

**USO DO 'BOOTSTRAP' NA ESTIMAÇÃO DE
PARÂMETROS EM MODELOS NÃO-
LINEARES - UMA APLICAÇÃO EM
MECANISMOS CINÉTICOS DE MICHAELIS-
MENTEN**

EUSTÁQUIO JOSÉ MACHADO

2006

EUSTÁQUIO JOSÉ MACHADO

**USO DO 'BOOTSTRAP' NA ESTIMAÇÃO DE PARÂMETROS EM
MODELOS NÃO-LINEARES - UMA APLICAÇÃO EM MECANISMOS
CINÉTICOS DE MICHAELIS-MENTEN**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-graduação em Agronomia, Área de Concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

Orientador
Prof. Dr. Joel Augusto Muniz

LAVRAS
MINAS GERAIS - BRASIL
2006

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Machado, Eustáquio José

Uso do 'Bootstrap' na estimação de parâmetros em modelos não lineares – uma aplicação em mecanismos cinéticos de Michaelis-Menten / Eustáquio José Machado. -- Lavras : UFLA, 2006.

131 p. : il.

Orientador: Joel Augusto Muniz.

Dissertação (Mestrado) - UFLA.

Bibliografia.

1. Regressão não linear. 2. Cinética enzimática. 3. Bootstrap.
4. Michaelis-Menten. I. Universidade Federal de Lavras. II. Título.

CDD - 519.536

EUSTÁQUIO JOSÉ MACHADO

**USO DO 'BOOTSTRAP' NA ESTIMAÇÃO DE PARÂMETROS EM
MODELOS NÃO-LINEARES - UMA APLICAÇÃO EM MECANISMOS
CINÉTICOS DE MICHAELIS-MENTE**

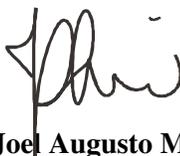
Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-graduação em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de "Mestre".

APROVADA em 22 de fevereiro de 2006

Prof. Dr. Renato Ribeiro de Lima - UFLA

Prof. Dr. Ruben Delly Veiga - UFLA

Prof. Dr. Mario Cesar Guerreiro - UFLA



Prof. Dr. Joel Augusto Muniz
UFLA
(Orientador)

**LAVRAS
MINAS GERAIS – BRASIL**

Aos meus pais João e Maria das Dores (*in memorian*),
pelo amor, bondade e confiança.

OFEREÇO

À minha esposa Francisca, pelo apoio,
confiança, paciência e amor que me dedica;
Ao meu filho Max, esperança de continuidade;
A minha irmã Dalva pela convivência, carinho e dedicação.

DEDICO

AGRADECIMENTOS

Ao professor Joel Augusto Muniz, pela paciência, apoio, estímulo, confiança e, sobretudo por ter me dado a honra de me orientar e acreditar no meu trabalho.

À Universidade Federal do Acre pela oportunidade que viabilizaram a realização desse estudo.

Ao Departamento de Matemática e Estatística da UFAC por terem confiado no meu aprendizado.

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Agronomia/Estatística e Experimentação Agropecuária.

Aos técnicos-administrativos do DEX-UFLA, pela amizade e companheirismo.

A todos os professores da Pós-Graduação em Agronomia/Estatística e Experimentação Agropecuária da UFLA.

A Pró-Reitoria de pós-graduação da UFAC, em especial ao professor Adailton e Coutinho pelo apoio e assistência.

Ao meu grande amigo Manoel, professor do Departamento de Matemática e Estatística da UFAC, pela oportunidade, incentivo e confiança, que mesmo longe, com certeza torce pelo meu sucesso.

As ex-alunas e amigas Clênia e Gabriela, pela amizade, pelos dias felizes e por tornarem a minha vida profissional um eterno prazer.

Ao meu amigo Michel meu ex-aluno e orientando pela colaboração, estímulo e dedicação durante todo o processo de seleção do mestrado “meu anjo da guarda”.

Ao meu amigo Diego ex-aluno e orientando, mestrando em ecologia da UFAC, pelo auxílio e contribuição no projeto de estudo em cinética enzimática e durante o desenvolvimento desse trabalho, grande guerreiro (Shogum).

Aos meus grandes amigos do mestrado, Charles, Claudiney, Elias, Josiane, Lívia, Nádia, Paulinho, Rejane e Vanessa, pela grande amizade, paciência e troca de conhecimentos.

À Verônica (também da minha turma), tenho que dedicar um parágrafo especial, por ser uma guerreira. Agradeço imensamente pelo agradável convívio e apoio nas horas difíceis, pelo seu espírito de solidariedade e amizade, durante todo o curso.

A CAPES pela concessão da bolsa de estudos.

A Deus, por ter me concedido a graça da vida!

SUMÁRIO

RESUMO	i
ABSTRACT	ii
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	4
2.1 Enzimas.....	4
2.1.1 Introdução às enzimas.....	4
2.1.2 Histórico.....	5
2.1.3 Cinética enzimática.....	9
2.2 Mecanismos cinéticos de Michaelis-Menten	13
2.2.1 Modelo matemático	13
2.2.2 Equação diferencial.....	21
2.3 Modelos de regressão.....	22
2.3.1 Considerações iniciais.....	22
2.3.2 Modelos de regressão linear	24
2.3.2.1 Conceito.....	24
2.3.2.2 Modelos de regressão linear múltipla (MRLM)	25
2.3.2.3 Modelo de regressão linear simples (MRLS)	26
2.3.2.4 O MRLS em forma matricial	27
2.3.2.5 Método de quadrados mínimos (MRLS)	29
2.3.2.6 Estimativas de variâncias e covariâncias (MRLS).....	32
2.3.2.7 Intervalos de confiança (MRLS).....	34
2.3.2.8 Diagnóstico das pressuposições básicas	35
2.3.3 Modelos de regressão não-linear	38
2.3.3.1 Conceito.....	38
2.3.3.2 Transformações.....	40
2.3.3.2.1 Lineweaver-Burk	41
2.3.3.2.2 Hanes-Wolf	43
2.3.3.2.3 Wolf-Augustinsson-Hofstee	45
2.3.3.2.4 Eadie-Scatchard	47
2.3.3.3 Aproximação linear pela série de Taylor	49
2.3.3.4 Método de quadrados mínimos ordinários.....	52
2.3.3.5 Propriedades assintóticas dos estimadores de mínimos quadrados.....	55
2.3.3.6 Processos numéricos iterativos	56
2.3.3.6.1 Generalidades.....	56
2.3.3.6.2 Método de Gauss Newton.....	58

2.4 Métodos empíricos: o ‘bootstrap’ na regressão	60
2.4.1 Introdução	60
2.4.2 O método ‘bootstrap’	62
2.4.3 O ‘bootstrap’ na regressão	63
2.5 Comparação dos modelos de regressão	66
2.5.1 Amplitude de um intervalo de confiança	66
2.5.2 Quadrado médio do erro (QME).....	67
2.5.3 Coeficiente de determinação (R^2)	69
2.5.4 Coeficiente de determinação ajustado ($R^2_{ajustado}$).....	71
3 MATERIAL E MÉTODOS	72
3.1 Material	72
3.2 Métodos	73
3.2.1 Derivadas parciais, matriz X e $X'X$ do modelo não-linear de Michaelis-Menten.....	73
3.2.2 Ajuste dos modelos não-lineares e lineares	74
3.2.3 Comparação dos modelos lineares e não-lineares.....	75
3.2.4 ‘Bootstrap’ não paramétrico na regressão.....	76
3.2.4.1 Modelo de Michaelis-Menten genérico ajustado.....	76
3.2.4.2 Detalhes do algoritmo de ‘bootstrap’ nos resíduos.....	77
4 RESULTADOS E DISCUSSÃO.....	79
4.1 Ajustes dos modelos não-lineares.....	79
4.1.1 Ajuste do modelo de Michaelis-Menten via aproximação assintótica.....	79
4.1.2 Ajuste do modelo de Michaelis-Menten via técnica de ‘bootstrap’	86
4.2 Ajuste dos modelos linearizados.....	91
4.2.1 Lineweaver-Burk	92
4.2.2 Hanes-Wolf.....	97
4.2.3 Woolf-Augustinson-Hofstee	100
4.2.4 Eadie-Scatchard	104
4.3 Comparação entre os modelos ajustados	107
5 CONCLUSÕES.....	111
REFERÊNCIAS BIBLIOGRÁFICAS.....	113
ANEXOS.....	117

RESUMO

MACHADO, Eustáquio José. **Uso do ‘Bootstrap’ na estimação de parâmetros em modelos não-lineares** – uma aplicação em mecanismos cinéticos de Michaelis-Menten. 2006. 131 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.¹

A equação hiperbólica, conhecida no contexto bioquímico como o modelo de Michaelis-Menten, é utilizada para descrever a velocidade de reações químicas envolvendo enzimas (cinética enzimática). Este estudo teve como objetivo comparar os ajustes do modelo de Michaelis-Menten (1913) que fez uso de dois modelos não-lineares e quatro modelos linearizados. Os dois modelos não-lineares (um utilizou o método clássico assintótico usual e o outro fez uso da abordagem “bootstrap”). Os modelos linearizados foram obtidos por meio de transformações do modelo não-linear que utilizou o método clássico assintótico usual. As estimações dos parâmetros para os modelos foram realizadas pelo método de mínimos quadrados ordinários utilizando rotinas apropriadas dos softwares estatísticos R e SAS. Para o diagnóstico da pressuposição de homocedasticidade dos resíduos foi utilizada a análise residual, e para a verificação da pressuposição da normalidade dos resíduos, foram utilizados o teste de probabilidade normal e o teste de Shapiro-Wilk. Para a comparação entre os modelos foram utilizados as amplitudes dos intervalos de confiança dos parâmetros, o quadrado médio dos resíduos (QME), o coeficiente de determinação ajustado (R^2_{ajustado}), indicando o modelo que melhor ajustou-se aos dados experimentais. Destaca-se que o modelo não-linear de Michaelis-Menten ajustado pela metodologia de “bootstrap” se mostrou adequada em estudos para amostras pequenas. O estudo mostra, em situações críticas, onde duas ou mais enzimas têm os K_m bastantes próximos em que a precisão da estimativa dos parâmetros pode ser um indicador importante nos mecanismos cinéticos enzimáticos.

¹ **Comitê orientador:** Joel Augusto Muniz – UFLA (Orientador).

ABSTRACT

MACHADO, Eustáquio José. **The use of ‘Bootstrap’ in parameter estimation for non-linear models** – Application in kinetic mechanisms of Michaelis-Menten. 2006. 131 p. Dissertation (Master in Agronomy/Statistics and Agricultural Experimentation) - Federal University of Lavras, Lavras, Minas Gerais, Brazil.¹

The hyperbolic equation, known as Michaelis-Menten’s model in the biochemical context, is used to describe the speed of the chemical reactions involving enzyme (enzymatic kinetic). The goal of this study was to compare the adjustments among Michaelis-Menten’s two non-linear models and four linear models (1913). For one of the two non-linear models it was applied the usual asymptotic classical method and for the other the bootstrap approach. The linear models were obtained by the transformations of the non-linear model, which used the usual asymptotic classical method. The parameter estimations for the methods were accomplished through the least squares method using suitable routines of R and SAS statistical software. For the diagnosis of homogeneity presupposition from the residues, it was used the residual graphic analyses, and for the verification of normality presupposition from the residues, it was used a normal probability graphic and Shapiro-Wilk test. For the comparison it was used the confidence interval extent of the parameters, the residue medium square (RMS), the adjusted determination coefficient (R^2_{adjusted}), pointing to the best model fitted to the experimental data. Michaelis-Menten’s non-linear model adjusted by the bootstrap methodology, proved to be suitable in studies for small samples. The study shows, in critical situations, where two or more enzymes have their Km close-set, pointing to the accurate estimate of the parameters as an important indicative for the enzymatic kinetic mechanisms.

¹ **Guidance Committee:** Joel Augusto Muniz – UFLA (Adviser).

1 INTRODUÇÃO

Como relata Cornish-Bowden (1995b), as enzimas são moléculas de proteína bastante grandes e complexas que agem como catalisadoras em reações bioquímicas e são fundamentais para qualquer processo bioquímico. Agindo em seqüências organizadas, elas catalisam as centenas de reações sucessivas pelas quais as moléculas e nutrientes são degradadas; nessas reações, a energia química é conservada e transformada e as macromoléculas são sintetizadas a partir de moléculas precursoras simples.

O complexo enzima-substrato é fundamental para a ação das enzimas, em contrapartida, sem catálise, as reações necessárias para digerir os alimentos, enviar sinais através dos nervos ou contrair um músculo simplesmente não ocorrem com uma velocidade útil. Ou seja, a regulação enzimática possui um papel chave na homeostase, que é a manutenção, no dia-a-dia, das atividades metabólicas balanceadas e produtivas, assim como na determinação da duração e da natureza dos processos de longo prazo, tais como a divisão e a diferenciação celular, que afetam o desenvolvimento e o crescimento de um organismo.

Múltiplas abordagens são comumente empregadas para estudar os mecanismos de ação de enzimas purificadas, entretanto, a velocidade da reação catalisada pode também revelar muito a respeito da enzima. O estudo das velocidades de reação e como elas mudam em resposta a mudanças nos parâmetros experimentais é conhecido como cinética enzimática.

A característica distintiva de uma reação catalisada por enzima é que ela ocorre no interior de uma cavidade chamada sítio ativo. Este sítio ativo é o ponto de ligação entre a enzima e o substrato, conforme Figura 1.

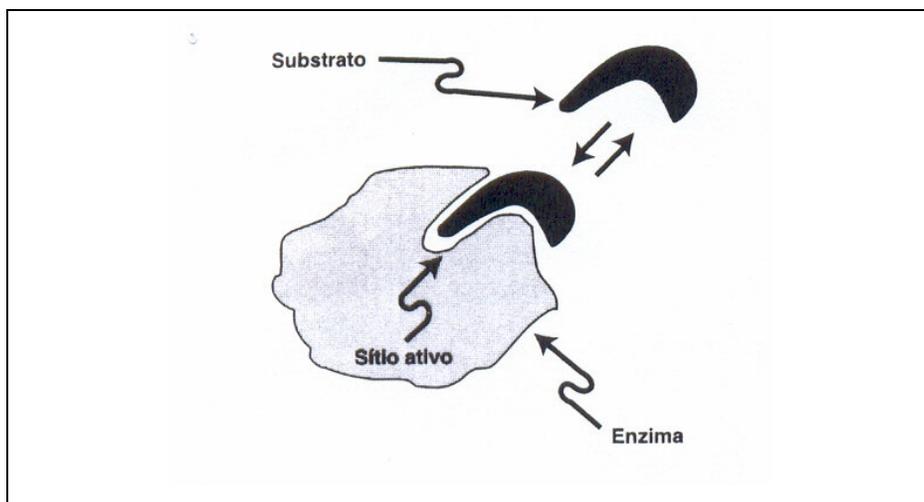


FIGURA 1 Enzima-substrato

O complexo enzima-substrato tem papel central na reação enzimática. Ele é o ponto de partida para os tratamentos matemáticos que definem o comportamento cinético das reações catalisadas enzimaticamente e para as descrições teóricas dos mecanismos enzimáticos.

O modelo teórico de Michaelis-Menten (1913) atravessou os limites da aplicação bioquímica e alcançou as áreas agronômicas a partir do estudo da dinâmica existente na relação entre a fertilidade do solo e produtividade de plantas. O agrônomo, neste caso, avalia entre a velocidade com que a planta absorve o nutriente e a disponibilidade deste elemento em solução. A velocidade máxima aqui é tida como influxo máximo e a constante de Michaelis-Menten como a concentração mínima.

Contudo, com pouco mais de 90 anos, a equação de Michaelis-Menten parece mesmo ter sua maior consolidação atuando como ponto de partida para tratamentos matemáticos que definem o comportamento cinético das reações

catalisadas por enzimas, bem como para descrições teóricas dos mecanismos enzimáticos.

Este estudo teve como objetivos comparar o ajuste de seis modelos estatísticos relacionados ao modelo de Michaelis-Menten (1913) e indicar o que melhor ajusta-se aos dados experimentais obtidos de Treolar (1974), utilizando para a comparação dos modelos estatísticos: as amplitudes dos intervalos de confiança dos parâmetros V_{\max} e K_m , o quadrado médio dos erros (QME) e o coeficiente de determinação ajustado (R^2_{ajustado}) e destacar a importância da técnica “bootstrap” como alternativa de ajuste para o modelo no qual a amostra é de tamanho pequeno.

2 REFERENCIAL TEÓRICO

2.1 Enzimas

2.1.1 Introdução às enzimas

Segundo Segel (1975), enzimas são proteínas altamente especializadas; são as catalisadoras das reações que ocorrem nos sistemas biológicos. Com exceção de um pequeno grupo de moléculas de RNA com propriedades catalíticas, todas as enzimas são proteínas.

As enzimas têm eficiência catalítica extraordinária, em geral muito maior que aquela dos catalisadores sintéticos. Elas têm um alto grau de especificidade por seus substratos, aceleram reações químicas específicas e, ainda, funcionam em soluções aquosas e em condições muito suaves de temperatura e pH. Devido aos altos graus de especificidade, é fator responsável pela rede ordenada de reação químicas que ocorrem nas células vivas e cuja soma constitui o metabolismo. A especificidade possibilita, ainda, a regulação das várias reações catalisadas enzimaticamente, alterando as atividades catalíticas por mecanismos especiais que possibilitam o controle do metabolismo pela célula. Poucos catalisadores não-biológicos mostram todas estas propriedades.

O estudo das enzimas tem imensa importância prática. Em algumas doenças, especialmente nas desordens genéticas herdadas, pode ocorrer, nos tecidos, a deficiência ou mesmo, a ausência total de uma ou mais enzimas. Em condições normais, doenças também podem ser causadas pelo excesso de atividade de uma enzima específica.

2.1.2 Histórico

De acordo com Cornish-Bowden (1995a), os estudos sobre catálise biológica iniciaram-se nos primeiros anos do século XIX, envolvendo a digestão da carne por secreções do estômago e a conversão do amido em açúcares simples pela saliva e por vários extratos vegetais.

Mahler & Cordes (1966) fazem um breve histórico sobre os estudos de enzimas. Segundo eles, a palavra enzima foi introduzida por Kuhne, em 1878, para designar a ocorrência, no levedo, de algo responsável pela atividade fermentativa. Berzelius, em 1828 (citado por MAHLER & CORDES, 1966), já tinha reconhecido a presença de fermentos de ocorrência natural que promoviam reações químicas e antecipou o conceito de catalisadores biológicos. Berzelius classificou os fermentos em “organizados” e “desorganizados”, com base na presença ou na ausência de microrganismos intactos. Kuhne aplicou a palavra enzima aos fermentos derivados de extratos de levedo, isto é, “fermentos não-organizados”.

Em 1850, Louis Pasteur (citado por MAHLER & CORDES, 1966), concluiu que a fermentação do açúcar em álcool pela levedura é catalisada por “fermentos”, e postulou que esses fermentos eram inseparáveis das estruturas das células vivas do levedo, o que ficou, por muitos anos, como uma hipótese verdadeira.

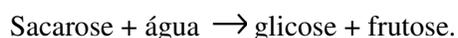
Em 1897, Eduard Buchner (citado por MAHLER & CORDES, 1966), provou que as enzimas envolvidas na fermentação continuavam fermentando mesmo quando removidas da estrutura das células vivas.

Em 1926, Sumner (citado por MAHLER & CORDES, 1966), foi o primeiro bioquímico a isolar e cristalizar a uréase e demonstrou que os seus cristais consistiam inteiramente de proteína, postulando que todas as enzimas eram proteínas. A princípio essa idéia não foi muito bem aceita, pois, não havia

outros exemplos; contudo, na década de 1930, John Howard Northrop e Meredith Stanley (citado por MARLER & CORDES 1966), cristalizaram pepsina e tripsina bovinas e concluíram que estas também eram proteínas. Isso fez com que a conclusão de Sumner (citado por MARLER & CORDES 1966), fosse aceita e lhe rendeu o Prêmio Nobel de Química, juntamente com Northrop e Stanley, em 1946.

O conceito de velocidade das reações de catálise-enzima foi primeiramente estudado na última metade do século XIX (citado por MARLER & CORDES 1966). Nesse período, o estudo com enzimas puras não eram eficazes, pois, os métodos de análises eram primitivos e o uso de agentes para o controle de pH não eram bem entendidos. Além disso, era costume seguir as reações um certo espaço de tempo em contraste com as práticas usuais modernas de mensuração inicial da velocidade para concentração de vários substratos diferentes, para os quais davam resultados que são mais facilmente interpretados.

Os primeiros estudos estavam preocupados com a fermentação de enzimas, particularmente a invertase, que catalisa a sacarose por meio da hidrólise, ou seja,



De acordo com Cornish-Bowden (1995a), O'Sullivan & Tompson estudaram esta reação e fizeram uma série de grandes achados: descobriram que a reação foi mais dependente sobre a acidez e determinaram que a acidez demonstrou proporção mais favorável, e a velocidade foi proporcional à soma de enzimas. Em baixas temperaturas, a enzima mostrou uma duplicação aproximadamente dobrada de velocidade com um aumento de temperatura de 10° C, entretanto, para a maioria das reações químicas ordinárias foi diferente. A

reação catálise-invertase mostrou uma aparente temperatura ótima, sobre a qual a velocidade derrubou rapidamente a velocidade zero.

A invertase provou ser o catalisador mais adequado, pois não foi destruída nem alterada na reação (exceto em altas temperaturas). Um exemplo foi que, após isso, a invertase mostrou-se ativa, mesmo depois de catalisar a hidrólise em mais de 100.000 vezes a massa de sacarose.

Finalmente, a estabilidade termal da enzima foi muito maior na presença do substrato do que na sua ausência.

Em 1880, Wurtz (citado por MAHLER & CORDES, 1966), tinha alcançado resultados similares anteriormente, quando estava estudando a papaína-catalisada e a hidrólise de fibrina. Ele observou um precipitado, que sugeriu ser, provavelmente, um composto papaína-fibrina que atuou como um intermediário na hidrólise.

