

**VERIFICAÇÃO POR SIMULAÇÃO MONTE
CARLO DA EFICIÊNCIA DA APLICAÇÃO DO
TESTE DE RAZÃO DE VEROSSIMILHANÇA
PARA IDENTIDADE DE MODELOS NÃO
LINEARES**

ROMULO BARBOSA VELOSO

2006

ROMULO BARBOSA VELOSO

**VERIFICAÇÃO POR SIMULAÇÃO MONTE CARLO
DA EFICIÊNCIA DA APLICAÇÃO DO TESTE DE
RAZÃO DE VEROSSIMILHANÇA PARA
IDENTIDADE DE MODELOS NÃO LINEARES.**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Orientador

Prof. Dr. Paulo César lima

LAVRAS
MINAS GERAIS - BRASIL
2006

ROMULO BARBOSA VELOSO

**VERIFICAÇÃO POR SIMULAÇÃO MONTE CARLO
DA EFICIÊNCIA DA APLICAÇÃO DO TESTE DE
RAZÃO DE VEROSSIMILHANÇA PARA
IDENTIDADE DE MODELOS NÃO LINEARES**

Dissertação apresentada à
Universidade Federal de Lavras,
como parte das exigências do curso
de Mestrado em Agronomia, área
de concentração em Estatística e
Experimentação Agropecuária,
para obtenção do título de
"Mestre".

APROVADA em quinta-feira, 26 de janeiro de 2006

Prof. José Higino Dias Filho

UNIMONTES

Prof. Daniel Furtado Ferreira

UFLA

Prof. Rubens Delly Veiga

UFLA

Prof. Paulo César Lima

UFLA

(Orientador)

LAVRAS

MINAS GERAIS - BRASIL

A Deus,

por ser a luz que me ilumina;

A Todos

que me desejam sucesso;

À gigantesca Fé

que sempre me impulsiona para frente;

OFEREÇO

Aos meus pais, Arthur e Edith, pois foram seus exemplos de dedicação e perseverança que me trouxeram até aqui;

A minha esposa Dora e meus filhos Denny, Lucy, Jeane e Thiago, que mais que eu mesmo, sempre acreditaram em mim;

A meus irmãos, Fátima, Roberto, Artur e Rogério, pois todas as páginas de minha vida têm a assinatura deles;

DEDICO

AGRADECIMENTOS

À Universidade Estadual de Montes Claros e ao Departamento de Ciências Exatas (DCEx) pela confiança.

À Universidade Federal de Lavras e ao Departamento de Ciências Exatas (DEX) pelos conhecimentos.

Ao Prof. Paulo César Lima pela orientação.

Ao Prof. Daniel Furtado Ferreira pela disponibilidade e pelo apoio.

Ao colega Marcelo Ângelo Cirillo pelo incentivo, pelo apoio e pela colaboração inestimáveis.

A todos aqueles que contribuíram direta ou indiretamente para que eu pudesse atingir este objetivo.

SUMÁRIO

RESUMO	i
ABSTRACT	ii
1 Introdução.....	1
2 Referencial teórico.....	3
2.1 Modelos determinísticos e modelos estatísticos.....	3
2.2 Métodos de estimação do vetor de parâmetros.....	4
2.2.1 Método dos mínimos quadrados.....	5
2.2.2 método da máxima verossimilhança	7
2.3 Solução do Sistema de equações normais para modelos não lineares	9
2.4 Teste de hipóteses.....	13
2.4.1 Erros em teste de hipóteses.....	14
2.4.2 O teste de razão de verossimilhança.....	17
2.4.2.1 O teste de razão de verossimilhança simples	17
2.5.2.2 O teste de razão de verossimilhança generalizado	18
2.6 Identidade de modelos de regressão	19
3 Metodologia.....	31
3.1 Modelos	31
3.2 Identidade de modelos não-lineares	34
3.3 Simulação Monte Carlo.....	36
4 Resultados e discussão	38
4.1 Controle de erro tipo I	40
4.1.1 Modelo Logístico.....	40
4.1.2 Modelo de Gompertz.....	42

4.1.3 Modelo Langmuir	43
4.2 Poder.....	44
4.2.1 Modelo Logístico.....	44
4.2.2 Modelo de Gompertz e modelo de Langmuir.....	45
5 Conclusões.....	48
Referências Bibliográficas.....	49
ANEXOS	52

RESUMO

VELOSO, Rômulo Barbosa. **VERIFICAÇÃO POR SIMULAÇÃO MONTE CARLO DA EFICIÊNCIA DA APLICAÇÃO DO TESTE DE RAZÃO DE VEROSSIMILHANÇA PARA IDENTIDADE DE MODELOS NÃO LINEARES**. LAVRAS: UFLA, 2006. 63p. (Dissertação – Mestrado em Estatística e Experimentação Agropecuária)*

Uma alternativa para comparar modelos não lineares é dada por meio do teste de identidade de modelos. A estatística utilizada mormente é dada por meio da razão de verossimilhanças. Objetivando avaliar por meio de simulação Monte Carlo a performance desse teste, mediante a diferentes tamanhos amostrais, número de equações, robustez e diferentes modelos, considerou-se como critério o controle do erro tipo I e o poder. Concluiu-se que o aumento no número de equações produziu um efeito conservativo no controle do erro tipo I para os modelos de Gompertz e Langmuir, para todos os tamanhos amostrais, e que o modelo de Gompertz mostrou-se mais poderoso e robusto para resíduos não normais.

* Comitê Orientador: Paulo César Lima-UFLA (Orientador).

ABSTRACT

VELOSO, Romulo Barbosa. **VERIFICATION BY MONTE CARLO SIMULATION OF THE EFFICICNY OF APPLICATION OF THE LIKELYHOOD RATIO TEST FOR IDENTITY OF NON-LINEAR MODELS.**LAVRAS: UFLA, 2006. 63p. (Dissertation - Master in Statistics and Agricultural Experimentation)*

An alternative to compare non-linear models is given by means of the test of identity of models. The statistics utilized normally is given by means of the likelihood ratio. Aiming to evaluate by means of the Monte Carlo simulation the performance of this test, by the aid of the different sample models, number of equations, robustness and different models, the control of the type I error and the power was considered as a criterion . it follows that the increase in the number of equations produced a conservative effect in the control of the type I error for the Gompertz and Langmuir models for all the sample sizes and that the Gompertz model proved more powerful and robust for non-normal residuo

* Guidance Committee: Paulo César Lima-UFLA (Major Professor).

1 INTRODUÇÃO

A construção de um modelo matemático que estabeleça uma relação funcional em um grupo de variáveis é uma das formas de descrever e prever fenômenos humanos e fenômenos naturais observáveis. Quanto maior o conhecimento sobre as variáveis envolvidas, maiores são as alternativas de controle, simulações, avaliações e maior é a possibilidade de adequação de um modelo matemático que explique sistemas biológicos, físicos, sociológicos, econômicos e outros.

Nos processos estocásticos os modelos determinísticos não são suficientes para que se estabeleça uma relação funcional satisfatória. Nestes casos, o modelo deve incluir componentes aleatórios associando-se, então, um modelo probabilístico ao fenômeno observado. Os modelos lineares e não-lineares são frequentemente utilizados nestes estudos.

Em razão de os modelos conterem variáveis aleatórias, podem existir situações em que modelos diferentes são usados para representar o mesmo fenômeno. Nesses casos é útil saber se estas funções podem ser representadas por uma única ou determinar se alguns parâmetros podem ter a mesma representação.

Alguns procedimentos estatísticos estão disponíveis para o estudo da identidade de modelos lineares e não-lineares. Para o caso de modelos não-lineares, Regazzi (2003) apresenta uma metodologia que utiliza o teste da razão de verossimilhança para testar as seguintes hipóteses relativas a um grupo de g equações de regressão não-lineares:

H_0 : as g equações são idênticas;

H_0 : as g equações têm uma constante de regressão comum;

H_0 : as equações g têm um ou mais coeficientes de regressão iguais.

As pressuposições básicas do método proposto por Regazzi (2003) são as de que os erros, além de serem normalmente distribuídos e independentes, tenham variância constante dentro do grupo estudado.

O objetivo deste trabalho é apresentar e analisar o método estatístico de identidades de modelos proposto por Regazzi (2003) aplicando simulação Monte Carlo para pequenas, médias e grandes amostras, assumindo as pressuposições propostas por Regazzi (2003) para os modelos não-lineares Logístico, de Gompertz e Langmuir.

2 REFERENCIAL TEÓRICO

2.1 - Modelos Determinísticos e Modelos Estatísticos

A representação geral de um modelo determinístico pode ser feita da forma:

$$Y = f(X_1, X_2, \dots, X_n, \theta_1, \theta_2, \dots, \theta_k) = f(X, \theta).$$

No modelo, $f(X, \theta)$ é a função que associa as variáveis explicativas à variável resposta Y . A utilização de $f(X, \theta)$ simplifica a notação, sendo X o vetor de n variáveis explicativas e θ , o vetor de k parâmetros que se deseja conhecer na função. As variáveis X_1, X_2, \dots, X_n representam os fatores que influenciam na resposta e são chamadas variáveis explicativas. Neste trabalho as variáveis explicativas foram consideradas constantes pré-fixadas, enquanto Bolfarini et al (1992) fazem uma abordagem do problema considerando-as variáveis aleatórias.

Os modelos estatísticos, de certa forma, podem ser considerados modelos determinísticos com algum componente extra, chamado de componente aleatório. Um modelo estatístico pode ser representado, de uma forma geral, por:

$$Y = f(X_1, X_2, \dots, X_n, \theta_1, \theta_2, \dots, \theta_k) + e_i = f(X, \theta) + e_i,$$

em que $f(X, \theta)$ é o componente determinístico e e_i , o componente aleatório. O componente aleatório é responsável pela parte não controlada ou não determinada das observações, associado a uma função probabilística; por exemplo, distribuição Normal, Exponencial, Weibull, Gama, Binomial, dentre outras.

Um modelo estatístico de qualidade depende da melhor escolha do componente determinístico, assim como da melhor caracterização do componente aleatório. O pesquisador deverá, a partir dos dados observados, fazer a escolha de um modelo apropriado.

