

À Kelly, Arthur, e a todos que contribuíram para este trabalho.

Agradecimentos

Agradecer nem sempre é algo trivial, especialmente pelo risco de esquecimento a que estamos sujeitos. Mesmo assim, não seria justo com quem tornou a realização deste trabalho possível, pecar por omissão neste assunto.

Sendo assim, gostaria, em primeiro lugar, de agradecer à minha orientadora, Betsabé Blas Achic, pela dedicação, atenção e otimismo, desde o início, mesmo com minhas severas restrições de tempo para desenvolver este trabalho de forma adequada. À ela devo esta dissertação. Agradeço também ao Coordenador da Pós-graduação em Estatística, Prof. Francisco Cribari, e aos professores Patrícia Espinheira, Raydonal Ospina, Audrey Cysneiros e Alex Dias Ramos, pela compreensão em relação aos limites impostos à quem teve que cursar o Mestrado em Estatística sem o afastamento total das atividades profissionais. À Profa. Fernanda Pimentel, do Departamento de Engenharia Química da Universidade Federal de Pernambuco (UFPE), pelo esclarecimento de várias dúvidas referentes a aplicação de modelos de calibração em química analítica, um dos assuntos desta dissertação. Aos professores Francisco Cribari-Neto e Mônica Sandoval, por aceitarem o convite para compor a banca de avaliação deste trabalho, e pelos comentários e sugestões apresentadas.

À Valeria Bittencourt, pela dedicação, simpatia, atenção e competência.

À Petrobras, nas pessoas do meu gerente geral, Milton Vaz da Silva, e meu gerente imediato, Leonardo Gouveia, pela autorização para redução de parte de minha jornada de trabalho, o que possibilitou minha presença frequente nas atividades do programa de pós-graduação em estatística da UFPE. Aos demais colegas da Petrobras, em Recife e no Rio de Janeiro, pelo apoio e incentivo.

Claro que não poderia esquecer meus colegas da UFPE, sem os quais este período na Universidade não teria sido tão interessante e enriquecedor: Marcos, Sérgio, Jéssica, Danyele, Heloísa, Pedro, Diego e todos os demais que fazem a Pós-graduação em Estatística.

À Kelly, minha esposa, e Arthur, meu filho, agradeço às horas roubadas de vocês. Sem este tempo gentilmente cedido, ainda que sob alguns protestos, não chegaria até aqui. À vocês também devo este trabalho.

La logique qui peut seule donner la certitude est l'instrument de la démonstration:
l'intuition est l'instrument de l'invention.

HENRI POINCARÉ

Resumo

O presente trabalho discute o problema de calibração em química analítica no contexto de não linearidade dos dados. A hipótese principal é que a média da variável resposta está restrita ao intervalo $(0, 1)$ e pode ser modelada por uma distribuição beta, de modo similar ao modelo de regressão beta (Ferrari e Cribari-Neto, 2004). O objetivo é propor uma extensão do modelo de regressão beta a estudos de calibração e verificar as propriedades de seu estimador para a concentração do analito x comparativamente aos modelos linear e quadrático, que supõe resíduos normalmente distribuídos com variância constante. Aplicações a dados reais para os modelos considerados são apresentadas.

Palavras-chave: Modelos de calibração, regressão beta, transmitância, lei de Beer-Lambert.

Abstract

This work discusses the calibration problem for nonlinear data in the context of analytical chemistry. We propose a new calibration model where for each calibration stage the response variable is beta distributed using a reparameterization of the beta law in terms of its mean and dispersion parameter (Ferrari e Cribari-Neto, 2004). This new approach is suitable for applications in chemical analysis where the response variable is continuous and restricted to the interval (0,1) and is related to other variable through a regression structure. The new model properties are compared with linear and quadratic models in terms of x_0 , the estimator of analyte concentration, where these two models assume normally distributed residuals with constant variance. Applications to real data set are also presented and discussed.

Keywords: Calibration models, beta regression models, transmittance, Beer-Lambert law.

Lista de Figuras

2.1	Intensidades I_0 e I_1 de um raio de luz através de um recipiente com comprimento l	6
2.2	Transmitância e absorvância para diferentes concentrações de um cromóforo	7
2.3	Relação entre o domínio das amostras e o domínio dos sinais no caso de análise elementar	8
2.4	Representação da relação entre calibração qualitativa e quantitativa	8
3.1	Densidades beta para diferentes valores de (p, q)	18
4.1	Absorvância e transmitância para diferentes concentrações do analito x	33
4.2	Absorvância e transmitância para diferentes concentrações do analito x com $\lambda = 410$	37
4.3	Absorvância e transmitância para diferentes concentrações do analito x com $\lambda = 440$	38
4.4	Absorvância e transmitância para diferentes concentrações do analito x com $\lambda = 530$	38

Lista de Tabelas

4.1	Resultados de uma determinação de ferro (Fe) por absorvância molecular	33
4.2	Estimativas e erros-padrão para os modelos de calibração usual, quadrático e beta . . .	34
4.3	Erro de \hat{x}_0 e intervalos de confiança aproximados para x_0 ($\gamma = 95\%$)	35
4.4	Crerios de informação para os modelos usual, quadrático e beta	35
4.5	Determinação da concentração do íon MnO_4^+ em soluções-padrão por espectrofotometria	37
4.6	Estimativas e erros-padrão para os modelos usual, quadrático e beta ($\lambda = 410$) . . .	39
4.7	Estimativas e erros-padrão para os modelos usual, quadrático e beta ($\lambda = 440$) . . .	40
4.8	Estimativas e erros-padrão para os modelos usual, quadrático e beta ($\lambda = 530$) . . .	41
4.9	Erro de \hat{x}_0 e intervalos de confiança aproximados para x_0 ($\gamma = 95\%$)	42
4.10	Crerios de informação para os modelos usual, quadrático e beta com $\lambda = 410, 440,$ 530	43

Índice

1	Introdução	1
1.1	Calibração: Definições e Histórico	1
1.2	Organização da Dissertação	4
1.3	Suporte Computacional	4
2	Modelos de Calibração Linear e Quadrático	5
2.1	Espectroscopia em Química Analítica	5
2.2	O Problema de Calibração	7
2.3	Modelo de Calibração Usual	10
2.4	Modelo de Calibração Quadrático	13
3	Modelo de Calibração Beta	17
3.1	A Distribuição Beta	17
3.2	Modelo de Regressão Beta	19
3.3	Modelo de Calibração Beta	22

3.3.1	Estrutura do Modelo de Calibração Beta	22
3.3.2	Matriz de Informação de Fisher Esperada	25
4	Aplicações	31
4.1	Critérios de Informação	31
4.2	Concentração de Ferro em Águas Naturais	32
4.3	Concentração de Permanganato	36
5	Conclusões	44
5.1	Comentários Finais	44
5.2	Sugestões para Estudos Futuros	45
	Referências Bibliográficas	46
A	Códigos em <i>R</i>	51

Introdução

Neste capítulo faz-se uma breve introdução ao presente trabalho, e procura-se contextualizá-lo no âmbito da literatura já existente sobre modelos de calibração. Com este objetivo, além da apresentação das definições necessárias ao entendimento do tema, discute-se a evolução dos estudos de calibração ao longo dos anos e descreve-se a estrutura desta dissertação.

1.1 Calibração: Definições e Histórico

Os modelos de calibração têm uma ampla gama de aplicações nos mais diversos ramos de conhecimento humano, sendo possível encontrar-se exemplos neste sentido nas áreas de biologia, *marketing*, física e economia. Nesta última área, em particular, estes modelos podem ser usados para avaliar os efeitos de mudanças nas políticas do governo, como, por exemplo, a alteração da alíquota de um imposto ou a eliminação de uma quota de importação para um produto (Dawkins, Srinivazan, Whalley, 2001). Por outro lado, em engenharia e física, técnicas de calibração têm sido utilizadas para aferir instrumentos de medição de grandezas físicas (Guimarães-Filho e Helene, 2003). Já em química e biologia, a calibração pode ser útil na determinação da composição de materiais, além da concentração de substâncias (Thomas e Ge, 2000).

O problema de calibração pode ser descrito a partir da relação entre as variáveis x e y , que é estabelecida através de uma função conhecida $h(\cdot)$. Em geral, a relação entre x e y não é exata, e neste caso a mesma pode ser escrita como $y = h(x) + \zeta$, onde ζ é o erro do modelo.

Os modelos de calibração são, usualmente, compostos por dois estágios. No primeiro estágio, co-

nhcedido como *processo de calibração*, obtêm-se n observações de uma variável aleatória y , (y_1, y_2, \dots, y_n) , associadas a valores fixos e conhecidos de x , (x_1, x_2, \dots, x_n) . No segundo estágio, o *experimento de calibração* propriamente dito, observam-se k realizações de y , ou seja $(y_{01}, y_{02}, \dots, y_{0k})$, $k \geq 1$, associadas a uma quantidade desconhecida x_0 . Para os modelos de calibração, o interesse principal reside em estimar x_0 , o que é feito conjuntamente com a função $h(\cdot)$ por meio de técnicas estatísticas.

Em química analítica, tais modelos são comumente utilizados para estimar a concentração desconhecida de um ou mais elementos em uma determinada amostra, a partir do uso de soluções-padrão com concentrações conhecidas. Neste sentido, pode-se definir calibração em química analítica como a relação entre o domínio das amostras (x) e o domínio dos sinais (y) (Danzer, Otto e Currie, 2004).

Na literatura, há vários estudos que tratam o problema de calibração assumindo que a função $h(\cdot)$, ou seja, a relação entre as variáveis y e x , é linear, e que o erro do modelo, ζ , segue uma distribuição normal. Neste caso, há dois estimadores pontuais que devem ser considerados para x_0 : o *clássico* e o *inverso* (Shukla, 1972). O primeiro é obtido a partir da estimação conjunta da função $h(\cdot)$ e de x_0 pelo método dos mínimos quadrados ou de máxima verossimilhança, enquanto o segundo baseia-se na regressão inversa de x sobre y , ou seja, na estimação de $h^{-1}(\cdot)$ por mínimos quadrados. Com o objetivo de estabelecer qual destes estimadores seria o melhor, vários autores propuseram diversos critérios para compará-los, por exemplo Eisenhart (1939), Krutchkoff (1967), Williams (1969a), Berkson (1969), Shukla (1972) e Lwin (1981). Para o presente trabalho, somente o estimador clássico será considerado.

No Brasil, o problema de calibração começou a atrair o interesse de pesquisadores a partir da década de 1980, quando Ramos (1987) apresentou um estudo sobre o modelo de calibração linear simples (univariado). Segundo Marciano (2012), na década seguinte, Galea-Rojas (1995) discutiu um modelo de calibração comparativa estrutural e funcional, Francisconi (1996) comparou instrumentos de medição utilizando calibração comparativa e Lima (1996) estudou calibração absoluta com erros na variáveis.¹ Pimentel e Barros Neto (1996) discutiram o modelo de calibração no contexto da química analítica, supondo $h(\cdot)$ linear e erros normalmente distribuídos.

Branco (1997), em um contexto bayesiano, analisou o problema da calibração em duas situações:

¹De acordo com Williams (1969b), a calibração absoluta utiliza uma medida padrão conhecida, ou então assume que o erro experimental é desprezível. Já a calibração relativa baseia-se no uso de vários instrumentos ou técnicas de medição, sem considerar nenhum(a) deles(as) como padrão. Neste último caso, cabe ao experimentador determinar a relação entre diversos testes ou instrumentos que fornecem mediadas similares (ver Marciano (2012)).

i) assumindo que o erro ζ era proveniente de uma distribuição elíptica dependente; e ii) supondo que este erro era função de um modelo elíptico independente. Este trabalho generalizou o uso de diversas distribuições que poderiam ser utilizadas para modelar o erro aleatório segundo o enfoque bayesiano (Figueiredo, 2009).

Bolfarine, Lima e Sandoval (1997), trabalhando com os estimadores clássico e inverso de x_0 para um modelo de calibração funcional com razão de variâncias conhecida, obtiveram aproximações de primeira ordem para o viés e erro quadrático médio (EQM) destes estimadores. Em seguida, os autores realizaram um estudo similar considerando apenas uma das variâncias conhecidas para o modelo de calibração funcional (Bolfarine, Lima e Sandoval, 1999).

Nos anos 2000, diversos estudos discutiram os modelos de calibração supondo a presença de erros de medida nas variáveis independentes em vários contextos, como Domingos-Filho (2000), Blas (2005), Arellano-Valle, Ozan, Bolfarine e Lachos (2005), Blas, Sandoval e Yoshida (2007), Blas e Sandoval (2010) e Blas, Bolfarine e Lachos (2011). Uma outra linha de trabalhos interessante é a que discute os modelos de calibração considerando que os erros seguem uma distribuição normal-assimétrica, por exemplo Arellano-Valle e Branco (2004), Lachos (2004), Rodriguez (2005) e Figueiredo (2009). Em química analítica, Oliveira e Aguiar (2009) sugeriram o uso de um modelo polinomial de segundo grau para ajustar a relação entre y e x no caso de uma variável resposta não-linear, obtendo ainda uma expressão para a variância do estimador de x_0 para este caso.

Mais recentemente, Marciano (2012) estudou o problema de calibração linear com repetições na variável resposta, supondo erros simétricos para as distribuições normal, t -Student, exponencial potência e logística tipo II.

O presente estudo difere dos anteriores por propor a modelagem da média da variável resposta y do modelo de calibração a partir do uso de uma distribuição beta, ao invés da modelagem dos erros ζ deste modelo. A distribuição beta é conhecida por sua grande flexibilidade, sendo possível obter-se vários formatos para sua função densidade de probabilidade a partir da escolha de diferentes valores para seus parâmetros. O modelo de calibração proposto é derivado a partir do modelo de regressão beta, sugerido por Ferrari e Cribari-Neto (2004).

1.2 Organização da Dissertação

A presente dissertação é composta por cinco capítulos, incluindo esta introdução. No Capítulo 2 apresenta-se a lei de Beer-Lambert, descreve-se o problema da calibração em química analítica e discute-se este problema no contexto dos modelos linear e quadrático (ou polinomial de segundo grau), tomando como base os tipos de variáveis resposta coerentes com a lei de Beer-Lambert.

No Capítulo 3 discute-se brevemente o modelo de regressão beta sugerido por Ferrari e Cribari-Neto (2004) e propõe-se um modelo de calibração para uma variável resposta restrita ao intervalo unitário $(0, 1)$, cuja média segue uma distribuição beta. No modelo proposto, considera-se uma reparametrização desta distribuição seguindo a linha adotada por Ferrari e Cribari-Neto (2004), o que permite a obtenção de uma estrutura de regressão para sua média para o primeiro e segundo estágio. Obtêm-se ainda as funções escores para este modelo, discute-se a estimação de seus parâmetros pelo método da máxima verossimilhança e chega-se a sua matriz de informação de Fisher esperada.

Duas aplicações a dados reais da química para os modelos de calibração usual, quadrático e beta, discutidos nos Capítulos 2 e 3, são apresentadas no Capítulo 4. A primeira aplicação utiliza uma amostra para a concentração de íons de ferro em águas naturais (Pimentel e Barros Neto, 1996), enquanto a segunda baseia-se em dados sobre o teor de permanganato em soluções-padrão, com leituras espectrofotométricas para diferentes comprimentos de onda (Barros Neto, Scarminio e Bruns, 2010). Por fim, o Capítulo 5 apresenta as principais conclusões deste trabalho.

1.3 Suporte Computacional

Os códigos para estimação dos modelos apresentados nesta dissertação, bem como as figuras referentes aos dados das aplicações, foram produzidos no ambiente de programação R (versão 2.14.2) para o sistema operacional Windows. R é um sistema integrado para computação estatística e geração de gráficos e encontra-se disponível gratuitamente em <http://www.r-project.org>. Para mais detalhes, ver Ihaka e Gentleman (1996), Cribari-Neto e Zarkos (1999) e Venables *et al.* (2012).

Esta dissertação foi digitada utilizando-se o sistema tipográfico L^AT_EX (Lamport, 1994), que consiste em uma série de macros e rotinas baseadas no sistema T_EX, criado por Donald Knuth, que facilitam o desenvolvimento e a edição de textos científicos.

Modelos de Calibração Linear e Quadrático

O objetivo deste capítulo é discutir o problema da calibração em química analítica no contexto de dois tipos de modelos: linear e quadrático (ou polinomial de segundo grau). Com este intuito, apresenta-se a lei de Beer-Lambert e descreve-se o problema da calibração em química analítica. A partir daí, discute-se os modelos mencionados tomando como base os tipos de variáveis resposta coerentes com a lei de Beer-Lambert.

2.1 Espectroscopia em Química Analítica e a Lei de Beer-Lambert

Segundo Meier e Zünd (2000), define-se *espectroscopia* como o estudo das propriedades da matéria através de sua relação com diferentes frequências do espectro eletromagnético. Alternativamente, *espectroscopia* também denomina o conjunto de métodos utilizados para medir o grau de concentração de uma determinada substância (analito) em um solvente, através de métricas que quantificam a radiação transmitida (ou absorvida) por uma amostra para um dada intensidade de energia (radiação) incidente sobre a mesma.

Para o caso da radiação luminosa, ao atingir uma certa substância, a mesma está sujeita a reflexão, refração, espalhamento ou absorção, sendo que apenas uma parcela desta energia será transmitida através do material (substância + solvente) e captada pelo equipamento de medição

(espectrômetro ótico). A hipótese básica em medição espectroscópica é que entre a resposta do espectrômetro e a concentração de um analito na amostra sob análise vale a lei de Beer-Lambert (Workman Jr., 1998, p. 4).