Os conflitos entre os resultados de Brown e aqueles de O'Sullivan & Thompson não foram os primeiros observados sobre este importante assunto, porque catálise de enzimas isoladas foi observada como fundamentalmente diferente para a sobrevivência de organismos. Porém, Buchner's, em 1897 (citado por MARLER & CORDES 1966), descobriu que a célula livre (por exemplo, sem vida) pode extrair levedura da catálise de fermentação alcoólica, como proposto por Brown, em 1902, ao reexaminar seus primeiros resultados. Depois, confirmando que estavam corretos, mostrou que resultados similares puderam ser obtidos com invertases purificadas. Sugeriu, então, que o mecanismo complexo substrato-enzima mostrou um limite sobre a velocidade que pode ser alcançada.

Provado que um complexo existente por um curto espaço de tempo antes da quebra do produto, e depois que a velocidade máxima tiver sido alcançada quando a concentração for bem maior ao converter todas as enzimas dentro do

complexo, de acordo com a lei de ação das massas. As concentrações de substrato mais baixas, a velocidade com que o complexo é formado quando se deseja uma transformação significativa e também a velocidade da hidrólise serão dependentes da concentração de substrato.

Em 1902 e 1903 Henri (citado por MARLER & CORDES 1966), criticou o modelo de Brown de ação de enzima sobre detalhes predeterminados que ele assumiu por toda a sua vida para o complexo enzima-substrato entre a criação abrupta e decadente. Ele propôs um mecanismo que foi conceitualmente similar ao de Brown, mas, este foi expresso em termos matemático e químico mais preciso, de um equilíbrio entre a enzima livre e os complexos enzima-substrato e enzima-produto.

Embora Brown & Henri tenham alcançado, essencialmente, conclusões corretas, elas funcionaram também sobre bases experimentais que estavam sujeitas a sérias objeções. O'Sullivan & Tompson experimentaram grande dificuldade em alcançar resultados coerentes, até que entenderam a importância da concentração do ácido. Brown preparou a enzima em diferentes formas e entendeu ser desnecessária a adição de ácido (presumivelmente, suas soluções foram suficientemente defendidas por componentes naturais de levedura). Henri não discutiu o problema. O'Sullivan & Tompson, os primeiros investigadores da invertase, não fizeram a concessão para a mutarrotação (mudança do sentido do desvio da luz polarizada) da produção da glicose na reação, embora esta, certamente, tenha afetado o seu resultados porque usaram métodos polarimétricos.

Com a introdução do conceito de concentração de íon hidrogênio $[H^+]$, expresso pela escala logarítmica de pH por Sorensen, em 1909 (citado por MARLER & CORDES 1966), Michaelis e Menten (1913) realizaram experimentos com a invertase. Eles controlaram o pH da reação adicionando

acetato, permitindo a mutarrotação do produto e mediar a velocidade inicial da reação sob diferentes concentrações de substrato.

Quando a velocidade inicial foi usada, fatores de complicação, assim como reações reversas produziram inibição e inativação das enzimas, talvez prevenindo equações simples provavelmente não usadas. Ao contrário desse refinamento, Michaelis e Menten, com resultados bem ajustados com os de Henri, propuseram um mecanismo essencialmente semelhante.

Henri 1902-1903 (citado por MAHLER & CORDES, 1966), assumiu que o primeiro passo reversível foi bastante rápido para ser representado por uma equação de equilíbrio constante para dissociação de substrato. As concentrações instantâneas de enzima e substrato livre não são diretamente mensuráveis, entretanto, elas devem ser expressas em termos de marcas iniciais.

Michaelis-Menten (1913) descobriram que essa teoria pode ser considerada em seus resultados com invertase, definido na natureza de seus experimentos, os quais têm servido como um padrão para a maioria das medidas subsequentes da cinética enzimática. Estes autores são respeitados como fundadores da moderna enzimologia e a equação desenvolvida por eles é conhecida como a equação de Michaelis-Menten, apesar de uma equação equivalente ter sido derivada primeiramente por Henri (1902 -1903).

2.1.3 Cinética enzimática

Para Lenninger (2000), uma discussão sobre cinética enzimática básica deve começar pela definição e estudo de alguns conceitos fundamentais. Um dos fatores principais que afetam a velocidade de uma reação catalisada por uma enzima purificada *in vitro* é a concentração do substrato presente, representada por [S]. Estudar os efeitos da concentração de substrato é difícil pelo fato de a

[S] variar durante o curso de uma dada reação, à medida que o substrato é convertido em produto.

Uma maneira de simplificar é medir a velocidade inicial da reação, designada por v_0 , que é definida como a quantidade de moléculas de produto formadas por unidade de tempo, a uma concentração fixa de enzima. Em um experimento cinético, a [S] é, sempre, muito maior que a concentração da enzima, denotada por [E]. Portanto, se o tempo de reação é suficientemente curto, as mudanças na [S] serão negligenciáveis a [S] pode ser tomada como uma constante.

O efeito provocado em V_0 pela variação da [S], quando a concentração de enzima é mantida constante, está mostrado na Figura 2.

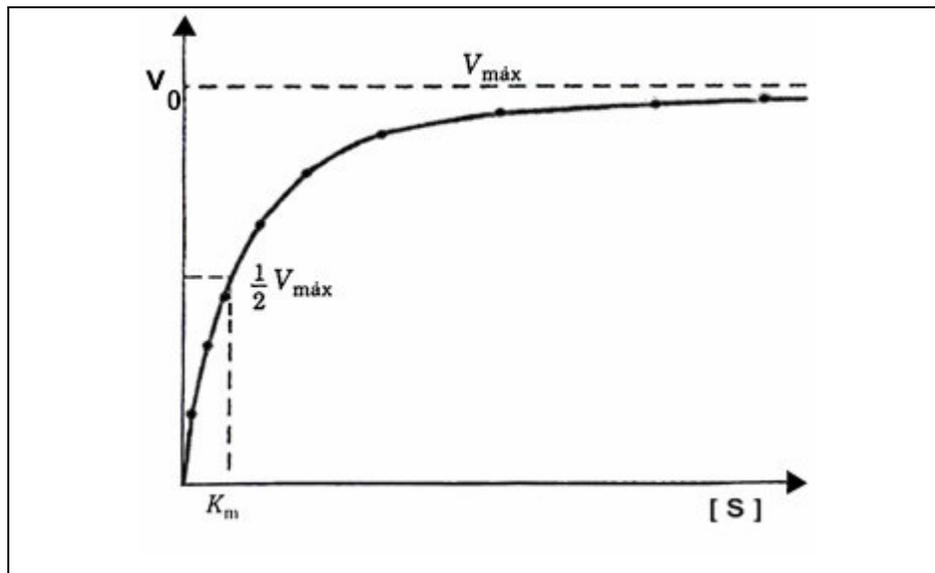


FIGURA 2 Efeito da concentração de substrato na velocidade inicial de uma reação catalisada enzimaticamente.

Por este tipo de gráfico, a velocidade máxima, representada por V_{\max} , pode ser determinada apenas aproximadamente, isso porque V_0 aproxima-se progressivamente de V_{\max} mas nunca chegará a atingi-la.

A concentração de substrato na qual V_0 é metade da V_{\max} é o K_m , a constante de Michaelis e Menten.

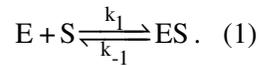
A concentração de enzima $[E]$, em um experimento como este, é tão baixa que $[S] \gg [E]$, ou seja, a concentração de substrato é muito, mais muito maior do que a concentração da enzima, mesmo quando a $[S]$ é descrita como relativamente pequena ou, mesmo, muito pequena.

A curva descreve parte de uma hipérbole retangular, com uma das assíntotas em V_{\max} . Se a curva continuasse abaixo da $[S]=0$, ela se aproximaria da assíntota vertical quando $[S] = -K_m$.

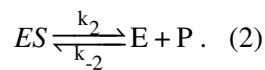
O complexo enzima-substrato, denotado por ES, é a chave para a compreensão deste comportamento cinético e representou o ponto de partida para a discussão da catálise enzimática.

O perfil cinético apresentado na Figura 2 levou Victor Henri (1903), (citado por MARLER & CORDES 1966), a propor que uma enzima liga-se à molécula de seu substrato para formar o complexo ES, sendo este um passo obrigatório no processo catalítico enzimático.

Este conceito foi expandido em uma teoria geral da ação das enzimas, especialmente por Leonor Michaelis e Maud Menten, em 1913. Eles propuseram que a enzima primeiro combina-se reversivelmente com o substrato, para formar o complexo enzima-substrato, em um passo reversível relativamente rápido.



Em um passo mais lento, o complexo ES se rompe com o reaparecimento da enzima livre e a formação do produto da reação P:



A segunda reação (equação 2) é mais lenta neste modelo e, portanto, limita a velocidade da transformação global do reagente em produto. Isto resulta que a velocidade da reação enzimática deve ser proporcional à concentração da substância reagente, isto é, ES.

Em qualquer instante de uma reação catalisada enzimaticamente, a enzima existe em duas formas, a não ligada ao substrato, ou forma livre E, e a forma ligada ao substrato ES. Em concentrações pequenas do substrato S, a maior quantidade da enzima estará na forma livre E. Nessas condições, a velocidade da reação será proporcional à [S] porque o equilíbrio na equação (1), à medida que a [S] aumentar, será deslocado na direção da formação de mais ES. A velocidade inicial máxima da reação (V_{\max}) será atingida quando, num dado instante, praticamente todas as moléculas da enzima estiverem na forma do complexo ES e a concentração da enzima livre E for insignificante.

Nestas condições, diz-se que a enzima está “saturada” com o substrato e a velocidade da reação não aumenta mais com ulteriores aumentos de [S]. Este efeito de saturação é uma característica distintiva dos catalisadores enzimáticos e é o responsável pelo patamar observado na Figura 2. Esta condição de saturação existirá sempre que [S] for suficientemente alta para manter todas as moléculas

de enzima na forma combinada com substrato ES. Depois que o complexo ES se resolve no produto da reação P e na enzima livre, esta enzima está apta a participar novamente da reação.

Quando a enzima é colocada em uma solução contendo um grande excesso de substrato, ocorre um período cinético inicial chamado de estado pré-estacionário durante o qual ocorre o crescimento da concentração de ES. O estado pré-estacionário é, usualmente, de duração muito curta e não pode ser observado facilmente. A reação atinge o estado estacionário muito rapidamente e, neste estado, a concentração do complexo, [ES], permanecerá praticamente constante durante o decorrer do tempo.

A velocidade inicial (V_0) medida, em geral, reflete o estado estacionário, ainda que V_0 esteja limitada aos tempos iniciais da reação. Em seus estudos, Michaelis e Menten ocuparam-se com a velocidade da reação enzimática durante o estado estacionário e o tipo de análise que desenvolveram é chamado de cinética do estado estacionário.

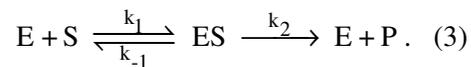
2.2 Mecanismos cinéticos de Michaelis-Menten

2.2.1 Modelo matemático

A relação entre a [S] e a V_0 para uma reação enzimática encontra-se na Figura 2. A curva que expressa esta relação tem a mesma forma geral para a maioria das enzimas (ela se aproxima de uma hipérbole retangular). A forma hiperbólica desta curva pode ser expressa algebricamente pela equação de Michaelis-Menten, partindo de sua hipótese básica de que, nas reações enzimáticas, o passo limitante da velocidade é a quebra do complexo ES para formar o produto e a enzima livre.

Os termos importantes são $[S]$, v_0 , V_{\max} e uma constante chamada de constante de Michaelis-Menten, ou K_m . Todos estes termos são facilmente medidos experimentalmente.

De acordo com Leninger (2000), a lógica básica e os passos algébricos de uma dedução moderna da equação de Michaelis-Menten, uma característica crítica do seu tratamento, é de que um complexo ES específico é um intermediário necessário na catálise. O modelo proposto, que é o mais simples, explica as propriedades cinéticas de muitas enzimas:



Uma enzima E se combina com uma molécula de S para formar um complexo ES, com uma constante de velocidade k_1 . O complexo ES tem dois destinos possíveis. Ele pode se dissociar a E e S, com uma constante de velocidade k_{-1} ou pode prosseguir para formar o produto P, com uma constante de velocidade k_2 . Supõe-se que quase nada do produto reverte ao substrato inicial, uma condição que se mantém na etapa inicial da reação, antes de a concentração do produto ser apreciável.

É desejável ter uma expressão que relacione a velocidade da reação catalítica com as concentrações de substrato e de enzima e com as velocidades das etapas elementares. O ponto de partida é que a velocidade da reação catalítica é igual ao produto da concentração do complexo ES por k_2 .

$$v = k_2[ES]. \quad (4)$$

É interessante expressar [ES] em termos de quantidades conhecidas. As taxas de formação e quebra de ES são dadas por:

$$\text{Velocidade de formação de ES} = k_1[E][S]. \quad (5)$$

$$\text{Velocidade de consumo de ES} = (k_{-1} + k_2)[ES]. \quad (6)$$

É importante verificar a velocidade da reação catalítica sob condições de estado estacionário (*steady state*). Em um estado estacionário, as concentrações dos intermediários permanecem constantes, pois as velocidades de formação e de consumo do complexo ES são iguais. Ressalta-se que o modelo proposto não apresenta uma solução algébrica trivial, sendo somente possível uma solução iterativa para o modelo. Porém, pode-se supor que a concentração do complexo [ES], após algum tempo permanece constante, aproximando-se de um estado estacionário. Com essa aproximação, as velocidades de formação e consumo do complexo ES devem ser iguais, assim:

$$\frac{d[ES]}{dt} = 0. \quad (7)$$

ou seja, velocidade de formação menos velocidade de consumo é,

$$k_1[E][S] - (k_{-1} + k_2)[ES] = 0. \quad (8)$$

portanto

$$k_1[E][S] = (k_{-1} + k_2)[ES]. \quad (9)$$

reorganizando a equação (9)

$$[ES] = \frac{[E][S]}{(k_{-1} + k_2)/k_1}. \quad (10)$$

A equação acima pode ser simplificada, definindo-se K_m , tal que

$$K_m = \frac{k_{-1} + k_2}{k_1}. \quad (11)$$

substituindo na equação (10) que, então, se torna,

$$[ES] = \frac{[E][S]}{K_m}. \quad (12)$$

A partir da equação (9) pode-se definir a concentração de substrato não combinado [S] como sendo igual à concentração do substrato total. A concentração de enzima não combinada [E], igual à concentração de enzima total $[E_T]$ menos a concentração do complexo ES, ou seja,

$$[E] = [E_T] - [ES]. \quad (13)$$

Aplicando-se esta expressão para [E] na equação (9)

obtem-se:

$$[ES] = ([E_T] - [ES])[S] / K_m \quad (14)$$

resolvendo-se a equação (10) para [ES], tem-se:

$$[ES] = [E_T] \frac{[S] / K_m}{1 + [S] / K_m} \quad (15)$$

ou seja:

$$[ES] = [E_T] \frac{[S]}{[S] + K_m} \quad (16)$$

substituindo-se [ES] na equação (4), tem-se:

$$[ES] = k_3 [E_T] \frac{[S]}{[S] + K_m} \quad (17)$$

A velocidade máxima V_{\max} é obtida quando os centros catalíticos na enzima estão saturados com substrato, ou seja, quando a [S] é muito maior do que K_m , de tal forma que $\frac{[S]}{[S] + K_m}$ se aproxima de 1. Assim,

$$V_{\max} = k_3[E_T]. \quad (18)$$

Para se obter a equação de Michaelis-Menten aplica-se a equação (18) na equação (17), encontrando-se:

$$v_0 = \frac{V_{\max}[S]}{[S] + K_m}. \quad (19)$$

é a equação da velocidade para uma reação catalisada enzimaticamente e com um único substrato. É uma expressão da reação quantitativa entre a velocidade inicial v_0 , a velocidade máxima V_{\max} , e a concentração inicial de substrato $[S]$, todas relacionadas por meio da constante de Michaelis-Menten K_m .

O K_m pode variar muito de enzima e mesmo para diferentes substratos de uma mesma enzima. O valor de K_m é, muitas vezes, empregado (inapropriadamente) como uma indicação da afinidade da enzima para seu substrato.

O significado real de K_m depende de aspectos específicos do mecanismo de reação, como o número e as velocidades relativas dos passos individuais da reação. Consideram-se, a seguir, as reações de dois passos. O K_m está definido pela equação (11), ou seja,

$$K_m = \frac{k_2 + k_{-1}}{k_1}.$$

Para a reação de Michaelis-Menten, k_2 é o limitante da velocidade; assim, quando $k_2 \ll k_{-1}$, K_m se reduz ao valor $\frac{k_{-1}}{k_1}$, o qual define a constante de dissociação, K_s , para o complexo ES. Onde essas condições prevalecem, o K_m representa uma medida da afinidade da enzima pelo substrato no complexo ES.

A equação (19) está de acordo com os fatos experimentais, confirmados pela consideração das situações limite, em que $[S]$ é ou muito mais alto ou muito baixo, como mostrado na Figura 3.

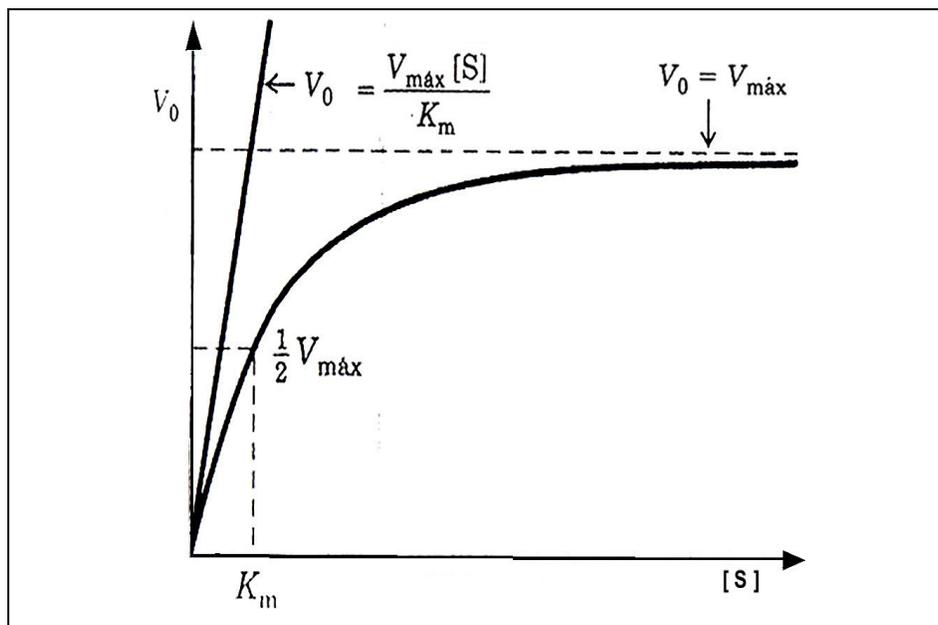


FIGURA 3 Dependência da velocidade inicial da reação da concentração de substrato.

Os parâmetros cinéticos que definem os limites de curvas em concentração de substrato grandes e pequenas, encontram-se na Figura 3.

Em [S] pequeno, $K_m \gg [S]$ e o termo [S], no denominador da equação de Michaelis-Menten (equação 19), torna-se insignificante; a equação pode, então, ser simplificada para

$$v_0 = \frac{V_{\max} [S]}{K_m + [S]} \xrightarrow{K_m \gg [S]} v_0 = \frac{V_{\max} [S]}{K_m}. \quad (20)$$

ou seja, a velocidade é diretamente proporcional à concentração do substrato e V_0 passa a ter uma dependência linear de [S], como observado na Figura 3.

Em [S] grande, quando $[S] \gg K_m$, o termo K_m , no denominador da equação de Michaelis-Menten, torna-se insignificante e a equação pode, assim, ser simplificada para

$$v_0 = \frac{V_{\max} [S]}{K_m + [S]} \xrightarrow{[S] \gg K_m} \frac{V_{\max} [S]}{[S]}. \quad \text{ou } v_0 = V_{\max}. \quad (21)$$

isto é, consistente como o patamar observado em [S] grandes.

A equação de Michaelis-Menten é, portanto, consistente com a dependência observada de V_0 em relação à [S], com o formato da curva definido pelos termos $\frac{V_{\max}}{K_m}$, quando [S] é pequena e apenas por V_{\max} , quando [S] é grande, o significado de K_m é evidente a partir da equação 19.

Quando $[S] = K_m$, então:

$$v_0 = \frac{V_{\max}}{2}. \quad (22)$$

assim, K_m é igual à concentração do substrato na qual a velocidade de reação é metade da sua velocidade máxima.

2.2.2 Equação diferencial

Na visão de Segel (1975), na forma diferencial, a equação da lei de velocidade assume a seguinte forma:

$$v_0 = -\frac{d[S]}{dt} = \frac{V_{\max}[S]}{K_m + [S]}. \quad (23)$$

reorganizando e invertendo:

$$V_{\max} dt = -\frac{K_m + [S]}{[S]} d[S]. \quad (24)$$

Integrando entre dois tempos, ou seja, entre tempo zero, t_0 , e outro tempo qualquer, t , entre a correspondente concentração dos dois substratos ($[S]_0$ e $[S]$), tem-se:

$$V_{\max} \int_{t_0}^t dt = -\int_{[S]_0}^{[S]} \frac{K_m + [S]}{[S]} d[S]. \quad (25)$$

Separando-se os termos no lado direito da expressão:

$$V_{\max} \int_{t_0}^t dt = -K_m \int_{[S]_0}^{[S]} \frac{d[S]}{[S]} - \int_{[S]_0}^{[S]} d[S]. \quad (26)$$

$$V_{\max} t = -K_m \ln \frac{[S]}{[S]_0} - ([S] - [S]_0). \quad (27)$$

Em que $([S] - [S]_0)$ é a concentração de substrato utilizado até o tempo t .

A equação não apresenta solução trivial, por outro lado, se $[S]_0$ e $[S]$ para qualquer dois ou mais tempos conhecidos, K_m e V_{\max} podem ser estimados por modelo não-linear (BATES & WATTS, 1988).

Para a aplicação da equação de Michaelis-Menten é essencial a determinação da constante K_m e da velocidade máxima, V_{\max} , que podem ser prontamente obtidas a partir das velocidades de reação medidas em diferentes concentrações de substrato, se uma enzima operar de acordo com o esquema simples, dado na equação 3.