Um modelo é chamado linear se o espaço caracterizado por seus parâmetros for linear; caso contrário, é dito não-linear. Frequentemente, o efeito das variáveis independentes X_i na variável resposta Y é adequadamente aproximado por uma função que não é linear nos parâmetros. Como exemplo, há o modelo estatístico não-linear apresentado por Gallant (1987), com a seguinte representação, em que $x^t = (x_1, x_2, x_3)$ e o vetor de parâmetros é $\theta^t = (\theta_1, \theta_2, \theta_3, \theta_4)$:

$$f(x, \theta) = \theta_1 x_1 + \theta_2 x_2 + \theta_4 e^{\theta_3 x_3}$$

Uma forma prática de verificar se um modelo é ou não linear consiste em aplicar as derivadas parciais do modelo a cada parâmetro e verificar se nos resultados obtidos o parâmetro utilizado como variável na derivação parcial continua presente; caso isto aconteça com pelo menos um dos parâmetros do modelo, então o modelo em estudo é não-linear.

2.2 - Métodos de estimação do vetor de parâmetros

O uso da análise de regressão é uma ferramenta muito útil para estimação dos parâmetros. Os métodos dos mínimos quadrados e da máxima verossimilhança são os mais utilizados para a estimação do vetor de parâmetros nos estudos de regressão.

2.2.1 - O método dos mínimos quadrados

O método dos mínimos quadrados tem variações conforme as pressuposições assumidas quanto aos erros, podendo ser o método dos quadrados mínimos ordinários, ponderados ou generalizados.

Em um modelo de regressão linear simples, o ajuste utilizando o método dos quadrados mínimos ordinários exige as seguintes pressuposições:

- i) $E(\varepsilon_i)=0$, em que ε_i é o erro associado a Y_i ;
- ii) Os erros sejam não correlacionados;
- iii) Os erros devem ser variáveis aleatórias com distribuição normal.

Com estas pressuposições, segundo Hoffmann e Vieira (1977), obtêm-se estimadores não tendenciosos e de variância mínima.

Conforme Draper e Smith (1998), pode ocorrer que algumas das observações usadas sejam menos fidedignas que outras. Isto geralmente acontece porque as variâncias das observações não são todas iguais, ou seja, não se tem a matriz não singular $V(\varepsilon)$ de variâncias e covariâncias na forma $I\sigma^2$. Quando isto ocorre, a diagonal desta matriz possui valores de variância diferentes. Pode ainda ocorrer que, em alguns fenômenos observados, os elementos fora da diagonal principal da matriz $V(\varepsilon)$ de variâncias e covariâncias não sejam todos nulos, isto é, as observações são correlacionadas. Nestes dois casos citados, a estimação pelo método dos mínimos quadrados ordinários não se aplica. Busca-se, então, trocar a variável de observações Y por outra variável de forma a satisfazer as suposições que garantam o uso do método dos quadrados mínimos ordinários.

Em modelos de regressão não linear, apesar da ampla aplicação, apenas assintoticamente o estimador de mínimos quadrados possui as propriedades do mesmo estimador em regressão linear, ou seja, quando o tamanho da amostra tende para infinito, as propriedades destes estimadores se aproximam. Para

amostras finitas, deve-se assumir que Y_i é normalmente distribuída em torno de sua média, com variância finita σ^2 para todo $i = 1, 2, \dots, n$. Ainda segundo Hoffmann e Vieira (1977), o estimador de parâmetros não é uma combinação linear de Y_i e, em geral, não é normalmente distribuído, como também é um estimador tendencioso e não é de variância mínima. Por estes motivos, as propriedades do estimador de mínimos quadrados nos modelos não lineares são desconhecidas para amostras de tamanho finito.

A partir desses aspectos, a teoria assintótica mostra que o estimador de mínimos quadrados torna-se cada vez menos viesado, cada vez mais normalmente distribuído e aproxima-se da variância mínima possível à medida que o tamanho da amostra tende para infinito. No entanto, não existem regras para saber se o tamanho da amostra é grande o suficiente para que sejam válidas as propriedades assintóticas.

Como no modelo de regressão linear, o critério de mínimos quadrados para o modelo de regressão não linear é:

$$S(\theta) = \sum_{i=1}^n (Y_i - f(\mathbf{X}_i, \theta))^2$$

em que Y_i é o valor observado, $f(\mathbf{X}_i, \theta)$ é o valor predito pelo modelo adotado e θ é o vetor de parâmetros.

A função $S(\theta)$ deve ser minimizada com respeito aos parâmetros de regressão não linear $\theta_0, \theta_1, \dots, \theta_{p-1}$, considerando um modelo de p parâmetros para se obterem as estimativas de mínimos quadrados de forma análoga ao método dos mínimos quadrados em regressão linear. A diferença é que a solução das equações normais em modelos não lineares usualmente requer um método numérico iterativo, pois a solução analítica geralmente não pode ser encontrada.

Seja o modelo não-linear com a notação:

$$Y_u = f(X_{1u}, X_{2u}, \dots, X_{nu}, \theta_1, \theta_2, \dots, \theta_p) + e_u,$$

com $\varepsilon_u = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k)'$. Este modelo pode ser escrito de forma generalizada:

$$Y_{ij} = f(X, \theta) + \varepsilon_{ij},$$

em que X é o vetor das variáveis explicativas ou dos valores observados, θ é o vetor dos parâmetros e ε_u é o u-ésimo erro, com $u = 1, 2, \dots, n$. Conforme afirmam Mood et al (1974), a pressuposição de normalidade e independência dos resíduos está implícita e $\varepsilon_u \sim N(0, \sigma^2)$. Draper e Smith (1998) definem a soma de quadrados do resíduo para o modelo não linear como:

$$S(\theta) = \sum_{u=1}^n \{Y_u - f(X_u, \theta)\}^2.$$

Os mesmos autores afirmam ainda que para um modelo de p parâmetros, as p equações normais podem ser encontradas derivando $S(\theta)$ com relação a cada um dos parâmetros do vetor θ . Desta maneira obtêm-se as estimativas de mínimos quadrados, $\hat{\theta}$:

$$\frac{\partial S(\theta)}{\partial \theta_i} = -2 \sum_{u=1}^n \{Y_u - f(X_u, \hat{\theta})\} \left[\frac{\partial f(X_u, \theta)}{\partial \theta_i} \right]_{\theta=\hat{\theta}} = 0, \quad \text{com } i=1, 2, \dots, p.$$

Entretanto, na maioria dos casos não é possível obter soluções para o sistema de equações normais não lineares, sendo necessário que se recorra a algum método iterativo.

2.2.2 - O método da Máxima verossimilhança

Mood et al (1974) definem da seguinte forma uma função de verossimilhança: “a função de n variáveis aleatórias X_1, X_2, \dots, X_n é uma densidade conjunta de n variáveis, dita $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n; \theta)$, a qual é considerada uma função de θ . Em particular, se X_1, X_2, \dots, X_n é uma amostra

aleatória de uma densidade $f(x;\theta)$, então a função de verossimilhança é $f(x_1;\theta), f(x_2;\theta), \dots, f(x_n;\theta)$ ". Mood et al (1974) apresentam a função de verossimilhança como $L(\theta) = f(x_1;\theta) f(x_2;\theta) \dots f(x_n;\theta)$. Muitas vezes a função de verossimilhança satisfaz as condições de regularidade e o estimador de máxima verossimilhança é a solução da equação $\frac{dL(\theta)}{d\theta} = 0$; assim são determinados os pontos críticos da função de verossimilhança.

Conforme Bearzoti (1998), o método da máxima verossimilhança consiste em, dentre os infinitos valores que $\hat{\theta}$ pode assumir, tomar aquele que mais provavelmente gerou os dados. $L(\theta)$ e $\log(L(\theta))$ têm seus máximos no mesmo ponto e, portanto, pode-se utilizar tanto uma quanto a outra para se determinar os valores destes máximos.

Com erros $\varepsilon \sim \text{NID}(0, \sigma^2)$, ou seja, independentes, normalmente distribuídos com variância constante σ^2 , a função de verossimilhança é dada por:

$$L(X, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n [Y_i - f(X_i, \hat{\theta})]^2\right]$$

para todo $i = 1, 2, \dots, n$. Nesta função, $f(X_i, \hat{\theta})$ é a estimativa da variável resposta para cada i ; Y_i é o valor observado da variável resposta e σ^2 é a variância. Esta função é maximizada em relação aos parâmetros minimizando

$$\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n [Y_i - f(X_i, \hat{\theta})]^2\right],$$

o que conduz aos mesmos estimadores obtidos pelos métodos dos mínimos quadrados.

2.3 - Solução do Sistema de Equações Normais para Modelos Não Lineares

Para a obtenção das equações normais referentes a um modelo não-linear

$$Y_i = f(X_i, \theta) + \varepsilon_i$$

utilizando os quadrados mínimos ordinários, é necessário minimizar a soma dos quadrados dos erros:

$$S(\theta) = \sum_{u=1}^n \{Y_i - f(X_i, \theta)\}^2 .$$

Assim, num modelo com p parâmetros, as p equações normais são encontradas derivando $S(\theta)$ com relação a cada um dos parâmetros do vetor θ , obtendo-se as estimativas de mínimos quadrados $\hat{\theta}$:

$$\frac{\partial S(\theta)}{\partial \theta_i} = -2 \sum_{u=1}^n (Y_i - f(\mathbf{X}_i, \theta)) \left[\frac{\partial f(\mathbf{X}_i, \theta)}{\partial \theta_p} \right] = 0 \quad \text{com } i = 1, 2, \dots, p.$$

Dessa forma é obtido o vetor $\hat{\theta}$, que é o vetor das estimativas de mínimos quadrados dos p parâmetros :

$$\underset{(p \times 1)}{\hat{\theta}} = \begin{bmatrix} \hat{\theta}_0 \\ \hat{\theta}_1 \\ \cdot \\ \cdot \\ \hat{\theta}_{p-1} \end{bmatrix}$$

As equações normais obtidas são não-lineares nas estimativas dos parâmetros $\hat{\theta}$, em geral difíceis de serem resolvidas. Dessa forma, são necessários métodos numéricos para a obtenção de uma solução das equações normais iterativamente. Dentre os muitos métodos iterativos usados para solucionar sistemas não lineares destaca-se o método de Gauss-Newton.