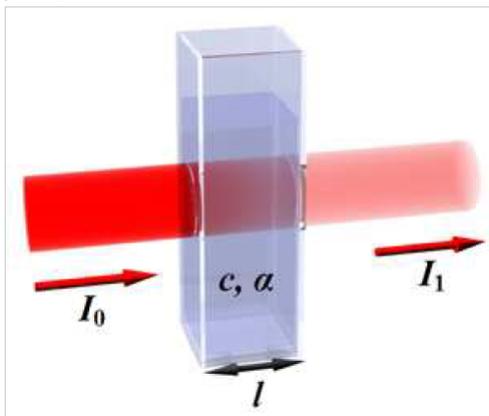
É possível se determinar a concentração de um *chromóforo* (i.e. uma substância que absorve radiação luminosa) em uma solução a partir desta lei (também conhecida como lei de Bouguer-Lambert-Beer), que estabelece uma relação entre as intensidades da radiação incidente (I_0) e emergente (I_1) e a concentração de uma dada substância nesta solução.

Isso posto, define-se a razão entre estas intensidades como *transmitância* (T), ou seja,

$$T = \frac{I_1}{I_0}, \quad T \in (0,1), \quad (2.1)$$

em que I_0 é a intensidade da luz incidente e I_1 a intensidade da luz emergente da solução (Figura 2.1).

Figura 2.1: Intensidades I_0 e I_1 de um raio de luz através de um recipiente com comprimento l



A partir de (2.1), pode-se escrever a lei de Beer-Lambert em seu formato padrão, ou seja

$$A = \alpha l = \epsilon c \quad (2.2)$$

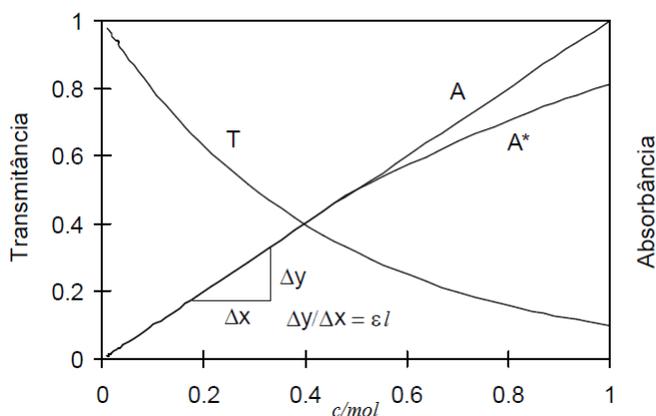
$$A = -\log_{10} T, \quad (2.3)$$

onde ϵ é o coeficiente de absorção molar (em $L \cdot mol^{-1} \cdot cm^{-1}$), l o comprimento interno (em cm) do

recipiente onde está a solução, c a concentração do material sob análise (em mol/L), α o coeficiente de absorção da substância (com $\alpha = \epsilon c$), A a absorvância (ou absorbância) e T a transmitância (L denota litro e mol representa molar).

Das equações (2.2) e (2.3) é fácil ver que a relação entre a concentração (c) e A é linear, enquanto entre c e T é não-linear, como mostrado na Figura 2.2. A curva A^* representa um desvio da lei de Beer-Lambert, em razão da elevação do índice de refração da luz e conseqüente redução da intensidade I_1 detectada pelo espectômetro, fato que eventualmente ocorre em concentrações mais elevadas do cromóforo em uma solução.

Figura 2.2: Transmitância e absorvância para diferentes concentrações de um cromóforo



As condições para validade da lei de Beer-Lambert são: i) os efeitos da reflexão, refração e espalhamento são desprezíveis; e ii) a radiação incidente deve ser monocromática (i.e. possuir um único comprimento de onda). É importante notar que, muitas vezes, I_1 decai exponencialmente com o comprimento l e com a concentração c , o que naturalmente torna T uma grandeza decrescente e não-linear.

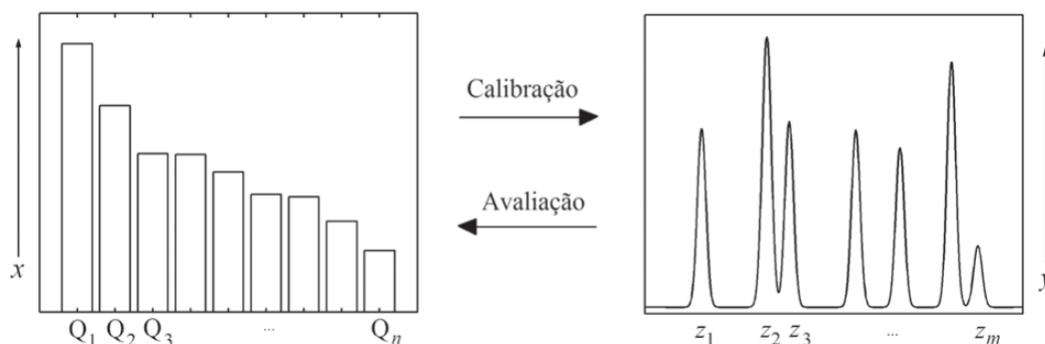
2.2 O Problema de Calibração

Segundo Danzer, Otto e Currie (2004), a calibração em química analítica refere-se à relação entre o domínio das amostras e o domínio das medidas (domínio dos sinais). Este domínios são expressos, respectivamente, através de uma função analítica $x = f(Q)$, onde Q são os analitos e x suas quantidades ou concentração em uma dada amostra-teste (Figura 2.3, lado esquerdo), e uma

função de medida $y = g(z)$ (onde z são os sinais e y a leitura dos sinais (absorvância)), que pode ser representada por um espectro, cromatograma, etc. (Figura 2.3, lado direito).

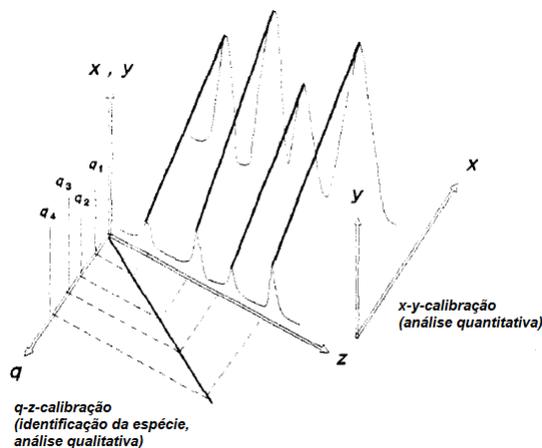
No processo de medição química, as quatro quantidades mencionadas (Q , x , z e y) devem estar relacionadas entre si, situação mostrada na Figura 2.4. As relações entre Q e x , e entre z e y , quando tomadas em conjunto estabelecem a composição da amostra sob análise.

Figura 2.3: Relação entre o domínio das amostras e o domínio dos sinais no caso de análise elementar



Fonte: Pimentel, Galvão e Araújo (2008)

Figura 2.4: Representação da relação entre calibração qualitativa e quantitativa



Fonte: Danzer e Currie (1998)

Nesse sentido, Danzer e Currie (1998) definem calibração em química analítica como a operação que determina a relação funcional entre valores medidos (intensidades do sinal y para certas posições

do sinal z) e as quantidades analíticas que caracterizam certos tipos de analitos Q e sua concentração x . A calibração também inclui a especificação do modelo (sua forma funcional), a estimação de seus parâmetros e erros, e sua validação por meio de experimentos de validação (Danzer e Currie, 1998).

É possível estabelecer uma relação entre a concentração x e o sinal (ou medida) y , que devido a erros de medição (ζ) tem a forma

$$y = h(x) + \zeta, \quad (2.4)$$

com $h(x)$ conhecida como função ou curva de calibração no sentido mais estrito.

Naturalmente, se o contexto da calibração é o mais favorável possível, têm-se que $h(\cdot)$ é uma função linear. Além disso, para o caso mais simples, com $x = [x_j] = [x_1]$, e onde outros fatores ou componentes não são considerados, y é um escalar. Neste último caso, tem-se um modelo univariado, que é aplicado a situações onde o interesse é medir a concentração de uma única substância química ($j = 1$). No entanto, é possível, sem maiores dificuldades, generalizar estas relações para o caso de multianalitos ($j = m$ substâncias químicas) (Danzer *et al.*, 2004; Blas, 2005). Neste trabalho, o interesse restringe-se ao caso univariado ou unidimensional.

Em problemas de calibração, para se estimar a concentração de uma substância (analito) presente em uma amostra é necessário encontrar (ou estimar) a função $h(\cdot)$. Este processo em geral é composto por duas etapas (ou estágios), onde a primeira corresponde a etapa de calibração propriamente dita:

- **Primeiro estágio:** são pré-fixadas n quantidades do analito x_i , $i = 1, 2, \dots, n$, (soluções-padrão) e, para cada uma, realizam-se medidas com um espectromômetro, que fornece as respostas y_i em termos de intensidade (ou absorvância).
- **Segundo estágio:** preparam-se k soluções-amostra com o mesmo analito, mas com a mesma concentração desconhecida (x_0) e obtêm-se as respostas y_{0i} na forma de absorvância, $i = n + 1, n + 2, \dots, n + k$.

A partir dos dois estágios do modelo, estima-se conjuntamente a função (2.4) e x_0 .

2.3 Modelo de Calibração Usual

Para o caso de respostas y_i na forma de absorvância, pode-se considerar que a relação entre esta e a concentração do analito (x) é linear, o que nos leva ao modelo de calibração usual (Shukla, 1972; Blas, Sandoval e Yoshida, 2007), ou seja,

$$y_i = \alpha_U + \beta_U x_i + \zeta_i, \quad i = 1, 2, \dots, n, \quad (2.5)$$

$$y_{i0} = \alpha_U + \beta_U x_0 + \zeta_i, \quad i = n + 1, n + 2, \dots, n + k, \quad (2.6)$$

onde a equação (2.5) representa o primeiro estágio enquanto (2.6) corresponde ao segundo estágio. O subscrito U denota os parâmetros associados a este modelo.

Para este modelo, valem as seguintes suposições: os x_i 's são fixos; $\zeta_i \sim N(0, \sigma_\zeta^2)$; e os erros ζ_i são não correlacionados ($i = 1, 2, \dots, n + k$). Os parâmetros são α_U , β_U , x_0 e σ_ζ^2 , sendo que a estimativa de maior interesse é a do parâmetro x_0 , com sua variância também sendo relevante.

Pode-se encontrar dois tipos estimadores para x_0 na literatura quando se considera o modelo usual (ver por exemplo, Shukla (1972)): o *estimador clássico*, obtido pela aplicação conjunta às equações (2.5) e (2.6) do método da máxima verossimilhança ou dos mínimos quadrados; e o *estimador inverso*, que é obtido a partir da regressão (por mínimos quadrados) de x sobre y , apesar de x não ser uma variável aleatória. Como mencionado no Capítulo 1, este último estimador não será utilizado neste trabalho.

A partir da função de densidade conjunta para o modelo (2.5)–(2.6),

$$\begin{aligned} f(y_i, y_{i0} \mid \alpha_U, \beta_U, x_0, x_i) &= \frac{1}{\sqrt{2\pi\sigma_\zeta^2}} \exp \left[-\frac{(y_i - \alpha_U - \beta_U x_i)^2}{2\sigma_\zeta^2} \right] \\ &\times \frac{1}{\sqrt{2\pi\sigma_\zeta^2}} \exp \left[-\frac{(y_{i0} - \alpha_U - \beta_U x_0)^2}{2\sigma_\zeta^2} \right], \end{aligned} \quad (2.7)$$

obtêm-se, pelo método de máxima verossimilhança, os estimadores para α_U , β_U e x_0 , respecti-

vamente, dados por

$$\hat{\alpha}_U = \bar{y} - \hat{\beta}_U \bar{x}, \quad (2.8)$$

$$\hat{\beta}_U = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{\beta}_U \neq 0, \quad (2.9)$$

$$\hat{x}_{0U} = \frac{\bar{y}_0 - \hat{\alpha}_U}{\hat{\beta}_U}, \quad (2.10)$$

com $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$ e $\bar{y}_0 = \frac{\sum_{i=n+1}^{n+k} y_{i0}}{k}$.

Os estimadores (2.8), (2.9) e (2.10) também podem ser obtidos por mínimos quadrados. Note que como $\hat{\alpha}_U$ e $\hat{\beta}_U$ dependem apenas da primeira etapa do modelo de calibração, vale a teoria para regressão linear simples e, como resultado, estes estimadores são não-viesados com variância mínima.

No entanto, o estimador de máxima verossimilhança (EMV) para σ_ζ^2 é

$$\hat{\sigma}_\zeta^2 = \frac{1}{n+k} \left[\sum_{i=1}^n (y_i - \hat{\alpha}_U - \hat{\beta}_U x_i)^2 + \sum_{i=n+1}^{n+k} (y_{i0} - \bar{y}_0)^2 \right], \quad (2.11)$$

que, conforme Blas (2005), é viciado, pois

$$\mathbb{E}(\hat{\sigma}_\zeta^2) = \frac{n+k-3}{n+k} \sigma_\zeta^2.$$

Corrigindo o viés deste estimador, tem-se

$$\hat{\sigma}_{\zeta_C}^2 = \frac{n+k}{n+k-3} \hat{\sigma}_\zeta^2 = \frac{1}{n+k-3} \left[\sum_{i=1}^n (y_i - \hat{\alpha}_U - \hat{\beta}_U x_i)^2 + \sum_{i=n+1}^{n+k} (y_{i0} - \bar{y}_0)^2 \right]. \quad (2.12)$$

Note que os estimadores $\hat{\sigma}_\zeta^2$ e $\hat{\sigma}_{\zeta_C}^2$ dependem dos dois estágios do modelo de calibração desde que $k > 1$. Quando $k = 1$, ambos dependem apenas do primeiro estágio do modelo.

A partir da matriz de informação de Fisher para este modelo, obtém-se a variância de \hat{x}_{0U} (Blas, 2005)

$$V_1(\hat{x}_{0U}) = \frac{\sigma_\zeta^2}{\beta_U^2} \left[\frac{1}{k} + \frac{1}{n} + \frac{(\bar{x} - x_0)^2}{nS_{xx}} \right], \quad (2.13)$$

onde $V_1(\hat{x}_{0U})$ é a aproximação de ordem n^{-1} para a variância de \hat{x}_{0U} e $S_{xx} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$.

Um estimador para $V_1(\hat{x}_{0U})$ pode ser obtido a partir da substituição dos respectivos estimadores para os parâmetros em (2.13).

Shukla (1972), tomando k fixo e expandindo (2.10) por série de Taylor no ponto (α_U, β_U) , e ignorando os termos de ordem menor que n^{-2} , obteve aproximações para o valor esperado, a variância, o viés e o erro quadratico médio (EQM) para o estimador de MV de x_0 . No caso,

$$\mathbb{E}(\hat{x}_{0U}) = x_0 + \frac{\sigma_\zeta^2(x_0 - \bar{x})}{n\beta_U^2 S_{xx}}, \quad (2.14)$$

$$V_2(\hat{x}_{0U}) = \frac{\sigma_\zeta^2}{\beta_U^2} \left[\frac{1}{n} + \frac{1}{k} + \frac{(\bar{x} - x_0)^2}{nS_{xx}} + \frac{3\sigma_\zeta^2}{kn\beta_U^2 S_{xx}} \right], \quad (2.15)$$

$$Vies(\hat{x}_{0U}) = \frac{\sigma_\zeta^2(x_0 - \bar{x})}{n\beta_U^2 S_{xx}}, \quad (2.16)$$

$$EQM(\hat{x}_{0U}) = \frac{\sigma_\zeta^2}{\beta_U^2} \left[\frac{1}{n} + \frac{1}{k} + \frac{(\bar{x} - x_0)^2}{nS_{xx}} + \frac{3\sigma_\zeta^2}{kn\beta_U^2 S_{xx}} \right]. \quad (2.17)$$

Sobre os resultados acima, pode-se observar que: (i) embora o estimador do modelo usual para x_0 seja viesado, o mesmo é assintoticamente não viesado, pois $Vies(\hat{x}_{0U}) \rightarrow 0$ quando $n \rightarrow \infty$; (ii) para x_0 muito próximo de \bar{x} pode-se ter $Vies(\hat{x}_{0U})$ pequeno, com o mesmo desaparecendo quando $x_0 = \bar{x}$; e (iii) para $k = qn$, $q \in \mathbb{Q}^+$, desprezando-se as expressões de ordem inferior a n^{-1} , as expressões (2.13) e (2.15) coincidem. Além disso, é importante mencionar que enquanto (2.13) considera tamanhos de amostras grandes para n e k (primeiro e segundo estágios, respectivamente), (2.15) considera apenas n grande mantendo k fixo no segundo estágio.