2.3 Modelos de regressão

2.3.1 Considerações iniciais

Mazucheli & Achcar (2002) relatam que a possível relação entre uma variável dependente com uma ou mais variáveis independentes é uma das tarefas mais comuns em análise estatística. Pode-se atingir este objetivo por meio dos bem conhecidos modelos de regressão.

Na visão de Draper & Smith (1998), os modelos de regressão são classificados como lineares e não-lineares essa classificação é baseada no conceito de derivadas parciais em conformidade com Stewart (2002).

Dentre as muitas diferenças existentes entre essas duas classes de modelos, a principal está relacionada às suas formulações. No caso linear, a partir de um conjunto de observações, busca-se o modelo que melhor explique a relação entre as variáveis inerentes a um dado fenômeno (MAZUCHELI & ACHCAR, 1997).

A classe de modelos lineares é bastante flexível, uma vez que muitos modelos podem ser formulados (DRAPER & SMITH, 1998; SEBER, 1977).

Já no caso não-linear, na maioria das vezes, as formulações de possíveis modelos são baseadas em considerações teóricas inerentes ao fenômeno que se tem interesse em modelar. Modelos formulados dessa forma são chamados de modelos científicos ou mecânicos (SEBER & WILD 1989; BATES & WATTS, 1988).

Estes modelos descrevem fenômenos em várias áreas de aplicação, como por exemplo, em engenharia, química, biologia, física, agricultura, economia, etc. Os modelos não-lineares são muito utilizados e a análise estatística desses modelos é de grande interesse prático.

Em conformidade com Ferreira (2005), um modelo é linear ou não-linear nos parâmetros e, com isso, não se está interessado no tipo de curva que a função representa.

2.3.2 Modelos de regressão linear

2.3.2.1 Conceito

Segundo Draper & Smith (1998), formalmente, pode-se dizer que um modelo é linear se as derivadas parciais da variável dependente em relação a cada parâmetro não forem funções dos próprios parâmetros.

Exemplificando, seja o modelo de regressão:

$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_n x_n + \varepsilon. \quad (28)$$

Em que:

$\theta_0, \theta_1, \theta_2, \dots, \theta_n$ são constantes, ou seja, os parâmetros a serem estimados;

x uma constante, a variável regressora;

y a variável aleatória, isto é, a variável resposta;

$\varepsilon \sim N(0; \sigma^2)$.

O cálculo das derivadas parciais de y em relação a θ_0, θ_1 e $\theta_2, \dots, \theta_n$,

$$\frac{\partial y}{\partial \theta_0} = 1, \quad \frac{\partial y}{\partial \theta_1} = x, \quad \frac{\partial y}{\partial \theta_2} = x^2, \quad \dots, \quad \frac{\partial y}{\partial \theta_n} = x^n$$

mostra que nenhuma das derivadas parciais calculadas depende de algum parâmetro do modelo. Portanto, o modelo é dito linear.

2.3.2.2 Modelos de regressão linear múltipla (MRLM)

Charnet et al. (1999), descrevem:

- a) modelo polinomial com uma variável regressora, que poderá ser estendido a k variáveis regressoras, sendo k um inteiro positivo.

$$Y = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_k x^k + \varepsilon. \quad (29)$$

- b) modelo de duas variáveis regressoras e interação, que pode ser estendido a k variáveis, sendo k um inteiro positivo

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \varepsilon. \quad (30)$$

- c) modelo de hiperplano com três variáveis regressoras, que pode ser estendido a k variáveis, sendo k um inteiro positivo

$$Y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \varepsilon \quad (31).$$

Dos modelos de regressão linear múltiplo (MRLM) enfoca-se o modelo de hiperplano, particularizado para uma variável regressora, o que será abordado nas secções subseqüentes e que recebe o nome especial de modelo de regressão linear simples (MRLS). Os resultados que serão obtidos para o MRLS podem ser estendidos ao MRLM de forma geral, embora esse não seja o propósito deste trabalho.

2.3.2.3 Modelo de regressão linear simples (MRLS)

De acordo com Charnet et al. (1999), podem-se considerar duas maneiras para a obtenção de uma amostra: valores de X prefixados e, para estes valores, obtenção de observações independentes de Y ou obtenção de uma amostra de (X, Y) bivariada. Em ambos os casos, tem-se uma amostra de tamanho n, sendo x_1, x_2, \dots, x_n os valores prefixados de X, ou os valores observados de X, e y_1, y_2, \dots, y_n os correspondentes valores observados de Y (variáveis aleatórias).

Neste estudo, o valor da variável independente foi prefixado. Um MRLS é expresso, na forma geral, pela equação,

$$Y_i = \theta_0 + \theta_1 x_i + \varepsilon_i. \quad (32)$$

em que:

$1 \leq i \leq n$, tamanho da amostra;

Y_i : variável resposta associada à i-ésima observação;

x_i : variável independente associada à i-ésima observação;

θ_0 : coeficiente linear ou intercepto;

θ_1 : coeficiente de regressão;

ε_i : erro aleatório associado à i-ésima observação.

2.3.2.4 O MRLS em forma matricial

De acordo com Graybill (1983), será muito mais simples resolver as equações normais se elas forem expressas em notação matricial e também a linguagem matricial possibilitará estabelecer uma analogia entre o modelo de regressão linear e o modelo de regressão não-linear que será mais facilmente compreendido.

$$Y_1 = \theta_0 + \theta_1 x_1 + \varepsilon_1$$

$$Y_2 = \theta_0 + \theta_1 x_2 + \varepsilon_2$$

...

$$Y_n = \theta_0 + \theta_1 x_n + \varepsilon_n$$

$$\varepsilon_i \sim N(0 ; \sigma^2)$$

$$\text{Cov}[\varepsilon_i, \varepsilon_j] = 0, \quad i \neq j \quad i, j = 1, \dots, n$$

θ_0 e θ_1 , constantes desconhecidas

x_1, \dots, x_n , constantes conhecidas.

Expressando-se matricialmente:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad e \quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}.$$

e seja a matriz **X**:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix},$$

denominada matriz do modelo.

Então:

$$\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \begin{bmatrix} \theta_0 + \theta_1 x_1 + \varepsilon_1 \\ \theta_0 + \theta_1 x_2 + \varepsilon_2 \\ \vdots \\ \theta_0 + \theta_1 x_n + \varepsilon_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \mathbf{Y}. \quad (33)$$

O vetor aleatório $\boldsymbol{\varepsilon}$ é composto de variáveis independentes, com distribuição $N(0; \sigma^2)$. Dessa forma, o vetor esperança dos elementos de $\boldsymbol{\varepsilon}$ é o vetor nulo de dimensão $(n \times 1)$ e a matriz, cuja a diagonal é formada pelas variâncias e os demais elementos são as covariâncias, é dada por:

$$\begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 \mathbf{I},$$

sendo \mathbf{I} a matriz identidade de ordem 2.

Assim, MRLS amostral assume a seguinte forma, (equação 33):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim \mathbf{N}(\mathbf{0}; \sigma^2 \mathbf{I})$$

2.3.2.5 Método de quadrados mínimos (MRLS)

Como relata Gallant (1987), o método de quadrados mínimos determina o estimador de $\boldsymbol{\theta}$, denotado por $\hat{\boldsymbol{\theta}}$ e representado por $\hat{\boldsymbol{\theta}} = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$ que minimiza a soma de quadrados dos erros (SQE).

Seja:

$$SQE = \sum_{i=1}^n \varepsilon_i^2$$

em que ε é dado por:

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\theta}$$

$$\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} y_1 - (\theta_0 + \theta_1 x_1) \\ y_2 - (\theta_0 + \theta_1 x_2) \\ \vdots \\ y_n - (\theta_0 + \theta_1 x_n) \end{bmatrix}$$

então:

$$SQE = \sum_{i=1}^n [y_i - (\theta_0 - \theta_1 x_n)]^2 = (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})^2. \quad (34)$$

tendo-se que o produto escalar entre dois vetores \mathbf{X} e \mathbf{Y} quaisquer, de mesma dimensão, é dado por

$$\mathbf{X}'\mathbf{Y} = [x_1, x_2, \dots, x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i$$

note-se que $\mathbf{X}'\mathbf{Y} = \mathbf{Y}'\mathbf{X}$

Também se considera que o produto escalar de um vetor por ele mesmo é a soma de quadrados de seus elementos,

$$\mathbf{Y}'\mathbf{Y} = [y_1, y_2, \dots, y_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n y_i y_i = \sum_{i=1}^n y_i^2$$

aplicando-se esses resultados na equação (34):

$$\begin{aligned}
SQE &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta})' (\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) \\
&= (\mathbf{Y}' - \boldsymbol{\theta}'\mathbf{X}')(\mathbf{Y} - \mathbf{X}\boldsymbol{\theta}) \\
&= \mathbf{Y}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\theta} - \boldsymbol{\theta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\theta}'\mathbf{X}'\mathbf{X}\boldsymbol{\theta}
\end{aligned}$$

percebe-se que todas as parcelas da expressão acima são escalares. Note também que a segunda parcela é igual a terceira, porque uma é o transposto da outra..

Assim:

$$SQE = \mathbf{Y}'\mathbf{Y} - 2\mathbf{Y}'\mathbf{X}\boldsymbol{\theta} + \boldsymbol{\theta}'\mathbf{X}'\mathbf{X}\boldsymbol{\theta}.$$

derivando, simultaneamente em termos de $\boldsymbol{\theta}$, utilizando-se resultados demonstrados por Charnet et al. (1999):

$$\begin{aligned}
\frac{\partial \mathbf{Y}'\mathbf{Y}}{\partial \boldsymbol{\theta}} - 2\frac{\partial \mathbf{Y}'\mathbf{X}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{\theta}'\mathbf{X}'\mathbf{X}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= 0 - 2(\mathbf{Y}'\mathbf{X})' + 2\mathbf{X}'\mathbf{X}\boldsymbol{\theta} \\
\frac{\partial \mathbf{Y}'\mathbf{Y}}{\partial \boldsymbol{\theta}} - 2\frac{\partial \mathbf{Y}'\mathbf{X}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{\theta}'\mathbf{X}'\mathbf{X}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} &= -2(\mathbf{Y}'\mathbf{X})' + 2\mathbf{X}'\mathbf{X}\boldsymbol{\theta}.
\end{aligned}$$

denominando por $\hat{\boldsymbol{\theta}}$ o vetor que anula a derivada, pode-se escrever:

$$\begin{aligned}
-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\theta}} &= 0 \\
\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\theta}} &= \mathbf{X}'\mathbf{Y} \quad (35)
\end{aligned}$$

denominado sistema de equações normais (SEN), desde que a inversa de $X'X$ exista.

2.3.2.6 Estimativas de variâncias e covariâncias (MRLS)

Em conformidade com Charnet et al. (1999), seja (SEN):

$$X'X\hat{\theta} = X'Y$$

desenvolvendo:

$$(X'X)^{-1} X'X\hat{\theta} = (X'X)^{-1} X'Y$$

$$\hat{\theta} = (X'X)^{-1} X'Y \quad (36)$$

As esperanças, variância e covariâncias de $\hat{\theta}$ serão obtidas simultaneamente utilizando-se notação matricial e alguns resultados da álgebra matricial.

Assim:

$$\begin{aligned} E(\hat{\theta}) &= E\left[(X'X)^{-1} X'Y\right] \\ &= (X'X)^{-1} X'E[Y] \\ &= (X'X)^{-1} X'X\theta \\ &= \mathbf{I}\theta \\ &= \theta \end{aligned}$$

$$\begin{aligned}
\text{Var}[\hat{\theta}] &= \text{Var}[(X'X)^{-1} X'Y] \\
&= (X'X)^{-1} X' \text{Var}[Y] [(X'X)^{-1} X']', \text{ onde } \text{Var}[Y] = \sigma^2 I \\
&= \sigma^2 (X'X)^{-1} X' [(X'X)^{-1} X']' \\
&= \sigma^2 (X'X)^{-1} X'X (X'X)^{-1}, \text{ onde } X'X (X'X)^{-1} = I \\
&= \sigma^2 (X'X)^{-1}
\end{aligned}$$

portanto, sob MRLS, a distribuição de probabilidade do vetor, correspondente ao valor prefixado da variável regressora dado em \mathbf{X} , é $Y \sim N(X\theta; \sigma^2 I)$ e, portanto, a distribuição de $\hat{\theta} = (X'X)^{-1} X'Y$ é normal, p-variada, sendo $p=2$, a dimensão de θ , com vetor de esperanças e matriz de variâncias e covariâncias dados acima. Ou seja,

$$\hat{\theta} \sim N_p(\theta; \sigma^2 (X'X)^{-1}).$$

Para $\hat{\theta}_j$, elemento de $\hat{\theta}$, $j=0,1$

$$\hat{\theta}_j \sim N(\theta_j; \sigma^2 C_{j+1, j+1}),$$

sendo $C_{j+1, j+1}$ o $(j+1)$ -ésimo elemento da diagonal principal de $(X'X)^{-1}$.

A covariância entre dois elementos de $\hat{\theta}$, θ_k e θ_j , para $k, j = 0,1$ $k \neq j$, é dada por $\sigma^2 C_{k+1, j+1}$, sendo $C_{k+1, j+1}$ o elemento da $(k+1)$ -ésima linha e $(j+1)$ -ésima coluna de $(X'X)^{-1}$.

2.3.2.7 Intervalos de confiança (MRLS)

Na visão de Neter et al. (1990), um estimador não viciado de σ^2 é dado pela soma de quadrados dos resíduos, dividida por $(n-p)$, neste caso o número de colunas de X , ou ainda, a dimensão de θ , em que $p=2$.

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\theta})'(Y - X\hat{\theta})}{n-p} \quad (37)$$

Definem-se intervalos de confiança para os parâmetros, utilizando-se as seguintes quantidades pivotais:

$$\frac{\hat{\theta}_j - \theta_j}{\sqrt{\hat{\sigma}^2 C_{j+1, j+1}}} \sim t_{(n-p)} \quad j=0,1$$

e

$$\frac{(n-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(n-p)}$$

sendo $\hat{\sigma}^2$ não viciado para σ^2 , com variância igual a $\frac{2(\sigma^2)^2}{(n-p)}$.

Os intervalos de confiança $(1-\alpha)100\%$ para os parâmetros θ_j , $j=0,1$ e σ^2 são dados, respectivamente, por

$$\hat{\theta}_j \pm t_{\left(\frac{\alpha}{2}, n-p\right)} \sqrt{\hat{\sigma}^2 C_{j+1, j+1}}, \quad j=0,1$$

e

$$\left[\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\chi^2_{\left(\frac{\alpha}{2}, n-p\right)}}; \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\chi^2_{\left(1-\frac{\alpha}{2}, n-p\right)}} \right],$$

sendo $t_{(\alpha, n)}$ o quantil $(1-\alpha)$ de distribuição t de Student com n graus de liberdade, $\chi^2_{(\alpha, n)}$ o quantil α de distribuição χ^2 com n graus de liberdade e $C_{(j+1, j+1)}$ o (j+1)-ésimo elemento da diagonal principal de $(X'X)^{-1}$, $j = 0, 1$

2.3.2.8 Diagnóstico das pressuposições básicas

Segundo Mood et al. (1974), se inferências forem realizadas no MRLS, e que poderão ser estendidas para os modelos de regressão não-linear, os erros (variável aleatória ε_i) deverão assumir um modelo específico de probabilidade e devem ser homocedásticos (ter variância constante).

Neste estudo, foi pressuposto que o modelo de probabilidade do erro era o modelo normal, com:

$$E(\varepsilon_i) = 0 \text{ e } \text{Var}(\varepsilon_i) = \sigma^2, \text{ isto é, } \varepsilon_i \sim N(0, \sigma^2);$$

por conseguinte, supõe-se, também, que os erros não eram correlacionados

$$E[\varepsilon_i, \varepsilon_j] = 0 \text{ para } i \neq j.$$

O diagnóstico das pressuposições sobre os erros pode ser feito por meio de seus estimadores, denominados resíduos, em que:

$$\hat{\varepsilon}_i = \hat{e}_i = Y_i - \hat{Y}_i,$$

utilizando-se a análise visual do esboço de alguns gráficos e ou testes estatísticos descritos a seguir:

(i) Pressuposição I: os erros têm variância constante

Para um dado valor de x , a variância do erro ε é sempre σ^2 , denominado variância residual, isto é, $E[\varepsilon] = \sigma^2$. Nesse caso, o erro é dito homocedástico.

Para a verificação dessa pressuposição, pode-se fazer análise gráfica, esboçando-se os valores de y ajustados versus os resíduos padronizados, pois, se a variância for constante, os resíduos se distribuem aleatoriamente em torno da média zero.

(ii) Pressuposição II: normalidade dos erros

Os erros devem ter distribuição normal $\varepsilon \sim N(0, \sigma^2)$, sendo essa pressuposição necessária para inferências. Para se testar a suposição de normalidade de uma variável aleatória por meio de uma amostra, pode ser utilizado o teste de Shapiro-Wilk, proposto em 1965, com a estatística W , (FERREIRA, 2005), representado por:

$$W = \frac{\left(\sum_{i=1}^n a_i x_i \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (38)$$

A equação (38) é dependente do coeficiente a_i , que é denominado por melhor estimador linear não viesado normalizado do valor esperado das estatísticas de ordem da distribuição normal padrão. A determinação exata desses coeficientes é dependente da determinação exata dos valores esperados e da matriz de variâncias e covariâncias das estatísticas de ordem da distribuição normal padrão.

Para avaliar ainda essa pressuposição, pode ser esboçado o gráfico dos quantis teóricos da distribuição normal padrão versus os resíduos padronizados. Então, se os pontos estiverem próximos da reta da 1ª bissetriz (sendo as escalas dos 2 eixos cartesianos as mesmas), esse fato é indicativo para a normalidade dos resíduos.

(iii) Pressuposição III: os erros são independentes

O erro de uma observação não é correlacionado com o erro em outra observação, isto é,

$$E[\varepsilon_i, \varepsilon_j] = 0 \text{ para } i \neq j.$$

Em 1950, foi desenvolvido o teste de Durbin & Watson, que avaliou a presença ou não de correlação serial ou autocorrelação, utilizando a estatística

descrita em Draper & Smith (1998). A hipótese de nulidade testada é a de que a autocorrelação dos resíduos é igual a zero:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (39)$$

que não será utilizada neste estudo, pois a literatura a respeito de experimentos de cinética enzimática indica que os erros não estão correlacionados.

Assim, após a verificação das pressuposições básicas, é possível construir intervalos estatísticos de forma confiável, que podem ser estendidos a modelos de regressão não linear. Em regressão não-linear é utilizada a aproximação a teoria assintótica.

2.3.3 Modelos de regressão não-linear

2.3.3.1 Conceito

Draper & Smith (1998) afirmam que um modelo é não-linear se as derivadas parciais da variável dependente em relação a cada parâmetro não forem funções dos próprios parâmetros. Bastaria uma de estas derivadas ser função dos parâmetros para que o modelo seja classificado como não-linear.

Exemplificando, seja o modelo:

$$y = \theta_1 + \theta_2 x e^{-\theta_2 x} . \quad (40)$$

em que θ_1 e θ_2 são os parâmetros a serem estimados, calculam-se as derivadas parciais

$$\frac{\partial y}{\partial \theta_1} = 1 \quad \frac{\partial y}{\partial \theta_2} = (x - \theta_2 x^2) e^{-\theta_2 x}.$$

endo que uma das derivadas parciais depende do parâmetro θ_2 , indicando que o modelo em questão é não-linear.

Um modelo de regressão não-linear é considerado “intrinsecamente linear” se este pode ser transformado em um modelo linear por meio de uma reparametrização apropriada. Pode-se, ainda, usar o termo “intrinsecamente linear” para referir-se a modelos que podem ser linearizados via alguma transformação.

Seja o modelo:

$$y = \theta_0 x^{\theta_1} \varepsilon, \quad (41)$$

esse modelo é intrinsecamente linear, já que pode ser linearizado por meio de uma transformação logarítmica:

$$\ln y = \ln \theta_0 + \theta_1 \ln x + \ln \varepsilon, \quad (42)$$

em que:

$$\begin{array}{l} y' = \ln y \quad \theta'_0 = \ln \theta_0 \\ x' = \ln x \quad \varepsilon' = \ln \varepsilon \end{array},$$

tem-se:

$$y' = \theta'_0 + \theta'_1 x' + \varepsilon'. \quad (43)$$

A partir do modelo linearizado estima-se os parâmetros θ'_0 e θ'_1 e através da operação inversa obtem-se as estimativas dos parâmetros θ_0 e θ_1 .

Caso não seja possível obter uma reparametrização ou uma transformação apropriada, que reduza o modelo à forma não linear tem-se os chamados modelos “intrinsecamente não-lineares”.

2.3.3.2 Transformações

Como relata (Bates & Watts, 1988; Guimarães, 2001) em geral, na prática, um modelo não-linear é linearizado para facilitar a obtenção das estimativas dos parâmetros.

De acordo com Currie (1982), o inconveniente de uma transformação é que, além do parâmetro perder sua interpretação intrínseca, pode-se alterar a estrutura e distribuição do erro, ou seja, se os erros do modelo original satisfizerem as suposições usuais de normalidade, independência e homogeneidade da variância, os erros do novo modelo, em geral, não satisfarão tais suposições porque tais transformações podem mudar a forma da distribuição dos erros ou fazer com que a variância desta distribuição deixe de ser constante.

Segundo Montgomery & Peck.(1992), apesar de, em muitos casos ser possível selecionar as transformações empiricamente, é desejável dispor de técnicas mais formais e objetivas para a determinação de uma transformação

apropriada. Uma técnica deste tipo é o procedimento de Box-Cox o qual não será apresentado neste estudo.

O modelo de Michaelis-Menten é intrinsecamente linear, e suas linearizações serão apresentadas nas secções subseqüentes, de acordo com Dowd & Riggs (1965).