O método de Gauss-Newton, também chamado método da linearização, usa uma expansão em série de Taylor e, através de termos lineares, aproxima o modelo não-linear. Com isto, torna possível a aplicação de mínimos quadrados ordinários para a estimação dos parâmetros.

O método de Gauss-Newton é iniciado dando-se valores iniciais aos parâmetros $\theta_0, \theta_1, \dots, \theta_{p-1}$, denotados por: $g_0^{(0)}, g_1^{(0)}, \dots, g_{p-1}^{(0)}$. Esses valores iniciais podem ser obtidos de estudos anteriores, conhecimentos teóricos ou por uma grade de valores que minimizem:

$$S(\theta) = \sum_{i=1}^n (Y_i - f(\mathbf{X}_i, \theta))^2 .$$

Com os valores iniciais dos parâmetros, aproxima-se a função esperada $f(\mathbf{X}_i, \theta)$ para os n casos por termos lineares da expansão em série de Taylor, de primeira ordem, em torno dos valores iniciais $g_p^{(0)}$. Obtem-se, para o i -ésimo caso:

$$f(X_i, \theta) \cong f(X_i, g_p^{(0)}) + \sum_{k=0}^{p-1} \left[\frac{\partial f(X_i, \theta)}{\partial \theta_p} \right]_{\theta = g^{(0)}} (\theta - g_p^{(0)}) \quad 2.3.1,$$

sendo $g_p^{(0)}$ o vetor dos valores iniciais dos parâmetros, ou tentativas iniciais, assim como f são avaliadas em $\theta_p = g_p^{(0)}$. Fazendo:

$$\begin{aligned} f_i^{(0)} &= f(X_i, g^{(0)}) \\ \beta_p^{(0)} &= (\theta_k - g_k^{(0)}) \\ D_{ip}^{(0)} &= \left[\frac{\partial f(X_i, \theta)}{\partial \theta_p} \right]_{\theta = g^{(0)}} , \end{aligned}$$

pode-se reescrever a aproximação 2.3.1 como:

$$f(X_i, \theta) \cong f_i^{(0)} + \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} .$$

Uma aproximação para o modelo $Y_i = f(X_i, \theta) + \varepsilon_i$ é dada por:

$$Y_i \cong f_i^{(0)} + \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} + \varepsilon_i$$

Passando $f_i^{(0)}$ para o lado esquerdo e denotando a diferença $Y_i - f_i^{(0)}$ por $Y_i^{(0)}$, tem-se:

$$Y_i^{(0)} \cong \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} + \varepsilon_i \quad i = 1, 2, \dots, n.$$

Obtem-se, assim, uma aproximação para um modelo de regressão linear. Cada coeficiente de regressão $\beta_k^{(0)}$ representa a diferença entre os verdadeiros parâmetros da regressão e as estimativas iniciais dos mesmos. Assim, os coeficientes de regressão representam uma correção que deve ser feita nos coeficientes de regressão iniciais. O propósito de ajustar o modelo de regressão linear é estimar os coeficientes de regressão $\beta_k^{(0)}$ e usar essas estimativas para corrigir as estimativas iniciais dos parâmetros de regressão. O modelo $Y_i^{(0)} \cong \sum_{k=0}^{p-1} D_{ik}^{(0)} \beta_k^{(0)} + \varepsilon_i \quad i = 1, 2, \dots, n$ na forma matricial fica é expresso por:

$$Y^{(0)} \cong D^{(0)} \beta^{(0)} + \varepsilon$$

$$\underset{(p \times 1)}{\beta^{(0)}} = \begin{bmatrix} \beta_0^{(0)} \\ \cdot \\ \cdot \\ \cdot \\ \beta_{p-1}^{(0)} \end{bmatrix} \quad \underset{(n \times 1)}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

$$\mathbf{Y}^{(0)} = \begin{bmatrix} Y_1 - f_1^{(0)} \\ \cdot \\ \cdot \\ \cdot \\ Y_n - f_n^{(0)} \end{bmatrix} \quad \mathbf{D}^{(0)} = \begin{bmatrix} D_{10}^{(0)} & \dots & D_{1,p-1}^{(0)} \\ \cdot \\ \cdot \\ \cdot \\ D_{n0}^{(0)} & \dots & D_{n,p-1}^{(0)} \end{bmatrix}$$

similar ao modelo de regressão linear: $Y = X\beta + \varepsilon$. A matriz \mathbf{D} faz o papel da matriz \mathbf{X} : $\mathbf{D} \rightarrow \mathbf{X}$. Pode-se, portanto, estimar os parâmetros $\beta^{(0)}$ pelo método de mínimos quadrados ordinários:

$$\mathbf{b}^{(0)} = (\mathbf{D}^{(0)'} \mathbf{D}^{(0)})^{-1} \mathbf{D}^{(0)'} \mathbf{Y}^{(0)}.$$

Estas estimativas de mínimos quadrados são usadas para a obtenção dos coeficientes de regressão estimados (sem intercepto) corrigidos $g_k^{(1)}$ por meio de:

$$\beta_k^{(\theta)} = (\theta_k - g_k^{(\theta)})$$

$$g_k^{(1)} = g_k^{(0)} + b_k^{(0)},$$

em que $g_k^{(1)}$ representa a estimativa corrigida de θ_k no fim da primeira iteração. Na forma matricial, tem-se:

$$\mathbf{g}^{(1)} = \mathbf{g}^{(0)} + \mathbf{b}^{(0)}.$$

Neste ponto, procura-se verificar se os coeficientes de regressão corrigidos representam uma melhoria na direção apropriada. Denotando o critério $S(\theta)$, calculado nos coeficientes de regressão iniciais $\mathbf{g}^{(0)}$, por $S(\theta)^{(0)}$, tem-se:

$$S(\theta)^{(0)} = \sum_{i=1}^n (Y_i - f(X_i, \mathbf{g}^{(0)}))^2 = \sum_{i=1}^n (Y_i - f_i^{(0)})^2.$$

No final da primeira iteração, os coeficientes de regressão corrigidos são $\mathbf{g}^{(1)}$. O critério $S(\theta)$, calculado nos coeficientes de regressão $\mathbf{g}^{(1)}$, será denotado por $S(\theta)^{(1)}$, ou seja, espera-se $S(\theta)^{(1)}$ menor que $S(\theta)^{(0)}$, pois os coeficientes de regressão no passo um deverão ser melhores:

$$S(\theta)^{(1)} = \sum_{i=1}^n (Y_i - f(\mathbf{X}_i, \mathbf{g}^{(1)}))^2 = \sum_{i=1}^n (Y_i - f_i^{(1)})^2$$

O método de Gauss-Newton repete o procedimento como foi descrito, com $\mathbf{g}^{(1)}$, sendo, agora, usado como valores iniciais. Isto resulta em um novo conjunto de estimativas corrigidas, representadas por $\mathbf{g}^{(2)}$, e tem-se um novo critério, $S(\theta)^{(2)}$. O processo iterativo continua até que as diferenças entre sucessivas estimativas dos coeficientes $\mathbf{g}^{(s+1)} - \mathbf{g}^{(s)}$ e/ou a diferença entre sucessivas somas de quadrados de erros $S(\theta)^{(s-1)} - S(\theta)^{(s)}$ tornem-se desprezíveis. As estimativas finais dos coeficientes de regressão são representadas por \mathbf{g} e a soma de quadrado dos erros, por $S(\theta)$.

2.4 - Teste de Hipóteses

A comparação de procedimentos, processos, tratamentos, ou ainda a verificação de que se uma afirmativa sobre determinada ocorrência é de fato verossímil, é realizada no campo da estatística através dos testes de hipóteses.

O teste de hipóteses geralmente tem início com base em alguma teoria, declaração ou afirmativa em relação a um determinado parâmetro de uma população. Esta afirmativa é chamada hipótese estatística.

Mood et al (1974) definem uma hipótese estatística como uma afirmação ou suposição sobre a distribuição de uma ou mais variáveis. Um teste de hipótese estatística, ainda conforme Mood et al (1974), é uma regra para decidir sobre a aceitação ou a rejeição desta hipótese. Os mesmos autores afirmam que um teste estatístico pode ser randomizado (casualizado) ou não-randomizado (não-casualizado) e definem estes testes como segue.

Seja \mathcal{X} o espaço amostral das observações. Um teste não-randomizado define que a hipótese \mathcal{H}_0 deve ser rejeitada se e somente se $(x_1, \dots, x_n) \in C_Y$, em que C_Y é um subconjunto de \mathcal{X} e (x_1, \dots, x_n) , uma amostra da população em estudo; Y é chamado teste não-casualizado e C_Y é chamada região crítica de teste Y . Ou seja, o teste fica determinado pela região crítica, pois ao obter a amostra, verifica-se se esta está na região crítica, e caso esteja, rejeita-se a hipótese nula.

Um teste Y de hipótese \mathcal{H}_0 é definido como um teste casualizado se Y é definido como a função $\Psi_Y(x_1, \dots, x_n) = P[\mathcal{H}_0 \text{ é rejeitado} \mid (x_1, \dots, x_n) \text{ observado}]$. A função $\Psi_Y(x_1, \dots, x_n)$ é chamada de função crítica do teste Y . Ou seja, a amostra é observada, a função $\Psi_Y(x_1, \dots, x_n)$ é aplicada assumindo-se que se for maior que um valor k , vale 0 (zero), e se menor que este valor k , vale 1 (um); é feito então um ensaio de Bernoulli com $p = \Psi_Y(x_1, \dots, x_n)$, e se o resultado obtido for sucesso (1), a hipótese H_0 é rejeitada.

2.4.1 - Erros de um teste de hipóteses

Um teste de hipóteses é formulado sob duas hipóteses possíveis: a hipótese nula e a hipótese alternativa. Levine (2005) afirma que a hipótese nula (H_0) é a hipótese a ser testada e a hipótese alternativa representa a conclusão

apoiada se a hipótese nula for rejeitada. Estes autores estabelecem três pontos chave para a metodologia clássica do teste de hipóteses:

- 1) A hipótese nula sempre se refere a um valor especificado da população, e não a uma estatística da amostra.
- 2) A afirmativa da hipótese nula sempre contém o sinal de igualdade com relação ao valor especificado do valor do parâmetro da população.
- 3) A afirmativa da hipótese alternativa nunca contém o sinal de igualdade com relação ao valor especificado do parâmetro da população.