Como, sob condições de regularidade, o estimador de máxima verossimilhança \hat{x}_{0U} é assintoticamente normal com média x_0 e variância $V(\hat{x}_{0U})$, então pode-se construir um intervalo de confiança para x_0 a partir do seguinte resultado assintótico $\frac{\hat{x}_{0U} - x_0}{\sqrt{V(\hat{x}_{0U})}} \xrightarrow{\mathcal{D}} N(0, 1)$, onde $V(\hat{x}_{0U})$ é a variância

dada por (2.13) ou (2.15). Então, um intervalo de confiança aproximado para x_{0U} , com nível de confiança igual a $\gamma = (1 - \alpha)$, é

$$\left(\hat{x}_{0U} - z_{\frac{\alpha}{2}} \sqrt{V(\hat{x}_{0U})}, \hat{x}_{0U} + z_{\frac{\alpha}{2}} \sqrt{V(\hat{x}_{0U})} \right), \quad (2.18)$$

onde $z_{\frac{\alpha}{2}}$ é o quantil de ordem $(1 - \frac{\alpha}{2})$ da distribuição normal padrão e $\widehat{V}(\hat{x}_{0U})$ é a estimativa da variância de \hat{x}_{0U} .

2.4 Modelo de Calibração Quadrático

Embora o modelo usual descrito na seção anterior seja o mais utilizado na prática, nem sempre este tipo de ajuste é o mais adequado. Por exemplo, quando a relação entre a variável resposta y e x , a concentração, é quadrática, a formulação mais apropriada para a função ou curva de calibração (2.4) é um ajuste polinomial de segundo grau (Oliveira e Aguiar, 2009), isto é,

$$y_i = \alpha_Q + \beta_Q x_i + \delta_Q x_i^2 + \xi_i, \quad i = 1, 2, \dots, n, \quad (2.19)$$

$$y_{i0} = \alpha_Q + \beta_Q x_0 + \delta_Q x_0^2 + \xi_i, \quad i = n + 1, n + 2, \dots, n + k, \quad (2.20)$$

com a equação (2.19) representando o primeiro estágio e (2.20) o segundo estágio. O subscrito Q denota os parâmetros deste modelo.

Este modelo pode ser útil quando se deseja ajustar uma relação entre y e x no caso onde a variável resposta está expressa em uma grandeza que apresenta um comportamento não-linear e monótono, como a transmitância (Figura 2.2). As hipóteses para este caso são as mesmas do modelo usual: x_i 's fixos; $\xi_i \sim N(0, \sigma_\xi^2)$; e erros ξ_i não correlacionados ($i = 1, 2, \dots, n + k$). Os parâmetros são α_Q , β_Q , δ_Q , x_0 e σ_ξ^2 , com a estimativa para x_0 como a que apresenta maior interesse nesse caso, e o estimador para sua variância ou incerteza padrão combinada também sendo relevante.

Os estimadores para α_Q , β_Q , e δ_Q podem ser obtidos por máxima verossimilhança ou por mínimos quadrados a partir do primeiro estágio, o que resulta em

$$\hat{\alpha}_Q = \bar{y} - \hat{\beta}_Q \bar{x} - \hat{\delta}_Q \bar{x}^2, \quad (2.21)$$

$$\hat{\beta}_Q = \begin{bmatrix} \hat{\beta}_Q \\ \hat{\delta}_Q \end{bmatrix} = (X_1^T X_1)^{-1} X_1^T Y, \quad (2.22)$$

com $X_1 = [x_i, x_i^2]$ e $Y = [y_i]$, $i = 1, 2, \dots, n$.

No caso de x_0 , seu estimador é obtido a partir do segundo estágio do modelo de calibração, ou seja,

$$\hat{x}_{0Q} = \frac{-\hat{\beta}_Q}{2\hat{\delta}_Q} \pm \sqrt{\left(\frac{\hat{\beta}_Q}{2\hat{\delta}_Q}\right)^2 - \frac{\hat{\alpha}_Q - \bar{y}_0}{\hat{\delta}_Q}}, \quad (2.23)$$

com a raiz sendo positiva (negativa) quando a função y_i é crescente (decrecente).

Assim como no caso do modelo usual, os estimadores para α_Q , β_Q , e δ_Q dependem apenas do primeiro estágio, o que garante que os mesmos sejam não-viesados e possuam variância mínima. No caso do estimador de σ_ξ^2 , diferentemente de Oliveira e Aguiar (2009), para obtê-lo considerou-se um procedimento análogo ao utilizado no modelo usual. Desta forma, o estimador de MV de σ_ξ^2 corrigido para viés é dado por

$$\hat{\sigma}_{\xi c}^2 = \frac{n+k}{n+k-4} \hat{\sigma}_\xi^2 = \frac{1}{n+k-4} \left[\sum_{i=1}^n (y_i - \hat{\alpha}_Q - \hat{\beta}_Q x_i - \hat{\delta}_Q x_i^2)^2 + \sum_{i=n+1}^{n+k} (y_{i0} - \bar{y}_0)^2 \right]. \quad (2.24)$$

Pela teoria da regressão múltipla, a matriz de variâncias e covariâncias para os estimadores dos parâmetros α_Q , β_Q e δ_Q é dada por (Rencher e Schaalje, 2008)

$$V(\hat{\mathbf{B}}_Q) = \sigma_\xi^2 (X^T X)^{-1}, \quad (2.25)$$

com $\hat{\mathbf{B}}_Q = (\hat{\alpha}_Q, \hat{\beta}_Q, \hat{\delta}_Q)$ e $X = [\mathbf{1}, X_1]$. Para se obter um estimador para esta matriz, basta substituir (2.24) em (2.25).

A variância (ou incerteza padrão combinada) de \hat{x}_{0Q} pode ser obtida a partir da expansão por série de Taylor da equação (2.23) em torno do ponto $\mathbf{A}_Q = (\alpha_Q, \beta_Q, \delta_Q, \bar{y}_0^*)$, onde $\bar{y}_0^* = \mathbb{E}(\bar{y}_0)$. Para ver isso, considere $\hat{x}_{0Q} = f(\hat{\alpha}_Q, \hat{\beta}_Q, \hat{\delta}_Q, \bar{y}_0)$. Então, tomando-se a expansão de primeira ordem por série de Taylor, tem-se (Leite e Singer, 1990)

$$\begin{aligned} f(\hat{\alpha}_Q, \hat{\beta}_Q, \hat{\delta}_Q, \bar{y}_0) &= f(\alpha_Q, \beta_Q, \delta_Q, \bar{y}_0^*) + \frac{\partial \hat{x}_{0Q}}{\partial \bar{y}_0} (\bar{y}_0 - \bar{y}_0^*) + \frac{\partial \hat{x}_{0Q}}{\partial \hat{\alpha}_Q} (\hat{\alpha}_Q - \alpha_Q) \\ &+ \frac{\partial \hat{x}_{0Q}}{\partial \hat{\beta}_Q} (\hat{\beta}_Q - \beta_Q) + \frac{\partial \hat{x}_{0Q}}{\partial \hat{\delta}_Q} (\hat{\delta}_Q - \delta_Q) + o_P(\|\hat{\mathbf{A}}_Q - \mathbf{A}_Q\|), \\ &\text{quando } \hat{\mathbf{A}}_Q \xrightarrow{P} \mathbf{A}_Q, \end{aligned} \quad (2.26)$$

onde $\hat{\mathbf{A}}_Q = (\hat{\alpha}_Q, \hat{\beta}_Q, \hat{\delta}_Q, \bar{y}_0)$ e o termo $o_P(\|\hat{\mathbf{A}}_Q - \mathbf{A}_Q\|)$ como o resto na fórmula de Taylor.

Note que $\mathbb{E}(\hat{\mathbf{A}}_Q) = \mathbf{A}_Q$, pois $\mathbb{E}(\hat{\alpha}_Q) = \alpha_Q$, $\mathbb{E}(\hat{\beta}_Q) = \beta_Q$, $\mathbb{E}(\hat{\delta}_Q) = \delta_Q$ e $\mathbb{E}(\bar{y}_0) = \bar{y}_0^*$, como mencionado anteriormente. Além disso, considere agora que $V(\bar{y}_0) = V(\bar{y}) = \frac{1}{n} \sigma_\xi^2$. Então, desprezando-se o resto de Taylor em (2.26), manipulando a expressão resultante e aplicando-se valor esperado chega-se a

$$\begin{aligned} V(\hat{x}_{0Q}) &= \left(\frac{\partial \hat{x}_{0Q}}{\partial \bar{y}_0} \right)^2 V(\bar{y}_0) + \left(\frac{\partial \hat{x}_{0Q}}{\partial \hat{\alpha}_Q} \right)^2 V(\hat{\alpha}_Q) + \left(\frac{\partial \hat{x}_{0Q}}{\partial \hat{\beta}_Q} \right)^2 V(\hat{\beta}_Q) \\ &+ \left(\frac{\partial \hat{x}_{0Q}}{\partial \hat{\delta}_Q} \right)^2 V(\hat{\delta}_Q) + 2 \left(\frac{\partial \hat{x}_{0Q}}{\partial \hat{\alpha}_Q} \frac{\partial \hat{x}_{0Q}}{\partial \hat{\beta}_Q} \right) Cov(\hat{\alpha}_Q, \hat{\beta}_Q) \\ &+ 2 \left(\frac{\partial \hat{x}_{0Q}}{\partial \hat{\alpha}_Q} \frac{\partial \hat{x}_{0Q}}{\partial \hat{\delta}_Q} \right) Cov(\hat{\alpha}_Q, \hat{\delta}_Q) + 2 \left(\frac{\partial \hat{x}_{0Q}}{\partial \hat{\beta}_Q} \frac{\partial \hat{x}_{0Q}}{\partial \hat{\delta}_Q} \right) Cov(\hat{\beta}_Q, \hat{\delta}_Q), \end{aligned} \quad (2.27)$$

onde as variâncias e covariâncias desta expressão, à exceção de $V(\bar{y}_0)$, são os elementos de (2.25).

O estimador para esta variância é obtido pela substituição de σ_ξ^2 por (2.24) quando necessário.

Os termos entre parênteses em (2.27) correspondem às derivadas parciais de (2.23) em relação a cada um dos parâmetros desta equação. Por exemplo, para o caso de y_i decrescente, tem-se

$$\frac{\partial \hat{x}_{0Q}}{\partial \bar{y}_0} = -\frac{1}{\sqrt{\hat{\beta}_Q^2 - 4\hat{\delta}_Q(\hat{\alpha}_Q - \bar{y}_0)}}, \quad (2.28)$$

$$\frac{\partial \hat{x}_{0Q}}{\partial \hat{\alpha}_Q} = \frac{1}{\sqrt{\hat{\beta}_Q^2 - 4\hat{\delta}_Q(\hat{\alpha}_Q - \bar{y}_0)}}, \quad (2.29)$$

$$\frac{\partial \hat{x}_{0Q}}{\partial \hat{\beta}_Q} = \frac{-1 - \frac{\hat{\beta}_Q}{\sqrt{\hat{\beta}_Q^2 - 4\hat{\delta}_Q(\hat{\alpha}_Q - \bar{y}_0)}}}{2\hat{\delta}_Q}, \quad (2.30)$$

$$\begin{aligned} \frac{\partial \hat{x}_{0Q}}{\partial \hat{\delta}_Q} &= \frac{\hat{\beta}_Q}{\sqrt{\hat{\beta}_Q^2 - 4\hat{\delta}_Q(\hat{\alpha}_Q - \bar{y}_0)}} \left(\frac{\hat{\beta}_Q + \sqrt{\hat{\beta}_Q^2 - 4\hat{\delta}_Q(\hat{\alpha}_Q - \bar{y}_0)}}{2\hat{\delta}_Q^2} \right) \\ &\quad - \frac{(\hat{\alpha}_Q - \bar{y}_0)}{\hat{\delta}_Q} \frac{1}{\sqrt{\hat{\beta}_Q^2 - 4\hat{\delta}_Q(\hat{\alpha}_Q - \bar{y}_0)}}, \end{aligned} \quad (2.31)$$

com a expressão (2.31) corrigida para um erro tipográfico em Oliveira e Aguiar (2009, p. 1572).

Note que quando y_i é crescente, as expressões (2.28)–(2.31) são alteradas apenas pela mudança no sinal de $\sqrt{\hat{\beta}_Q^2 - 4\hat{\delta}_Q(\hat{\alpha}_Q - \bar{y}_0)}$ nas três primeiras derivadas e no sinal de $\hat{\beta}_Q$ no termo entre parênteses em (2.31).

Para se construir um intervalo de confiança aproximado para x_0 a partir do modelo quadrático, utiliza-se o mesmo procedimento apresentado para o caso do modelo usual.

Modelo de Calibração Beta

Neste capítulo propõe-se um modelo de calibração para uma variável resposta restrita ao intervalo unitário $(0, 1)$. A suposição básica é que esta variável segue uma distribuição beta, que é reparametrizada de forma a permitir a obtenção de uma estrutura de regressão para a média da mesma, de forma similar a Ferrari e Cribari-Neto (2004). Isto posto, discute-se a construção de um modelo de calibração para o qual a variável resposta do primeiro e segundo estágio segue a distribuição beta. Além disso, obtêm-se as funções escores relevantes para este modelo, discute-se a estimação de seus parâmetros pelo método da máxima verossimilhança e deriva-se a matriz de informação de Fisher esperada.

3.1 A Distribuição Beta

Uma variável aleatória y segue a distribuição beta, $y \sim \mathcal{B}(p, q)$, com $p > 0$ e $q > 0$, quando sua função densidade de probabilidade (f.d.p.) é dada por

$$f(y | p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad y \in (0, 1), \quad (3.1)$$

com $\Gamma(\cdot)$ representando a função gama, i.e., $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$.

Para a f.d.p. (3.1), o r -ésimo momento não-central é

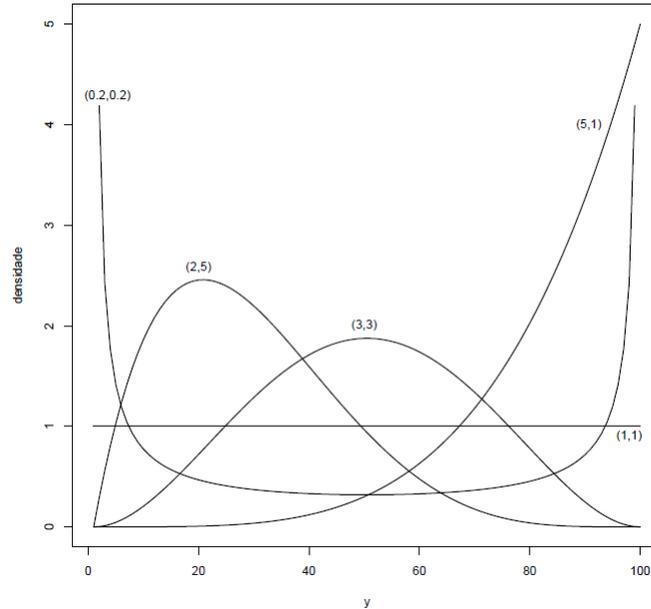
$$\mu_{(r)} = \mathbb{E}(y^r) = \int_0^1 y^r \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1} = \frac{\Gamma(p+r)\Gamma(p+q)}{\Gamma(p)\Gamma(p+q+r)} = \frac{p_{(r)}}{(p+q)_{(r)}}, \quad (3.2)$$

com $k_{(r)} = k \times (k+1) \times (k+2) \times \dots \times (k+r-1)$, pois $\Gamma(x) = (x-1)!$ para $x \in \mathbb{Z}^+$.

A partir de (3.2), obtêm-se a média e a variância de y , $\mathbb{E}(y) = \frac{p}{p+q}$ e $V(y) = \frac{pq}{(p+q)^2(p+q+1)}$, respectivamente. Note que por meio da escolha de diferentes valores para os parâmetros p e q da densidade (3.1) é possível obter diferentes f.d.p. para $y \in (0, 1)$.

A Figura 3.1 apresenta alguns exemplos de densidades beta para diferentes valores dos parâmetros (p, q) . Nestes exemplos, pode-se observar que quando $p = q$ a densidade é simétrica, tornando-se assimétrica quando $p \neq q$. Com $p > q$ tem-se assimetria à esquerda, e para $p < q$ tem-se assimetria à direita. Em particular, quando $p = q = 1$, tem-se a distribuição uniforme padrão. Isto demonstra a grande flexibilidade desta distribuição para a modelagem das mais diversas aplicações práticas.

Figura 3.1: Densidades beta para diferentes valores de (p, q)



3.2 Modelo de Regressão Beta

Como visto na seção anterior, a distribuição beta é conhecida em razão de sua flexibilidade para a modelagem de várias situações práticas. Neste sentido, Ferrari e Cribari-Neto (2004) propuseram um modelo de regressão beta para modelar o comportamento de variáveis contínuas restritas ao intervalo unitário como função de um conjunto de variáveis exógenas. Neste caso, como a variável dependente está restrita a $(0, 1)$, a mesma pode ser interpretada como uma razão ou proporção. Note que o modelo proposto pelos autores é semelhante à classe de modelos lineares generalizados discutida em McCullagh e Nelder (1989), exceto pelo fato que a distribuição da variável resposta não pertence à família exponencial.

Para definir o modelo, Ferrari e Cribari-Neto (2004) utilizaram uma reparametrização para a distribuição beta que permitiu a obtenção de uma estrutura de regressão para a média da variável resposta e um parâmetro de precisão. Sejam $\mu = \frac{p}{p+q}$ e $\phi = p+q$, o que leva a $p = \mu\phi$ e $q = (1-\mu)\phi$.