2.3.3.2.1 Lineweaver-Burk $\left(\frac{1}{v_0} \text{ x } \frac{1}{[S]} \right)$

Seja o modelo matemático de Michaelis-Menten, equação (19)

$$v_0 = \frac{V_{\max} [S]}{[S] + K_m}$$

invertendo-se:

$$\frac{1}{v_0} = \frac{K_m + [S]}{V_{\max} \cdot [S]},$$

Separando-se os componentes do numerador do lado direito da equação:

$$\frac{1}{v_0} = \frac{K_m}{V_{\max} [S]} + \frac{[S]}{V_{\max} [S]}.$$

pode ser simplificada para:

$$\frac{1}{v_0} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}} \cdot \frac{1}{[S]}. \quad (44)$$

A equação (44) corresponde uma equação linear da forma $y = a + bx$,
em que:

$$a = \frac{1}{V_{\max}} \text{ e } b = \frac{K_m}{V_{\max}}, \text{ representada pela Figura 4.}$$

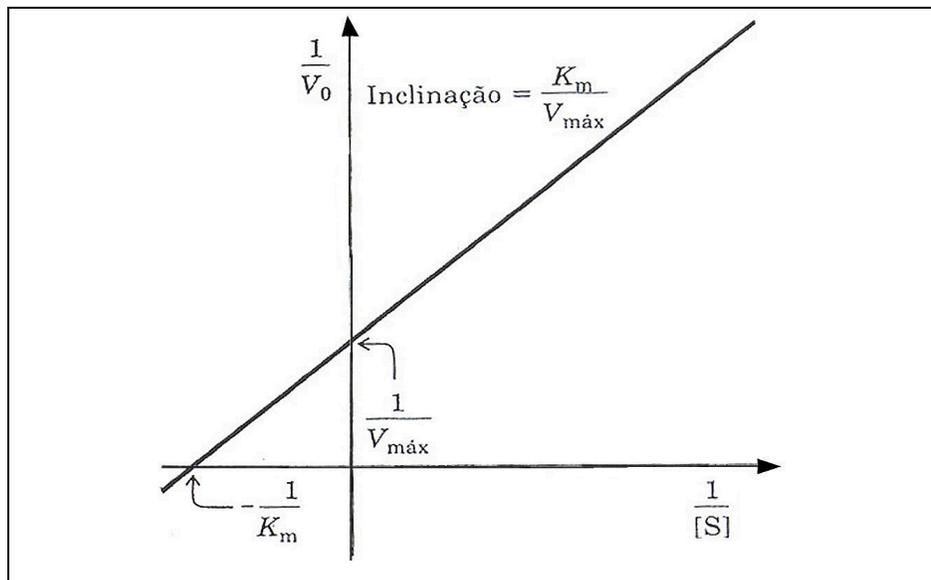


FIGURA 4 Gráfico Lineweaver-Burk

O gráfico Lineweaver-Burk permite calcular o V_{\max} por meio do intercepto com a ordenada e o K_m pelo prolongamento da reta e o intercepto com a abscissa. Posteriormente, incorpora-se o erro ao modelo e, por meio de um modelo de regressão linear, estimam-se os parâmetros.

2.3.3.2.2 Hanes-Wolf $\left(\frac{[S]}{v_0} \times [S] \right)$

Utilizando-se a equação (44), ou seja, Lineweaver-Burk:

$$\frac{1}{v_0} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}} \cdot \frac{1}{[S]}.$$

Multiplicando-se por $[S]$, obtém-se:

$$\frac{1}{v_0} [S] = \frac{1}{V_{\max}} \cdot [S] + \frac{K_m}{V_{\max}} \cdot \frac{1}{[S]} \cdot [S].$$

Cancelando e organizando a equação:

$$\frac{[S]}{v_0} = \frac{K_m}{V_{\max}} + \frac{1}{V_{\max}} [S]. \quad (45)$$

A equação (45) é uma equação linear da forma $y = a + bx$, em que

$$a = \frac{K_m}{V_{\max}} \text{ e } b = \frac{1}{V_{\max}}, \text{ representada pela Figura 5}$$

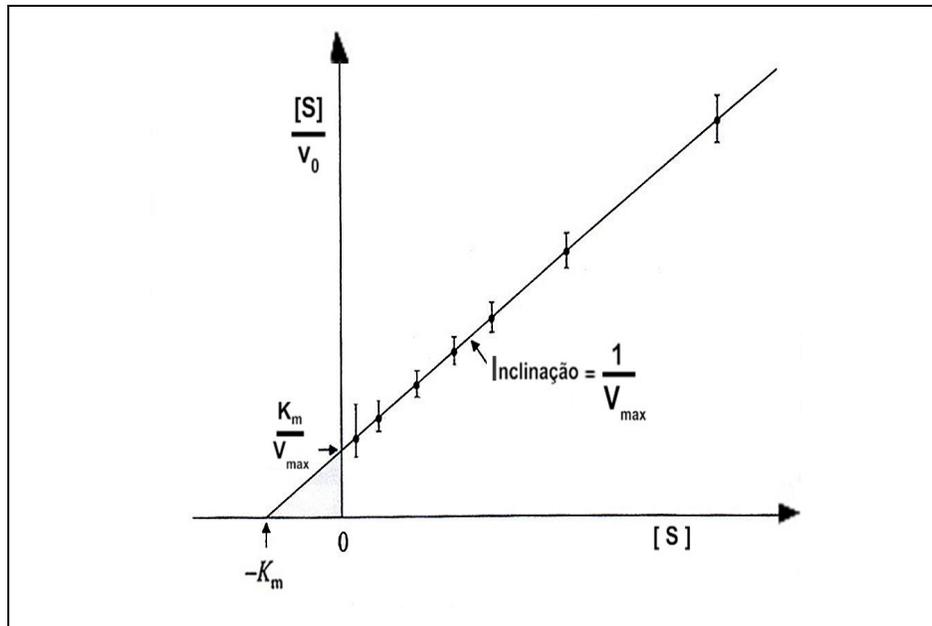


FIGURA 5 Gráfico Hanes-Woolf

O gráfico Hanes-Woolf permite calcular o V_{\max} pela inclinação e o K_m pelo prolongamento da reta e o intercepto com a abscissa. Posteriormente, incorpora-se o erro ao modelo e, por meio de um modelo de regressão linear, estimam-se os parâmetros.

2.3.3.2.3 Wolf-Augustinsson-Hofstee $\left(v_0 \times \frac{v_0}{[S]} \right)$

Utilizando-se a equação (19)

$$v_0 = \frac{V_{\max} [S]}{[S] + K_m}$$

divide-se o numerador e o denominador por $[S]$

$$v_0 = \frac{\frac{V_{\max} \cdot [S]}{[S]}}{\frac{K_m + [S]}{[S]}}$$

cancelando $[S]$ no numerador e separando-se os termos pelos sinal de adição no denominador,

$$v_0 = \frac{V_{\max}}{\frac{K_m}{[S]} + 1}$$

Multiplicando-se invertido,

$$v_0 \left(\frac{K_m}{[S]} + 1 \right) = V_{\max} .$$

$$\frac{K_m}{[S]} v_0 + v_0 = V_{\max} .$$

Isolando-se o valor de v_0 ,

$$v_0 = V_{\max} - \frac{K_m \cdot v_0}{[S]} .$$

Organizando-se a equação,

$$v_0 = V_{\max} - K_m \frac{v_0}{[S]} . \quad (46)$$

Esta equação (46) é linear da forma $y = a + bx$, em que $a = V_{\max}$ e $b = -K_m$, representada pela Figura 6.

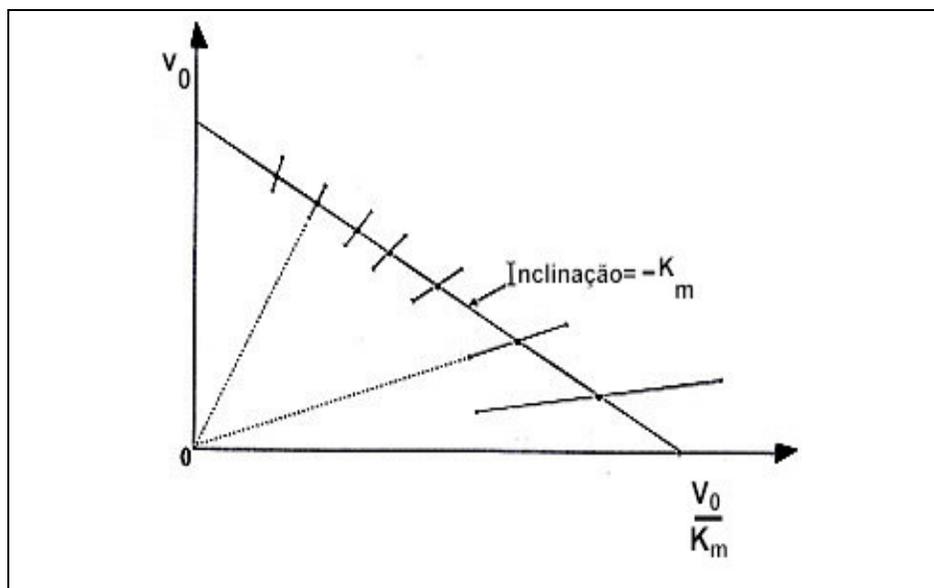


FIGURA 6 Gráfico Wolf-Augustinsson-Hofstee

O valor de K_m é obtido pela inclinação da reta e o valor de V_{\max} pelo intercepto da reta com a ordenada. Posteriormente, incorpora-se o erro ao modelo e, por meio de um modelo de regressão linear, estimam-se os parâmetros.

2.3.3.3.4 Eadie-Scatchard $\left(\frac{v_0}{[S]} \times v_0 \right)$

A partir da equação Wolf-Augustinsson-Hofstee, equação (46)

$$v_0 = -K_m \frac{v_0}{[S]} + V_{\max}$$

isolando-se V_{\max} e rearranjando-se:

$$V_{\max} = v_0 + \frac{K_m}{[S]}.$$

Dividindo-se tudo por K_m ,

$$\frac{V_{\max}}{K_m} = \frac{v_0}{K_m} + \frac{K_m \cdot v_0}{[S] \cdot K_m}.$$

Cancelando-se K_m , isolando-se $\frac{v}{[S]}$ e rearranjando-se:

$$\frac{v_0}{[S]} = \frac{V_{\max}}{K_m} - \frac{1}{K_m} \cdot v_0 \quad (47)$$

Esta equação linear (47) é da forma $y = a + bx$, em que $a = \frac{V_{\max}}{K_m}$ e

$b = -\frac{1}{K_m}$, representada pela Figura 7.

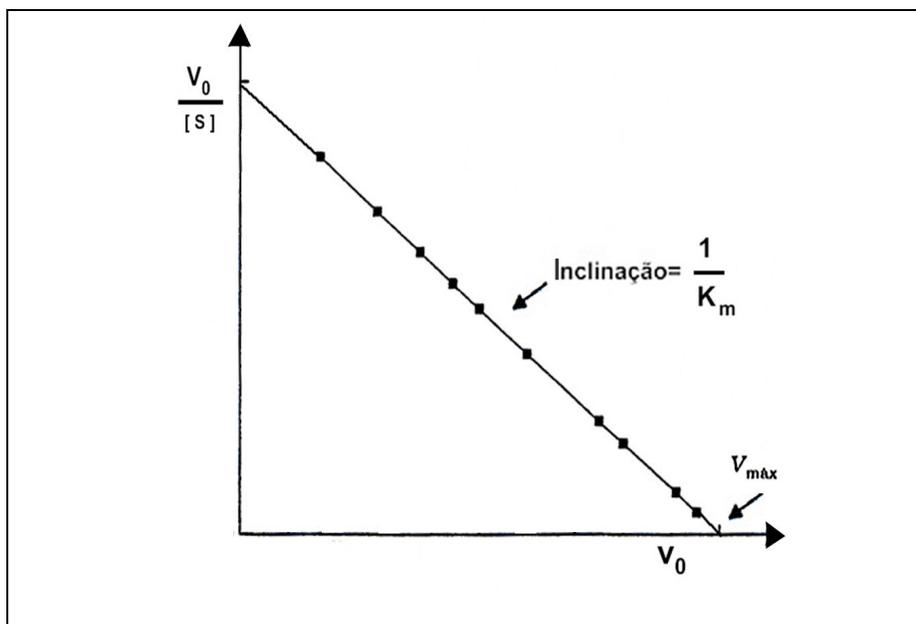


FIGURA 7 Gráfico Eadie-Scatchard

O valor de $V_{m\acute{a}x}$ é obtido pelo prolongamento da reta e intercepto com a abscissa e o K_m é obtido pelo inverso da inclinação da reta. Posteriormente, incorpora-se o erro ao modelo e, por meio do modelo de regressão linear, estimam-se os parâmetros.

2.3.3.3 Aproximação linear pela série de Taylor

De conformidade com (Burden & Faires, 2003; Seber & Wild, 1989), seja o modelo de regressão não-linear,

$$y_i = f(x_i, \theta) + \varepsilon_i, \quad i=1,2,\dots,n \quad (48)$$

$$\theta^0 = \begin{bmatrix} \theta_1^0 \\ \theta_2^0 \\ \vdots \\ \theta_p^0 \end{bmatrix}.$$

tem-se:

$$f(x_i, \theta) \cong f(x_i, \theta^0) + \left[\frac{\partial f(x_i, \theta)}{\partial \theta_1} \right]_{\theta=\theta_1^0} + \dots + \left[\frac{\partial f(x_i, \theta)}{\partial \theta_p} \right]_{\theta=\theta_p^0} (\theta_p - \theta_p^0)$$

$$f(x_i, \theta) \cong f(x_i, \theta^0) + \sum_{j=1}^p \left[\frac{\partial f(x_i, \theta)}{\partial \theta_j} \right]_{\theta=\theta_j^0} (\theta_j - \theta_j^0)$$

ou

$$f(\theta) \cong f(\theta^0) + F(\theta^0)(\theta - \theta^0) \quad (49)$$

em que $F(\theta^0)$ é uma matriz $n \times p$ de derivadas parciais no ponto $\theta = \theta^0$.

$F(\theta^0)$ tem a forma:

$$F(\theta^0) = \begin{bmatrix} \frac{\partial f(x_1, \theta^0)}{\partial \theta_1} & \frac{\partial f(x_1, \theta^0)}{\partial \theta_2} & \dots & \frac{\partial f(x_1, \theta^0)}{\partial \theta_p} \\ \frac{\partial f(x_2, \theta^0)}{\partial \theta_1} & \frac{\partial f(x_2, \theta^0)}{\partial \theta_2} & \dots & \frac{\partial f(x_2, \theta^0)}{\partial \theta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f(x_n, \theta^0)}{\partial \theta_1} & \frac{\partial f(x_n, \theta^0)}{\partial \theta_2} & \dots & \frac{\partial f(x_n, \theta^0)}{\partial \theta_p} \end{bmatrix}$$

2.3.3.4 Método de quadrados mínimos ordinários

Como relata Jennrich & Ralston (1979), o princípio dos mínimos quadrados é usado para estimar os parâmetros em modelos não-lineares, da mesma forma que em modelos lineares. A estimativa para θ é dada pelo vetor $\hat{\theta}$ que minimiza a soma de quadrados do resíduo.

$$SQE = \sum_{i=1}^n [y_i - f(x_i, \theta)]^2 \quad (50)$$

ou, em notação matricial

$$SQE(\theta) = [Y - \mathbf{f}(\theta)]' [Y - \mathbf{f}(\theta)] \quad (51)$$

derivando-se a $SQE(\theta)$ em relação a θ , tem-se:

$$\begin{aligned}
\frac{\partial SQE(\theta)}{\partial \theta'} &= [\mathbf{Y} - \mathbf{f}(\theta)]' \frac{\partial}{\partial \theta'} [\mathbf{Y} - \mathbf{f}(\theta)] + [\mathbf{Y} - \mathbf{f}(\theta)] \frac{\partial}{\partial \theta'} [\mathbf{Y} - \mathbf{f}(\theta)]' \\
&= [\mathbf{Y} - \mathbf{f}(\theta)]' \frac{\partial}{\partial \theta'} [\mathbf{Y} - \mathbf{f}(\theta)] + [\mathbf{Y} - \mathbf{f}(\theta)]' \frac{\partial}{\partial \theta'} [\mathbf{Y} - \mathbf{f}(\theta)] \\
&= -2[\mathbf{Y} - \mathbf{f}(\theta)]' \left[\frac{\partial}{\partial \theta'} \mathbf{f}(\theta) \right], \text{ onde } \left[\frac{\partial}{\partial \theta'} \mathbf{f}(\theta) \right] = \mathbf{F}(\theta) \\
&= -2[\mathbf{Y} - \mathbf{f}(\theta)]' \mathbf{F}(\theta)
\end{aligned}$$

igualando-se a zero tem-se:

$$\begin{aligned}
-2[\mathbf{Y} - \mathbf{f}(\hat{\theta})]' \mathbf{F}(\hat{\theta}) &= 0 \\
[\mathbf{Y} - \mathbf{f}(\hat{\theta})]' \mathbf{F}(\hat{\theta}) &= 0 \\
\mathbf{F}(\hat{\theta}) [\mathbf{Y} - \mathbf{f}(\hat{\theta})]' &= 0 \quad (\text{S.E.N}) \quad (52)
\end{aligned}$$

Sabendo que $\mathbf{F}(\hat{\theta})$ é uma matriz de derivadas parciais, a qual será denotada por \mathbf{X} . Então, (SEN) não-linear pode ser reescrito com:

$$\begin{aligned}
\mathbf{X}'[\mathbf{Y} - \mathbf{f}(\hat{\theta})] &= 0 \\
\mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{f}(\hat{\theta}) &= 0 \\
\mathbf{X}'\mathbf{f}(\hat{\theta}) &= \mathbf{X}'\mathbf{Y} \quad (53)
\end{aligned}$$

faz-se em equação (49) $[\mathbf{Y} - \mathbf{f}(\hat{\theta})] = \boldsymbol{\varepsilon}$ tem-se:

$$\mathbf{X}'\boldsymbol{\varepsilon} = 0 \quad (54)$$

é, portanto, ortogonal às colunas da matriz jacobiana $F(\theta)$ calculada em $\theta = \hat{\theta}$.

Em regressão linear:

$$F = F(\hat{\theta}) = X,$$

a identificação entre X , no caso linear e F , no caso não-linear, vale, em geral, isto é, as expressões utilizadas no estudo da inferência do modelo linear e com erros normais têm uma relação para o caso não-linear, obtida por intermédio da substituição da matriz X por F .

A razão disso é que a teoria de inferência estatística que se desenvolve para o modelo de regressão não-linear baseia-se, essencialmente, na aproximação linear por série de Taylor, dada pela equação (49):

$$f(\theta) \approx f(\theta^0) + F(\theta - \theta^0).$$

em regressão linear, o erro é ortogonal às colunas da matriz X , e, em regressão não-linear, o erro é ortogonal às colunas do jacobiano de $f(\theta)$, estimado em $\theta = \hat{\theta}$.

Analogamente ao modelo linear, o estimador de $\hat{\sigma}_0^2$ é dado pela expressão:

$$\hat{\sigma}_0^2 = \frac{SQE(\hat{\theta})}{n - p} \quad (55)$$

De acordo com (Box, 1971; Bates & Watts 1988), não existe uma solução explícita para o Sistema de Equações Normais (SEN) não-lineares; as soluções são obtidas por intermédio de métodos iterativos que serão apresentados posteriormente.

2.3.3.5 Propriedades assintóticas dos estimadores de mínimos quadrados

Para Souza (1998), considerando como verdadeiras as suposições de que os erros são variáveis aleatórias independentes e identicamente distribuídas, normais com média zero e variância constante σ^2 , podem-se fazer certas considerações a respeito dos estimadores de mínimos quadrados.

Em modelos de regressão lineares, os estimadores de mínimos quadrados são não viesados, normalmente distribuídos e, ainda, possuem variância mínima entre qualquer outra classe de estimadores. Essas propriedades são aceitas como as melhores propriedades que uma classe de estimadores podem apresentar (Searle, 1971). Podem existir outros estimadores, também não viesados, mas esses são menos eficientes no sentido de suas variâncias excederem a variância dos estimadores de mínimos quadrados.

Já no caso não-linear, essas propriedades somente são válidas assintoticamente, isto é, quando o tamanho da amostra é suficientemente grande. Em geral, em pequenas amostras, essas propriedades não são satisfeitas (Jennrich, 1969).

Segundo Ratkowsky (1983), o tamanho da amostra necessário para os resultados assintóticos terem validade depende, fundamentalmente, do modelo em estudo.

Podem existir modelos que, mesmo com amostras consideradas muito grandes, os resultados assintóticos não são válidos. Entretanto, Souza (1998)

afirma que, à medida que o tamanho da amostra aumenta, os resultados assintóticos vão se tornando mais aplicáveis.

De acordo com (Guttman & Meeter, 1965; Seber & Wild, 1989), quando os estimadores de mínimos quadrados apresentarem um pequeno viés, distribuição próxima da normal e verdadeira variância próxima daquelas dadas pela matriz de variâncias-covariâncias assintóticas, pode-se afirmar que os estimadores de mínimos quadrados exibem um comportamento próximo do comportamento linear. Quanto mais próximo do linear for o comportamento de um modelo, mais precisos serão os resultados assintóticos e, conseqüentemente, mais confiáveis serão as inferências.

Mazucheli & Achcar (2002) avaliam a extensão do comportamento não-linear por meio das bem conhecidas medidas de não-linearidade, expressões usadas para avaliar a adequabilidade à aproximação linear e seus efeitos nas inferências que não serão abordadas nesse texto.

2.3.3.6 Processos numéricos iterativos

2.3.3.6.1 Generalidades

Conforme Raaijmakers (1987), vários métodos iterativos são propostos na literatura para a obtenção das estimativas de mínimos quadrados dos parâmetros de um modelo de regressão não-linear. Os mais utilizados são o método de Gauss-Newton ou método da linearização, o método de Steepest-Descent ou método do gradiente e o método de Marquardt. Esses métodos utilizam as derivadas parciais da função esperança: $f(X_n; \theta)$ em relação aos parâmetros.

Segundo Gallant (1987), a maioria dos métodos de estimação, como, por exemplo a estimação de quadrados mínimos, requer um estimador $\hat{\theta}$ obtido por meio da minimização de uma função apropriada $SQE(\theta)$ de θ . Por exemplo, na estimação por mínimos quadrados, minimiza-se, conforme a equação (51)

$$SQE(\theta) = [Y - \mathbf{f}(\theta)]' [Y - \mathbf{f}(\theta)] \quad (56)$$

A função $SQE(\theta)$ raramente pode ser minimizada analiticamente. Assim, um procedimento iterativo é usualmente necessário para encontrar $\hat{\theta}$. Inicialmente tem-se um valor inicial $\theta^{(1)}$ e, seqüencialmente, encontram-se $\theta^2, \theta^3, \dots, \theta^k$, até que $\theta^k - \theta^{k-1} < \lambda$, em que λ seja o nível de tolerância escolhido para precisão necessária para o problema proposto, isto é, atinge-se a convergência, em que obtém-se uma aproximação para $\hat{\theta}$.