Ao se realizar um teste de hipótese é possível que se cometam erros com relação à decisão tomada. Estes erros são denominados Erro Tipo I e Erro Tipo II. Seja H_0 a hipótese nula e H_1 a hipótese alternativa, o erro tipo I consiste em rejeitar a hipótese H_0 quando esta é verdadeira e o erro tipo II é o erro que se comete ao aceitar ou não rejeitar a hipótese H_0 quando esta é falsa. O tamanho de um erro é a probabilidade de que esse erro seja cometido. Desta forma, estão associadas aos erros probabilidades denominadas α e β para os erros tipo I e tipo II, respectivamente. Neyman e Pearson (1928), em um trabalho pioneiro, conceituaram os erros estatísticos e definiram poder de teste como a probabilidade $(1-\beta)$ de que seja rejeitada a hipótese de nulidade quando esta é falsa.

Se $f(\cdot, \theta)$ é uma função de distribuição de probabilidade parametrizada por θ , em que $\theta \in \Theta$ (Θ - espaço paramétrico), Mood et al (1974) definem função poder do teste como segue: “ Seja Υ um teste de hipótese nula H_0 . A função poder de teste Υ , denotada por $\pi_{\Upsilon}(\theta)$, é definida como a probabilidade que H_0 seja rejeitada quando a distribuição da qual a amostra foi obtida é parametrizada por θ ”.

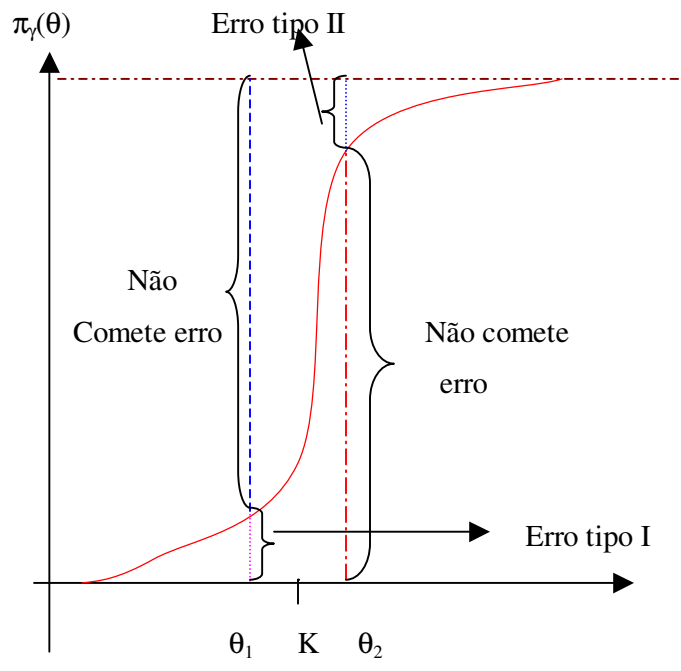
A função poder de teste ideal pode ser descrita como tendo $\pi_T(\theta)$ como a probabilidade de se rejeitar $H_0 = 0$ se $\theta \in H_0$ e como a probabilidade de se rejeitar $H_0 = 1$ se $\theta \notin H_0$.

Erros tipo I e tipo II não podem ser evitados completamente, segundo Meyer (1983), mas é necessário que se procure mantê-los com probabilidades relativamente pequenas de que ocorram. Por este motivo, a utilização de um teste de hipóteses exige que o pesquisador esteja disposto a assumir um risco ao fazer uma tomada de decisão. A diminuição da probabilidade de um tipo de erro acarreta o aumento da probabilidade do outro para um determinado tamanho de amostra, o que torna extremamente difícil conseguir controlar os erros de forma a garantir pequenas probabilidades de ocorrência de ambos. Na prática, um erro pode ser mais importante que o outro, e este fato deve ser considerado quando buscamos a situação mais favorável no fenômeno para ocorrência dos dois erros.

A redução simultânea destes erros pode ser alcançada pelo aumento do tamanho da amostra, que não depende do valor de significância adotado, conforme Carmer (1976). Ainda segundo Carmer (1976), quando se trabalha com delineamentos experimentais, a utilização de delineamento experimental mais apropriado ou de técnicas amostrais mais cuidadosas contribui para o aumento da precisão, ou seja, a diminuição dos erros.

Considera-se um bom teste aquele que rejeita com alta probabilidade a hipótese H_0 quando esta não é verdadeira, e aceita a mesma hipótese com alta probabilidade quando o parâmetro θ pertence ao espaço paramétrico Θ estabelecido por H_0 .

A escolha de um teste para atender uma situação específica é baseada na curva de poder do teste. Esta curva dá a probabilidade de rejeição de H_0 para os diversos valores do parâmetro. Como exemplo, o gráfico da Figura 2.4.1 representa a curva de poder de um teste no qual se deseja rejeitar uma hipótese nula se o parâmetro $\theta > k$, com k constante.



Representação gráfica da interpretação da curva de poder
 Para hipótese $H_0 : \theta > k$

Uma maneira racional de decidir se uma amostra vem de uma entre duas populações é, ainda segundo Mood et al (1974), optar pela população que tenha maior probabilidade de acontecer para aquele determinado valor do parâmetro.

Desta idéia intuitiva surge o método da razão de verossimilhança desenvolvido por Neyman e Pearson (1928), método usual de obtenção de testes segundo Farias et al (2003).

2.4.2 - O teste da razão de verossimilhança

2.4.2.1 - O teste simples da razão de verossimilhança

Mood et al (1974) definem o teste da razão de verossimilhança simples da seguinte maneira. Seja a amostra aleatória X_1, \dots, X_n , oriunda de uma dentre duas distribuições $f_0(\cdot)$ ou $f_1(\cdot)$. Um teste γ de $H_0: X_1 \sim f_0(\cdot)$ versus $H_1: X_1 \sim f_1(\cdot)$ é definido como um teste simples de razão de verossimilhança se γ é definido por :

$$\begin{cases} \text{Rejeita-se } H_0 & \text{se } \lambda < k; \\ \text{Não se rejeita } H_0 & \text{se } \lambda > k; \end{cases}$$

em que

$$\lambda = \lambda(x_1, x_2, \dots, x_n) = \frac{\prod_{i=1}^n f_0(x_i)}{\prod_{i=1}^n f_1(x_i)} = \frac{L_0(x_1, x_2, \dots, x_n)}{L_1(x_1, x_2, \dots, x_n)} = \frac{L_0}{L_1}$$

e k é uma constante não negativa. L_j é a função de verossimilhança para uma amostra de uma população de densidade $f_j(\cdot)$.

2.4.2.2 – O teste da razão de verossimilhança generalizado

Um teste de razão de verossimilhança é um procedimento para testar uma hipótese com nível α de significância, tem a região de rejeição (Azzalini, 1996):

$$R = \{ y : \lambda(y) \leq \lambda(\alpha) \text{ onde } \lambda(y) = \frac{\sup_{\theta \in \Theta_0} L(\theta, y)}{\sup_{\theta \in \Theta} L(\theta, y)} \}$$

e λ_α é escolhido tal que

$$\sup_{\theta \in \Theta_0} \Pr\{\lambda(Y) \leq \lambda_\alpha; \theta\} \leq \alpha ;, \text{ em geral } \lambda(Y) \in [0,1],$$

mas em muitos casos, $\lambda(Y) \in (0,1]$.

2.5 - Identidade de modelos de regressão

A comparação de equações para modelos lineares é um estudo já apresentado por diversos autores. Battisti (2001) apresenta trabalho relevante sobre o assunto, em que é mostrado com detalhes o teste proposto por Chow (1960) para testar a igualdade de duas regressões lineares, o qual dá como resultado se duas equações de regressão são ou não iguais. Gujarati (1970b), também em 4 etapas, apresenta o teste de Chow (1960), ilustrado com o seguinte procedimento:

- São mostrados dois conjuntos de informações ou saídas (S) com N_1 e N_2 observações, respectivamente modelados pelo mesmo modelo linear e observados em momentos diferentes para um mesmo fenômeno, cuja representação algébrica pode ser:

$$S_{1i} = a_1 + b_{1i} + u_{1i} \quad \text{com } i = 1, 2, \dots, N_1;$$

$$S_{2i} = a_2 + b_{2i} + u_{2i} \quad \text{com } i = 1, 2, \dots, N_2.$$

- São combinadas todas N_1 e N_2 observações e são feitas as estimativas de mínimos quadrados de a e b na regressão combinada $S = a + bx$. Desta equação é obtida a soma do quadrado dos resíduos SQ_R , com $(N_1 + N_2 - k)$ graus de liberdade ($k =$ número de parâmetros a serem estimados).

- De forma análoga são obtidas as somas de quadrados SQ_1 e SQ_2 da primeira e da segunda equações separadamente, com $(N_1 - k)$ e $(N_2 - k)$ graus de liberdade, respectivamente. Chega-se a $SQ_C = (SQ_1 + SQ_2)$, com $(N_1 + N_2 - 2k)$ graus de liberdade.
- Tem-se $SQ_P = SQ_R - SQ_C$ para a aplicação do teste F com a seguinte estatística:

$$F = \frac{\frac{SQ_P}{k}}{\frac{SQ_C}{N_1 + N_2 - 2k}}.$$

Se $F > F_\alpha$ para o nível de significância α , rejeita-se a hipótese de que os parâmetros a 's e b 's sejam iguais.

O teste verifica se todos os parâmetros a e b envolvidos nos modelos são respectivamente iguais a um parâmetro a e um parâmetro b . Gujarati(1970b), com a utilização de variáveis dummy, apresentou um teste com a seguinte proposta para o modelo:

$$S_i = a_0 + a_1D + a_2Y_i + a_3(Dx_i) + u_i \quad \text{com } i = 1, 2, \dots, (N_1 + N_2),$$

em que $D = 1$ se a observação pertence ao conjunto de observações N_1 (primeiro conjunto de observações) e $D = 0$ se a observação pertence ao conjunto de observações N_2 (segundo conjunto de observações).