Logo, a partir desta reparametrização, a densidade da variável resposta (y), dada por (3.1), torna-se

$$f(y | \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad y \in (0, 1), \quad \mu \in (0, 1), \quad \phi > 0, \quad (3.3)$$

com a média e a variância da variável resposta (y), definidas na Seção 3.1, tornando-se $\mathbb{E}(y) = \mu$ e $V(y) = \frac{\mu(1-\mu)}{1+\phi}$, respectivamente. Note que ϕ pode ser interpretado como um parâmetro de precisão, de forma que, para μ fixo, quanto maior ϕ , menor a variância de y .

Suponha agora que $y_i \sim \mathcal{B}(\mu_i, \phi)$, com $i = 1, 2, \dots, n$, y_i 's independentes, e ϕ desconhecido e constante $\forall i$. Adicionalmente, assuma que a média de y_i pode ser escrita como

$$g(\mu_i) = \eta_i = \sum_{j=1}^r x_{ij} \beta_j, \quad (3.4)$$

onde $\beta = (\beta_1, \dots, \beta_r)^T$ é o vetor de parâmetros desconhecidos do modelo de regressão ($\beta \in \mathbb{R}^r$) e $\mathbf{x}_i = (x_{i1}, \dots, x_{ir})$ é o vetor de observações de r variáveis explicativas ($r < n$), que possuem valores fixos e conhecidos.

A função $g(\cdot)$ é uma função de ligação estritamente monótona, duas vezes continuamente diferenciável, que mapeia o intervalo $(0, 1)$ em \mathbb{R} . Além disso, como a variância de y_i é função de μ_i , que por sua vez depende dos valores de x_i , a mesma não necessariamente é constante, o que indica que o modelo é naturalmente heterocedástico.

Como mostrado por vários autores, há diferentes escolhas possíveis para a função de ligação $g(\cdot)$ (Ferrari e Cribari-Neto, 2004; Ospina, Cribari-Neto e Vasconcellos, 2006; Simas, Barreto-Souza e Rocha, 2010). Neste trabalho, optou-se por discutir os modelos de regressão e calibração beta no contexto das funções de ligação logit e log-log, cujas inversas são, respectivamente,

$$\mu_i = \frac{\exp(x_i\beta)}{1 + \exp(x_i\beta)} \quad (3.5)$$

$$\mu_i = \frac{1}{\exp(1/\exp(x_i\beta))}. \quad (3.6)$$

Note que a função logit permite interpretar os parâmetros β_j como razões de probabilidades. Para ver isso, basta mostrar que o impacto sobre μ_i da mudança em um dos x'_j s, mantendo-se os demais constantes, é igual a $e^{\Delta x_j \beta_j}$ (Ferrari e Cribari-Neto, 2004).

Para estimar os parâmetros do modelo de regressão beta, utiliza-se o método da máxima verossimilhança, com a função de verossimilhança dada por

$$L(\mathbf{y}, \mu, \phi) = \prod_{i=1}^n f(y_i | \mu_i, \phi) = \left(\frac{\Gamma(\phi)}{\Gamma(\mu_i\phi)\Gamma((1-\mu_i)\phi)} \right)^n \prod_{i=1}^n y_i^{\mu_i\phi-1} (1-y_i)^{(1-\mu_i)\phi-1},$$

onde $\mathbf{y} = (y_1, \dots, y_n)$, $\mu = (\mu_1, \dots, \mu_n)$ e $f(\cdot)$ é a f.d.p. para a variável resposta y_i .

Logo, a log-verossimilhança é dada por

$$\begin{aligned} \ell(\beta, \phi) = \log L(\cdot) &= \sum_{i=1}^n \ell_i(\mu_i, \phi) = \sum_{i=1}^n [\log(\Gamma(\phi)) - \log(\Gamma(\phi\mu_i)) - \log(\Gamma((1-\mu_i)\phi))] \\ &+ (\mu_i\phi - 1) \log y_i + ((1-\mu_i)\phi - 1) \log(1-y_i) \end{aligned} \quad (3.7)$$

Os componentes do vetor escore, obtidos a partir das primeiras derivadas de (3.7) com relação

a cada um dos parâmetros β_j , $j = 1, \dots, r$, e ϕ , são dados, respectivamente, por

$$U_{\beta}(\beta, \phi) = \frac{\partial \ell(\cdot)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i(\mu_i, \phi)}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \phi \sum_{i=1}^n (y_i^* - \mu_i^*) \frac{1}{g'(\mu_i)} \frac{\partial \eta_i}{\partial \beta_j}, \quad (3.8)$$

$$U_{\phi}(\beta, \phi) = \frac{\partial \ell(\cdot)}{\partial \phi} = \sum_{i=1}^n [\mu_i(y_i^* - \mu_i^*) + \log(1 - y_i) - \psi((1 - \mu_i)\phi) + \psi(\phi)], \quad (3.9)$$

onde $y_i^* = \log\left(\frac{y_i}{1-y_i}\right)$, $\mu_i^* = \psi(\mu_i\phi) - \psi((1 - \mu_i)\phi)$, com $\psi(x) = \frac{d \log(\Gamma(x))}{dx}$ e $x > 0$.

Os estimadores de MV para este modelo são obtidos a partir da solução das equações $U_{\beta}(\beta, \phi) = \mathbf{0}$ e $U_{\phi}(\beta, \phi) = 0$, onde $\mathbf{0}$ é um vetor nulo $r \times 1$. Estes estimadores não possuem forma fechada, logo, é necessária a utilização de métodos numéricos para maximizar a função (3.7), como, por exemplo, algoritmos de otimização não-lineares Newton ou quasi-Newton (Nocedal e Wright, 2006).

A matriz de informação de Fisher para este modelo é obtida a partir das derivadas de segunda ordem de (3.7) com relação a cada um dos parâmetros do modelo. Então, para $w_i = \phi[\psi'(\mu_i\phi) + \psi'((1 - \mu_i)\phi)] \frac{1}{[g'(\mu_i)]^2}$, $c_i = \phi[\psi'(\mu_i\phi)\mu_i - \psi'((1 - \mu_i)\phi)(1 - \mu_i)]$ e $d_i = -\psi'(\phi) + \mu_i^2\psi'(\phi\mu_i) + (1 - \mu_i)^2\psi'((1 - \mu_i)\phi)$, com $\psi'(x) = \frac{\partial^2 \psi(x)}{\partial x^2}$, tem-se $W = \text{diag}\{w_1, w_2, \dots, w_n\}$, $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$, $C = (c_1, \dots, c_n)$ e $D = \text{diag}\{d_1, \dots, d_n\}$.

Logo, tomando-se o valor esperado das segundas derivadas em (3.7), sob certas condições de regularidade, e após alguma manipulação algébrica, tem-se a matriz de informação de Fisher

$$K = \begin{bmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{bmatrix} = \begin{bmatrix} \phi X^T W X & X^T T C \\ C^T T^T X & \text{tr}(D) \end{bmatrix} \quad (3.10)$$

onde X é uma matriz $n \times r$, cuja i -ésima linha é \mathbf{x}_i .

Ferrari e Cribari-Neto (2004) ressaltam que, ao contrário do caso dos modelos lineares generalizados (McCullagh e Nelder, 1989), os parâmetros β e ϕ não são ortogonais, pois $K_{\beta\phi} = K_{\phi\beta} = X^T T C \neq \mathbf{0}$. A matriz de variância assintótica dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta é dada pela inversa de K (i.e. K^{-1}).

Maiores detalhes sobre o processo de estimação dos parâmetros deste modelo, bem como sobre a derivação da matriz de informação de Fisher (3.10), podem ser encontrados em Ferrari e Cribari-

Neto (2004). Na Seção 3.3 deste capítulo, será utilizado um procedimento análogo ao destes autores para se obter a matriz de Fisher para um modelo de calibração cuja variável resposta segue a distribuição beta.

3.3 Modelo de Calibração Beta

No Capítulo 2 deste trabalho apresentou-se dois tipos de modelos para o problema de calibração em química analítica: usual e quadrático. Como visto, estes modelos consideram que a relação entre a variável resposta y e a concentração do analito (x) é sempre linear, mesmo para casos nos quais y apresenta um comportamento não-linear (i.e. geralmente quando y é a transmitância).

Como mostrado nas Seções 2.1 e 2.2, nem sempre a relação entre y e x será linear, o que naturalmente sugere o uso de modelos não-lineares para tratar o problema de calibração. Neste sentido, o modelo proposto nesta seção assume que a variável resposta segue uma distribuição beta (i.e. $y \sim \mathcal{B}(p, q)$, com $p > 0$, $q > 0$), que, quando reparametrizada nos moldes da discussão da Seção 3.2, permite propor uma abordagem para o problema em questão no contexto do modelo de regressão beta, onde a relação entre y e x é claramente não-linear.

A abordagem proposta é adequada para casos nos quais a variável resposta está restrita ao intervalo $(0, 1)$, a exemplo da transmitância (T), grandeza relacionada a absorvância (A) através da transformação logarítmica presente na lei de Beer-Lambert (equação (2.3)). Além disso, o ajuste sugerido, como é o caso para o modelo de regressão beta, evita a ocorrência de valores ajustados para y fora do intervalo de interesse para esta variável, como poderia ocorrer para o modelo quadrático.

3.3.1 Estrutura do Modelo de Calibração Beta

Para se obter uma estrutura para modelagem da variável resposta (y) adequada ao contexto da calibração, supondo que esta variável é função de uma única variável independente x (fixa e conhecida) no primeiro estágio (e y_0 função de um parâmetro x_0 desconhecido no segundo estágio), modificou-se a parametrização utilizada por Ferrari e Cribari-Neto (2004), de forma a permitir a identificabilidade do modelo proposto. Para o primeiro estágio de calibração tem-se que $y \sim \mathcal{B}(p, q)$, com $p > 0$, $q > 0$, enquanto para o segundo estágio, $y_0 \sim \mathcal{B}(p_0, q_0)$, com $p_0 > 0$, $q_0 > 0$, e/ou $p \neq p_0$,

$q \neq q_0$.

Neste caso, a reparametrização das densidades beta é feita para as duas etapas do modelo, isto é, para o primeiro estágio têm-se $\mu = \frac{p}{p+q}$, $\phi = p + q$, e para o segundo estágio, $\mu_0 = \frac{p_0}{p_0+q_0}$ e $\phi_0 = p_0 + q_0$. Note que $\mathbb{E}(y) = \mu$, $\mathbb{E}(y_0) = \mu_0$, $\mu \in (0, 1)$, $\mu_0 \in (0, 1)$, $\phi > 0$ e $\phi_0 > 0$.

Conseqüentemente as variâncias das variáveis resposta y e y_0 tornam-se $Var(y) = \frac{\mu(1-\mu)}{1+\phi}$ e $Var(y_0) = \frac{\mu_0(1-\mu_0)}{1+\phi_0}$, respectivamente, com ϕ e ϕ_0 (fixos e desconhecidos) podendo ser interpretados como parâmetros de precisão (para μ (μ_0) fixo, quanto maior o valor de ϕ (ϕ_0), menor a variância de y (y_0)).

Seja $y_i \sim \mathcal{B}(\mu_i, \phi)$, com $i = 1, 2, \dots, n$, e $y_{i0} \sim \mathcal{B}(\mu_0, \phi_0)$, para $i = n + 1, n + 2, \dots, n + k$. Adicionalmente, considere uma função de ligação $g(\cdot)$, tal que $g(\cdot) : (0, 1) \rightarrow \mathbb{R}$, com $g(\cdot)$ estritamente monótona e duas vezes continuamente diferenciável. Então, considerando-se a reparametrização acima, $g(\mu_i) = \eta_i = \alpha_B + \beta_B x_i$ para o primeiro estágio da calibração, e $g(\mu_0) = \eta_0 = \alpha_B + \beta_B x_0$ para o segundo estágio, onde o subscrito B denota o modelo de calibração beta, tem-se

$$\mu_i = \mathbb{E}(y_i) = g^{-1}(\alpha_B + \beta_B x_i), \quad (3.11)$$

$$\mu_0 = \mathbb{E}(y_{i0}) = g^{-1}(\alpha_B + \beta_B x_0), \quad (3.12)$$

com $Var(y_i) = \frac{\mu_i(1-\mu_i)}{1+\phi}$ e $Var(y_{i0}) = \frac{\mu_0(1-\mu_0)}{1+\phi_0}$. Note que a transformação acima restringe a média condicional das variáveis de interesse ao intervalo $(0, 1)$, e como a variância de y_i neste caso é uma função de μ_i (e ϕ), tem-se então um modelo naturalmente heteroscedástico. Além disso, têm-se $\mu_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)}$ e $\mu_0 = \frac{\exp(\eta_0)}{1+\exp(\eta_0)}$, quando $g(\cdot)$ é uma função de ligação logit, $\mu_i = \frac{1}{\exp(1/\exp(\eta_i))}$ e $\mu_0 = \frac{1}{\exp(1/\exp(\eta_0))}$, quando $g(\cdot)$ é uma função de ligação log-log.

Assim como no caso do modelo de regressão beta, para se estimar os parâmetros do modelo de calibração beta $(\alpha_B, \beta_B, x_0, \phi, \phi_0)$ utiliza-se o método da máxima verossimilhança. No entanto, o processo de estimação destes parâmetros é feito considerando-se simultaneamente os dois estágios do modelo de calibração através da função de verossimilhança conjunta, dada por

$$\begin{aligned}
L(\mathbf{y}, \mathbf{y}_0, \mu, \mu_0, \phi, \phi_0) &= \prod_{i=1}^n f_1(y_i | \mu_i, \phi) \prod_{i=n+1}^{n+k} f_0(y_{i0} | \mu_0, \phi_0) \\
&= \left(\frac{\Gamma(\phi)}{\Gamma(\mu_i \phi) \Gamma((1 - \mu_i) \phi)} \right)^n \prod_{i=1}^n y_i^{\mu_i \phi - 1} (1 - y_i)^{(1 - \mu_i) \phi - 1} \\
&\times \left(\frac{\Gamma(\phi_0)}{\Gamma(\mu_0 \phi_0) \Gamma((1 - \mu_0) \phi_0)} \right)^k \prod_{i=n+1}^{n+k} y_{i0}^{\mu_0 \phi_0 - 1} (1 - y_{i0})^{(1 - \mu_0) \phi_0 - 1},
\end{aligned}$$

onde $\mathbf{y} = (y_1, \dots, y_n)$, $\mathbf{y}_0 = (y_{(n+1)0}, \dots, y_{(n+k)0})$, $\mu = (\mu_1, \dots, \mu_n)$ e $f_1(\cdot)$ e $f_0(\cdot)$ são as f.d.p.s das variáveis resposta y_i e y_{i0} , respectivamente. Logo, a função de log-verossimilhança é

$$\begin{aligned}
\ell(\mathbf{y}, \mathbf{y}_0, \mu, \mu_0, \phi, \phi_0) &= \log L(\cdot) = \sum_{i=1}^n \ell_{i1}(\mu_i, \phi) + \sum_{i=n+1}^{n+k} \ell_{i0}(\mu_0, \phi_0) = \\
&= \sum_{i=1}^n [\log(\Gamma(\phi)) - \log(\Gamma(\phi \mu_i)) - \log(\Gamma((1 - \mu_i) \phi))] + \\
&+ (\mu_i \phi - 1) \log y_i + ((1 - \mu_i) \phi - 1) \log(1 - y_i) + \\
&+ \sum_{i=n+1}^{n+k} [\log(\Gamma(\phi_0)) - \log(\Gamma(\phi_0 \mu_0)) - \log(\Gamma((1 - \mu_0) \phi_0))] + \\
&+ (\mu_0 \phi_0 - 1) \log y_{i0} + ((1 - \mu_0) \phi_0 - 1) \log(1 - y_{i0}). \tag{3.13}
\end{aligned}$$

Tomando as primeiras derivadas de (3.13) em relação a cada um dos cinco parâmetros de interesse $(\alpha_B, \beta_B, x_0, \phi, \phi_0)$, com $\psi(x) = \frac{d \log(\Gamma(x))}{dx}$, $x > 0$, $\Theta = (\theta_1, \theta_2, \theta_3) = (\alpha_B, \beta_B, x_0)$ e $j = 1, 2, 3$, tem-se as funções escore, que são

$$\begin{aligned}
U_{\Theta}(\Theta, \phi, \phi_0) &= \frac{\partial \ell(\cdot)}{\partial \theta_j} = \sum_{i=1}^n \frac{\partial \ell_{i1}(\mu_i, \phi)}{\partial \mu_i} \frac{d \mu_i}{d \eta_i} \frac{\partial \eta_i}{\partial \theta_j} + \sum_{i=n+1}^{n+k} \frac{\partial \ell_{i0}(\mu_0, \phi_0)}{\partial \mu_0} \frac{d \mu_0}{d \eta_0} \frac{\partial \eta_0}{\partial \theta_j} \\
&= \phi \sum_{i=1}^n (y_i^* - \mu_i^*) \frac{1}{g'(\mu_i)} \frac{\partial \eta_i}{\partial \theta_j} + \phi_0 \sum_{i=n+1}^{n+k} (y_{i0}^* - \mu_0^*) \frac{1}{g'(\mu_0)} \frac{\partial \eta_0}{\partial \theta_j}, \tag{3.14}
\end{aligned}$$

$$U_{\phi}(\Theta, \phi, \phi_0) = \frac{\partial \ell(\cdot)}{\partial \phi} = \sum_{i=1}^n [\mu_i(y_i^* - \mu_i^*) + \log(1 - y_i) - \psi((1 - \mu_i)\phi) + \psi(\phi)], \quad (3.15)$$

$$U_{\phi_0}(\Theta, \phi, \phi_0) = \frac{\partial \ell(\cdot)}{\partial \phi_0} = \sum_{i=n+1}^{n+k} [\mu_0(y_{i0}^* - \mu_0^*) + \log(1 - y_{i0}) - \psi((1 - \mu_0)\phi_0) + \psi(\phi_0)], \quad (3.16)$$

onde $y_i^* = \log\left(\frac{y_i}{1-y_i}\right)$, $\mu_i^* = \psi(\mu_i\phi) - \psi((1-\mu_i)\phi)$, $y_{i0}^* = \log\left(\frac{y_{i0}}{1-y_{i0}}\right)$, $\mu_0^* = \psi(\mu_0\phi_0) - \psi((1-\mu_0)\phi_0)$, $\frac{d\mu_i}{d\eta_i} = \frac{1}{g'(\mu_i)}$ e $\frac{d\mu_0}{d\eta_0} = \frac{1}{g'(\mu_0)}$.