A maioria dos algoritmos que são utilizados para a estimação de θ é de adaptações ou aproximações do método clássico de Newton.

De acordo com (Gallant, 1987; Burden & Faires 2003), não existe um procedimento iterativo que seja superior a todos os outros. O que existe são certas classes de algoritmos que são, em geral, mais eficientes em relação a certas características ou, equivalentemente, existem certas limitações para as quais determinados algoritmos são, em geral, mais eficientes.

Desde que as funções de θ sejam aproximadamente quadráticas, os algoritmos usados para a estimação funcionam bem; caso contrário, podem haver problemas. Algumas vezes, é encontrado um mínimo local em um lugar do mínimo global. Nenhum dos métodos existentes pode garantir convergência

para um mínimo global, entretanto, uma maneira de tentar prevenir possíveis soluções locais é pela escolha de bons valores iniciais, já que o ponto para o qual um procedimento converge depende também da escolha do valor inicial.

Na prática, para Mazucheli & Achcar (2002), não se deve contentar, pelo menos de imediato, com o valor obtido pelo processo iterativo, mas, sim, realizar vários testes partindo de vários iniciais diferentes e, então, comparar os resultados. Esse procedimento ajuda a prevenir possíveis soluções equivocadas, ou seja, que se referem a mínimos locais.

O primeiro passo de um processo iterativo, obviamente, é escolher segundo algum critério, o algoritmo a ser utilizado e também uma estimativa inicial para o processo, que podem ser obtidos por meio de experiências anteriores. Se, por acaso, o algoritmo escolhido não convergir satisfatoriamente, pode-se tentar reprocessá-lo usando melhores valores iniciais ou, então, utilizar uma modificação desse algoritmo ou ainda a utilização de um outro método diferente.

Podem-se usar transformações nos parâmetros que, dependendo do problema, possam facilitar a convergência nos métodos iterativos.

2.3.3.6.2 Método de Gauss Newton

Gallant (1987) afirma que o método de Gauss-Newton consiste em linearizar o modelo não linear por meio do desenvolvimento da série de Taylor da função $f(X_i; \theta)$ até o termo de primeira ordem, na vizinhança de θ^0 .

Seja o seguinte modelo não linear:

$$Y_i = f(X_i; \theta) + \varepsilon_i \quad (57)$$

o desenvolvimento da série de Taylor da função $f(\mathbf{X}_i; \boldsymbol{\theta})$ é dado por

$$f(\boldsymbol{\theta}) \approx f(\boldsymbol{\theta}^0) + \mathbf{F}(\boldsymbol{\theta}^0)(\boldsymbol{\theta} - \boldsymbol{\theta}^0).$$

logo, o (SEN) não-linear, equação (53)

$$\mathbf{X}'\mathbf{f}(\hat{\boldsymbol{\theta}}) = \mathbf{X}'\mathbf{Y}.$$

pode ser reescrito como:

$$\mathbf{X}'[\mathbf{f}(\boldsymbol{\theta}^0) + \mathbf{F}(\boldsymbol{\theta}^0)(\boldsymbol{\theta} - \boldsymbol{\theta}^0)] = \mathbf{X}'\mathbf{Y}.$$

Sabendo-se que $\mathbf{F}(\boldsymbol{\theta}^0)$ é uma matriz de derivadas parciais, a qual será denotada por \mathbf{X} , então, (SEN) não linear pode ser reescrito como:

$$\mathbf{X}'[\mathbf{f}(\boldsymbol{\theta}^0) + \mathbf{X}(\boldsymbol{\theta} - \boldsymbol{\theta}^0)] = \mathbf{X}'\mathbf{Y}.$$

Segue-se que:

$$\mathbf{X}'\mathbf{f}(\boldsymbol{\theta}^0) + \mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) = \mathbf{X}'\mathbf{Y}$$

$$\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) = \mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{f}(\boldsymbol{\theta}^0)$$

$$\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) = \mathbf{X}'[\mathbf{Y} - \mathbf{f}(\boldsymbol{\theta}^0)]$$

$$\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) = \mathbf{X}'\boldsymbol{\varepsilon}$$

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

$$\mathbf{I}(\boldsymbol{\theta} - \boldsymbol{\theta}^0) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

$$\boldsymbol{\theta} - \boldsymbol{\theta}^0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

$$\boldsymbol{\theta} = \boldsymbol{\theta}^0 + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

Logo, o algoritmo de Gauss-Newton será:

$$\boldsymbol{\theta}^j = \boldsymbol{\theta}^{j-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}. \quad (58)$$

A convergência ocorrerá conforme secção 2.3.3.6.1.

2.4 Métodos empíricos: o ‘bootstrap’ na regressão

2.4.1 Introdução

Segundo Souza (1998), quando as hipóteses de normalidade e da aproximação linear assintótica no modelo de regressão não-linear se tornam questionáveis devido, possivelmente, ao tamanho reduzido da amostra ou à curvatura excessiva da superfície resposta, a técnica “bootstrap” de estimação se torna atrativa, não somente como alternativa para o processo inferencial, como também como ferramenta de diagnóstico.

O princípio básico da técnica “bootstrap” é o seguinte: suponha que,

$$\hat{\boldsymbol{\theta}} \text{ e } \mathbf{s}^2 = \hat{\sigma}_0^2$$

sejam tratados como os valores verdadeiros dos parâmetros desconhecidos

$$\theta^0 \text{ e } \sigma^2$$

do modelo de regressão não-linear,

$$Y_t = f(x_t, \theta^0), \quad t = 1, \dots, n \quad (59)$$

Considere, então, M amostras de tamanho n , simuladas do modelo amostral,

$$y_t^* = f(x_t, \hat{\theta}) + \delta_t, \quad t = 1, \dots, n \quad (60)$$

em que os

$$\delta_t \text{ são iid } N(0, \sigma^2).$$

Para cada uma destas M amostras, podem-se calcular os estimadores de mínimos quadrados.

$$\hat{\theta}_m \text{ e } s_m^2, \quad m = 1, \dots, M$$

de

$$\hat{\theta} \text{ e } s^2$$

O princípio “bootstrap” estabelece que

$$\hat{\theta}_m \text{ e } s_m^2,$$

em repetidas amostras do modelo amostral, se comportam em relação a

$$\hat{\theta} \text{ e } s^2$$

do mesmo modo que

$$\hat{\theta} \text{ e } s^2,$$

em repetidas amostras do modelo original, se comportam em relação aos elementos desconhecidos

$$\theta^0 \text{ e } \sigma^2.$$

Desse modo, torna-se possível observar as propriedades distribucionais dos estimadores de mínimos quadrados a partir de simulações do modelo de regressão não-linear definido por y^* .

O método descrito acima é conhecido como “bootstrap” paramétrico, sendo possível atenuar o conceito e gerar um “bootstrap” não paramétrico. O uso do “bootstrap” está indicado no contexto da presença de não-linearidade para o modelo de regressão não-linear.

2.4.2 O método ‘bootstrap’

De acordo com Efron & Tibshiran (1993), a estimação “bootstrap” pode ser vista como um princípio de substituição, assim como o método de momentos ou o princípio de substituição de frequências relativas.

A definição seguinte captura a essência do princípio “bootstrap” em teoria da estimação.

Seja $A = (x_1, \dots, x_n)$ uma amostra aleatória de uma população com distribuição $F(x)$ definida em R^k .

Seja $\hat{F}(x)$ uma estimativa de $F(x)$ com base na amostra A .

Quando não se especifica $F(x)$, por razões óbvias, refere-se ao processo como “bootstrap” não paramétrico.

2.4.3 O ‘bootstrap’ na regressão

Souza (1998) afirma que é importante, em regressão não-linear, o estabelecimento de intervalos de confiança para os parâmetros do modelo. Segundo ele, a técnica “bootstrap” pode fornecer de modo automático a escala de mensuração adequada relativamente à qual é possível determinar intervalos de confiança confiáveis.

Considere-se o modelo de regressão geral

$$y_t = f(x_t, \theta^0) + \varepsilon_t. \quad (61)$$

em que os erros são independentes, homoscedásticos e têm média nula. Suponha que as hipóteses suficientes para garantir os resultados assintóticos envolvendo o estimador de mínimos quadrados se verifiquem.

Nesse contexto, o princípio “bootstrap” funciona em problemas de regressão desde que alguns cuidados sejam tomados no processo de seleção das amostras.

A forma mais simples do “bootstrap” paramétrico para o modelo de regressão não-linear envolve a hipótese de normalidade. Tomam-se amostras \mathcal{E}_t^* da distribuição $N(0, \sigma^2)$ em que σ^2 é o erro médio quadrático da regressão não-linear e geram-se respostas y_t^* , satisfazendo

$$y_t^* = f(x_t, \hat{\theta}) + \mathcal{E}_t^*. \quad (62)$$

As repetições “bootstrap” $\hat{\theta}_*$ e σ_*^2 são obtidas minimizando-se a soma de quadrados residuais

$$\sum_{t=1}^n (y_t^* - f(x_t, \theta))^2.$$

Ratkowsky (1983) e Seber & Wild (1989) exploram esta forma do “bootstrap” paramétrico no diagnóstico de regressão não-linear.

Efron & Tibshirani (1993) descrevem dois métodos de “bootstrap” não paramétricos para gerar amostras para o modelo de regressão não-linear. No primeiro, de “bootstrap” não paramétrico, envolve amostragem aleatória com repetição dos pares (x_t, y_t) , $t = 1, \dots, n$ e o cálculo do estimador de mínimos quadrados para os pares (x_t^*, y_t^*) , $t = 1, \dots, n$ gerados nessas amostras.

Esta abordagem explora menos a estrutura do modelo de regressão, mas, em contrapartida, permite uma análise conjunta do problema de regressão (X aleatório!).

Na segunda forma, determina-se o estimador de mínimo quadrados de

$$\hat{\theta} \text{ e os resíduos } \hat{e}_t = y_t - f(x_t, \hat{\theta}).$$

As regressões “bootstrap” se obtêm fazendo

$$y_t^* = f(x_t, \hat{\theta}) + \hat{e}_t^* \quad (63)$$

em que os erros

$$\hat{e}_t^*, \quad t = 1, \dots, n,$$

são escolhidos com probabilidade uniforme $1/n$ na seqüência $\hat{e}_1, \dots, \hat{e}_n$. É importante que os resíduos de mínimos quadrados \hat{e}_t tenham sido centrados em relação a sua média.

Essa técnica de “bootstrap” nos resíduos é utilizada em qualquer situação, tendo conhecimento ou não da distribuição dos erros, também em situações em que se torna difícil à obtenção de expressões analíticas, qualquer algoritmo de “bootstrap” tem de contemplar sua hipótese de interesse.

A técnica “bootstrap” inicia-se quando se retira uma amostra de tamanho n da população e, a partir dessa amostra, obtêm-se M subamostras de tamanho n , que serão denominadas reamostras, denotadas por R_1, \dots, R_M .

Manly (1998) recomenda que M seja maior que 1000. A partir dessas subamostras calculam-se a(s) estimativas dos k parâmetros de interesse e, para cada parâmetro, constrói-se o histograma baseado nas M estimativas, denominado distribuição empírica.

2.5 Comparação dos modelos de regressão

2.5.1 Amplitude de um intervalo de confiança

Na visão de Draper & Smith (1998), a amplitude de um intervalo de confiança é dada pela diferença entre o limite superior (LS) e o limite inferior do mesmo (LI).

A comparação entre os intervalos de confiança dos parâmetros dos modelos não-lineares, um utilizando a metodologia “bootstrap” e o outro utilizando a estatística assintótica usual, pode ser realizada considerando sua amplitude. Aquele que possuir uma amplitude menor apresentará uma maior precisão na estimação dos parâmetros do modelo. Para isso, torna-se necessário obter o erro padrão dos parâmetros do modelo não-linear.

Nesta visão, Draper & Smith (1998) apresentam as estimativas da matriz assintótica de variâncias e covariâncias da seguinte forma:

$$\begin{aligned}\hat{V}(\hat{\theta}) &= (X'X)^{-1} \hat{\sigma}^2 \\ \hat{V}(\hat{\theta}) &= (X'X)^{-1} \text{QME}\end{aligned}\quad (64)$$

em que o quadrado médio do erro (QME) é uma estimativa da variância populacional dada por σ^2 .

$$\hat{V}(\hat{\theta}) = \begin{pmatrix} \hat{V}(\hat{\theta}_1) & \dots & Cov(\hat{\theta}_1, \hat{\theta}_j) \\ \vdots & \ddots & \vdots \\ Cov(\hat{\theta}_1, \hat{\theta}_j) & \dots & \hat{V}(\hat{\theta}_j) \end{pmatrix}$$

com $j = 1, 2, \dots, p$, em que p é o número de parâmetros.

Logo, o erro padrão da estimativa do parâmetro θ_j é dada por:

$$s(\hat{\theta}_j) = \sqrt{\hat{V}(\hat{\theta}_j)} \quad (65)$$

Dessa forma, define-se intervalo de confiança para θ_j como:

$$IC(\theta_j) = \hat{\theta}_j \pm t_{(g.l.erro, \frac{\alpha}{2})} \sqrt{\hat{V}(\hat{\theta}_j)} \quad (66)$$

2.5.2 Quadrado médio do erro (QME)

De acordo com Ferreira (2005), utilizando-se o resultado demonstrado em Charnet et al. (1999),

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2. \quad (67)$$

Pode-se especificar as parcelas da seguinte forma:

- a) $\sum_{i=1}^n (y_i - \bar{y})^2$ é a soma de quadrados total (ajustada), denominada por SQT, representando a variação total de Y em torno de sua média;
- b) $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ é a soma de quadrados do erro, denominada por SQE, representando a variação de Y em torno da reta;
- c) $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ é a soma de quadrados de regressão, denominada por SQR_{eg} , representando a variação das esperanças específicas de Y, dado x, em torno de sua média.

Expressando-se a equação (67) tem-se:

$$SQT = SQE + SQR_{eg} \quad (68)$$

utilizando-se o resultado abaixo:

$$\frac{\hat{\theta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\hat{\sigma}^2} = \frac{SQR_{eg}}{\frac{SQE}{n-2}} \quad (69)$$

devidamente demonstrado por Chanet et al. (1999), obtém-se:

$$\hat{\sigma}^2 = \frac{SQE}{n-2} \quad (70)$$

$$\hat{\sigma}^2 = \text{QME}$$

QME é o quadrado médio dos erros.

O resultado acima além de estimar σ^2 , é um bom estimador para σ^2 , por meio dos componentes de variância utilizando-se o método dos momentos, que pode ser representado em uma tabela de análise de variância. Quanto menor QME, maior será qualidade do ajuste do modelo, sendo um bom indicador para comparações de modelos ajustados.

2.5.3 Coeficiente de determinação (R^2)

O coeficiente de determinação ajustado é definido por Draper & Smith (1998):

$$R^2 = \frac{SQR_{eg}}{SQT} = 1 - \frac{SQE}{SQT} \quad (71)$$

sendo $0 \leq R^2 \leq 1$, já que

$$0 \leq SQR_{eg} \leq SQT.$$

O coeficiente de determinação R^2 pode ser interpretado como a proporção da variabilidade presente nas observações da variável resposta y , que é explicada pela variável regressora x , no modelo de regressão ajustado aos dados.

De acordo com Souza (1998), o coeficiente de determinação deve ser empregado com cuidado, devido aos seguintes motivos:

- i) é possível aumentar o valor de R^2 por meio da adição de novas variáveis regressoras ao modelo. No entanto, apesar do maior valor para R^2 , nem sempre o novo modelo será melhor do que o primeiro;
- ii) a magnitude de R^2 depende da faixa de variação da variável regressora. É possível que R^2 seja pequeno, simplesmente porque as observações em x estão assumindo um campo de variação muito pequeno para que o relacionamento entre y e x possa ser detectado. Por outro lado, um valor elevado para R^2 pode ocorrer porque x assumiu um campo de variação irrealisticamente grande, o qual não ocorre em condições usuais;
- iii) R^2 não mede a adequação do modelo já que, mesmo que x e y não apresentem um relacionamento linear, é possível que R^2 assumam valores elevados;
- iv) R^2 não mede a magnitude da inclinação da reta de regressão. Um grande valor para R^2 não implica em um valor elevado para a inclinação da reta.

2.5.4 Coeficiente de determinação ajustado ($R^2_{ajustado}$)

Hofmam & Vieira (1998) afirmam que o coeficiente de determinação é uma medida descritiva da qualidade do ajustamento obtido.

Entretanto, o valor do coeficiente de determinação depende do número de observações da amostra, tendendo a crescer quando n diminui; no limite, para $n=2$, teria-se sempre $R^2 = 1$, pois dois pontos determinam uma reta e os desvios são, portanto, nulos.

Uma tentativa de superar esse inconveniente é utilizar o coeficiente de determinação corrigido para graus de liberdade indicado por $R^2_{ajustado}$, para o número de parâmetros da regressão, da seguinte forma:

$$R^2_{ajustado} = 1 - \left[\frac{n-1}{n-2} \right] \left[1 - \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right] \quad (72)$$

3 MATERIAL E MÉTODOS

3.1 Material

Para ilustrar a metodologia utilizaram-se dados sobre a velocidade da reação enzimática obtidas de Treolar (1974) em sua tese.

O número de “counts” por minuto de produto radioativo da reação foi medido como uma função da concentração de substrato em partes por milhão (ppm) e destes “counts”, foi calculada a razão inicial ou velocidade da reação ($counts/min^2$).

O experimento foi conduzido com as enzimas tratadas com o substrato Puromycin. A velocidade é assumida como dependente da concentração de substrato, de acordo com a equação de Michaelis-Menten.

TABELA 1 Concentração do substrato Puromycin (ppm) e velocidade inicial da reação ($counts/min^2$).

Concentração de substrato (ppm) [S]	Velocidade da reação (counts/min) v_0
0,02	76
0,02	47
0,06	97
0,06	107
0,11	123
0,11	139
0,22	159
0,22	152
0,56	191
0,56	201
1,10	207
1,10	200

Fonte: Treolar (1974).

3.2 Métodos

3.2.1 Derivadas parciais, matriz X e $X'X$ do modelo não-linear de Michaelis-Menten

Nesta etapa, será utilizado o modelo hiperbólico de Michaelis-Menten (1913), que fornece uma estimativa para a constante de afinidade do substrato para com a enzima (K_m) e o valor máximo (V_{\max}), quando a enzima está completamente saturada com o substrato descrito pelo modelo na equação (19)

$$v_o = \frac{V_{\max} [S]}{K_m + [S]}$$

As derivadas parciais deste modelo são dadas por:

$$\frac{\partial v_o}{\partial V_{\max}} = \frac{[S]}{K_m + [S]}$$
$$\frac{\partial v_o}{\partial K_m} = -\frac{V_{\max} [S]}{(K_m + [S])^2}$$

Portanto, a matriz de derivadas parciais $F = F(\theta) = {}_n X_2$, sendo n o número de concentrações de substrato e $p=2$, o número de parâmetros do modelo, é dada por:

$$X = \begin{bmatrix} \frac{[S]_1}{K_m + [S]_1} & -\frac{V_{\max} [S]_1}{(K_m + [S]_1)^2} \\ \vdots & \vdots \\ \frac{[S]_n}{K_m + [S]_n} & -\frac{V_{\max} [S]_n}{(K_m + [S]_n)^2} \end{bmatrix}$$

Sendo,

$$X'X = \begin{bmatrix} \frac{([S]_1)^2}{(K_m + [S]_1)^2} + \dots + \frac{([S]_n)^2}{(K_m + [S]_n)^2} & -\frac{V_{\max} ([S]_1)^2}{(K_m + [S]_1)^3} - \dots - \frac{V_{\max} ([S]_n)^2}{(K_m + [S]_n)^3} \\ -\frac{V_{\max} ([S]_1)^2}{(K_m + [S]_1)^3} - \dots - \frac{V_{\max} ([S]_n)^2}{(K_m + [S]_n)^3} & \frac{V_{\max}^2 ([S]_1)^2}{(K_m + [S]_1)^4} + \dots + \frac{V_{\max}^2 ([S]_n)^2}{(K_m + [S]_n)^4} \end{bmatrix}$$

3.2.2 Ajuste dos modelos não-lineares e lineares

Com o modelo não-linear de Michaelis-Menten, foram obtidas estimativas dos parâmetros para seis modelos, sendo dois modelos não-lineares e quatro linearizados.

A estimação dos parâmetros dos modelos linearizados foi avaliada pelo método de quadrados mínimos ordinários e utilizou como recurso computacional o software estatístico R[®] v 2.1.1 (R DEVELOPMENT CORE TEAM, 2005), por meio da função lm, anexo 1A.

A estimação dos parâmetros do modelo não-linear, via matriz de variâncias e covariâncias, foi avaliada pelo método de quadrados mínimos ordinários, considerando o algoritmo iterativo de Gauss-Newton. Utilizou-se como recurso computacional o software estatístico R[®] v 2.1.1 (R development Core Team, 2005), por meio da funções Ssmicmen (modelo de Michaelis-Menten) e nls (ajuste de modelos não-lineares) cuja a rotina é apresentada no anexo 1A.

A estimação dos parâmetros do modelo não-linear, fazendo uso da metodologia “bootstrap” e a Figura 8, utilizou SAS[®] v.3 (PROCEDURES GUIDES FOR COMPUTERS, 2004), por meio do Proc Nlin , anexo 2A.

3.2.3 Comparação dos modelos lineares e não-lineares

Para comparar os modelos não-lineares foram utilizados o valor do quadrado médio do erro (QME), o coeficiente de determinação ajustado ($R^2_{ajustado}$) e as amplitudes dos intervalos de confiança dos parâmetros.

Os modelos lineares foram comparados utilizando-se o coeficiente de determinação ajustado ($R^2_{ajustado}$) e as amplitudes dos intervalos de confiança dos parâmetros.