Com a ajuda de variáveis dummy aditivas e multiplicativas torna-se possível verificar se duas equações de regressão lineares diferem uma da outra no intercepto, na inclinação ou em ambos os parâmetros.

Graybill (1976) apresenta um teste para verificar se os modelos lineares são paralelos, se têm o mesmo intercepto ou se são idênticos. Partindo de H modelos lineares simples:

$$\left\{ \begin{array}{ll} Y_{1j} = x_1 + \beta_1 x_{1j} + \varepsilon_{1j} & j = 1, 2, \dots, n_1 \\ Y_{2j} = x_2 + \beta_2 x_{2j} + \varepsilon_{2j} & j = 1, 2, \dots, n_2 \\ \dots & \dots \\ Y_{Hj} = x_H + \beta_H x_{Hj} + \varepsilon_{Hj} & j = 1, 2, \dots, n_H \end{array} \right.$$

com $\sum_{h=1}^H n_h = N$, $n_h > 2$ para todo h , $\varepsilon_{ij} \sim N(\varepsilon : 0, \sigma^2)$, o autor formulou as seguintes hipóteses, propondo as respectivas estatísticas:

1) $H_0 : \beta_1 = \beta_2 = \dots = \beta_H$ (as H linhas são paralelas) e

$H_a : \beta_h \neq \beta_{h'}$ para pelo menos um $h \neq h'$ ($h, h' = 1, 2, \dots, H$).

Rejeita-se H_0 se e somente se $W_p \geq F_{\alpha; H-1, N-2H}$, tal que:

$$W_p = \frac{\sum_{h=1}^H \left[\hat{\beta}_h - \frac{\sum_{j=1}^H \hat{\beta}_j b_{jj}}{\sum_{i=1}^H b_{ii}} \right] b_{hh}}{(H-1)\sigma^2}, \text{ em que } b_{hh} = \sum_{t=1}^{n_h} (x_{ht} - \bar{x}_h)^2.$$

2) $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_H$ (as H linhas possuem o mesmo intercepto) e

$H_a : \alpha_{h'} \neq \alpha_{h''}$ para pelo menos um $h' \neq h''$ ($h, h' = 1, 2, \dots, H$).

Rejeita-se H_0 se e somente se $W_t \geq F_{\alpha; H-1, N-2H}$, em que:

$$W_t = \frac{\sum_{k=1}^H \left[\hat{\alpha}_k - \frac{\sum_{j=1}^H \hat{\alpha}_j \alpha_{jj}}{\sum_{i=1}^H \alpha_{ii}} \right]^2 \alpha_{kk}}{(H-1)\sigma^2} \quad \text{em que } \alpha_{kk} = \frac{n_k \sum_{t=1}^{n_k} (x_{kt} - \bar{x}_k)^2}{\sum_{s=1}^{n_k} x_{ks}^2}$$

3) $H_0 : \alpha_1 + \beta_1 x_0 = \alpha_2 + \beta_2 x_0 = \dots = \alpha_H + \beta_H x_0$ (as H linhas possuem intercepto no ponto x_0 conhecido) e

H_a : pelo menos uma linha não tem intercepto no ponto x_0 conhecido.

Rejeita-se H_0 se e somente se $W_o \geq F_{\alpha; H-1, N-2H}$, em que

$$W_o = \frac{\sum_{h=1}^H \left[\left(\hat{\alpha}_h + \hat{\beta}_h x_0 \right) - \frac{\sum_{j=1}^H \left(\hat{\alpha}_j + \hat{\beta}_j x_0 \right) c_{jj}}{\sum_{i=1}^H c_{ii}} \right]^2 c_{hh}}{(H-1)\sigma^2},$$

$$\text{em que } c_{kk} = \frac{n_k \sum_{t=1}^{n_k} (x_{kt} - x_0)^2}{\sum_{s=1}^{n_k} (x_{ks} - x_0)^2}$$

Draper e Smith (1998) apresentam um teste de comparação de modelos lineares utilizando teste da razão de Verossimilhança. Bates e Watts

(1988) apresentam teste de razão de verossimilhança aplicando a aproximação dada pela estatística F para comparar modelos não lineares. Considerando uma notação tal que f denota o modelo completo, p denota o modelo reduzido, S é a soma de quadrados residuais, v são os graus de liberdade e P denota o número de parâmetros, Bates e Watts (1988) sumarizam os cálculos como na Tabela 2.5.1.

Tabela 2.5.1 - Soma de quadrados na análise de modelos aninhados

Fontes de Variação	Soma de Quadrados	GL	Quadrados Médios	F _{calc}
Parâmetros Extras	$S_e = S_p - S_f$	$V_e = P_f - P_p$	$S_e^2 = S_e / v_e$	S_e^2 / S_f^2
Modelo Completo	S_{ef}	$V_e = N - P_f$	$S_f^2 = S_f / v_f$	
Modelo Reduzido	S_p	$N - P_p$		

GL = Graus de Liberdade

Rao (1973) apresenta a estatística:

$$\Lambda_e = \frac{\sup_{R_i(\theta)=0} L(\theta)}{\sup_{\theta} L(\theta)},$$

em que o numerador é o supremo quando θ está sujeito às restrições $R_i(\theta)=0$, $i = 1, \dots, k$ e o denominador é o supremo sobre todos os valores admissíveis de θ . Rao (1973) aplica esta estatística em teste de regressão não linear e, conforme Regazzi (2004), tem aproximação dada pela estatística qui-quadrado.

Regazzi (2004) utiliza o teste da razão de Verossimilhança com aproximação da estatística qui-quadrado em teste para verificar a identidade de parâmetros e a identidade de modelos não lineares, utilizando o modelo de crescimento logístico com erro aditivo apresentado abaixo, e aplica a seguinte metodologia:

$$y_{ij} = \frac{a_i}{1 + b_i \exp(-c_i x_{ij})} + \varepsilon_{ij},$$

com

$j = 1, 2, \dots, n_i$;

$i = 1, 2, \dots, g$;

$a_i, b_i, c_i > 0$, sendo:

y_{ij} o valor observado na j -ésima unidade experimental do i -ésimo grupo;

x_{ij} o valor da variável independente;

a_i, b_i, c_i os parâmetros a serem estimados;

ε_{ij} o erro aleatório em que $\varepsilon_{ij} \sim \text{NID}(0, \sigma^2)$ e

$\sum_{i=1}^g n_i = n$ o número total de observações.

O autor considerou as seguintes hipóteses:

a) $H_0: a_1 = \dots = a_g$ versus H_1 : nem todos a_i são iguais;

b) $H_0: b_1 = \dots = b_g$ versus H_1 : nem todos b_i são iguais;

c) $H_0: c_1 = \dots = c_g$ versus H_1 : nem todos c_i são iguais;

d) $H_0: a_1 = \dots = a_g$ e $c_1 = \dots = c_g$ versus H_1 : pelo menos uma igualdade é uma desigualdade;

e) $H_0: a_1 = \dots = a_g, b_1 = \dots = b_g$ e $c_1 = \dots = c_g$ versus H_1 : pelo menos uma igualdade é uma desigualdade.

Com auxílio de variáveis dummy para identificação dos grupos, o modelo fica escrito como segue:

$$y_{ij} = \sum_{u=1}^g D_u \left[\frac{a_i}{1 + b_i \exp(-c_i x_{ij})} \right] + \varepsilon_{ij},$$

com $D_i = 1$ caso a observação y_{ij} pertença ao grupo i e 0 caso contrário.

Fazendo:

$$\mu(a_i, b_i, c_i, x_{ij}) = \frac{a_i}{1 + b_i \exp(-c_i x_{ij})} \text{ ou seja } \mu(a_i, b_i, c_i, x_{ij}) = y_{ij}$$

e dadas as observações y_{ij} , a função de verossimilhança é:

$$f(\theta) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^g \sum_{j=1}^{N_i} [y_{ij} - \mu(a_i, b_i, c_i, x_{ij})]^2\right\},$$

em que maximizar $f(\theta)$ com respeito aos parâmetros a_i, b_i, c_i , $i=1, \dots, g$ é o mesmo que minimizar

$$S(a_1, \dots, a_g, b_1, \dots, b_g, c_1, \dots, c_g) = \sum_{i=1}^g \sum_{j=1}^{N_i} [y_{ij} - \mu(a_i, b_i, c_i, x_{ij})]^2$$

Regazzi (2003) considera aqui o problema geral de se testar hipótese de nulidade $H_0: \theta \in w$ versus $H_1: \theta \in w^c$, em que w é um sub conjunto do espaço paramétrico Ω e w^c é o complementar de w em Ω . Regazzi (2003) afirma ainda que a estatística do teste da razão de verossimilhança para este problema é

$$L = \left(\frac{\sigma_\Omega^2}{\sigma_w^2} \right) \text{ sendo } \sigma_\Omega^2 \text{ a estimativa de máxima verossimilhança de } \sigma^2 \text{ para}$$

quando não ocorre nenhuma restrição no espaço paramétrico e σ_w^2 é a estimativa de máxima verossimilhança para σ^2 quando as restrições lineares colocadas em H_0 são impostas no espaço paramétrico Ω . Do exposto, tem-se que:

$$L = \frac{\sigma_\Omega^2}{\sigma_w^2} = \frac{(2\pi)^{-\frac{n}{2}} (\hat{\sigma}_0^2)^{-\frac{n}{2}} \exp\left\{-\frac{n}{2}\right\}}{(2\pi)^{-\frac{n}{2}} (\hat{\sigma}^2)^{-\frac{n}{2}} \exp\left\{-\frac{n}{2}\right\}} = \frac{(\hat{\sigma}_0^2)^{-\frac{n}{2}}}{(\hat{\sigma}^2)^{-\frac{n}{2}}};$$

$$\ln(L) = \ln \left(\frac{\left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right)^{-\frac{n}{2}}}{\left(\frac{\hat{\sigma}^2}{\sigma} \right)^{-\frac{n}{2}}} \right) = \frac{n}{2} \ln \left(\frac{\left(\frac{\hat{\sigma}^2}{\sigma} \right)}{\left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right)} \right);$$

$$2 \ln(L) = n \ln \left(\frac{\left(\frac{\hat{\sigma}^2}{\sigma} \right)}{\left(\frac{\hat{\sigma}^2}{\sigma_0^2} \right)} \right) \xrightarrow[n \rightarrow \infty]{d} \chi_v^2;$$

em que v é grau de liberdade calculado pelo número de parâmetros estimados em Ω menos o número de parâmetros estimados em w , conforme Rao (1973).