As estimativas de MV dos parâmetros α_B , β_B , x_0 , ϕ e ϕ_0 são resultado da solução das equações $U_{\Theta}(\Theta, \phi, \phi_0) = \mathbf{0}$, $U_{\phi}(\Theta, \phi, \phi_0) = 0$ e $U_{\phi_0}(\Theta, \phi, \phi_0) = 0$ onde $\mathbf{0}$ é o vetor nulo de dimensão 3×1 . Como a solução destas equações não leva a expressões analíticas, deve-se utilizar métodos numéricos para se obter as estimativas de MV para os parâmetros em questão. Neste caso, assume-se um conjunto de valores iniciais para $(\alpha_B, \beta_B, x_0, \phi, \phi_0)$, segundo alguma regra para o estabelecimento deste "chute" inicial. A estimativa de x_0 obtida desta forma será denotada por \hat{x}_{0B} .

Além disso, observe que uma vez obtidos os estimadores de MV para $(\alpha_B, \beta_B, x_0, \phi, \phi_0)$, chega-se facilmente a μ , μ_0 , $Var(y_i)$ e $Var(y_{i0})$, e a p , q , p_0 e q_0 , dado que o modelo é identificável. Note também que α_B e β_B podem ser interpretados de forma semelhante a de um modelo linear e x_0 é um parâmetro desconhecido, por exemplo, em química analítica, é a concentração (ou teor do produto na solução) a ser determinada.

3.3.2 Matriz de Informação de Fisher Esperada

Para se obter a matriz de informação de Fisher para o modelo de calibração beta, é necessário calcular as segundas derivadas de (3.13). Então, derivando-se novamente as expressões (3.14), (3.15) e (3.16) em relação a cada um dos parâmetros $(\alpha_B, \beta_B, x_0, \phi, \phi_0)$, com $\Theta = (\theta_1, \theta_2, \theta_3) = (\alpha_B, \beta_B, x_0)$, $\psi'(x) = \frac{\partial^2 \psi(x)}{\partial x^2}$ e $j, s = 1, 2, 3$, tem-se

$$\begin{aligned}
\frac{\partial^2 \ell(\cdot)}{\partial \theta_j \partial \theta_s} &= \sum_{i=1}^n \frac{\partial}{\partial \mu_i} \left(\frac{\partial \ell_{i1}(\mu_i, \phi)}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \theta_s} \frac{\partial \eta_i}{\partial \theta_j} + \sum_{i=n+1}^{n+k} \frac{\partial}{\partial \mu_0} \left(\frac{\partial \ell_{i0}(\mu_0, \phi_0)}{\partial \mu_0} \frac{d\mu_0}{d\eta_0} \right) \times \\
&\times \frac{d\mu_0}{d\eta_0} \frac{\partial \eta_0}{\partial \theta_s} \frac{\partial \eta_0}{\partial \theta_j} \\
&= \sum_{i=1}^n \left(\frac{\partial \ell_{i1}^2(\mu_i, \phi)}{\partial \mu_i^2} \frac{d\mu_i}{d\eta_i} + \frac{\partial \ell_{i1}(\mu_i, \phi)}{\partial \mu_i} \frac{\partial}{\partial \mu_i} \frac{d\mu_i}{d\eta_i} \right) \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \theta_s} \frac{\partial \eta_i}{\partial \theta_j} + \\
&+ \sum_{i=n+1}^{n+k} \left(\frac{\partial \ell_{i0}^2(\mu_0, \phi_0)}{\partial \mu_0^2} \frac{d\mu_0}{d\eta_0} + \frac{\partial \ell_{i0}(\mu_0, \phi_0)}{\partial \mu_0} \frac{\partial}{\partial \mu_0} \frac{d\mu_0}{d\eta_0} \right) \frac{d\mu_0}{d\eta_0} \frac{\partial \eta_0}{\partial \theta_s} \frac{\partial \eta_0}{\partial \theta_j}, \tag{3.17}
\end{aligned}$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \theta_j \partial \phi} = \sum_{i=1}^n \left[(y_i^* - \mu_i^*) - \phi \frac{\partial \mu_i^*}{\partial \phi} \right] \frac{1}{g'(\mu_i)} \frac{\partial \eta_i}{\partial \theta_j}, \tag{3.18}$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \theta_j \partial \phi_0} = \sum_{i=n+1}^{n+k} \left[(y_{i0}^* - \mu_0^*) - \phi_0 \frac{\partial \mu_0^*}{\partial \phi_0} \right] \frac{1}{g'(\mu_0)} \frac{\partial \eta_0}{\partial \theta_j}, \tag{3.19}$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \phi^2} = \sum_{i=1}^n [\psi'(\phi) - \mu_i^2 \psi'(\phi \mu_i) - (1 - \mu_i)^2 \psi'((1 - \mu_i)\phi)], \tag{3.20}$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \phi_0^2} = \sum_{i=n+1}^{n+k} [\psi'(\phi_0) - \mu_0^2 \psi'(\phi_0 \mu_0) - (1 - \mu_0)^2 \psi'((1 - \mu_0)\phi_0)], \tag{3.21}$$

$$\frac{\partial^2 \ell(\cdot)}{\partial \phi \partial \phi_0} = 0 = \frac{\partial^2 \ell(\cdot)}{\partial \phi_0 \partial \phi}. \tag{3.22}$$

Note que as expressões (3.17)–(3.22) são semelhantes às obtidas por Ferrari e Cribari-Neto (2004) no processo de obtenção da matriz (3.10), distinguindo-se apenas pela inclusão dos termos referentes à segunda etapa do modelo de calibração (derivadas em relação a x_0 , segundo termo da soma em (3.17), e expressões (3.19), (3.21) e (3.22)).

Então, como $\mathbb{E}\left(\frac{\partial \ell_{i1}(\mu_i, \phi)}{\partial \mu_i}\right) = 0$, $\mathbb{E}\left(\frac{\partial \ell_{i0}(\mu_0, \phi_0)}{\partial \mu_0}\right) = 0$, $\mathbb{E}(y_i^*) = \mu_i^*$ e $\mathbb{E}(y_{i0}^*) = \mu_0^*$, tomando-se o valor esperado de (3.17)–(3.22), para $j, s = 1, 2, 3$, tem-se,

$$\begin{aligned} \mathbb{E}\left(\frac{\partial^2 \ell(\cdot)}{\partial \theta_j \partial \theta_s}\right) &= \sum_{i=1}^n \mathbb{E}\left(\frac{\partial \ell_{i1}^2(\mu_i, \phi)}{\partial \mu_i^2}\right) \left(\frac{d\mu_i}{d\eta_i}\right)^2 \frac{\partial \eta_i}{\partial \theta_s} \frac{\partial \eta_i}{\partial \theta_j} + \\ &+ \sum_{i=n+1}^{n+k} \mathbb{E}\left(\frac{\partial \ell_{i0}^2(\mu_0, \phi_0)}{\partial \mu_0^2}\right) \left(\frac{d\mu_0}{d\eta_0}\right)^2 \frac{\partial \eta_0}{\partial \theta_s} \frac{\partial \eta_0}{\partial \theta_j}, \end{aligned} \quad (3.23)$$

$$\mathbb{E}\left(\frac{\partial^2 \ell(\cdot)}{\partial \theta_j \partial \phi}\right) = - \sum_{i=1}^n \phi [\psi'(\mu_i \phi) \mu_i - \psi'((1 - \mu_i) \phi) (1 - \mu_i)] \frac{1}{g'(\mu_i)} \frac{\partial \eta_i}{\partial \theta_j}, \quad (3.24)$$

$$\mathbb{E}\left(\frac{\partial^2 \ell(\cdot)}{\partial \theta_j \partial \phi_0}\right) = - \sum_{i=n+1}^{n+k} \phi_0 [\psi'(\mu_0 \phi_0) \mu_0 - \psi'((1 - \mu_0) \phi_0) (1 - \mu_0)] \frac{1}{g'(\mu_0)} \frac{\partial \eta_0}{\partial \theta_j}, \quad (3.25)$$

$$\mathbb{E}\left(\frac{\partial^2 \ell(\cdot)}{\partial \phi^2}\right) = \sum_{i=1}^n [\psi'(\phi) - \mu_i^2 \psi'(\phi \mu_i) - (1 - \mu_i)^2 \psi'((1 - \mu_i) \phi)], \quad (3.26)$$

$$\mathbb{E}\left(\frac{\partial^2 \ell(\cdot)}{\partial \phi_0^2}\right) = \sum_{i=n+1}^{n+k} [\psi'(\phi_0) - \mu_0^2 \psi'(\phi_0 \mu_0) - (1 - \mu_0)^2 \psi'((1 - \mu_0) \phi_0)], \quad (3.27)$$

$$\mathbb{E}\left(\frac{\partial^2 \ell(\cdot)}{\partial \phi \partial \phi_0}\right) = 0. \quad (3.28)$$

A partir das expressões acima, é fácil obter a matriz de informação de Fisher para o modelo de calibração com distribuição beta. Como

$$\begin{aligned} \frac{\partial \ell_{i1}^2(\mu_i, \phi)}{\partial \mu_i^2} &= -\phi^2 [\psi'(\mu_i \phi) + \psi'((1 - \mu_i) \phi)] \\ \frac{\partial \ell_{i0}^2(\mu_0, \phi_0)}{\partial \mu_0^2} &= -\phi_0^2 [\psi'(\mu_0 \phi_0) + \psi'((1 - \mu_0) \phi_0)] \end{aligned}$$

então, (3.23) pode ser reescrita de forma mais compacta,

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \theta_j \partial \theta_s} \right) = -\phi \sum_{i=1}^n w_i x_{ij} - \phi_0 \sum_{i=n+1}^{n+k} w_0 x_0, \quad (3.23')$$

onde $w_i = \phi[\psi'(\mu_i \phi) + \psi'((1 - \mu_i)\phi)] \frac{1}{[g'(\mu_i)]^2}$ e $w_0 = \phi_0[\psi'(\mu_0 \phi_0) + \psi'((1 - \mu_0)\phi_0)] \frac{1}{[g'(\mu_0)]^2}$. Na forma matricial, tem-se

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \Theta \partial \Theta^T} \right) = -\phi X^T W X - \phi_0 X_0^T W_0 X_0, \quad (3.23'')$$

onde $W = \text{diag}\{w_1, w_2, \dots, w_n\}$, $W_0 = \text{diag}\{w_0, w_0, \dots, w_0\}$, $X = [\mathbf{1}, x_i, \mathbf{0}]$ e $X_0 = [\mathbf{0}, \mathbf{0}, x_0]$.

Além disso, como

$$\begin{aligned} \frac{\partial \mu_i^*}{\partial \phi} &= \psi'(\mu_i \phi) \mu_i - \psi'((1 - \mu_i)\phi)(1 - \mu_i), \\ \frac{\partial \mu_0^*}{\partial \phi_0} &= \psi'(\mu_0 \phi_0) \mu_0 - \psi'((1 - \mu_0)\phi_0)(1 - \mu_0), \end{aligned}$$

para $j, s = 1, 2, 3$, chega-se a

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \theta_j \partial \phi} \right) = -\sum_{i=1}^n c_i \frac{1}{g'(\mu_i)} x_{ij}, \quad (3.24')$$

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \theta_j \partial \phi_0} \right) = -\sum_{i=n+1}^{n+k} c_0 \frac{1}{g'(\mu_0)} x_0, \quad (3.25')$$

com $c_i = \phi[\psi'(\mu_i \phi) \mu_i - \psi'((1 - \mu_i)\phi)(1 - \mu_i)]$ e $c_0 = \phi_0[\psi'(\mu_0 \phi_0) \mu_0 - \psi'((1 - \mu_0)\phi_0)(1 - \mu_0)]$.

Na forma matricial, com $C = (c_1, \dots, c_n)$, $C_0 = (c_0, \dots, c_0)$, $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$ e $T_0 = \text{diag}\{1/g'(\mu_0), \dots, 1/g'(\mu_0)\}$, tem-se

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \Theta \partial \phi} \right) = -X^T T C, \quad (3.24'')$$

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \Theta \partial \phi_0} \right) = -X_0^T T_0 C_0. \quad (3.25'')$$

Para $d_i = -\psi'(\phi) + \mu_i^2 \psi'(\phi \mu_i) + (1 - \mu_i)^2 \psi'((1 - \mu_i)\phi)$ e $d_0 = -\psi'(\phi_0) + \mu_0^2 \psi'(\phi_0 \mu_0) + (1 - \mu_0)^2 \psi'((1 - \mu_0)\phi_0)$, pode-se reescrever (3.26)–(3.27) como

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \phi^2} \right) = -\sum_{i=1}^n d_i, \quad (3.26')$$

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \phi_0^2} \right) = -\sum_{i=n+1}^{n+k} d_0, \quad (3.27')$$

que em notação matricial é

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \phi^2} \right) = -\text{tr}(D), \quad (3.26'')$$

$$\mathbb{E} \left(\frac{\partial^2 \ell(\cdot)}{\partial \phi_0^2} \right) = -\text{tr}(D_0), \quad (3.27'')$$

onde $D = \text{diag}\{d_1, \dots, d_n\}$ e $D_0 = \text{diag}\{d_0, \dots, d_0\}$.

Por fim, sob condições de regularidade, a matriz de informação de Fisher para o modelo de calibração beta é dada por

$$K_B = \begin{bmatrix} K_{\Theta\Theta} & K_{\Theta\phi} & K_{\Theta\phi_0} \\ K_{\phi\Theta} & K_{\phi\phi} & K_{\phi\phi_0} \\ K_{\phi_0\Theta} & K_{\phi_0\phi} & K_{\phi_0\phi_0} \end{bmatrix} = \begin{bmatrix} \phi X^T W X + \phi_0 X_0^T W_0 X_0 & X^T T C & X_0^T T_0 C_0 \\ C^T T^T X & \text{tr}(D) & 0 \\ C_0^T T_0^T X_0 & 0 & \text{tr}(D_0) \end{bmatrix}. \quad (3.29)$$

Note a semelhança entre as matrizes (3.29) e (3.10), com a primeira diferenciando-se da segunda pela inclusão dos termos correspondentes à segunda etapa do modelo de calibração beta (no caso, $\phi_0 X_0^T W_0 X_0$ na posição $K_{\Theta\Theta}$, e os elementos incluídos na terceira linha e terceira coluna de (3.29)).

Além disso, assim como no caso do modelo de regressão beta, para o modelo proposto tem-se que, à exceção de ϕ e ϕ_0 , os parâmetros $(\alpha_B, \beta_B, x_0, \phi, \phi_0)$ são não ortogonais dois a dois.

Sob condições de regularidade, o EMV de x_0 (\hat{x}_{0B}) é assintoticamente normal com média x_0 e variância $V(\hat{x}_{0B})$. Logo, a construção de um intervalo de confiança aproximado para o parâmetro x_0 do modelo proposto, com nível de confiança γ , pode ser feita utilizando-se o resultado assintótico $\frac{\hat{x}_{0B} - x_0}{\sqrt{V(\hat{x}_{0B})}} \xrightarrow{\mathcal{D}} N(0, 1)$, com a estimativa da variância de \hat{x}_{0B} ($\widehat{V}(\hat{x}_{0B})$) obtida a partir da inversa de K_B (i.e. K_B^{-1}). Este intervalo é dado por

$$\left(\hat{x}_{0B} - z_{\frac{\alpha}{2}} \sqrt{\widehat{V}(\hat{x}_{0B})}, \hat{x}_{0B} + z_{\frac{\alpha}{2}} \sqrt{\widehat{V}(\hat{x}_{0B})} \right), \quad (3.30)$$

onde $z_{\frac{\alpha}{2}}$ é o quantil de ordem $(1 - \frac{\alpha}{2})$ da distribuição normal padrão e $\widehat{V}(\hat{x}_{0B})$ é a estimativa da variância de \hat{x}_{0B} .

Por fim, observe que neste capítulo não se considerou o uso de uma estrutura de regressão para modelar o comportamento do parâmetro de dispersão ϕ , como discutido em Simas, Barreto-Souza e Rocha (2010). No entanto, vale ressaltar que a referida modelagem não se aplica a ϕ_0 , pois para o segundo estágio a única variável explicativa é uma constante (x_0) a ser determinada.

Além disso, casos nos quais os dados assumem valores contínuos nos intervalos $[0, 1]$, $[0, 1)$ e $(0, 1]$, o que conduz a abordagem da distribuição beta inflacionada (Ospina e Ferrari, 2010, 2012), também não são tratados neste estudo em razão da variável resposta y para o modelo de calibração proposto (i.e. transmitância) não assumir tipicamente os valores 0 ou 1.