Os modelos lineares e não-lineares foram comparados utilizando o coeficiente de determinação ajustado ($R^2_{ajustado}$) e comparam-se as amplitudes dos intervalos de confiança dos parâmetros.

3.2.4 'Bootstrap' não paramétrico na regressão

3.2.4.1 Modelo de Michaelis-Menten genérico ajustado

Foi usado o resultado apresentado na secção 4.1.1, ou seja, o modelo de Michaelis-Menten ajustado (MM), que será generalizado e assumirá a seguinte forma:

$$\hat{v}_i^* = \frac{V_{\max}^* [S]_i^*}{K_m^* + [S]_i^*} \quad (73)$$

Em que:

$1 \leq i \leq n$: tamanho da amostra; \hat{v}_i^* : o valor predito;

$[S]_i^*$ a concentração de substrato associada à i -ésima observação;

V_{\max}^* parâmetro representando o valor máximo que v_0^* pode assumir;

K_m^* parâmetro representando a metade da V_{\max}^*

O erro será representado por E_i^* , dado por $E_i^* = \hat{v}_i^* - v_i^*$, deve-se evidenciar que os valores assumidos genericamente por $[S]_i^*$, v_i^* , V_{\max}^* e K_m^* podem assumir valores em \Re , desde que estejam inseridos no espaço de resultados possíveis para um experimento de cinética enzimática.

QUADRO 1 Modelo MM genérico ajustado pela equação (73)

\mathbf{v}_i^*	$[S]_i^*$	$\hat{\mathbf{v}}_i^*$	E_i^*
v_1^*	$[S]_1^*$	\hat{v}_1^*	E_1^*
v_2^*	$[S]_2^*$	\hat{v}_2^*	E_2^*
\vdots	\vdots	\vdots	\vdots
v_n^*	$[S]_n^*$	\hat{v}_n^*	E_n^*

3.2.4.2 Detalhes do algoritmo de ‘bootstrap’ nos resíduos

1º) Fixa \hat{v}_1^* .

2º) Sorteia um valor do E_i^* .

3º) Compõem a nova observação fixando \hat{v}_1^*

$$v_1^{**} = \hat{v}_1^* + E_i^* .$$

4º) Repetir os passos de 1 a 3 até completar uma nova amostra dos n novos valores observados. Ao final, tem-se a 1ª subamostra.

5º) Repetir o processo M vezes de forma, cada subamostra será denominada de reamostragem e denotada por R_i com $i = 1, \dots, M$.

Para cada reamostragem variando de R_1, \dots, R_M serão feitas as estimativas dos parâmetros envolvidos no modelo utilizando-se o método de mínimos quadrados ordinários, considerando o algoritmo iterativo de Gauss-Newton. A estimativa final dos parâmetros é a média da distribuição empírica será representada por histogramas.

4 RESULTADOS E DISCUSSÃO

4.1 Ajustes dos modelos não-lineares

4.1.1 Ajuste do modelo de Michaelis-Menten via aproximação assintótica

Na equação (19) incorpora-se o erro e considerando-o aditivo, obtém-se o modelo de regressão estatístico não-linear

$$v_{0_i} = \frac{V_{\max} [S]_i}{K_m + [S]_i} + \varepsilon_i \quad (74).$$

Na Figura 8, observa-se a relação entre a velocidade inicial da reação (v_0) versus a concentração de substrato $[S]$, dados obtidos por meio dos dados experimentais.

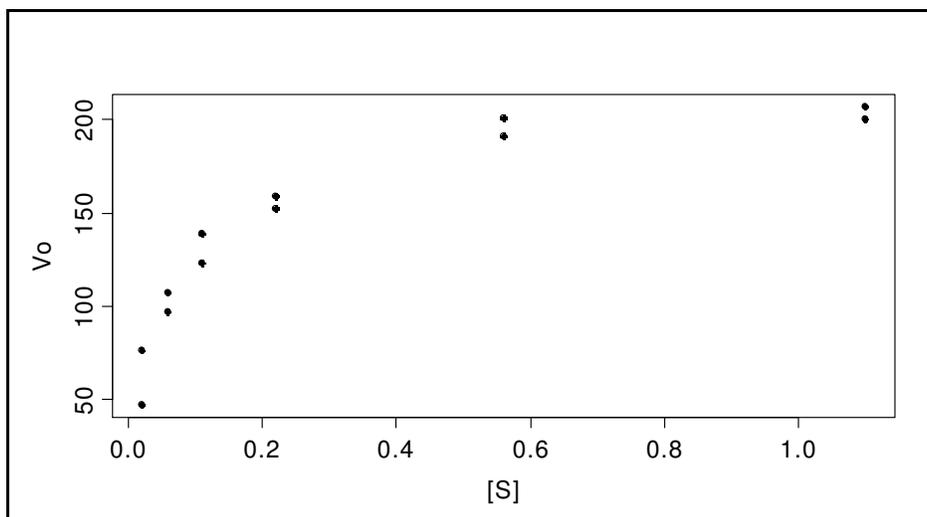


FIGURA 8 V_0 versus a [S]

Utilizando-se as estimativas dos parâmetros K_m e V_{max} de um dos modelos linearizados, obtidas por meio do método de mínimos quadrados por meio do sistema de equações normais (SEN), o modelo de Lineawer-Burk, que será posteriormente ajustado na sub-seção 4.1.1 têm-se,

$$\hat{K}_m = 0,04838$$

$$\hat{V}_{max} = 195,69$$

que são utilizadas como valores iniciais para a estimação dos parâmetros K_m e V_{max} do modelo não-linear de Michaelis-Menten, pelo método de Gauss-Newton, representado na Tabela 2.

TABELA 2 Resultados do algoritmo do método de Gauss-Newton

Iteração	QME	V_{\max}	K_m
1 ^a	1920,643	195,802	0,04840
2 ^a	1207,887	210,888	0,06136
3 ^a	1195,604	212,490	0,06380
4 ^a	1195,450	212,664	0,06409
5 ^a	1195,449	212,682	0,06411
6 ^a	1195,449	212,683	0,06412

Fonte: Resultados do software estatístico R

Foram necessárias seis iterações para a convergência do algoritmo de Gauss-Newton, dando as estimativas dos parâmetros de acordo com os critérios estabelecidos para a convergência na secção 2.3.3.6.

Calculou-se a matriz de variâncias e covariâncias do modelo cujo resultado é dado por,

$$\begin{matrix} & V_{\max} & K_m \\ V_{\max} & \begin{bmatrix} 48,26285 & 0,044014 \end{bmatrix} \\ K_m & \begin{bmatrix} 0,044014 & 0,000068 \end{bmatrix} \end{matrix}$$

também as matrizes X e $X'X$,

$$X = \begin{bmatrix} 0,000094027 & -0,000000028 \\ 0,000094027 & -0,000000028 \\ 0,000282029 & -0,000000085 \\ 0,000282029 & -0,000000085 \\ 0,000516932 & -0,000000155 \\ 0,000516932 & -0,000000155 \\ 0,001033333 & -0,000000311 \\ 0,001033333 & -0,000000311 \\ 0,002621050 & -0,000000789 \\ 0,002621050 & -0,000000789 \\ 0,005145391 & -0,000001543 \\ 0,005145391 & -0,000001543 \end{bmatrix} \text{ e } X'X = \begin{bmatrix} 0,0000695897 & -0,000000021 \\ -0,000000021 & 0,6268814000 \end{bmatrix}$$

Os estimadores de mínimos quadrados ordinários para os parâmetros K_m e V_{\max} são apresentados na Tabela 3.

TABELA 3 Estimativa dos parâmetros com os seus respectivos erros padrões e intervalos de confiança

	Estimativas	Erro Padrão	IC (95%)	
K_m	0,0641	0,00828	0,0457	0,0826
V_{\max}	212,7	6,9472	197,2	228,2

Fonte: Resultados do software Estatístico R

A soma de quadrados de erros é dada por $SQE = 1.195,45$. Um estimador da variância é dado pelo quadrado médio do erro (QME),

$$QME = S^2 = \frac{SQE(\hat{\theta})}{n-p} \quad \text{onde} \quad QME = \frac{1.195,45}{10} = 119,4$$

Verificou-se a adequabilidade do modelo, baseada na secção 2.3.2.7, que verifica as pressuposições básicas. Considerando-se o modelo de Michaelis-Menten (74) e os dados da tabela 1, o modelo ajustado em que $i=1, \dots, 12$ é dado por,

$$\hat{v}_i = \frac{212,7[S]_i}{0,0641 + [S]_i} \quad (\text{MM})$$

Na Figura 9 observa-se o gráfico normal probabilístico ou quantil teórico (a) que apresenta boa linearidade, indicando, portanto, boa normalidade para os resíduos padronizados. Observando-se o gráfico dos resíduos padronizados versus os valores ajustados (b), nota-se que os pontos estão distribuídos aleatoriamente em torno do zero, implicando que a variância dos erros é constante e a média igual a zero.

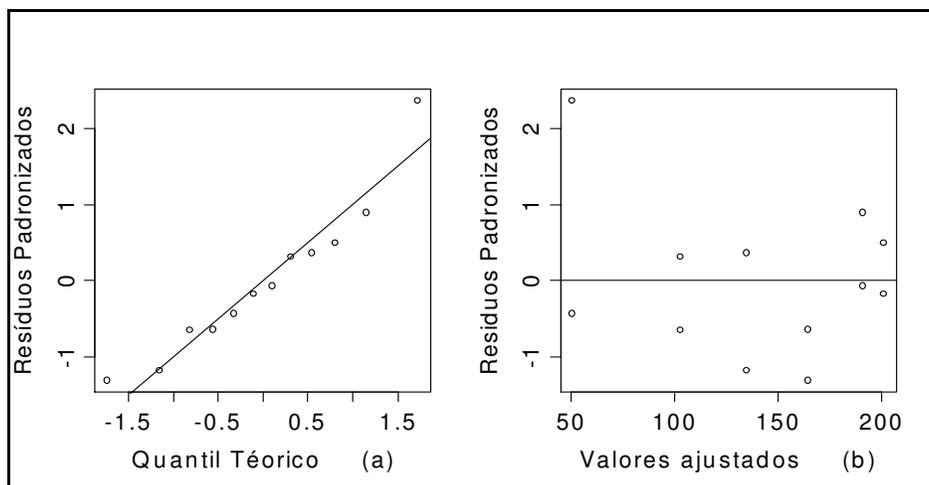


FIGURA 9 Gráfico do quantil teórico e dos resíduos padronizados versus os valores ajustados do MM

Para a verificação da pressuposição da normalidade dos resíduos, também foi realizado o teste Shapiro-Wilk, apresentando como resultado um p – valor = 0,3698

O coeficiente de determinação ajustado do modelo

$$R_{ajustado}^2 = 0,9574$$

explica 95,7% da variabilidade dos dados, o que é muito satisfatório. Por esse critério evidencia-se que o modelo proposto é adequado para os dados.

O modelo ajustado (MM) representado na Figura 10

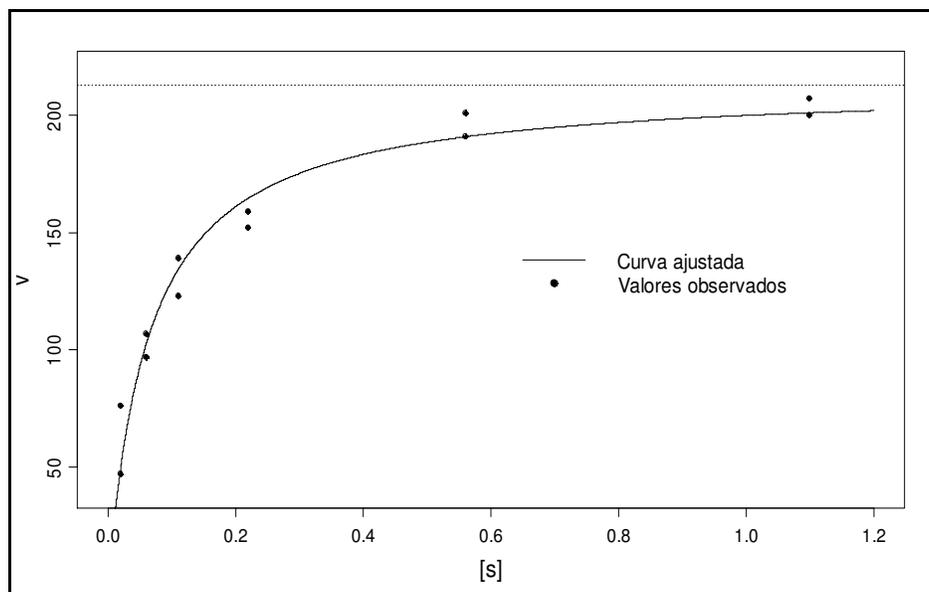


FIGURA 10 Gráfico do ajuste da equação de regressão para o modelo de Michaelis-Menten (MM)

4.1.2 Ajuste do modelo de Michaelis-Menten via técnica de ‘bootstrap’

QUADRO 2 Modelo ajustado de Michaelis-Menten (MM)

Observação	Concentração de substrato [S]	Velocidade inicial V_0	Valor predito \hat{v}_0	Resíduo ε
1	0,02	76	50,566	25,4340
2	0,02	47	50,566	-3,5660
3	0,06	97	102,811	-5,8110
4	0,06	107	102,811	4,1890
5	0,11	123	134,362	-11,3616
6	0,11	139	134,362	4,6384
7	0,22	159	164,685	-5,6847
8	0,22	152	164,685	-12,6847
9	0,56	191	190,833	0,1671
10	0,56	201	190,833	10,1671
11	1,10	207	200,969	6,0312
12	1,10	200	200,969	-0,9688

Seguindo o método descrito na secção 3.2.4.1, tem-se o Quadro 3.

QUADRO 3 1ª reamostragem de ‘bootstrap’

Observação	Concentração de substrato [S]	Velocidade inicial V_0
1	0,02	39,204401
2	0,02	39,204401
3	0,06	90,126301
4	0,06	112,97809
5	0,11	134,52873
6	0,11	128,67694
7	0,22	159,00000
8	0,22	169,32306
9	0,56	196,86410
10	0,56	195,02191
11	1,10	195,15780
12	1,10	226,40277

Dando prosseguimento à metodologia foram obtidas 2.000 reamostras, que produziram as estimativas para os parâmetros K_m e V_{max} , indicadas Tabela 4.

TABELA 4 Estimativas dos parâmetros com os seus respectivos erros-padrão e intervalos de confiança via técnica de ‘bootstrap’

	Estimativas	Erro Padrão	IC (95%)	
K_m	0,0630	0,007	0,049	0,080
V_{max}	213,27	6,738	201,96	228,31

Fonte: Resultados obtidos do software estatístico SAS

A soma de quadrados de erros é dada por $SQE = 1.195,45$. Um estimador da variância é dado por,

$$QME = S^2 = \frac{SQE(\hat{\theta})}{n - p}$$
$$QME = \frac{1.195,45}{10} = 119,45$$

A partir dos resultados da Tabela 4 e da metodologia descrita, o modelo ajustado pela técnica de 'bootstrap' por meio da simulação é dado:

$$\hat{V}_{0_i} = \frac{213,27[S]_i}{0,0630 + [S]_i}, \quad i = 1, \dots, 12 \quad (\text{MB})$$

Foi realizada no modelo ajustado de Michaelis-Menten (MB) via técnica de 'bootstrap' a verificação da pressuposição da normalidade dos resíduos, realizando-se o teste Shapiro-Wilk, que apresentou como resultado:

$$p - \text{valor} = 0,3698$$

O coeficiente de determinação ajustado do modelo

$$R^2_{\text{ajustado}} = 0,9574$$

explica 95,7% da variabilidade dos dados, o que é considerado muito satisfatório. Por esse critério, pode-se evidenciar que o modelo proposto é adequado para os dados.

Com os dados simulados via técnica de 'bootstrap' foram obtidos os histogramas das distribuições empíricas de

$$V_{\max}, K_m \text{ e } QME$$

confirmando-se a normalidade do ajuste dado pelo teste de Shapiro-Wilk.

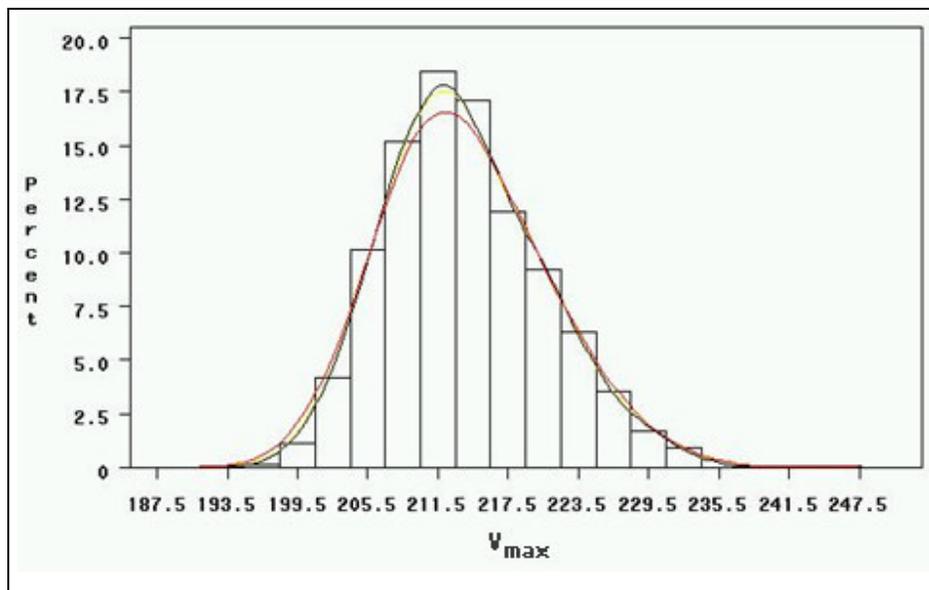


FIGURA 11 Histograma de V_{\max}

Observa-se, na Figura 11 que a distribuição empírica de \hat{V}_{\max} , além do propósito de informar a respeito do V_{\max} , independente da distribuição dos erros ou da forma da distribuição, se aproxima da distribuição normal, validando a estimativa do parâmetro dado pelo modelo de regressão não-linear de Michaelis-Menten (MM).

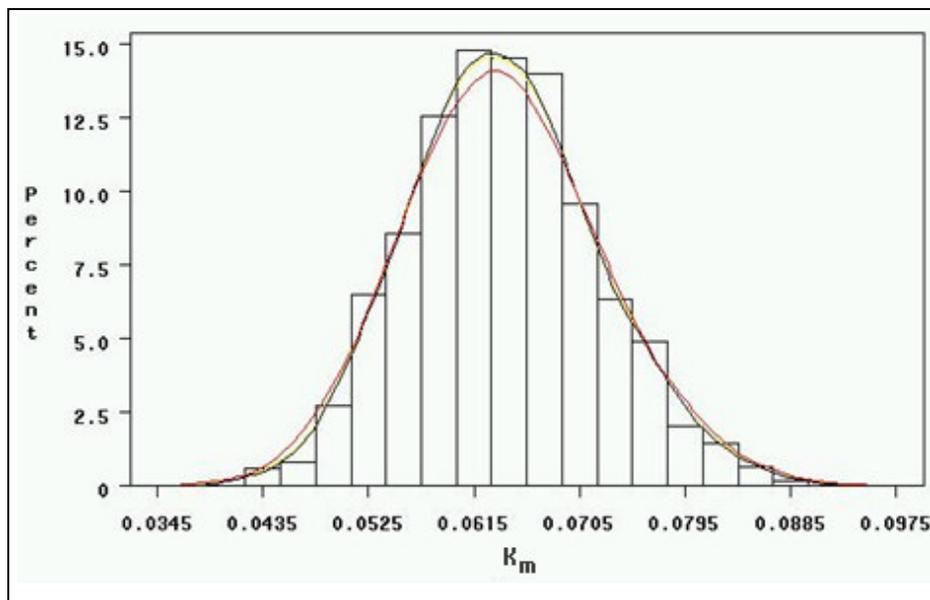


FIGURA 12 Histograma de K_m

Nota-se, na Figura 12 que a distribuição empírica de \hat{K}_m , além do propósito de informar a respeito do K_m , independente da distribuição dos erros ou da forma da distribuição, se aproxima da distribuição normal, validando a estimativa do parâmetro dado pelo modelo de regressão não-linear de Michaelis-Menten (MM).

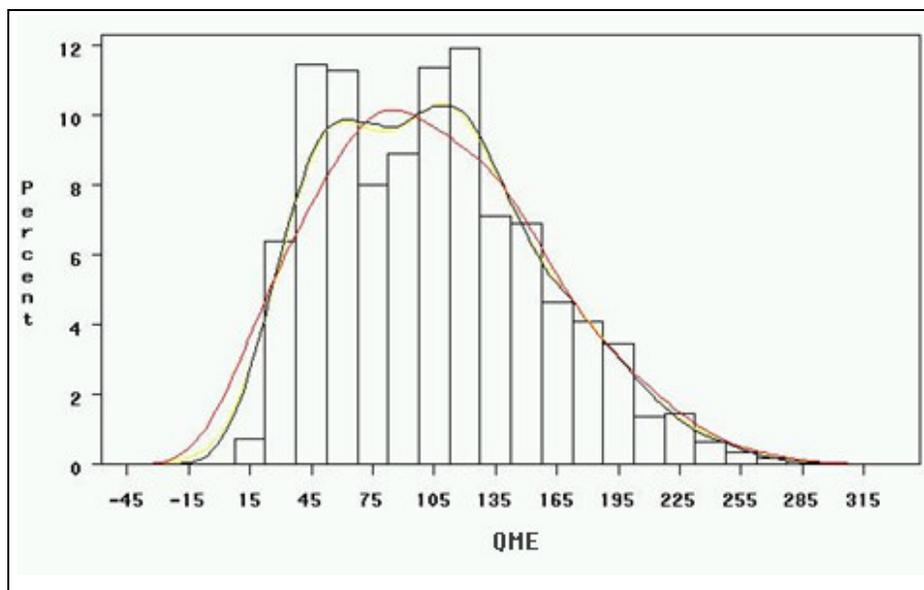


FIGURA 13 Histograma de *QME*

Nota-se, na Figura 13, que a distribuição empírica de *QME*, além do propósito de informar a respeito do *QME*, independente da distribuição dos erros ou da forma da distribuição, se aproxima da distribuição normal, validando a estimativa do parâmetro dado pelo modelo de regressão não-linear de Michaelis-Menten (MM).