Regazzi(2004) acrescenta ainda que o teste para as hipóteses apresentadas contra as hipóteses alternativas correspondentes e as estatísticas do teste da razão de verossimilhança com distribuição qui-quadrado são, aproximadamente, $-n \ln \left(\frac{\sigma_\Omega^2}{\sigma_w^2} \right) = -n \ln \left(\frac{SQR_\Omega}{SQR_w} \right)$ para cada hipótese testada.

2.6 - Simulação e o Método Monte Carlo

Os matemáticos são geralmente classificados como adeptos da matemática pura ou da matemática aplicada. Todavia, uma forma de classificação que vai se tornando cada dia mais presente é a do adepto da matemática experimental. O que difere o matemático experimental dos demais é que este trabalha com aparatos em laboratório ou em um local que permita reproduzir, experimentalmente, o fenômeno em estudo. A principal diferença

entre os teóricos e os experimentais é que os primeiros deduzem conclusões a partir de postulados, enquanto os últimos inferem conclusões de observações.

Segundo Bussab e Morettin (2004), a necessidade de se trabalhar com situações que envolvam aleatoriedade é muito comum, mas mesmo construindo o modelo probabilístico, algumas questões não podem ser resolvidas analiticamente e torna-se necessário recorrer a estudos de simulação para a obtenção de aproximações das quantidades de interesse. O método Monte Carlo preenche a lacuna existente na Estatística Experimental quando se precisa trabalhar com variáveis deste tipo. Este é um método numérico proposto para resolver problemas matemáticos pela simulação de variáveis aleatórias.

Aceita-se que o surgimento do método Monte Carlo data de 1949, quando foi publicado o artigo intitulado “The Monte Carlo Method”- Metrópolis (1949). Porém, mesmo antes de 1949 diversos problemas foram muitas vezes resolvidos por meio de amostragem aleatória, que é de fato o princípio do método Monte Carlo. Contudo, como a simulação é um trabalho pesado, este método só se tornou uma técnica de utilização mais ampla a partir da popularização dos computadores.

Antes dos computadores a geração de variáveis aleatórias acontecia, segundo Ross (1990), manualmente ou por meios mecânicos, como, por exemplo, através do uso de recursos como a roleta, o lançamento de dados e a retirada de cartas, entre outros, critérios que ainda hoje estão intimamente relacionados a jogos. Por este motivo, em homenagem aos famosos cassinos do Principado de Mônaco, este é chamado de Método Monte Carlo.

Segundo Sobol (1994), a expressão variável aleatória é usada com a conotação de que não se sabe o valor exato que poderá assumir. O mesmo autor afirma ainda que, para os matemáticos, o termo variável aleatória significa que, mesmo não sabendo o valor desta variável em determinada ocorrência, são

conhecidos os diversos valores ela poderá assumir e as probabilidades destes valores.

A simulação é definida por Ross (1990) como a habilidade de gerar números aleatórios. Contudo, computacionalmente, o que se pode gerar, na verdade, são os chamados números pseudo-aleatórios.

Números pseudos-aleatórios constituem uma seqüência de valores deterministicamente gerados e, assim, existe uma seqüência de observações a partir da qual haverá uma repetição previsível, o que descaracteriza a aleatoriedade. Portanto, as variáveis aleatórias geradas computacionalmente são, de fato, aproximações de variáveis aleatórias, ainda segundo Ross (1990).

Bussab e Morettin (2004) afirmam que, de maneira particular, a simulação pode ser entendida como a aplicação de uma variável aleatória de um modelo (Normal, Poisson, Binomial e outros). Assim, os valores simulados podem ser considerados como elementos de uma amostra.

Uma forma de gerar números pseudo-aleatórios é partir de um valor inicial x_0 , chamado semente e, recursivamente, computar valores sucessivos x_n , $n \geq 1$, fazendo:

$$x_n = ax_{n-1} \text{ módulo } m,$$

em que a e m são valores inteiros positivos conhecidos. O módulo m condiciona a divisão de ax_n por m e o resto é tomado como valor de x_n . Assim, x_n é um valor entre $0, 1, 2, 3, 4, 5, 6, \dots, m-1$ e a quantidade x_n/m é chamada de número pseudo-aleatório e é tomada como uma aproximação para o valor da variável aleatória uniforme entre $(0,1)$. Esta função, usada para gerar números aleatórios, estabelece o chamado Método Congruencial Multiplicativo.

Desde que cada valor x_n assume um valor $0, 1, 2, 3, 4, 5, 6, \dots, m-1$, segue que após um número finito de valores gerados, a seqüência de números deverá novamente repetir. Por este motivo, as constantes a e m deverão ser tais que a

quantidade de números gerados deverá ser grande antes que haja repetição. Estas constantes deverão, pois, satisfazer três condições:

1. Para qualquer semente, a seqüência resultante deve ter a aparência de uma seqüência de variáveis aleatórias uniformes independentes (0,1).
2. Para um valor inicial de semente, o número de variáveis aleatórias que pode ser gerado antes da repetição da seqüência é grande.
3. Os valores podem ser computados eficientemente em um computador digital.

Outra forma de gerar números pseudo-aleatórios é utilizando recursos do tipo:

$$X_n = (ax_{n-1} + c) \text{ módulo } m.$$

Este tipo de gerador é chamado gerador Congruencial Misto por envolver termos aditivos e multiplicativos. A escolha de m para este tipo de gerador é mais eficiente quando este é tomado com tamanho igual ao da palavra no computador (32 bits, 16 bits), ainda conforme Ross (1990).

O método Monte Carlo baseia-se no fato de que a partir de uma variável aleatória uniforme (0,1) pode-se gerar outra com distribuição de probabilidade qualquer, usando a correspondente função de distribuição acumulada. Este procedimento é formalizado no resultado chamado de Teorema da Probabilidade Integral.

2.6.1 O Teorema da Probabilidade Integral

A geração de uma variável aleatória de uma distribuição qualquer é obtida aplicando-se a inversa da função de distribuição acumulada da distribuição desejada a uma variável aleatória uniforme (0,1). O resultado

matemático que garante que a imagem da função de distribuição acumulada de qualquer variável aleatória tenha distribuição uniforme (0,1) é chamado de Teorema da Probabilidade Integral, com o seguinte enunciado e demonstração:

Enunciado:

“Seja X uma variável aleatória (v.a.) com função de densidade de probabilidade (f.d.p.) e seja a função de densidade acumulada (f.d.a.) dada por F . Então $U=F(X)$ tem distribuição uniforme no intervalo (0,1).”

Demonstração:

Seja F uma f.d.a. de uma v.a. contínua, F é estritamente crescente, ou seja, F é bijetora; portanto, se $u=F(X)$ existe, $x=F^{-1}(u)$. Seja $G(u)$ a f.d.a. de u .

$$G(u)=P(U \leq u)= P(F(X) \leq u) = P(X \leq F^{-1}(u))= F(F^{-1}(u)) = FoF^{-1}(u) = u.$$

Mas X é v.a. contínua; portanto, $F(X) = \int_{-\infty}^x f(x)dx =G(u)$ é a f.d.a

de X . Daí,

$$g(u) = \frac{d}{du} G(u) = \frac{d}{du} u = 1.I_{(0,1)}(u) ,$$

em que $I_{(0,1)}(u)$ é a função indicadora, logo: $u \sim U(0,1)$.

3 METODOLOGIA

A metodologia proposta nesse trabalho visa avaliar o método proposto por Regazzi (2003) para a verificação da identidade de modelos não lineares. Para tanto foram consideradas diferentes tamanhos de amostra ($n = 15, 50$ e 100) e diferentes números de equações ($k = 2, 3$ e 4), com k representando o número de populações.

3.1 - Modelos

Os modelos analisados foram: Logístico, de Gompertz (Hoffmann e Vieira, 1977) e de Langmuir (Ratkowsky, 1990). Para cada modelo e para cada combinação do tamanho amostral com o número de equações, foram realizadas 2000 simulações Monte Carlo.

Para cada amostra simulada foram estimados os parâmetros do modelo completo e do modelo reduzido, referenciados aqui por Ω e ω , respectivamente. Essa notação foi utilizada no modelo Logístico (expressões 3.1.1 e 3.1.2), no modelo de Langmuir (expressões 3.1.3 e 3.1.4) e no modelo de Gompertz (expressões 3.1.5 e 3.1.6).

$$y_{ij} = \Omega_{\log} = \sum_{u=1}^k D_u \left[\frac{\alpha_u}{1 + e^{(\beta_u - \gamma_u)x}} \right] + \varepsilon_{ij}, j=1, \dots, n ; n=1, \dots, k \quad (3.1.1)$$

$$y_j = \omega_{\log} = \frac{\alpha}{1 + e^{(\beta - \gamma)x}} + \varepsilon_j, j=1, \dots, nk \quad (3.1.2)$$

$$y_{ij} = \Omega_{Lang} = \sum_{u=1}^k D_u \left[\frac{\alpha_u \beta_u X^{1-\gamma_u}}{1 + \beta_u X^{1-\gamma_u}} \right] + \varepsilon_{ij}, j=1, \dots, n ; n=1, \dots, k \quad (3.1.3)$$

$$y_j = \omega_{Lang} = \frac{\alpha \beta X^{1-\gamma}}{1 + \beta X^{1-\gamma}} + \varepsilon_j, \quad j=1, \dots, nk \quad (3.1.4)$$

$$y_{ij} = \Omega_{gomp} = \sum_{u=1}^k D_u [\exp(\alpha_u + \beta_u y^x)] + \varepsilon_{ij},$$

(3.1.5)

$$y_i = \overline{\omega}_{gomp} = \exp(\alpha + \beta y^{x_j}) + \varepsilon_j, \quad j = 1, 2, \dots, nk \quad (3.1.6).$$

De acordo com os modelos estabelecidos, definiu-se a seguinte hipótese:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_g = \alpha \quad \text{e} \quad \beta_1 = \beta_2 = \dots = \beta_g = \beta \quad \text{e} \quad \gamma_1 = \gamma_2 = \dots = \gamma_g = \gamma$$

H_1 : que pelo menos uma igualdade é uma desigualdade.