Aplicações

Neste capítulo são apresentadas duas aplicações para os modelos de calibração usual, quadrático e beta, discutidos nos Capítulos 2 e 3 deste trabalho. A primeira aplicação utiliza uma amostra para a concentração de íons de ferro em águas naturais, extraída de Pimentel e Barros Neto (1996). A segunda aplicação baseia-se em dados sobre o teor de permanganato em soluções-padrão, com leituras espectrofotométricas em diferentes comprimentos de onda (Barros Neto, Scarminio e Bruns (2010)).

4.1 Critérios de Informação

Uma questão natural que surge quando se ajusta um modelo a um conjunto de dados é quão bom é esse ajuste. Quando se tem dois ou mais modelos, a questão relevante passa a ser qual o melhor deles (i.e. qual apresenta melhor ajuste relativo?). Dentre as diversas ferramentas existentes na literatura sobre seleção de modelos, algumas são baseadas na teoria da informação e são construídas a partir de funções perda que consideram o número de parâmetros (r) do modelo e o tamanho da amostra (n).

Nesse sentido, para n fixo, há um *trade-off* entre a qualidade do ajuste de um modelo e r , com um modelo com mais parâmetros sofrendo uma maior penalidade que um mais parcimonioso. Este princípio é utilizado por ferramentas conhecidas como critérios de informação, que consideram o melhor modelo aquele que apresenta menor valor para a função perda. Três critérios de informação utilizados em aplicações práticas são o AIC (Akaike Information Criterion), BIC (Bayesian Informa-

tion Criterion), também conhecido como critério de informação de Schwarz, e HQ (Hannan-Quinn Information Criterion), que são definidos da seguinte forma

$$AIC = -\frac{2\ell(\hat{\Theta})}{n} + \frac{2r}{n}, \quad (4.1)$$

$$BIC = -\frac{2\ell(\hat{\Theta})}{n} + \frac{r \log(n)}{n}, \quad (4.2)$$

$$HQ = -\frac{2\ell(\hat{\Theta})}{n} + \frac{2r \log(\log(n))}{n}, \quad (4.3)$$

onde $\ell(\hat{\Theta})$ é o valor máximo da função de log-verossimilhança do modelo sob análise, $\hat{\Theta}$ o valor estimado para o vetor de parâmetros Θ deste modelo, n é o número de observações e r o número de parâmetros em Θ .

Segundo estes critérios, o modelo mais adequado para um certo conjunto de dados é aquele que possui o menor valor para a estatística em questão. Neste trabalho, os mesmos serão utilizados para indicar, dentre os modelos considerados nas aplicações das Seções 4.2 e 4.3, qual apresenta o melhor ajuste relativo.

4.2 Concentração de Ferro em Águas Naturais

Os dados da amostra sob análise foram extraídos de Pimentel e Barros Neto (1996) e compreendem um conjunto de 8 concentrações-padrão de ferro (Fe) distribuídas em cinco níveis diferentes (com quatro replicatas para a concentração 0.2), para a determinação espectrofotométrica da concentração deste elemento em águas naturais por absorção molecular (com ortofenantrolina como agente complexante). A absorvância desses padrões foi lida em um espectrofotômetro Micronal B-242 na região ultravioleta visível (UV-VIS), em cubetas de 4 cm. O teor de concentração de ferro na água (em mg/L) é denotada por x , que é a única variável explicativa do modelo.

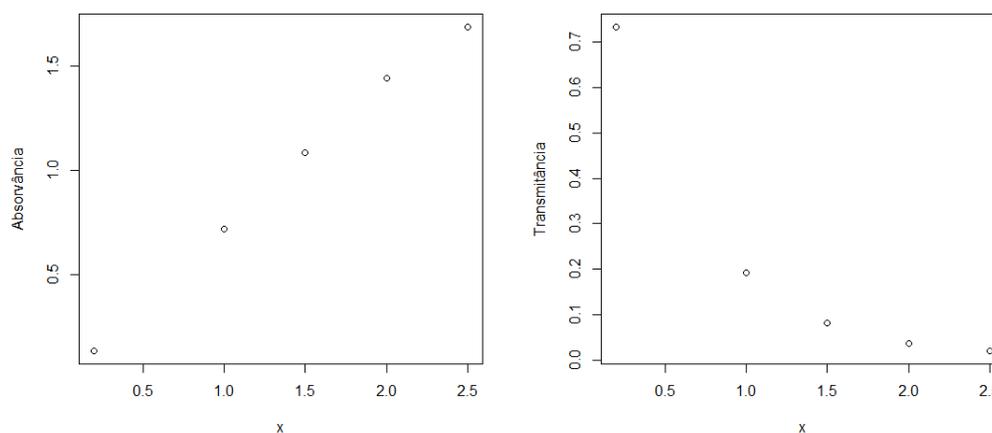
Para a presente aplicação, foram descartadas três das quatro observações correspondentes às replicatas para a concentração 0.2, e aplicou-se a transformação (2.3) da lei de Beer-Lambert às observações remanescentes (em absorvância), de forma a obter seu equivalente em transmitância. Deste modo, obteve-se aos dados apresentados na Tabela 4.1

Tabela 4.1: Resultados de uma determinação de ferro (Fe) por absorvância molecular

Concentração do analito, mg/L (x)	Absorvância	Transmitância
0.2	0.1351	0.7327
1.0	0.7169	0.1919
1.5	1.0846	0.0823
2.0	1.4416	0.0362
2.5	1.6849	0.0207

Fonte: Pimentel e Barros Neto (1996)

Este conjunto de dados foi utilizado para ajustar o primeiro estágio dos modelos discutidos nas Seções 2.3, 2.4 e 3.3, para a variável resposta (y) em absorvância ou transmitância, de acordo com o intervalo de variação permitido para y e com o comportamento desta variável (linear ou não-linear). Note que, para o presente caso, o modelo quadrático foi usado para ajustar os dados da variável resposta em transmitância em razão de seu comportamento não-linear (Figura 4.1).

Figura 4.1: Absorvância e transmitância para diferentes concentrações do analito x 

Os parâmetros dos modelos usual e quadrático foram estimados pelo método de máxima verossimilhança. Para o modelo de calibração beta, seus parâmetros foram estimados utilizando-se o algoritmo L-BFGS-B (Nocedal e Wright (2006)) através da maximização numérica da função de

log-verossimilhança (3.13), limitando-se o intervalo de variação para $(\phi_B, \phi_{0B}, x_{0B})$ a \mathbb{R}_3^+ e considerando as funções escore analíticas (3.14)–(3.16). Os valores iniciais para α_B , β_B , ϕ e ϕ_0 foram obtidos das estimativas resultantes da aplicação do modelo de regressão beta (Seção 3.2) aos dados da Tabela 4.1, tomando-se $\phi_{inicial} = \phi_{0inicial}$. Para x_0 , seu valor inicial foi dado por (2.10), que é o estimador de máxima verossimilhança do modelo usual para este parâmetro (calculado a partir do mesmo conjunto de dados).

Isto posto, para $n = 5$ e $k = 1$, considerou-se conhecido para a segunda etapa apenas o valor da variável resposta, com $y_0 = 0.1519$ (absorvância (A)) ou $y_0 = 0.70486$ (transmitância (T)), tratando-se o valor de x_0 neste estágio como desconhecido. No entanto, como o valor de y para esta etapa foi convenientemente escolhido a partir dos pontos amostrais descartados anteriormente, isto significa que o mesmo está associado a um x_0 verdadeiro de 0.2, o que tornou possível medir o erro da estimativa pontual deste parâmetro para cada modelo, isto é $Erro(\hat{x}_0) = \hat{x}_0 - x_0$. Além disso, com o intuito de avaliar a qualidade do ajuste relativo para estes modelos, calculou-se os valores para os critérios de informação de Akaike, BIC e Hannan-Quinn. Os resultados obtidos são apresentados nas Tabelas 4.2, 4.3 e 4.4 (na Tabela 4.2 os erros-padrão estão entre parênteses).

Tabela 4.2: Estimativas e erros-padrão para os modelos de calibração usual, quadrático e beta

Modelo	y	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\beta}_2$	$\hat{\phi}$	$\hat{\phi}_0$	\hat{x}_0
Usual	A	0.02505	0.68582	-	-	-	0.18497
		(0.04376)	(0.02659)				(0.08996)
Quadrático	T	0.88171	-0.85450	0.20735	-	-	0.21857
		(0.05672)	(0.09543)	(0.03436)			(0.07489)
Beta (logit)	T	1.32830	-2.45476	-	144.292	142882	0.18649
		(0.20001)	(0.21535)		(92.0246)	(203268)	(0.06870)
Beta (log-log)	T	1.05043	-1.05965	-	18.5878	181173	0.00000
		(0.29542)	(0.16883)		(16.9185)	(255546)	(0.27881)

Tabela 4.3: Erro de \hat{x}_0 e intervalos de confiança aproximados para x_0 ($\gamma = 95\%$)

Modelo	$Erro(\hat{x}_0)$	Intervalos de confiança
Usual	-0.0150	(0.00865; 0.36129)
Quadrático	0.0186	(0.07179; 0.36535)
Beta (logit)	-0.0135	(0.05183; 0.32115)
Beta (log-log)	-0.2000	(-0.54648; 0.54648)

Tabela 4.4: Critérios de informação para os modelos usual, quadrático e beta

Modelo	$\ell(\hat{\Theta})$	AIC	BIC	HQ
Usual	9.43170	-2.97268	-3.12891	-3.39197
Quadrático	10.8328	-3.13312	-3.36746	-3.76206
Beta (logit)	17.1553	-4.86211	-5.25267	-5.91034
Beta (log-log)	12.0957	-2.83827	-3.22883	-3.88649

A partir dos resultados apresentados na Tabela 4.2, observa-se alguma variação entre as estimativas para x_0 , incluindo seus erros-padrão, para os modelos de calibração usual, quadrático e beta, este último considerando funções de ligação logit e log-log. Nota-se também que o erro-padrão de \hat{x}_0 para o modelo beta com função de ligação logit é o menor dentre os modelos considerados. Com relação aos demais parâmetros, em geral, as estimativas de α e β para os três modelos, e de δ para o modelo quadrático, são todas significativas a 5%, a exceção de $\hat{\alpha}$ para o modelo usual, que é não significativa. Esta situação se repete para as estimativas de ϕ e ϕ_0 para o modelo beta (logit e log-log), com o elevado valor de $\hat{\phi}_0$ podendo ser explicado pelo fato que $k = 1$.

A Tabela 4.3 mostra o erro do estimador de x_0 para cada modelo ($Erro(\hat{x}_0) = \hat{x}_0 - x_0$), com $x_0 = 0.2$, além dos intervalos de confiança aproximados para este parâmetro, considerando um

nível de confiança de 95%. Para a presente aplicação, o modelo beta com função de ligação logit apresentou um estimador para x_0 mais exato que os modelos usual e quadrático, além de, como mencionado anteriormente, ter a menor variância (desvio-padrão), e um intervalo de confiança com menor amplitude para este parâmetro. Os resultados da Tabela 4.4 mostram que o modelo beta (logit) apresenta melhor ajuste segundo os três critérios de informação (AIC, BIC e HQ), quando comparado aos modelos usual e quadrático. Isso demonstra que o modelo proposto, quando utiliza uma função de ligação logit, é o mais adequado para modelar os dados em questão.

4.3 Concentração de Permanganato

Para esta aplicação, utilizou-se uma amostra apresentada em Barros Neto, Scarminio e Bruns (2010), obtida a partir de um estudo sobre a determinação simultânea do teor de permanganato (MnO_4^+), sais de dicromato ($\text{Cr}_2\text{O}_7^{2-}$) e cobalto (Co^{2+}). Várias amostras de soluções-padrão foram preparadas e analisadas por espectrofotometria ultra-violeta visível (UV-VIS), em três comprimentos de onda (λ) diferentes, 410, 440 e 530 *nm* (nanômetros). As leituras estão expressas em absorvância, com o teor da concentração de permanganato nas soluções em *mL* (mililitros), denotado por x .

Assim como no caso anterior, aplicou-se a transformação logarítmica da lei de Beer-Lambert às leituras em absorvância (A) de forma a se obter seu equivalente em transmitância (T). Os dados estão na Tabela 4.5, agrupados de acordo com $\lambda = 410, 440$ e 530 .

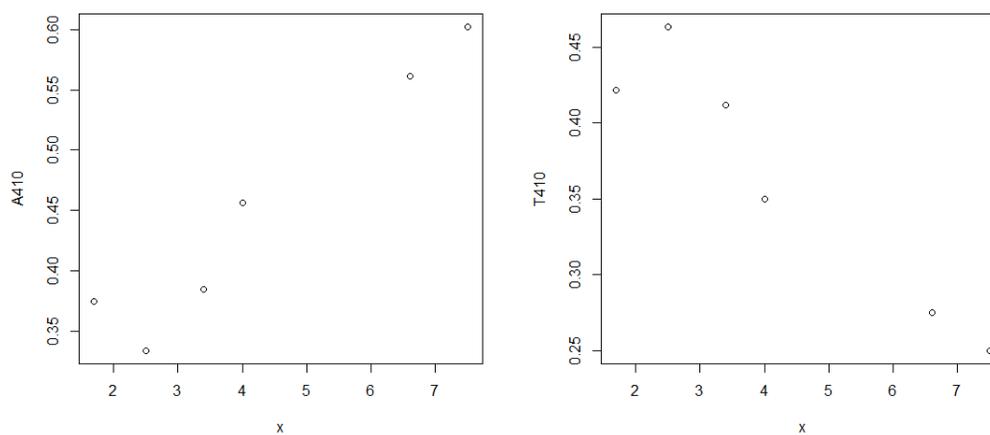
Na presente aplicação, para efeito de estimação dos modelos de calibração usual, quadrático e beta (logit e log-log), as observações referentes à concentração $x = 5.0$ na Tabela 4.5 foram utilizadas apenas no segundo estágio destes modelos (variável y_0), na forma de absorvância ou transmitância. As Figuras 4.2, 4.3 e 4.4 apresentam o conjunto de dados utilizado no primeiro estágio da calibração.

Os modelos de calibração citados foram ajustados ao conjunto de dados em questão, para $n = 6$ e $k = 3$. A variável resposta (y) foi utilizada na forma de absorvância ou transmitância de acordo com o intervalo de variação permitido para a mesma e com seu comportamento (linear ou não-linear) (Figuras 4.2, 4.3 e 4.4). No presente caso, como a variável resposta não apresentou um padrão claro de não-linearidade quando expressa em transmitância, optou-se por utilizar o modelo quadrático para ajustar y em ambos os casos (absorvância e transmitância), reportando-se apenas os resultados que possam ser interpretados de acordo com a prática usual em calibração.

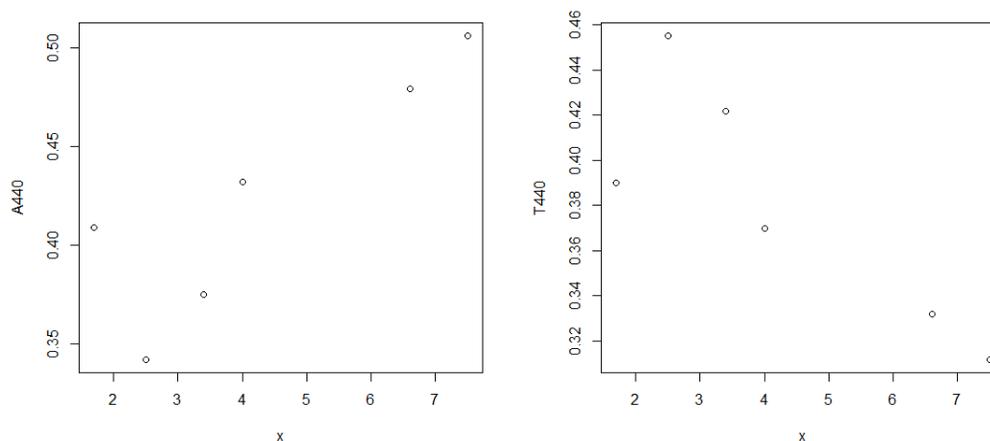
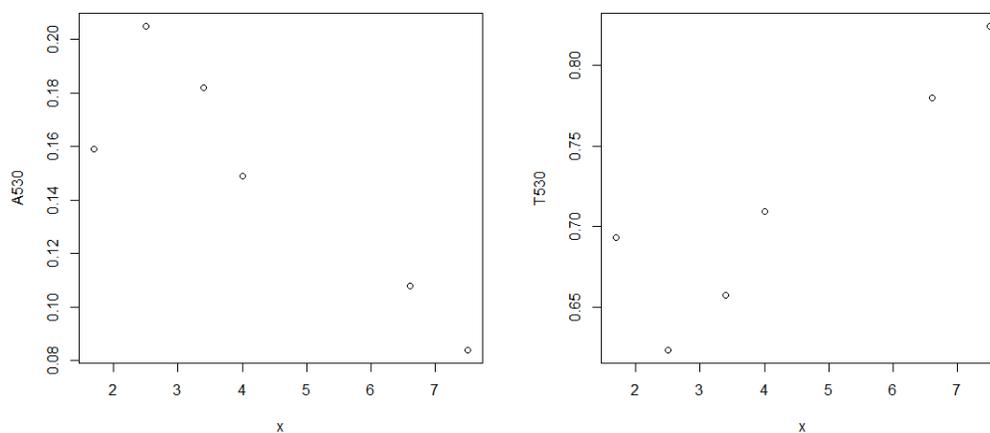
Tabela 4.5: Determinação da concentração do íon MnO_4^+ em soluções-padrão por espectrofotometria

Concentração do analito, mL x	410		440		530	
	A	T	A	T	A	T
1.7	0.3750	0.4217	0.4090	0.3899	0.1590	0.6934
2.5	0.3340	0.4634	0.3420	0.4550	0.2050	0.6237
3.4	0.3850	0.4121	0.3750	0.4217	0.1820	0.6577
4.0	0.4560	0.3499	0.4320	0.3698	0.1490	0.7096
5.0	0.4720	0.3373	0.4260	0.3750	0.1490	0.7096
5.0	0.4720	0.3373	0.4260	0.3750	0.1490	0.7096
5.0	0.5230	0.2999	0.4880	0.3251	0.1130	0.7709
6.6	0.5610	0.2748	0.4790	0.3319	0.1080	0.7798
7.5	0.6020	0.2500	0.5060	0.3119	0.0840	0.8241

Fonte: Barros Neto, Scarminio e Bruns (2010)

Figura 4.2: Absorvância e transmitância para diferentes concentrações do analito x com $\lambda = 410$ 

Os referidos modelos foram estimados utilizando os mesmos procedimentos descritos na Seção 4.2. Para o segundo estágio dos mesmos, considerou-se conhecidos apenas os valores das variáveis

Figura 4.3: Absorvância e transmitância para diferentes concentrações do analito x com $\lambda = 440$ Figura 4.4: Absorvância e transmitância para diferentes concentrações do analito x com $\lambda = 530$ 

resposta y_0 , tratando-se os valores de x_0 neste estágio como desconhecidos. No entanto, como mencionado anteriormente, os valores de y_0 escolhidos (em absorvância ou transmitância), estão associados a um x_0 verdadeiro igual a 5.0 (ver Tabela 4.5). Isto tornou possível medir o erro da estimativa pontual deste parâmetro para cada modelo. Para se avaliar a qualidade do ajuste relativo para cada um dos modelos em questão, obteve-se os valores para os critérios de informação de Akaike, BIC e Hannan-Quinn.