4.2 Ajuste dos modelos linearizados

De conformidade com a secção 2.3.3.2, sobre transformações e nas subsecções subsequentes 2.3.3.2.1; 2.3.3.2.2; 2.3.3.2.3 e 2.3.3.2.4, acrescentando-se os erros relacionados aos modelos, têm-se os modelos estatísticos lineares.

Estes modelos serão colocados na forma $y = ax + b$, e as estimativas dos parâmetros a e b serão obtidas pelo método de mínimos quadrados ordinários (SEN), com seus respectivos intervalos de confiança (IC).

Por intermédio das operações inversas referenciadas por Bussab (1988), obter-se-ão as estimativas, e, os IC para K_m e V_{\max} .

4.2.1 Lineweaver-Burk

Na equação (44), tem-se o modelo, ao qual será incorporado o erro, denotado por ϵ_{LB}

$$\frac{1}{v_{0_i}} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}} \frac{1}{[S]_i} + \epsilon_{LB} \quad (75)$$

por intermédio de transformações adequadas,

$y = \frac{1}{v_{0_i}}$	e	$x = \frac{1}{[S]_i}$
$a = \frac{1}{V_{\max}}$	e	$b = \frac{K_m}{V_{\max}}$

obtêm-se o modelo linear estatístico de Lineweaver-Burk (LB)

$$y = a + bx + \epsilon_{LB} \quad (76)$$

Os estimadores de mínimos quadrados ordinários para os parâmetros a e b são apresentados na Tabela 5.

TABELA 5 Estimativas dos parâmetros a e b com os seus respectivos erros padrões e intervalos de confiança

	Estimativas	Erro padrão	IC(95%)	
a	0,0051072	0,0007040	0,0035385755	0,0066757878
b	0,0002472	0,0000321	0,0001756875	0,0003187544

Fonte: Resultados obtidos do software estatístico R

A soma de quadrados de erros é dada por:

$$SQE = 0,001892.$$

Um estimador da variância é dado por

$$QME = S^2 = \frac{SQE(\hat{\theta})}{n - p}$$

$$QME = \frac{0,001892}{10} = 0,000189$$

Considerando o modelo transformado (76) e os dados da Tabela 1, o modelo ajustado pelo método de mínimos quadrados ordinários utilizando o (SEN)

$$\hat{Y}_i = 0,0051072 + 0,0002472x_i. \text{ (LB)}$$

em que $i = 1, \dots, 12$

Na Figura 14 observa-se o gráfico normal probabilístico ou quantil teórico (c) que não apresenta boa linearidade, indicando, portanto, não normalidade para os resíduos padronizados. Observando-se o gráfico dos resíduos padronizados versus os valores ajustados (d), nota-se que os pontos não estão distribuídos aleatoriamente em torno do zero, implicando que a variância dos erros não é constante e a média não é igual a zero.

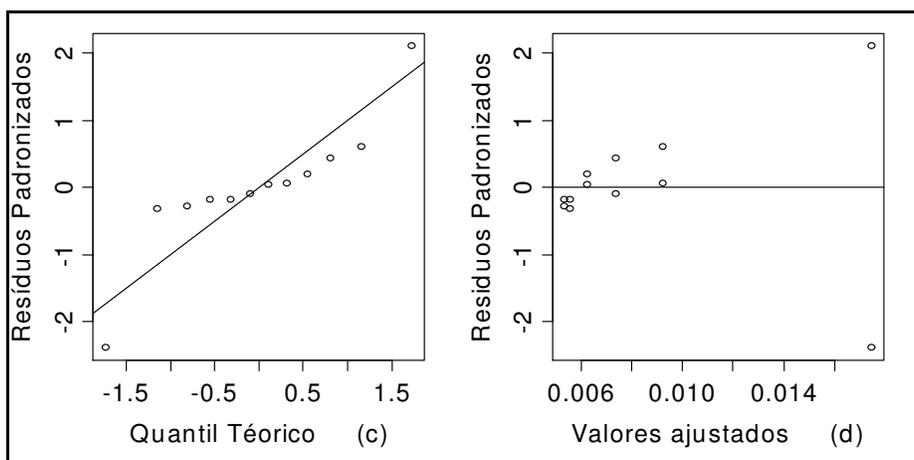


FIGURA 14 Gráfico do quantil teórico e dos resíduos padronizados versus os valores ajustados do LB

Para verificação da pressuposição da normalidade dos resíduos, também foi realizado o teste Shapiro-Wilk, apresentando

p – valor = 0,0250

o coeficiente de determinação ajustado do modelo

$$R^2_{ajustado} = 0,8413$$

explica 84,13% da variabilidade dos dados, o que seria razoavelmente satisfatório por esse critério, se não ocorresse a violação das pressuposições básicas, sendo o modelo ajustado (LB) representado na Figura 15.

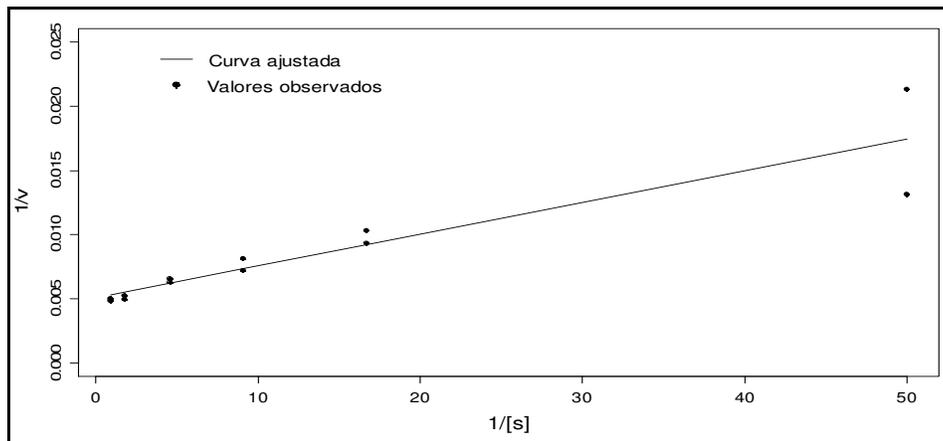


FIGURA 15 Gráfico do ajuste da equação de regressão para o modelo de Lineweaver-Burk (LB)

Sejam os intervalos de confiança de a e b,

$$IC(a : 95\%) = [0,0035385755; 0,0066757878]$$

$$IC(b : 95\%) = [0,0001756875; 0,0003187544]$$

Utilizando as transformações necessárias, obtêm-se as estimativas para K_m e V_{\max} ,

$$a = \frac{1}{V_{\max}} \quad \therefore V_{\max} = \frac{1}{a}$$

$$V_{\max} = \frac{1}{0,0051072} = 195,802$$

$$K_m = V_{\max} \cdot b$$

$$K_m = 195,802 \cdot 0,0002472 = 0,048402$$

Por meio de operações inversas obtêm-se os IC para V_{\max} e K_m , onde

$$L_{i(V_{\max})} = \frac{1}{0,0066757878} = 149,79$$

$$L_{s(V_{\max})} = \frac{1}{0,0035385755} = 282,59$$

$$L_{i(K_m)} = L_{i(V_{\max})} \cdot L_{s(b)}$$

$$L_{i(K_m)} = 149,79 \cdot 0,0003187544 = 0,047$$

$$L_{s(K_m)} = L_{s(V_{\max})} \cdot L_{i(b)}$$

$$L_{s(K_m)} = 282,59 \cdot 0,0001756875 = 0,049$$

$IC(V_{\max} : 95\%) = [149,79 ; 282,59]$, cuja amplitude é igual a 134,80

$IC(K_m : 95\%) = [0,0477 ; 0,0496]$, cuja amplitude é igual a 0,002

4.2.2 Hanes-Wolf

Na equação (45) tem-se o seguinte modelo ao qual será incorporado o erro que será denominado por ε_{HW}

$$\frac{[S]}{v_0} = \frac{K_m}{V_{\max}} + \frac{1}{V_{\max}} \cdot [S] + \varepsilon_{HW} \quad (77)$$

em conseqüência,

$$y = \frac{[S]_i}{v_{o_i}} \quad \text{e} \quad x = [S]_i$$
$$a = \frac{K_m}{V_{\max}} \quad \text{e} \quad b = \frac{1}{V_{\max}}$$

obté-m-se,

$$y = a + bx + \varepsilon_{HW} \quad (78)$$

Os estimadores de mínimos quadrados ordinários para os parâmetros a e b são apresentados na Tabela 6.

TABELA 6 Estimativas dos parâmetros a e b com os seus respectivos erros padrões e intervalos de confiança.

	Estimativas	Erro Padrão	IC (95%)	
a	0,000314	0,000034	0,0002380796	0,0003901309
b	0,004625	0,000066	0,0044772118	0,0047727590

Fonte: Resultados obtidos do software estatístico R

A soma de quadrados de erros é dada por $SQE = 0,000088$, um estimador da variância é dado por,

$$QME = S^2 = \frac{SQE(\hat{\theta})}{n - p}$$

$$QME = \frac{0,000088}{10} = 0,000009$$

considerando o modelo transformado (78) e os dados da tabela 1, o modelo ajustado pelo método de mínimos quadrados (SEN), em que $i = 1, \dots, 12$ e HW significa Hanes-Wolf é dado por:

$$\hat{Y}_i = 0,000314 + 0,004625x_i + \varepsilon_{HW} \quad (\text{HW})$$

Na Figura 16 observa-se o gráfico do quantil teórico (e) indicando, normalidade para os resíduos padronizados, no gráfico dos resíduos padronizados versus os valores ajustados (f), os pontos estão distribuídos aleatoriamente em torno do zero, implicando que a variância dos erros é constante e a média é igual a zero.

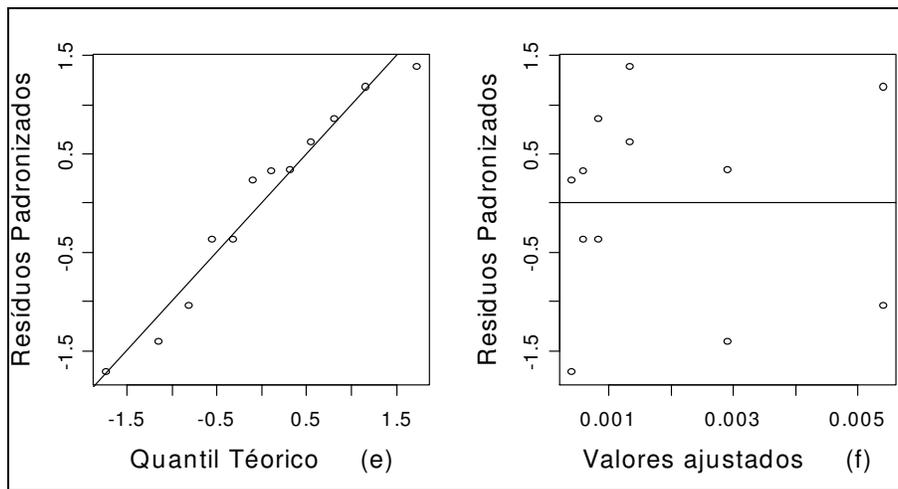


FIGURA 16 Gráfico do ajuste do modelo de regressão de HW

Para verificação da pressuposição da normalidade dos resíduos, também foi realizado o teste Shapiro-Wilk, apresentando como resultado um p -valor = 0,6234.

O coeficiente de determinação ajustado do modelo

$$R^2_{ajustado} = 0,9977$$

explica 99,777% da variabilidade dos dados, o que é muito satisfatório, assim, por esse critério, evidencia-se que o modelo proposto é adequado para os dados.

Analogamente ao modelo LB, utilizou-se o software Estatístico R e obtiveram-se as estimativas de,

$K_m = 0,0679$
$V_{\max} = 216,22$

e os intervalos de confiança

$IC(V_{\max} : 95\%) = [209,52 ; 223,35]$
cuja amplitude é do IC é igual a 13,83
$IC(K_m : 95\%) = [0,053 ; 0,082]$
cuja amplitude é do IC é igual a 0,029

4.2.3 Woolf-Augustinson-Hofstee

Na equação (46) tem-se o seguinte modelo ao qual será incorporado o erro que será denominado por ε_{WAH}

$$v_0 = V_{\max} - K_m \frac{v_0}{[S]} + \varepsilon_{WAH} \quad (79)$$

em consequência,

$$y = v_{0_i} \quad \text{e} \quad x = \frac{v_{0_i}}{[S]_i}$$
$$a = V_{\max} \quad \text{e} \quad b = -K_m$$

Obtêm-se,

$$y = a + bx + \varepsilon_{WAH} \quad (80)$$

Os estimadores de mínimos quadrados ordinários para os parâmetros a e b são apresentados na Tabela 7.

TABELA 7 Estimativas dos parâmetros a e b com os seus respectivos erros padrões e intervalos de confiança.

	Estimativas	Erro padrão	IC (95%)	
a	193,867711	11,761537	167,6613256	220,07404893
b	-0,043524	0,007449	-0,06012097	-0,02692623

Fonte: Resultados obtidos do software estatístico R

A soma de quadrados de erros é dada por $SQE = 26,44$, um estimador da variância é dado por,

$$QME = S^2 = \frac{SQE(\hat{\theta})}{n - p}$$

$$QME = \frac{26,44}{10} = 2,644$$

considerando o modelo transformado (80) e os dados da Tabela 1, o modelo ajustado pelo método de mínimos quadrados (SEN), em que $i = 1, \dots, 12$ e WAF significa Woolf-Augustinson-Hofstee, é dado por:

$$\hat{Y}_i = 193,86771 - 0,043524x_i + \varepsilon_{WAF} \quad (\text{WAF})$$

Na Figura 17 observa-se o gráfico do quantil teórico (g), indicando não normalidade para os resíduos padronizados, no gráfico dos resíduos padronizados versus os valores ajustados (h), os pontos não estão distribuídos aleatoriamente em torno do zero, implicando que a variância dos erros é não constante e a média não é igual a zero.

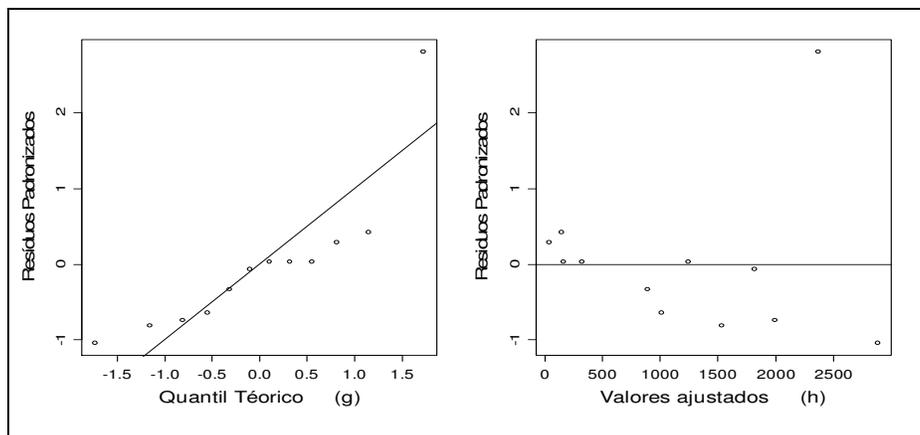


FIGURA 17 Gráfico do ajuste do modelo de regressão de WAF

Para a verificação da pressuposição da normalidade dos resíduos, foi realizado o teste Shapiro-Will, apresentando como resultado um

$$p - \text{valor} = 0,9997$$

O coeficiente de determinação ajustado do modelo,

$$R^2_{\text{ajustado}} = 0,7508$$

explica 75,08% da variabilidade dos dados, o que não é satisfatório, por esse critério evidencia-se que o modelo proposto não é adequado para os dados.

Analogamente ao modelo LB, utiliza-se o software Estatístico R e obtêm-se as estimativas de,

$K_m = 0,043524$
$V_{\max} = 193,86771$

e os intervalos de confiança

$IC(V_{\max} : 95\%) = [167,66 ; 220,07]$ cuja amplitude é do IC é igual a 52,41
$IC(K_m : 95\%) = [0,027 ; 0,060]$ cuja amplitude é do IC é igual a 0,033

4.2.4 Eadie-Scatchard

Na equação (47) tem-se o seguinte modelo ao qual será incorporado o erro que será denominado por ε_{ES}

$$\frac{v_0}{[S]} = \frac{V_{\max}}{K_m} - \frac{1}{K_m} \cdot v_0 + \varepsilon_{ES} \quad (81)$$

em conseqüência,

$$\begin{array}{l} y = \frac{v_{0_i}}{[S]_i} \quad \text{e} \quad x = v_{0_i} \\ a = \frac{V_{\max}}{K_m} \quad \text{e} \quad b = -\frac{1}{K_m} \end{array}$$

obtêm-se,

$$y = a + bx + \varepsilon_{ES} \quad (82)$$

Os estimadores de mínimos quadrados ordinários para os parâmetros a e b são apresentados na Tabela 8.

TABELA 8 Estimativas dos parâmetros a e b com os seus respectivos erros padrões e intervalos de confiança.

	Estimativas	Erro Padrão	IC (95%)	
<i>a</i>	3717,323	11,761537	2698,16736	4736,47853
<i>b</i>	-17,771	3,041	-24,54741	-10,99399

Fonte: Resultados obtidos do software estatístico R

A soma de quadrados de erros é dada por $SQE = 534,6$, um estimador da variância é dado por,

$$QME = S^2 = \frac{SQE(\hat{\theta})}{n - p}$$

$$QME = \frac{534,6}{10} = 53,46$$

Considerando o modelo transformado (82) e os dados da tabela 1, o modelo ajustado pelo método de mínimos quadrados (SEN), onde $i = 1, \dots, 12$ e ES significa Eadie Scathard é dado por:

$$\hat{Y}_i = 3717,323 - 17,771x_i + \varepsilon_{ES} \quad (\text{ES})$$

Na Figura 18 observa-se o gráfico do quantil teórico (i) indicando, portanto, não normalidade para os resíduos padronizados. O gráfico dos resíduos padronizados versus os valores ajustados (j) não está distribuído aleatoriamente

em torno do zero, implicando que a variância dos erros não é constante e a média não é igual a zero.

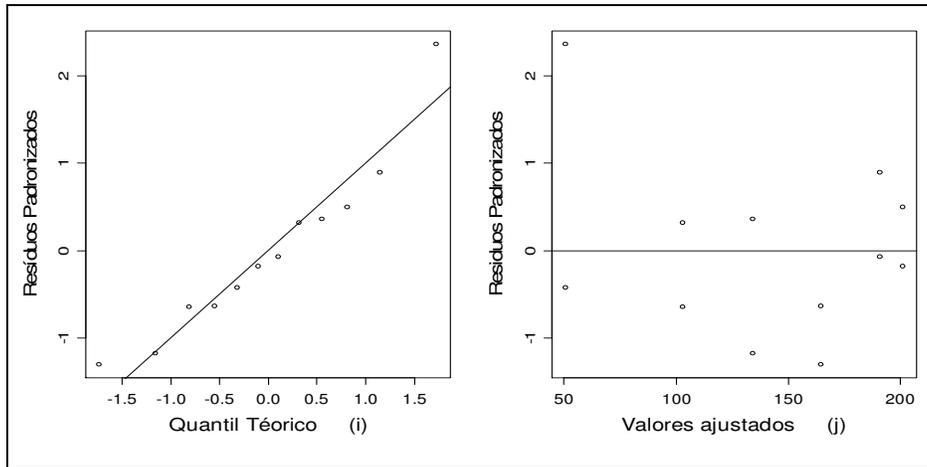


FIGURA 18 Gráfico do ajuste do modelo de regressão de ES

Para a verificação da pressuposição da normalidade dos resíduos, foi realizado o teste Shapiro-Wilk, apresentando como resultado um

$$p - \text{valor} = 0,0042$$

o coeficiente de determinação ajustado do modelo

$$R^2_{\text{ajustado}} = 0,7508$$

explica 75,08% da variabilidade dos dados, o que não é satisfatório. Por esse critério evidencia-se que o modelo proposto não é adequado para os dados.

Analogamente ao modelo LB, utiliza-se o software estatístico R obtendo-se as estimativas de

$K_m = 0,0562$
$V_{\max} = 209,18$

e os intervalos de confiança

$IC(V_{\max} : 95\%) = [100,92 ; 430,82]$ cuja amplitude é do IC é igual a 329,90
$IC(K_m : 95\%) = [0,040 ; 0,091]$ cuja amplitude é do IC é igual a 0,051

4.3 Comparação entre os modelos ajustados

Sejam os modelos estatísticos lineares e não-lineares representados nas Tabelas 9 e 10.

TABELA 9 Estimativas de V_{\max} , K_m , QME , R^2_{ajustado} e Shapiro-Wilk

Modelos	Estimativas			R^2_{ajustado}	Shapiro-Wilk
	V_{\max}	K_m	QME		
LB	195,80	0,0484	-	0,8413	0,0250
HW	216,22	0,0679	-	0,9977	0,6234
WAH	193,87	0,0484	-	0,7508	0,9997
ES	209,18	0,0562	-	0,7507	0,0042
MM	212,71	0,0641	119,45	0,9601	0,3698
MB	213,27	0,0630	119,45	0,9574	0,3698

Fonte: Resultados obtidos do software estatístico R

TABELA 10 Intervalos de confiança de V_{\max} , K_m e amplitude dos IC

Modelos	V_{\max}		K_m	
	IC (95%)	Amplitude	IC (95%)	Amplitude
LB	[149,79 ; 282,59]	134,80	[0,047 ; 0,049]	0,002
HW	[209,52 ; 223,35]	13,83	[0,053 ; 0,082]	0,029
WAH	[167,66 ; 220,07]	52,41	[0,027 ; 0,060]	0,033
ES	[100,92 ; 430,82]	329,90	[0,040 ; 0,091]	0,051
MM	[197,30 ; 229,29]	31,99	[0,047 ; 0,086]	0,039
MB	[201,96 ; 228,31]	26,35	[0,049 ; 0,080]	0,031

Fonte: Resultados obtidos do software estatístico R

Dentre os modelos linearizados, o que apresentou o pior ajuste foi o modelo de Eadie-Scatchard (ES), por ter a maior amplitude de intervalo de

confiança dos parâmetros de V_{\max} e K_m , tendo sido o que apresentou o menor coeficiente de determinação ajustado (Tabela 9 e 10).