Nos modelos completos, acrescentou-se uma variável dummy com o intuito de identificar a amostra perante a população, determinada pela equação, ou seja, valores assumidos por k . Assim, nas expressões 3.1.1, 3.1.3 e 3.1.5, para a variável D foi atribuído o valor 1 se a observação y_{ij} pertence à população ($i = 1, 2, \dots, k$) e o valor 0, caso contrário.

Com a finalidade de garantir a obtenção dos valores de X em todo o domínio do modelo, utilizou-se o seguinte recurso:

- 1 - Gerou-se uma variável u_0 uniforme no intervalo de existência da imagem do modelo, ou seja, para o modelo logístico $u_0 \sim U(0, \alpha)$, para o modelo de Gompertz $u_0 \sim U(0, e^\alpha)$ e para o modelo de Langmuir $u_0 \sim U(0, \alpha)$.
- 2 - Em seguida gerou-se a variável aleatória x , em que $x = f(u_0, \theta)$, com $f(u_0, \theta)$ dada pela inversa de cada um dos modelos estudados, como segue:

$$f(u_0, \theta) = \frac{\beta - \ln\left(\frac{\alpha - u_0}{u_0}\right)}{\gamma}$$

para o modelo logístico;

$$f(u_0, \theta) = e^{\left(\frac{\ln\left(\frac{u_0}{\alpha\beta - u_0\beta}\right)}{1-\gamma}\right)}$$

para o modelo de Langmuir e

$$f(u_0, \theta) = \frac{\ln\left(\frac{\ln(u_0) - \alpha}{\beta}\right)}{\ln(\gamma)}$$

para o modelo de Gompertz.

A estatística utilizada por Regazzi (2003), definida pela razão de verossimilhanças para cada modelo adotado, foi dada pelas expressões abaixo:

$$L_{R(\text{Log})} = N \ln \left(\frac{QM(\omega_{\text{Log}})}{QM(\Omega_{\text{Log}})} \right),$$

$$L_{R(\text{Gomp})} = N \ln \left(\frac{QM(\omega_{\text{Gomp}})}{QM(\Omega_{\text{Gomp}})} \right) e$$

$$L_{R(\text{Lang})} = N \ln \left(\frac{QM(\omega_{\text{Lang}})}{QM(\Omega_{\text{Lang}})} \right).$$

em que $N = nk$. $QM(\Omega.)$ refere-se ao quadrado médio residual obtido por meio da análise de variância de cada modelo completo e, analogamente, $QM(\omega.)$ refere-se ao modelo reduzido.

Para estas estatísticas a regra de decisão foi determinada por um quantil $(1-\alpha^*)$ de uma distribuição qui-quadrado com v graus de liberdade, calculado pela diferença entre o número de parâmetros do modelo completo e o número de parâmetros do modelo reduzido e com α^* correspondendo ao nível de significância.

3.2 - Identidade de modelos não-lineares

Para obtenção das somas de quadrados residuais utilizadas nas estatísticas provenientes da razão de verossimilhanças, em todos os modelos avaliados, inicialmente determinou-se a matriz das derivadas parciais. Para simplificar a notação, assumiu-se os modelos Logístico, Langmuir e Gompertz como uma função $f(X,\theta)$ em que o vetor paramétrico foi dado por $\theta=(\alpha,\beta,\gamma)$ e \mathbf{X} , pelo o vetor das variáveis explicativas. Assim, a matriz das derivadas parciais, definidas em relação ao modelo completo, foi dada por:

$$F_{\Omega} = \begin{bmatrix} \frac{\partial f}{\partial \alpha_{11}} & \frac{\partial f}{\partial \beta_{11}} & \frac{\partial f}{\partial \gamma_{11}} & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & & & & \vdots \\ \frac{\partial f}{\partial \alpha_{1n}} & \frac{\partial f}{\partial \beta_{1n}} & \frac{\partial f}{\partial \gamma_{1n}} & & & & & 0 \\ & & & & & & & \vdots \\ & & & & & & & 0 \\ & & & & & & \frac{\partial f}{\partial \alpha_{k1}} & \frac{\partial f}{\partial \beta_{k1}} & \frac{\partial f}{\partial \gamma_{k1}} \\ & & & & & & \vdots & \vdots & \vdots \\ & & & & & & \frac{\partial f}{\partial \alpha_{kn}} & \frac{\partial f}{\partial \beta_{kn}} & \frac{\partial f}{\partial \gamma_{kn}} \end{bmatrix},$$

com $k = 1, 2, \dots$, número de equações e $n = 1, 2, \dots$, número de observações.

Os elementos nas diagonais representam as derivadas em relação a cada parâmetro, e fora das diagonais, o valor assumido é zero. Em relação ao modelo reduzido, a matriz das derivadas parciais foi representada por:

$$F_w = \begin{bmatrix} \frac{\partial f}{\partial \alpha_1} & \frac{\partial f}{\partial \beta_1} & \frac{\partial f}{\partial \gamma_1} \\ \vdots & \vdots & \vdots \\ \frac{\partial f}{\partial \alpha_n} & \frac{\partial f}{\partial \beta_n} & \frac{\partial f}{\partial \gamma_n} \end{bmatrix},$$

em que $n = 1, 2, \dots$, ik para $i = 1, 2, \dots$, número de observações e $k = 1, 2, \dots$, número de equações.

Em todas as equações, ε_{ij} é o erro aleatório associado à j -ésima observação do i -ésimo modelo com $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, n$. Pela própria formalização do modelo, tem-se como suposição assumir a distribuição desses erros como normal com média zero e variância constante, ou seja, $\varepsilon_{ij} \sim N(0, \sigma^2)$. Entretanto, desejando avaliar a robustez deste teste perante as distribuições não normais, simularam-se os mesmos modelos considerando o erro dado por uma distribuição uniforme no intervalo (0,1) isto é, $\varepsilon_{ij} \sim U(0,1)$.

Pra garantir uma qualidade razoável, definida pelo valor do coeficiente de determinação $R^2=0,90$, foram realizadas as seguintes operações:

$$R^2 = 1 - \frac{SQR}{SQtotalc} \Rightarrow 1 - R^2 = \frac{SQR}{SQtotal}.$$

Como $SQR + SQMC = SQTotal$, então:

$$1 - R^2 = \frac{SQR}{SQR + SQMc} \Rightarrow SQR = (1 - R^2)(SQR + SQMc)$$

$$SQR - (1 - R^2)SQR = (1 - R^2)SQMc \Rightarrow (1 - 1 + R^2)SQR = (1 - R^2)SQMc .$$

Isolando-se a soma de quadrados residual (SQR), tem-se a seguinte expressão:

$$SQR = \frac{\left((1 - R^2) \times SQMc \right)}{R^2} ,$$

em que R^2 foi fixado em 0,9 e SQMc correspondeu à soma de quadrado do modelo completo. Importante ressaltar que o valor do SQMc assume valor diferente para cada modelo avaliado.

3.3 - Simulação Monte Carlo

O critério utilizado para avaliar a eficiência do teste de razão de verossimilhança proposto por Regazzi (2003) foi dado pelo controle do erro tipo I e poder do teste proposto. A determinação da probabilidade do erro tipo I foi dada pela proporção entre o número de vezes em que a H_0 foi rejeitada, considerando $\alpha^* = 5\%$, e o número total de simulações. Devido à possibilidade da obtenção de valores negativos, optou-se por considerar como simulações válidas apenas os casos em que estas situações não ocorreram. Determinou-se, como satisfatório, um mínimo de 1000 realizações amostrais válidas.

Em relação à obtenção dos valores de poder, em que sabidamente os parâmetros populacionais foram diferenciados, conforme evidenciou a hipótese alternativa H_1 , em que pelo menos uma desigualdade ocorreu.

Os parâmetros utilizados na simulação referente ao controle do erro tipo I foram dados por:

Modelo Logístico: $\alpha_i = 1.504, \beta_i = 1.025, \gamma_i = 0.090$;

Modelo de Gompertz: $\alpha_i = 1.504, \beta_i = -1.025, \gamma_i = 0.090$;

Modelo de Langmuir: $\alpha_i = 1.504, \beta_i = 1.025, \gamma_i = 0.090$;

em que $i = 1, 2, \dots, k$. Os valores de poder foram simulados sob H_1 , entretanto a desigualdade foi determinada apenas para o valor do parâmetro α_2 , sendo este fixado em 5,504.

Este procedimento foi repetido tanto para análise ideal do teste, sob a hipótese de normalidade dos erros, como na verificação da robustez.

Para a realização da metodologia proposta, implementou-se um programa utilizando Statistical Analysis System (1999) apresentado no Anexo .

4 RESULTADOS E DISCUSSÃO

Inicialmente, aplicou-se o teste de Regazzai(2003) em dois experimentos, sendo que, na primeira situação, a estatística obtida por meio dos valores amostrais foi negativa, sendo incoerente com a distribuição qui-quadrado. Os resultados são dados abaixo.

Tabela 4.1. Valores gerados para o modelo Logístico considerando duas equações e tamanho da amostra para n=20.