Os resultados obtidos para os três comprimentos de onda do conjunto de dados de Barros Neto *et al.* (2010) ($\lambda = 410, 440$ e 530) estão nas Tabelas 4.6 a 4.10.

Tabela 4.6: Estimativas e erros-padrão para os modelos usual, quadrático e beta ($\lambda = 410$)

Modelo	y	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\phi}$	$\hat{\phi}_0$	\hat{x}_0
Usual	A	0.25729	0.04550	-	-	-	5.09291
		(0.02946)	(0.00618)				(0.49360)
Quadrático	A	0.31535	0.01434	0.00331	-	-	5.39170
		(0.08481)	(0.04280)	(0.00450)			(1.76609)
Beta (logit)	T	0.09621	-0.15839	-	420.971	420.972	5.22223
		(0.09366)	(0.02052)		(242.783)	(343.353)	(0.48495)
Beta (log-log)	T	0.40157	-0.09765	-	440.113	695.346	5.31342
		(0.05944)	(0.01207)		(253.836)	(567.376)	(0.40299)

Como mencionado acima, os resultados reportados neste aplicação para o modelo quadrático desconsideram os casos para os quais as estimativas obtidas não apresentam interpretação coerente com a usual nos modelos de calibração, como valores negativos para \hat{x}_0 , mesmo que não significativos. Isto ocorreu quando o modelo em questão foi utilizado para ajustar dados da Tabela 4.5 para y expresso em transmitância com $\lambda = 410$ e 440 , e y em absorvância para $\lambda = 530$. Além de estimativas para x_0 fora do intervalo de variação de x , com $x > 0$, os valores para as estimativas da variância de \hat{x}_0 nestes casos foram os maiores dentre as situações consideradas para cada λ .

As Tabelas 4.6, 4.7 e 4.8 mostram uma certa variação entre as estimativas de x_0 , conforme o modelo utilizado e o comprimento de onda (λ), com o modelo quadrático apresentando os maiores desvios-padrão para este parâmetro para todos os λ 's. Por outro lado, o modelo beta (logit e log-log) possui o menor desvio-padrão estimado para \hat{x}_0 para todos os comprimentos de onda e para ambas as funções de ligação consideradas.

A Tabela 4.9 apresenta o erro do estimador de x_0 para os modelos usual, quadrático e beta, considerando $x_0 = 5.0$ como valor verdadeiro, e os intervalos de confiança aproximados para este

Tabela 4.7: Estimativas e erros-padrão para os modelos usual, quadrático e beta ($\lambda = 440$)

Modelo	y	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\phi}$	$\hat{\phi}_0$	\hat{x}_0
Usual	A	0.32380	0.02335	-	-	-	5.26108
		(0.03197)	(0.00670)				(1.09461)
Quadrático	A	0.39658	-0.01570	0.00415	-	-	5.84349
		(0.08922)	(0.04503)	(0.00473)			(3.64089)
Beta (logit)	T	-0.13072	-0.084551	-	346.049	409.466	5.34367
		(0.10194)	(0.02177)		(199.517)	(333.943)	(0.93204)
Beta (log-log)	T	0.26489	-0.05405	-	353.232	353.232	5.37662
		(0.06611)	(0.01357)		(203.662)	(288.036)	(0.94374)

parâmetro para um nível de confiança igual a 95%. O modelo usual apresentou o estimador de x_0 mais exato para todos os comprimentos de onda, seguido pelo modelo beta (logit e log-log). Note que o erro para as estimativas deste parâmetro varia conforme a função de ligação utilizada no modelo de calibração beta, com a função logit apresentando melhores resultados que a log-log segundo este critério. A amplitude dos intervalos de confiança para x_0 do modelo beta (logit e log-log) foi a menor dentre todos os modelos considerados.

De forma geral, os resultados mostrados na Tabela 4.10 evidenciam a maior qualidade do ajuste do modelo beta aos dados da Tabela 4.5, para todos os λ 's, em comparação aos modelos usual e quadrático, segundo os três critérios de informação considerados (AIC, BIC e HQ). Vale ressaltar que, na maioria dos casos, o modelo beta com a função de ligação log-log demonstra um melhor ajuste que seu similar com a função logit. Por fim, a análise da Tabela 4.10 sugere que o modelo quadrático apresentou o pior ajuste à amostra em questão em todos os casos.

Tabela 4.8: Estimativas e erros-padrão para os modelos usual, quadrático e beta ($\lambda = 530$)

Modelo	y	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	$\hat{\phi}$	$\hat{\phi}_0$	\hat{x}_0
Usual	A	0.22221	-0.01737	-	-	-	4.90721
		(0.02203)	(0.00462)				(0.94185)
Quadrático	T	0.70154	-0.03085	0.00639	-	-	5.62335
		(0.08439)	(0.04259)	(0.00448)			(3.37617)
Beta (logit)	T	0.30722	0.14732	-	203.345	203.351	4.66408
		(0.14016)	(0.03103)		(117.168)	(165.703)	(0.76636)
Beta (log-log)	T	0.57966	0.12449	-	194.848	194.849	4.62691
		(0.11821)	(0.02707)		(112.260)	(158.762)	(0.79499)

Tabela 4.9: Erro de \hat{x}_0 e intervalos de confiança aproximados para x_0 ($\gamma = 95\%$)

λ	Modelo	$Erro(\hat{x}_0)$	Intervalos de confiança
410	Usual	0.09291	(4.12544; 6.06037)
	Quadrático	0.39170	(1.93016; 8.85324)
	Beta (logit)	0.22223	(4.27173; 6.17273)
	Beta (log-log)	0.31342	(4.52356; 6.10328)
440	Usual	0.26108	(3.11563; 7.40652)
	Quadrático	0.84349	(-1.29266; 12.9796)
	Beta (logit)	0.34367	(3.51688; 7.17047)
	Beta (log-log)	0.37662	(3.52688; 7.22636)
530	Usual	0.09279	(3.06118; 6.75324)
	Quadrático	0.62335	(-0.99394; 12.2407)
	Beta (logit)	0.33592	(3.16201; 6.16614)
	Beta (log-log)	-0.37309	(3.06874; 6.18508)

Tabela 4.10: Critérios de informação para os modelos usual, quadrático e beta com $\lambda = 410, 440, 530$

λ	Modelo	$\ell(\hat{\Theta})$	AIC	BIC	HQ
410	Usual	13.4020	-3.80067	-3.87008	-4.07853
	Quadrático	13.9008	-3.63358	-3.73770	-4.05038
	Beta (logit)	21.7875	-5.59584	-5.76938	-6.29051
	Beta (log-log)	22.0817	-5.6939	-5.86742	-6.38856
440	Usual	12.9107	-3.63689	-3.70630	-3.91476
	Quadrático	13.5962	-3.53208	-3.63620	-3.94888
	Beta (logit)	20.3735	-5.12449	-5.29802	-5.81916
	Beta (log-log)	20.4192	-5.13973	-5.31327	-5.83440
530	Usual	15.1469	-4.38229	-4.45170	-4.66016
	Quadrático	13.9304	-3.64347	-3.74759	-4.06028
	Beta (logit)	18.6441	-4.54802	-4.72155	-5.24269
	Beta (log-log)	18.5087	-4.50290	-4.67643	-5.19757

Conclusões

5.1 Comentários Finais

No contexto de calibração, o principal parâmetro de interesse é x_0 , que representa o valor ou a quantidade desconhecida a ser determinada com diversos objetivos, por exemplo, a avaliação de políticas em economia, a aferição de instrumentos em física e a verificação da eficácia de um medicamento em medicina, dentre outros. Em química analítica, em particular, x_0 é a concentração desconhecida de um analito em uma solução para a qual se conhece apenas a leitura espectrofotométrica associada a este parâmetro. A literatura sobre modelos de calibração apresenta diversas abordagens para o tratamento do problema de calibração, com vários enfoques e aplicações nas mais diversas áreas de conhecimento.

Nesta dissertação apresentou-se um modelo de calibração que considera uma variável resposta restrita ao intervalo $(0, 1)$ cuja média pode ser modelada com o uso de uma distribuição beta, que também é função de uma estrutura de regressão do tipo sugerida por Ferrari e Cribari-Neto (2004). Este modelo difere dos estudos anteriores por propor a modelagem da média da variável resposta, ao invés da modelagem dos erros do modelo, como é usual na literatura sobre modelos de calibração.

O modelo proposto foi derivado considerando-se uma reparametrização da distribuição beta para os dois estágios do mesmo, o que permitiu a obtenção de um modelo cujas médias das variáveis resposta para o primeiro e segundo estágios podem, eventualmente, seguir distribuições beta distintas (i.e., com valores distintos para seus parâmetros). Além disso, derivou-se sua matriz de informação de Fisher esperada, o que possibilita a realização de testes de hipóteses e a construção de intervalos

de confiança para seus parâmetros.

Os parâmetros do modelo de calibração beta, para funções de ligação logit e log-log, foram estimados pelo método da máxima verossimilhança, considerando-se as funções escore analíticas obtidas no Capítulo 3. As estimativas obtidas para este modelo para o parâmetro x_0 foram comparadas em aplicações com dados reais de química analítica, com as resultantes de um modelo linear (*modelo usual*) e um modelo quadrático, para variáveis resposta coerentes com a lei de Beer-Lambert. Mesmo considerando que as amostras das referidas aplicações são pequenas ($n \leq 10$), o modelo proposto (logit e log-log) apresentou melhor ajuste em todos os casos segundo os critérios de informação apresentados no Capítulo 4. O erro do estimador deste parâmetro para o modelo proposto foi menor que o do modelo quadrático em todos os casos considerados, e menor que o do estimador do modelo usual em apenas um caso. No entanto, a estimativa para o erro-padrão do estimador de x_0 apresentada pelo modelo de calibração beta foi a menor dentre os modelos estudados em ambas as aplicações.

5.2 Sugestões para Estudos Futuros

Dentro do âmbito dos modelos de regressão beta, uma extensão natural para o modelo aqui proposto é a modelagem do parâmetro de dispersão do primeiro estágio, ϕ , na linha sugerida por Simas *et. al.* (2010). Outra possibilidade é comparar o referido modelo, que usa a distribuição beta para modelar a média das variáveis resposta, com um modelo derivado de acordo com a abordagem da distribuição beta inflacionada (Ospina e Ferrari, 2010, 2012)). O objetivo neste caso seria verificar qual dos dois modelos apresenta melhor ajuste quando estas variáveis assumem valores muito próximos, ou eventualmente iguais, aos limites de seu domínio de variação (i.e. 0 ou 1), e o impacto sobre o valor estimado de x_0 e sua variância.

Uma sugestão no contexto dos modelos com erros de medida é considerar que a variável independente para o modelo de calibração beta apresenta erro de medida tipo Berkson, como discutido em Blas, Sandoval e Yoshida (2007). Por fim, pode-se derivar também, sem maiores dificuldades, uma versão do modelo proposto para o caso de calibração multivariada.

Referências Bibliográficas

- [1] ARELLANO-VALLE, R. B.; BRANCO, M. D. *Distribuições Elípticas Assimétricas*. Caxambu. XVI SINAPE. 2004.
- [2] ARELLANO-VALLE, R. B.; OZAN, S.; BOLFARINE, H.; LACHOS, V. H. Skew normal measurement error models. *Journal of Multivariate Analysis*, v. 96, n. 2, p. 265-281. 2005.
- [3] BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. *Como Fazer Experimentos: Pesquisa e Desenvolvimento na Ciência e na Indústria*. 4ª ed., Porto Alegre. Bookman, 2010.
- [4] BERKSON, J. Estimation of a linear function for a calibration line; consideration of a recent proposal. *Technometrics*, v. 11, n. 4, p. 649-660. 1969.
- [5] BLAS, B. G. *Calibração Controlada Aplicada na Química Analítica*. Dissertação de mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo. 2005.
- [6] BLAS, B. G.; SANDOVAL, M. C. Heteroscedastic controlled calibration model applied to analytical chemistry. *Journal of Chemometrics*, v. 24, p. 241-248. 2010.
- [7] BLAS, B. G.; BOLFARINE, H.; LACHOS, V. H. Statistical analysis of controlled calibration model with replicates. *Journal of Statistical Computation and Simulation*, p. 1 - 21. 2011.
- [8] BLAS, B. G.; SANDOVAL, M. C.; YOSHIDA, O. S. Homoscedastic controlled calibration model. *Journal of Chemometrics*, v. 21, p. 145-155, 2007.
- [9] BOLFARINE, H.; LIMA, C. R. O. P.; SANDOVAL, M. C. Linear calibration in functional regression models. *Communications in Statistics: Theory and Methods*, v. 26, n. 10, p. 2307-2328, 1997.

- [10] BOLFARINE, H.; LIMA, C. R. O. P.; SANDOVAL, M. C. Linear calibration in functional regression models with one of the variances known. *South African Statistical Journal*, v. 33, p. 95-116, 1999.
- [11] BRANCO, M. D. *Calibração: Uma Abordagem Bayesiana*. Tese de doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1997.
- [12] CRIBARI-NETO, F.; ZARKOS, S. G. R. Yet another econometric programming environment. *Journal of Applied Econometrics*, v. 14, n. 3, p. 319-329, 1999.
- [13] CRIBARI-NETO, F.; ZEILEIS, A. Beta regression in R. *Journal of Statistical Software*, v. 34, n. 2, 2010.
- [14] DANZER, K.; CURRIE, L.A. Guidelines for calibration in analytical chemistry: part 1. Fundamentals and single component calibration. *Pure and Applied Chemistry*, v. 70, n. 4, p. 993-1040, 1998.
- [15] DANZER, K.; OTTO, M.; CURRIE, L.A. Guidelines for calibration in analytical chemistry: part 2. Multispecies calibration. *Pure and Applied Chemistry*, v. 76, n. 6, p. 1215-1225, 2004.
- [16] DAWKINS, C.; SRINIVASAN, T. N.; WHALLEY, J. Calibration. In *Handbook of Econometrics*, J. J. Heckmans; E. Leamer, Eds., vol. 5, Amsterdam, Elsevier Science, p. 3653-3703, 2001.
- [17] DOMINGOS-FILHO, M. *Calibração Absoluta Estrutural sem a Suposição de Normalidade*. Dissertação de mestrado, Departamento de Estatística, Universidade Federal de Pernambuco, Recife, 2000.
- [18] EISENHART, C. The interpretation of certain regression methods, and their use in biological and industrial research. *Annals of Mathematical Statistics*, v. 10, p. 162-186, 1939.
- [19] FERRARI, S. L. P.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, v. 31, n. 7, 799-815, 2004.
- [20] FIGUEIREDO, C. C. *Calibração Linear Assimétrica*. Tese de doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2009.