O modelo que apresentou o melhor ajuste foi o de Hanes-Wolf (HW), pois alcançou a segunda maior precisão da amplitude do K_m , distanciando-se do modelo de Lineweaver-Burk (LB) em 27 milésimos (Tabela 10). Quanto ao V_{\max} , obteve a maior precisão da amplitude relacionada aos modelos linearizados. Em relação ao R^2_{ajustado} , obteve-se a maior porcentagem superando os outros modelos linearizados, (Tabela 9) portanto, foi escolhido entre os modelos lineares como o modelo que melhor se ajustou aos dados experimentais.

Em relação aos modelos não-lineares, a diferença percebida está relacionada às amplitudes dos intervalos de confiança dos parâmetros dos modelos ajustados MM e MB, conforme Tabela 11.

TABELA 11 Amplitude dos intervalos de confiança MM e MB

Modelos	K_m	V_{\max}
MM	0,039	31,99
MB	0,031	26,35

Fonte: Resultados obtidos do software estatístico R

Pelos resultados da Tabela 11 observa-se que as amplitudes do modelo MB, ou seja, via técnica de “bootstrap”, são menores, indicando que o mesmo, por esse critério, apresenta uma maior precisão na estimação dos parâmetros, mas com diferença mínima para o modelo MM.

Esses resultados podem ser um indicativo de que os dois modelos não-lineares se adequaram de forma satisfatória aos dados experimentais.

O modelo estatístico linear de Hanes-Wolf (HW) apresentou uma maior precisão, devido às amplitudes dos intervalos de confiança dos parâmetros V_{\max} e K_m serem menores do que as amplitudes dos intervalos de confiança dos parâmetros V_{\max} e K_m do modelo estatístico não-linear de Michaelis-Menten (MB) e um maior coeficiente determinado ajustado (R^2_{ajustado}).

Diante dessas evidências pode-se indicar que o modelo estatístico linear de Hanes-Wolf (HW) foi o mais adequável aos dados experimentais.

5 CONCLUSÕES

No presente estudo e de acordo com os resultados obtidos, pode-se concluir que:

- 1) as amplitudes dos intervalos de confiança dos parâmetros servem de comparação para os modelos de regressão, sendo que o que tiver a menor amplitude é considerado o de maior precisão para as estimativas dos parâmetros. O coeficiente de determinação ajustado é utilizado para comparar os modelos de regressão, sendo que o de maior valor é considerado o que explica melhor o modelo. Neste estudo, o modelo utilizado que teve a maior precisão nas estimativas dos parâmetros V_{\max} e K_m e o maior coeficiente de determinação ajustado (R^2_{ajustado}) foi o modelo estatístico linearizado de Hanes-Wolf (HW), modelo que melhor ajustou-se aos dados experimentais;
- 2) o modelo que apresentou o pior ajuste aos dados experimentais foi o modelo estatístico linearizado de Eadie-Scatchard (ES), por possuir a maior amplitude do intervalo de confiança dos parâmetros V_{\max} e K_m e o menor coeficiente de determinação ajustado (R^2_{ajustado});
- 3) a metodologia “bootstrap” validou o modelo de Michaelis-Menten (MM) que utilizou resultados da teoria assintótica, possibilitando resultados inferenciais mais confiáveis;
- 4) a metodologia ‘bootstrap’ serve como alternativa válida para a estimação dos parâmetros de um modelo de regressão não-linear, por ter-se uma amostra pequena, considerando-se que a teoria assintótica é válida para amostras grandes;

- 5) o modelo de Michaelis-Menten (MM) ajustado utilizando resultados da teoria assintótica e o modelo de Michaelis-Menten via técnica de 'bootstrap' ajustado são adequados aos dados experimentais de forma satisfatória, embora o que teve a maior precisão nas estimativas dos parâmetros foi o MB;
- 6) em situações críticas, em que o K_m de duas enzimas estão muito próximos, e, se a escolha da enzima for a que tenha maior eficiência, essa escolha será a da enzima mais eficiente, a de menor K_m , então, a estimativa do parâmetro terá que ser a mais precisa possível.
- 7) tendo em vista os dados utilizados é possível concluir que as funções SSmicmen e nls apresentam-se como uma forma prática e objetiva de se obter estimativas para os parâmetros do modelo de Michaelis-Menten.

REFERÊNCIAS BIBLIOGRÁFICAS

- BATES, D.M.; WATTS, D.G. **Nonlinear regression analysis and its applications**. New York: J. Wiley, 1988. 365p.
- BOX, M.J. Bias in nonlinear estimation. **Journal of the Royal Statistical Society**, Série B, v.33, p.171-201, 1971.
- BURDEN, R.L.; FAIRES, J. **Análise numérica**. São Paulo: Pioneira T. Learning, 2003. 736p.
- BUSSAB, W. de O. **Análise de variância e de regressão: uma introdução**. 2.ed. São Paulo: Atual, 1988. 147p.
- CHARNET, R. et al. **Análise de modelos de regressão linear com aplicação**. Campinas, SP: Unicamp, 1999. 356p.
- CORNISH-BOWDEN, A. **Fundamentals of enzyme kinetics**. 2.ed. London: Portland, 1995a.
- CORNISH-BOWSEN, A. **Analysis of enzyme kinetic data**. New York: Oxford University, 1995b. 198p.
- CURRIE, D.J. Estimating Michaelis-Menten Parameters: Bias, Variance e o Experimental Design. **Biometrics**, v.38, n.4, p.907-919, 1982.
- DOWD, F.E.; RIGGS, D.S. A comparacion of estimates of Michaelis-Menten kinectic constants fron various linear transformation. **Journal of Biological Chemistry**, v.240, p.863-869, 1965.
- DRAPER, N.R.; SMITH, H. **Applied regression analysis**. 3.ed. New York: J.Wiley, 1998. 706p.
- EFRON, B.; TBHIRANI, R. **An introduction to the bootstrap**. New York: Chapman & Hall, 1993.

FERREIRA, D.F. **Estatística básica**. Lavras: UFLA, 2005. 664p.

GALLANT, A.R. **Nonlinear statistical models**. New York: J.Wiley, 1987. 610p.

GRAYBILL, F.A. **Matrices with applications in statistics**. Belmont: Wadsworth International Group, 1983.

GUIMARÃES, P.S. **Ajuste de curvas experimentais**. Santa Maria, RG: UFSM, 2001. 233p.

GUTTMAN, I.; MEETER, D.A. On Beale's measures of nonlinearity. **Technometrics**, v.7, p.623-637, 1965.

HOFFMANN, R.; VIEIRA, S. **Análise de regressão: uma introdução à econometria**. 3.ed. São Paulo: HUCITEC, 1998. 379p.

JENNRICH, R.I. Asymptotic properties of nonlinear least-squares estimation. **The Annals of Mathematical Statistics**, v.20, p.633-643, 1969.

JENNRICH, R.I.; RALSTON, M. L. Fitting nonlinear models to data. **Annual Rev. Biophys. Bioeng**, v.8, p.195-238, 1979.

LEHNINGER, A.L. et al. **Princípios de bioquímica**. 2.ed. São Paulo: Savier, 2000.

MAHLER, H.R.; CORDES E.H. **Biological chemistry**. New York: A Harper International 1966.

MANLY, B.F.J. **Randomization, Bootstrap and Monte Carlo methods in biology**. London: Chapman & Hall, 1998. 399p.

MAZUCHELI, J.; ACHCAR, J.A. **Modelos de regressão não-linearidades: algumas considerações importantes**. São Carlos: USP. Instituto de Ciências Matemáticas, 1997. p.1-32 Notas Didáticas.

MAZUCHELI, J.; ACHCAR, J.A. Algumas considerações em regressão não-linear. **Acta Scientiarum**, Maringá, PR, v.24, n.6, p.1761-1770, 2002.

MICHAELIS, L.; MENTEN, M.L. Die Kinetik der invertinwirkung. **Brochen**, v.49, p.333-369, 1913.

MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the theory of statistics**. 3.ed. Singaporé: Mcgraw-Hill. International, 1974.

MONTEGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. 2.ed. New York: J.Wiley, 1992.

NETER, J.; WASSERMAN, W.; KUTNER, M.H. **Applied linear statistical models-regression: analysis of variance and experimental designs**. 3.ed. Boston: R.D. Irwin, 1990.

RAAIJMAKERS, J.G.W. Statistical analysis of the Michaelis-Menten equation. **Biometrics**, v.43, p.793-803, 1987.

RATKOWSKY, D.A. **Nonlinear regression modeling: a unified practical approach**. New York: M.Dekker, 1983. v.48. 276p.

R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Áustria. ISBN 3-900051-07-0. Disponível em: <http://www.R-project.org>. Acesso em 2006.

SAS Institute. **SAS Procedures guide for computers**. 6.ed. Cary, NC, v.5, 373p. 2005.

SEARLE, S.R **Linear models**. New York: Wiley, 1971.

SEBER, G.A.F. **Linear regression analysis**. New York: Wiley, 1977.

SEBER, G.A.F.; WILD, C.J. **Nonlinear regression**. New York: J.Wiley, 1989.

SEGEL, I.H. **Enzyme kinetics**. New York: J.Wiley, 1975. 465p.

SOUZA, G. da S. **Introdução aos modelos de regressão linear e não linear.** Brasília: EMBRAPA, SP/ SEA, 1998. 489p.

STEWART, J. **Cálculo.** São Paulo: Pioneira Thomson Learning, 2002. v.2, 1150p.

TREOLAR, M.A. **Effects of puromycin on galactosyltransferase of golgi membranes.** 1974. Master's (Thesis)-University of Toronto, Toronto.

WERKEMA, M.C.C. **Análise de regressão:** como entender o relacionamento entre as variáveis de um processo. Belo Horizonte, MG: Fundação Christiano Ottoni, Escola de Engenharia da UFMG, 1996. 302p.

ANEXOS

	Pag.
ANEXO A	
ANEXO 1A Rotina R, programa utilizado para obtenção das estimativas dos parâmetros do modelo não-linear via aproximação assintótica e as estimativas dos parâmetros dos modelos linearizados.	118
ANEXO 2A Rotina SAS, programa utilizado para obtenção das estimativas dos parâmetros do modelo não-linear, utilizando a técnica “bootstrap” e a Figura 8.	127

ANEXO 1A

Rotina R, programa utilizado para obtenção das estimativas dos parâmetros do modelo não-linear via aproximação assintótica e as estimativas dos parâmetros dos modelos linearizados.

```
#Lineaweaver- Burk#  
x=c(0.02,0.02,0.06,0.06,0.11,0.11,0.22,0.22,0.56,0.56,1.10,1.10)  
y=c(76,47,97,107,123,139,159,152,191,201,207,200)  
Dados=data.frame(x=x,y=y)  
c=1/Dados[,2]  
d=1/Dados[,1]  
Resultado=lm(c~d)  
summary(Resultado)  
respad=(Res-mean(Res))/sd(Res)  
par(mfrow=c(1,2))  
#qqnorm (respad,xlab="Quantil Teórico (a)", ylab="Resíduos Padronizados",  
main = "",cex.lab=1.5,cex.main=1.8,cex.axis=1.2)  
#abline(0,1)  
#plot(Pred,respad,xlab="Valores ajustados (b)",ylab="Resíduos Padronizados",  
main = "", cex.lab=1.5, cex.main=1.8, cex.axis=1.2)  
#abline(h=0)  
shapiro.test(Res)  
library(lmtest)
```

```

dwtest(z~w,data=Dados)
vcov(Resultado)
ValorInicial=Resultado$coef
Vm=1/ValorInicial[1]
Km=ValorInicial[2]/ValorInicial[1]
IC=matrix((confint(Resultado)),2,2)
ICVm=c((1/IC[1,2]),(1/IC[1,1]))
ICKm=c(IC[2,2]/IC[1,2],IC[2,1]/IC[1,1])

#Gráfico da regressão#
par(mfrow=c(1,1))
plot(d,Pred,xlab="1/[s]",ylab="1/v",cex.lab=1.5,cex.axis=1.2,type="l",lty=1,
ylim=c(0,0.025))
points(d,c,pch=19)
legend(3,0.025,'Curva ajustada',lty=1,cex=1.3,bty="n")
legend(4,0.023,'Valores observados ',cex=1.3,bty="n",pch=19,lty=0)
ICVm
ICKm

#Regressão não linear- Método de Gauss Newton#
Estimativas=nls(y~SSmicmen(x,Vm,Km),data=Dados,start=list(Vm=Vm,Km=
Km),
trace=T)

```

```

a=summary(Estimativas)

a

Preditos=predict(Estimativas)

Resíduos=resid(Estimativas)

R=(cor.test(Dados[,2],Preditos))

R2=R$est^2

Rajust=1-((11/10)*(1-R2))

respad=(Resíduos-mean(Resíduos))/sd(Resíduos)

par(mfrow=c(1,2))

qqnorm (respad,xlab="Quantil Téorico (i)", ylab="Resíduos Padronizados",
main = "",cex.lab=1.5,cex.main=1.8,cex.axis=1.2)

abline(0,1)

plot(Preditos,respad,xlab="Valores ajustados (j)",ylab="Residuos
Padronizados", main = "", cex.lab=1.5, cex.main=1.8, cex.axis=1.2)

abline(h=0)

shapiro.test(Resíduos)

dwtest(y~SSmicmen(x,Vm,K),data=Dados)

vcov(Estimativas)

kmv<-matrix(c(coefficients(Estimativas)),2,1)

kmv

X=matrix(0,12,2)

X[,1]=x/(kmv[1,]+x)

X[,2]=(-kmv[2,]*x)/((kmv[1,]+x)^2)

X

```

```
XTX=t(X)%*%X
```

```
XTX
```

```
#Gráfico da regressão#
```

```
par(mfrow=c(1,1))
```

```
abline(h=212.7,lty=3)
```

```
x1=seq(0,1.2,by=0.00001)
```

```
f=(kmv[1,]*(x1))/(kmv[2,]+x1)
```

```
plot(x1,f,xlab="[s]",ylab="v",cex.lab=1.5,cex.axis=1.2,type="l",lty=1,ylim=c(40  
,220))
```

```
points(x,y,pch=19)
```

```
abline(h=212.7,lty=3)
```

```
legend(0.6,150,'Curva ajustada',lty=1,cex=1.3,bty="n")
```

```
legend(0.65,140,'Valores observados',cex=1.3,bty="n",pch=19,lty=0)
```

```
#Hanes-Wolf#
```

```
e=Dados[,1]/Dados[,2]
```

```
f=Dados[,1]
```

```
Resultado=lm(e~f)
```

```
summary(Resultado)
```

```
Pred=predict(Resultado)
```

```
Res=resid(Resultado)
```

```

respad=(Res-mean(Res))/sd(Res)
par(mfrow=c(1,2))
qqnorm (respad,xlab="Quantil Teórico (c)", ylab="Resíduos Padronizados",
main = "",cex.lab=1.5,cex.main=1.8,cex.axis=1.2)
abline(0,1)
plot(Pred,respad,xlab="Valores ajustados (d)",ylab="Resíduos Padronizados",
main = "", cex.lab=1.5, cex.main=1.8, cex.axis=1.2)
abline(h=0)
shapiro.test(Res)
library(lmtest)
dwtest(z~w,data=Dados)
vcov(Resultado)
ValorInicial=Resultado$coef
Vm=1/ValorInicial[2]
K=ValorInicial[1]/ValorInicial[2]
IC=matrix((confint(Resultado)),2,2)
ICVm=c((1/IC[2,2]),(1/IC[2,1]))
ICKm=c(IC[1,1]/IC[2,1],IC[1,2]/IC[2,2])
ICVm
ICKm

#Gráfico da regressão#
par(mfrow=c(1,1))
plot(f,Pred,xlab="[s]",ylab="[s]/v",cex.lab=1.5,cex.axis=1.2,type="l",lty=1)

```

```

points(f,e,pch=19)
legend(0,0.005,'Curva ajustada',lty=1,cex=1.3,bty="n")
legend(0.02,0.0045,' Valores observados ',cex=1.3,bty="n",pch=19,lty=0)

#Woolf-Augustinson- Hofstee#
g=Dados[,2]
h=Dados[,2]/Datos[,1]
Resultado=lm(g~h)
summary(Resultado)
Pred=predict(Resultado)
Res=resid(Resultado)
respad=(Res-mean(Res))/sd(Res)
par(mfrow=c(1,2))
qqnorm (respad,xlab="Quantil Téorico (e)", ylab="Resíduos Padronizados",
main = "",cex.lab=1.5,cex.main=1.8,cex.axis=1.2)
abline(0,1)
plot(Pred,respad,xlab="Valores ajustados (f)",ylab="Residuos Padronizados",
main = "", cex.lab=1.5, cex.main=1.8, cex.axis=1.2)
abline(h=0)
shapiro.test(Res)
library(lmtest)
dwtest(z~w,data=Dados)
vcov(Resultado)
ValorInicial=Resultado$coef

```

```

Vm=ValorInicial[1]
K=-ValorInicial[2]
IC=matrix((confint(Resultado)),2,2)
ICVm=c((IC[1,1]),(IC[1,2]))
ICKm=c(-IC[2,2],-IC[2,1])
ICVm
ICKm

#Gráfico da regressão#
par(mfrow=c(1,1))
plot(h,Pred,xlab="v/[s]",ylab="v",cex.lab=1.5,cex.axis=1.2,type="l",lty=1,
ylim=c(50,220))
points(h,g,pch=19)
legend(2000,200,'Curva ajustada',lty=1,cex=1.3,bty="n")
legend(2095,180,'Valores observados ',cex=1.3,bty="n",pch=19,lty=0)

#Eadie-Scatchard
i=Dados[,2]/Dados[,1]
j=Dados[,2]
Resultado=lm(i~j)
Result=summary(Resultado)
Pred=predict(Resultado)
Res=resid(Resultado)

```

```

respad=(Res-mean(Res))/sd(Res)
par(mfrow=c(1,2))
qqnorm (respad,xlab="Quantil Teórico (g)", ylab="Resíduos Padronizados",
main = "",cex.lab=1.5,cex.main=1.8,cex.axis=1.2)
abline(0,1)
plot(Pred,respad,xlab="Valores ajustados (h)",ylab="Residuos Padronizados",
main = "", cex.lab=1.5, cex.main=1.8, cex.axis=1.2)
abline(h=0)
shapiro.test(Res)
library(lmtest)
dwtest(z~w,data=Dados)
vcov(Resultado)
ValorInicial=Resultado$coef
Vm=-ValorInicial[1]/ValorInicial[2]
K=-1/ValorInicial[2]
IC=matrix((confint(Resultado)),2,2)
ICVm=c((-IC[1,1]/IC[2,1]),(-IC[1,2]/IC[2,2]))
ICKm=c(-1/IC[2,1],-1/IC[2,2])
ICVm
ICKm
par(mfrow=c(1,1))
plot(j,Pred,xlab="v/[s]",ylab="1/v",cex.lab=1.5,cex.axis=1.2,type="l",lty=1,
ylim=c(0,4000))
points(j,i,pch=19)

```

```
legend(140,4000,'Curva ajustada',lty=1,cex=1.3,bty="n")
```

```
legend(148,3650,'Valores observados',cex=1.3,bty="n",pch=19,lty=0)
```

ANEXO 2A

Rotina SAS, programa utilizado para a obtenção das estimativas iniciais dos parâmetros do modelo não-linear, utilizando a técnica Bootstrap.

```
data teste;
```

```
input x y;
```

```
cards;
```

```
0.02 76
```

```
0.02 47
```

```
0.06 97
```

```
0.06 107
```

```
0.11 123
```

```
0.11 139
```

```
0.22 159
```

```
0.22 152
```

```
0.56 191
```

```
0.56 201
```

```
1.10 207
```

```
1.10 200
```

```
;
```

```
proc nlin data=teste ;
```

```
parms theta1=195.695 theta2=0.04838;
```

```
model y=(theta1*x)/(theta2+x);  
output out=saida P=PREDICT R=RESID;  
run;  
quit;
```

```
proc univariate data=saida normal;  
var resid;  
run; quit;
```

```
* proc print data=saida;  
* run; *quit;
```

```
proc iml;  
  
use saida;  
read all into dados;  
print dados;  
predboot=j(nrow(dados),3,0);  
nboot=2000;
```

```

/* ***** Procedimento Bootstrap ***** */

do ib=1 to nboot;
    obs=0;
    do i=1 to nrow(dados);
/* ***** Observação predita ***** */
        obs=dados[i,3];

/* ***** Sorteio do residuos ***** */
        nli=int(ranuni(0)*nrow(dados));
        do while (nli=0);
            nli=int(ranuni(0)*nrow(dados));
        end;
/* ***** Fim do Sorteio ***** */

        predboot[i,1]=ib;
        predboot[i,2]=dados[i,1];
        predboot[i,3]=obs+dados[nli,4];

    end;

    matboot=matboot//predboot;
end;

```

```

create resboot from matboot;

append from matboot;

/* ***** Listagem dos dados reamostrados ***** */

*proc print data=resboot;

*run; quit;

proc nlin data=resboot outest=saida maxiter=1000 converge=0.00000001
noprnt ;

parms theta1=212.684 theta2=0.06412;

model col3=(theta1 *col2)/(theta2+col2);

output out=dadosmod parms=theta1 parms=theta2;

by col1;

run;

data resumo ; set saida;

if _type_status_=3 then delete;

if _type_='ITER' then delete;

if _type_='COVB' then delete;

qmres=_SSE_/10;

*proc print data=resumo;

```

```

*run; *quit;

/* ***** ESTATÍSTICAS DESCRITIVAS DOS PARÂMETROS E
VARIANCIA **** */

proc univariate normal data=resumo PCTLDEF=2;

var theta1 theta2 qmres;

output out=resumo PCTLPRE=theta1 theta2 qmres PCTLPTS=97.5
PCTLNAME=p975
        PCTLPRE=theta1 theta2 qmres PCTLPTS=2.5 PCTLNAME=p025;

histogram theta1/kernel(c=1 2 3 k=normal quadratic);
histogram theta2/kernel(c=1 2 3 k=normal quadratic);
histogram qmres/kernel(c=1 2 3 k=normal quadratic);

run;

proc print data=resumo;

run;quit;

run;

quit;

```