- [21] FRANCISCONI, C. N. *Comparação de Instrumentos de Medição Usando Calibração Comparativa*. Dissertação de mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1996.
- [22] GALEA-ROJAS, M. *Calibração Comparativa Estrutural e Funcional*. Tese de doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1995.
- [23] GRÄTZER, G. *More Math into LaTeX*. 4th ed., New York, Springer, 2007.
- [24] GUIMARÃES-FILHO, Z. O.; HELENE, O. One step self-calibration procedure in gammaray energy measurements. *Brazilian Journal of Physics*, v. 33, n. 2, p. 280-281, 2003.
- [25] IHAKA, R.; GENTLEMAN, R. R. A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, v. 5, n. 3, p. 299–314, 1996.
- [26] KRUTCHKOFF, R. Classical and inverse regression methods of calibration. *Technometrics*, v. 9, n. 3, p. 425-439, 1967.
- [27] LACHOS, V. H. *Modelos Lineares Mistos Assimétricos*. Tese de doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2004.
- [28] LAMPORT, L. *LaTeX: A Document Preparation System*. 2nd ed., Reading, Massachusetts, Addison-Wesley Professional, 1994.
- [29] LEITE, J. G.; SINGER, J. M. *Métodos Assintóticos em Estatística: Fundamentos e Aplicações*. São Paulo, 9º Simpósio Braileiro de Probabilidade e Estatística, IME-USP, 1990.
- [30] LIMA, C. R. O. P. *Calibração Absoluta com Erros nas Variáveis*. Tese de doutorado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1996.
- [31] LWIN, T. Discussion of Hunter and Lamboy's 1981 paper. *Technometrics*, v. 23, p. 339-341, 1981.
- [32] MARCIANO, F. W. P. *Modelo de Calibração com Erros Simétricos*. Dissertação de mestrado, Departamento de Estatística, Universidade Federal de Pernambuco, Recife, 2012.
- [33] McCULLAGH, P.; NELDER, J. A. *Generalized Linear Models*. 2nd ed., London: Chapman and Hall, 1989.
- [34] MEIER, P. C.; ZÜND, R. E. *Statistical Methods in Analytical Chemistry*. 2nd ed., New York, Wiley, 2000.

- [35] NOCEDAL, J.; WRIGHT, S. J. *Numerical Optimization*. 2nd ed., New York: Springer-Verlag, 2006.
- [36] OLIVEIRA, E. C.; AGUIAR, P. F. Validação da metodologia da avaliação de incerteza em curvas de calibração melhor ajustadas por polinômios de segundo grau. *Química Nova*, vol. 32, n. 6, p. 1571-1575, 2009.
- [37] OSPINA, R.; FERRARI, S. L. P. Inflated beta distributions. *Statistical Papers*, vol. 51, n. 1, p. 111–126, 2010.
- [38] OSPINA, R.; FERRARI, S. L. P. A general class of zero-or-one inflated beta regression models. *Computational Statistics and Data Analysis*, vol. 56, n. 6, p. 1609–1623, 2012.
- [39] OSPINA, R.; CRIBARI-NETO, F.; VASCONCELLOS, K. L. P. Improved point and interval estimation for a beta regression model. *Computational Statistics & Data Analysis*, vol. 51, n. 2, p. 960–981, 2006.
- [40] PIMENTEL, M. F.; BARROS NETO, B. Calibração: uma revisão para químicos analíticos. *Química Nova*, vol. 19, n. 3, p. 268-277, 1996.
- [41] PIMENTEL, M. F.; GALVÃO, R. K. H.; ARAÚJO, M. C. U. Recomendações para calibração em química analítica. Parte 2. Calibração multianalito. *Química Nova*, vol. 31, n. 2, p.462-467, 2008.
- [42] RAMOS, P. G. A. *Calibração Linear*. Dissertação de mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 1987.
- [43] RENCHER, A. C.; SCHAALJE, G. B. *Linear Models in Statistics*. 2nd ed., Hoboken, Wiley, 2008.
- [44] RODRIGUEZ, C. L. B. *Inferência Bayesiana no Modelo Normal Assimétrico*. Dissertação de mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2005.
- [45] SHUKLA, G. K. On the problem of calibration. *Technometrics*, v. 14, n. 3, p. 547-553, 1972.
- [46] SIMAS, A. B.; BARRETO-SOUZA, W.; ROCHA, A. V. Improved estimators for a general class of beta regression models. *Computational Statistics & Data Analysis*, vol. 54, n. 2, p. 348–366, 2010.

- [47] THOMAS, E. V.; GE, N. Development of robust multivariate calibration models. *Technometrics*, v. 42, n. 2, 168-177, 2000.
- [48] VENABLES, W. N.; SMITH, D. M.; the R Development Core Team. *An Introduction to R*. 2nd ed., Network Theory Limited, 2012.
- [49] WILLIAMS, E. J. A note on regression methods in calibration. *Technometrics*, v. 11, n. 1, p. 189-192, 1969a.
- [50] WILLIAMS, E. J. Regression methods in calibration problems. *Bulletin of the international Statistical Institute*, v. 43, p. 17-28, 1969b.
- [51] WORKMAN, JR. Optical spectrometers. In *Applied Spectroscopy: A Compact Reference for Practitioners*, Jerry Workman, Jr., Art W. Springsteen, Eds., San Diego, Academic Press, 1998.

Códigos em *R*

Neste apêndice são apresentandos os códigos em *R* utilizados para estimação dos parâmetros dos modelos de calibração usual, quadrático e beta nas aplicações do Capítulo 4. A estimação do modelo proposto (beta) usou o pacote *betareg* para determinação dos valores iniciais para α_B , β_B , ϕ e ϕ_0 para o algoritmo L-BFGS-B, com $\phi_{inicial} = \phi_{0_{inicial}}$. O valor inicial para x_0 foi obtido a partir da substituição das estimativas de máxima verossimilhança (ou mínimos quadrados) dos parâmetros α_U e β_U do modelo usual, na expressão (2.10). O algoritmo L-BFGS-B foi implementado através da função *optim()*.

```
/******  
* CÓDIGOS EM R  
*  
* USO: Estimação dos parâmetros dos modelos de calibração usual e quadrático;  
* estimação dos parâmetros do modelo proposto (Beta) através do algoritmo L-BFGS-B  
*  
* AUTOR: Mileno Tavares Cavalcante  
*  
* DATA 1a. VERSÃO: Jul/2012  
*  
* DATA ÚLTIMA ATUALIZAÇÃO: Jan/2013  
*  
*****
```

```

# Código para estimação dos parâmetros do modelo usual, x0 e variância de x0_hat #

y<-c()
y0<-c()
x<-c()

gy_ls<-lm(y ~ x)
w<-coef(gy_ls)
x0_hat<-(mean(y0)-w[1])/w[2] #estimador para x0

alpha<-w[1]
beta<-w[2]

var_resid<-(1/(length(y)+length(y0)-3))*(sum((y-alpha-beta*x)^2)
+sum((y0-mean(y0))^2))
Sxx<-sum((x-mean(x))^2)/length(y)
var_x0_hat <-(var_resid/beta^2)*(1/length(y0)+1/length(y)
+((mean(x)-x0_hat)^2)/(length(y)*Sxx))
#estimador para a variância de x0_hat

summary(gy_ls) #estimativas dos parâmetros e erros-padrão

logLik(gy_ls) # valor máximo da função de log-verossimilhança
AIC<--2*logLik(gy_ls)/length(y)+2*2/length(y) # critério de informação de Akaike
BIC<--2*logLik(gy_ls)/length(y)+2*log(length(y))/length(y) # critério de
informação bayesiana
HQ<--2*logLik(gy_ls)/length(y)+2*2*log(log(length(y)))/length(y) # critério de
informação de Hannan-Quinn

```

```

# Código para estimação dos parâmetros do modelo quadrático, x0 e variância
de x0_hat #

i) para y crescente

y<-c()
y0<-c()
x<-c()
z<-x^2

gy_ls2<-lm(y ~ x + z)
w<-summary(gy_ls2)$coeff # estimativas dos parâmetros
cov_m<-vcov(gy_ls2) # estimativa da matriz de covariância dos parâmetros

# estimador de x0 para y crescente
x0_hat<--(w[2,1]/(2*w[3,1]))+sqrt((w[2,1]/(2*w[3,1]))^2-(w[1,1]-mean(y0))/w[3,1])

alpha<-w[1,1]
beta1<-w[2,1]
beta2<-w[3,1]

y_hat<-w[1,1]+w[2,1]*x+w[3,1]*z
a1<-1/((w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2))
a2<--1/((w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2))
a3<-(-1+w[2,1]/((w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2)))/(2*w[3,1])
a4<-(w[2,1]/((w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2)))*((-w[2,1]
+(w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2))/(2*w[2,1]^2))-((w[1,1]
-mean(y0))/w[3,1])*a2

uy_hat<-((1/(length(y)+length(y0)-4))*(sum((y-y_hat)^2)+sum((y0-mean(y0))^2)))

```

```

# estimador para a variância de x0_hat (y crescente)
var_x0_hat<-a1^2*uy_hat+a2^2*cov_m[1,1]+a3^2*cov_m[2,2]+a4^2*cov_m[3,3]
+2*a2*a3*cov_m[1,2]+2*a2*a4*cov_m[1,3]+2*a3*a4*cov_m[2,3]

logLik(gy_ls2) # valor máximo da função de log-verossimilhança
AIC<--2*logLik(gy_ls2)/length(y)+2*3/length(y) # critério de informação de Akaike
BIC<--2*logLik(gy_ls2)/length(y)+3*log(length(y))/length(y) # critério de informação
bayesiana
HQ<--2*logLik(gy_ls2)/length(y)+2*3*log(log(length(y)))/length(y) # critério de
informação de Hannan-Quinn

ii) para y decrescente

gy_lst2<-lm(y ~ x + z)
w<-summary(gy_lst2)$coeff # estimativas dos parâmetros
cov_m<-vcov(gy_lst2) # estimativa da matriz de covariância dos parâmetros

# estimador de x0 para y decrescente
x0_hat<--(w[2]/(2*w[3]))-sqrt((w[2]/(2*w[3]))^2-(w[1]-mean(y0))/w[3])

alpha<-w[1]
beta1<-w[2]
beta2<-w[3]

y_hat<-w[1,1]+w[2,1]*x+w[3,1]*z
a1<--1/((w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2))
a2<-1/((w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2))
a3<-(-1-w[2,1]/((w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2)))/(2*w[3,1])
a4<-(w[2,1]/((w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2)))*((w[2,1]
+(w[2,1]^2-4*w[3,1]*(w[1,1]-mean(y0)))^(1/2))/(2*w[2,1]^2))-((w[1,1]
-mean(y0))/w[3,1])*a2

```

```

uy_hat<-((1/(length(y)+length(y0)-4))*(sum((y-y_hat)^2)+sum((y0-mean(y0))^2)))

# estimador para a variância de x0_hat (y decrescente)
var_x0_hat<-a1^2*uy_hat+a2^2*cov_m[1,1]+a3^2*cov_m[2,2]+a4^2*cov_m[3,3]
+2*a2*a3*cov_m[1,2]+2*a2*a4*cov_m[1,3]+2*a3*a4*cov_m[2,3]

logLik(gy_lst2) # valor máximo da função de log-verossimilhança
AIC<--2*logLik(gy_lst2)/length(y)+2*3/length(y) # critério de informação de Akaike
BIC<--2*logLik(gy_lst2)/length(y)+3*log(length(y))/length(y) # critério de informação
bayesiana
HQ<--2*logLik(gy_lst2)/length(y)+2*3*log(log(length(y)))/length(y) # critério de
informação de Hannan-Quinn

# Código para estimação dos parâmetros do modelo proposto (Beta), x0 e variância
de x0_hat #

y<-c()
y0<-c()
x<-c()

gy_logit <- betareg(y ~ x)
v<-coef(gy_logit) # estimativas para os parâmetros do modelo de regressão Beta com
função de ligação logit (valores iniciais para alpha, beta, phi e phi0)

gy_loglog <- betareg(y ~ x,link="loglog")
v<-coef(gy_loglog) # estimativas para os parâmetros do modelo de regressão Beta com
função de ligação log-log (valores iniciais para alpha, beta, phi e phi0)

gy_ls<-lm(y ~ x)
w<-coef(gy_ls) # estimativas para os parâmetros do modelo usual

```

```

x0_hat<-(mean(y0)-w[1])/w[2] #valor inicial para x0

# função de log-verossimilhança
loglik<-function(z) {
alpha<-z[1]
bet<-z[2]
phi<-z[3]
phi0<-z[4]
x0<-z[5]

# inversas da função de ligação logit
mu<-exp(alpha+bet*x)/(1+exp(alpha+bet*x)) # 1o. estágio
mu0<-exp(alpha+bet*x0)/(1+exp(alpha+bet*x0)) # 2o. estágio

# inversas da função de ligação log-log
mu<-(1/(exp(1/exp(alpha+bet*x)))) # 1o. estágio
mu0<-(1/(exp(1/exp(alpha+bet*x0)))) # 2o. estágio

sum(lgamma(phi)-lgamma(phi*mu)-lgamma((1-mu)*phi)+(mu*phi-1)*log(y)
+((1-mu)*phi-1)*log(1-y))+sum(lgamma(phi0)-lgamma(phi0*mu0)-lgamma((1-mu0)*phi0)
+(mu0*phi0-1)*log(y0)+((1-mu0)*phi0-1)*log(1-y0))
}

# funções escore ou gradiente
grad<-function(z) {
alpha<-z[1]
bet<-z[2]
phi<-z[3]
phi0<-z[4]
x0<-z[5]

# inversas da função de ligação logit e suas derivadas

```

```

mu<-exp(alpha+bet*x)/(1+exp(alpha+bet*x)) # 1o. estágio
mu0<-exp(alpha+bet*x0)/(1+exp(alpha+bet*x0)) # 2o. estágio
mualp<-exp(alpha+bet*x)/((1+exp(alpha+bet*x))^2) # 1o. estágio (derivada)
mubet<-x*exp(alpha+bet*x)/((1+exp(alpha+bet*x))^2) # 1o. estágio (derivada)
mu0alp<-exp(alpha+bet*x0)/((1+exp(alpha+bet*x0))^2) # 2o. estágio (derivada)
mu0bet<-x0*exp(alpha+bet*x0)/((1+exp(alpha+bet*x0))^2) # 2o. estágio (derivada)
mu0x0<-bet*exp(alpha+bet*x0)/((1+exp(alpha+bet*x0))^2) # 2o. estágio (derivada)

# inversas da função de ligação log-log e suas derivadas
mu<-(1/(exp(1/exp(alpha+bet*x)))) # 1o. estágio
mu0<-(1/(exp(1/exp(alpha+bet*x0)))) # 2o. estágio
mualp<-(1/exp(1/exp(alpha+bet*x)))*(1/exp(alpha+bet*x)) # 1o. estágio (derivada)
mubet<-(1/exp(1/exp(alpha+bet*x)))*(x/exp(alpha+bet*x)) # 1o. estágio (derivada)
mu0alp<-(1/exp(1/exp(alpha+bet*x0)))*(1/exp(alpha+bet*x0)) # 2o. estágio (derivada)
mu0bet<-(1/exp(1/exp(alpha+bet*x0)))*(x0/exp(alpha+bet*x0)) # 2o. estágio (derivada)
mu0x0<-(1/exp(1/exp(alpha+bet*x0)))*(bet/exp(alpha+bet*x0)) # 2o. estágio (derivada)

c(-sum(phi*(digamma(phi*mu)-digamma((1-mu)*phi)-log(y/(1-y)))*mualp)-
sum(phi0*(digamma(phi0*mu0)-digamma((1-mu0)*phi0)-log(y0/(1-y0)))*mu0alp),
-sum(phi*(digamma(phi*mu)-digamma((1-mu)*phi)-log(y/(1-y)))*mubet)
-sum(phi0*(digamma(phi0*mu0)-digamma((1-mu0)*phi0)-log(y0/(1-y0)))*mu0bet),
sum(digamma(phi)-mu*digamma(phi*mu)-(1-mu)*digamma((1-mu)*phi)
+mu*log(y/(1-y))+log(1-y)),
sum(digamma(phi0)-mu0*digamma(phi0*mu0)-(1-mu0)*digamma((1-mu0)*phi0)
+mu0*log(y0/(1-y0))+log(1-y0)),
-sum(phi0*(digamma(phi0*mu0)-digamma((1-mu0)*phi0)-log(y0/(1-y0)))*mu0x0)
}

#algoritmo L-BFGS-B com limites inferiores para phi, phi0, x0_hat
opt<-optim(c(v[1],v[2],v[3],v[3],x0_hat),loglik,grad,method="L-BFGS-B",
lower=c(-Inf,-Inf,0,0,0),control=list(fnscale=-1),hessian=TRUE)

```

```
opt # valores estimados para alpha, beta, phi, phi0 e x0

H<-opt$hessian # matriz hessiana

H_inv<-solve(-H,LINPACK =TRUE) # inversa da matriz de informação de Fisher
observada

var_x0_hat<-H_inv[5,5] # variância estimada de x0_hat
var_alpha_hat<-H_inv[1,1] # variância estimada de alpha_hat
var_beta_hat<-H_inv[2,2] # variância estimada de beta_hat
var_phi_hat<-H_inv[3,3] # variância estimada de phi_hat
var_phi0_hat<-H_inv[4,4] # variância estimada de phi0_hat

opt$value # valor máximo da função de log-verossimilhança

AIC<--2*opt$value/length(y)+2*5/length(y) # critério de informação de Akaike
BIC<--2*opt$value/length(y)+5*log(length(y))/length(y) # critério de informação
bayesiana
HQ<--2*opt$value/length(y)+2*5*log(log(length(y)))/length(y) # critério de
informação de Hannan-Quinn

***** FIM *****
```