Experimento	X	Y	Experimento	X	y
1	2,76673	1,49011	2	1,02154	1,34741
1	0,88648	1,32629	2	2,60440	1,47935
1	-0,26113	1,01877	2	1,80239	1,42025
1	-0,53994	0,90855	2	-0,16645	1,07507
1	2,12958	1,45162	2	3,98946	1,49441
1	0,47346	1,23633	2	0,80776	1,32260
1	2,35914	1,47916	2	0,81532	1,30380
1	1,30463	1,39291	2	0,44452	1,24740
1	2,70835	1,48349	2	2,23010	1,48453
1	2,92016	1,48617	2	2,13991	1,47109
1	1,53532	1,41080	2	0,29616	1,19446
1	3,42515	1,50720	2	-1,65005	0,48948
1	-1,67151	0,46031	2	-2,10662	0,32108
1	-0,00056	1,10778	2	3,21365	1,48411
1	1,82336	1,45467	2	-0,49097	0,92638
1	1,80470	1,43541	2	1,99889	1,45809
1	0,29128	1,17933	2	0,45976	1,24393
1	0,87056	1,31025	2	3,60861	1,50016
1	2,88442	1,49556	2	-1,701300	0,45441
1	1,13898	1,34309	2	1,39635	1,40561

Para cada experimento, os resultados provenientes da análise de variância para o modelo completo e reduzido são apresentados na Tabela 4.2.

Tabela 4.2 Análise de variância para os modelos completo e reduzido referente aos dados amostrais dos dois experimentos.

EXP.	Modelo Completo			Modelo Reduzido		
	SQR	GL	QM	SQR	GL	QM
1	0,00339	34	0,0000997059	0,00351	37	0,000095
2	0,00340	34	9,99118E-05	0,00405	37	0,000110

Posteriormente à obtenção dos quadrados médios residuais, aplicou-se a estatística provenientes da razão de verossimilhança, conforme ilustrado na metodologia. Os resultados desta estatística são descritos na Tabela 4.3.

Os resultados encontrados na Tabela 4.3 revelaram uma situação que exemplificou uma crítica em relação à estatística de verossimilhança proposta pelo teste de identidade de modelos não lineares. Uma forma de contornar esse problema é considerar o valor absoluto do ln. Novamente, vale ressaltar que no processo de simulação, os casos em que a estatística do teste apresentou valores negativos foram excluídos da análise.

Tabela 4.3. Valores das estatística obtidas com a aplicação do teste da razão de verossimilhança nos experimentos avaliados.

Experimento	Valor da estatística da razão de verossimilhança
1	-0,050911523
2	0,091266804

4.1 - Controle do Erro Tipo I

4.1.1 - Modelo logístico

Todos os resultados apresentados foram obtidos considerando pelo menos 1000 simulações válidas e foi testada a hipótese nula de que todas as equações são idênticas, isto é, a equação comum pode ser usada como estimativa das equações envolvidas.

Os resultados encontrados na Tabela 4.4 para o modelo Logístico referem-se às probabilidades obtidas por meio do método Monte Carlo, sendo estas comparadas com o valor nominal fixado em 5%, com erros normais e uniformes. Conforme descrito na metodologia, as configurações avaliadas foram representadas pelo número de equações e tamanho amostral.

Tabela 4.4. Probabilidades obtidas via simulação Monte Carlo referente ao controle da taxa do Erro Tipo I para o modelo Logístico, em que todos os parâmetros são iguais.

Nº Equações (k)	Tamanho	$\epsilon \sim N(0, \delta^2)$	$\epsilon \sim u(0,1)$
	Amostra (n)	Prob.	Prob.
2	15	0,0825611	0,254902
2	50	0,0357143	0,2579801
2	100	0,0360515	0,2408994
3	15	0,0419977	0,0507246
3	50	0,0047923	0,2783172
3	100	0,0063694	0,5100671
4	15	0,063211	0,5352381
4	50	0,0114943	0,3300971
4	100	0,0032103	0,5786925

Considerando a normalidade nos resíduos, sendo esta a situação ideal

na construção do teste, constatou-se que para 2 equações, as probabilidades observadas para os tamanhos amostrais de $n = 50$ e $n = 100$, o método de identidade de modelos aplicado ao modelo Logístico controlou o erro tipo I. Entretanto, verificou-se que para amostras menores, isto é, $n = 15$, a eficiência no controle deste erro foi prejudicada, embora seu valor tenha sido relativamente próximo ao valor nominal fixado.

Analisando o caso em que os resíduos foram dados pela distribuição Uniforme (0,1), por meio dos resultados (Tabela 4.4) ficou evidente que o controle do erro tipo I não foi eficiente. Este fato foi perceptível devido aos valores das probabilidades para todas as amostras serem bastante superiores em relação ao nível nominal, com diferença de aproximadamente vinte pontos percentuais.

Uma outra característica observada nos resultados descritos por meio da Tabela 4.4 foi a constatação de que o modelo Logístico, submetido ao teste de identidade, mostrou uma sensibilidade em relação ao número de equações. Essa sensibilidade é detectada ao se compararem os resultados, provenientes de duas equações, com os resultados oriundos de 3 equações. Neste aspecto, o aumento do número de equações provocou um comportamento conservativo para amostras maiores ($n = 50$ e $n = 100$); isto é, as probabilidades empíricas para o controle do erro tipo I foram bem abaixo do valor nominal fixado em 5%. Este resultado conservativo também se confirmou aumentando o número de equações ($k = 4$).

Em se tratando dos resíduos não normais, de um modo geral o teste de identidade de modelos aplicado ao modelo Logístico não controlou o erro tipo I, produzindo probabilidades superiores ao nível nominal e inconsistentes com a teoria assintótica.

4.1.2 - Modelo de Gompertz

Os resultados encontrados na Tabela 4.5 corresponderam às probabilidades do erro tipo I, dadas por meio da aplicação do teste de identidade de modelos considerando a comparação das equações dadas pelo modelo Gompertz.

Considerando os resíduos normalmente distribuídos, observou-se que o teste de identidade de modelos aplicado a grandes amostras propiciou probabilidades próximas ao nível nominal. Este fato foi notório diante do aumento do número de equações, de tal forma que este aumento tornou o teste gradativamente conservativo.

Tabela 4.5. Probabilidades obtidas via simulação Monte Carlo referente ao controle da taxa de Erro Tipo I para o modelo de Gompertz, em que todos os parâmetros são iguais.

Nº Equações (k)	Tamanho Amostra (n)	$\varepsilon \sim N(0, \delta^2)$	$\varepsilon \sim u(0,1)$
		Prob.	Prob.
2	15	0,068259	0,0595581
2	50	0,0412882	0,0440816
2	100	0,037947	0,037553
3	15	0,0350872	0,0203804
3	50	0,0356870	0,014444
3	100	0,0150215	0,018393
4	15	0,0143678	0,0130112
4	50	0,0016051	0,0043924
4	100	0,0059701	0,0057887

No caso da análise da robustez, cujos resultados foram dados considerando os resíduos distribuídos uniformemente, verificou-se que o teste

foi robusto. Isto se deu pelo fato de os valores das probabilidades obtidas serem aproximados dos valores obtidos na situação de os resíduos serem distribuídos normalmente. Além dos mais, estas probabilidades mostraram-se condizentes com o nível nominal fixado.

Analisando o tamanho amostral, verificou-se que apenas para as amostras pequenas o teste não controlou o erro tipo I. No entanto, aumentando o número de equações, esta deficiência foi corrigida, de forma que o teste passou a ser conservativo. Este fato foi detectável inclusive na situação em que os resíduos foram dados pela distribuição uniforme.

4.1.3 - Modelo de Langmuir

Em relação à aplicação do teste de identidade de modelos sobre o modelo de Langmuir, as probabilidades do erro tipo I encontradas na Tabela 4.6 permitiram constatar que, sob normalidade a performance do teste mediante este modelo, novamente confirmou-se que o aumento do número de equações produziu um efeito conservativo. Isto se deu para todos os tamanhos amostrais avaliados.

No caso dos resultados obtidos mediante os resíduos não normais, verificou-se que para as amostras pequenas, independentemente do número de equações, o teste de identidade de modelos aplicados ao modelo Langmuir não controlou o erro tipo I. Contudo, aumentando o tamanho da amostra para $n = 50$ e $n = 100$, considerando o número de equações definido em $k = 2$, notou-se que o teste proporcionou probabilidades próximas ao valor nominal fixado, isto considerando o erro de Monte Carlo. Assim, detectou-se a robustez do teste mediante essas especificações.

Tabela 4.6. Probabilidades obtidas via simulação Monte Carlo referente ao controle da taxa de Erro Tipo I para o modelo de Langmuir, em que todos os parâmetros são iguais.

Nº Equações (k)	Tamanho Amostra (n)	$\varepsilon \sim N(0, \delta^2)$	$\varepsilon \sim u(0,1)$
		Prob.	Prob.
2	15	0,0694275	0,090991
2	50	0,041515	0,0517241
2	100	0,0352369	0,0569683
3	15	0,0355097	0,1071429
3	50	0,0092879	0,085329
3	100	0,0102041	0,119048
4	15	0,0196078	0,25000
4	50	0,0053476	0,22222
4	100	0,0107143	0,10000

4.2 - Poder

4.2.1 Modelo Logístico

Considerando as mesmas condições relativas ao número de equações, tamanho amostral e distribuição dos resíduos avaliados em relação ao controle do erro tipo I, verificou-se que, de um modo geral, para os erros normais o teste de identidade de modelos foi altamente poderoso (Tabela 4.7). No entanto, vale ressaltar que mediante as amostras menores, $n = 15$, esse teste não controlou o erro tipo I. Isto foi verificado em todas as situações contempladas pelo número de equações avaliadas ($k = 2, 3$ e 4), considerando o erro de Monte Carlo,

principalmente detectado para $k = 3$; portanto, não fizeram sentido os resultados propiciados em relação ao poder do teste para esse tamanho amostral.

Vale ressaltar que nas situações em que o teste foi conservativo, os resultados referentes aos valores de poder foram satisfatórios, uma vez que era esperado que o poder fosse relativamente baixo.

Tabela 4.7. Probabilidades obtidas via simulação Monte Carlo referentes ao poder do teste, para o modelo Logístico com desigualdade no parâmetro alfa.

Nº Equações	Tamanho Amostra	$\varepsilon \sim N(0, \delta^2)$	$\varepsilon \sim u(0,1)$
		Prob.	Prob.
2	15	1	1,0000000
2	50	1	0,7904762
2	100	1	0,9928450
3	15	1	0,9829440
3	50	1	0,7766990
3	100	1	0,9218770
4	15	1	0,9684211
4	50	1	0,9103520
4	100	1	0,8889000

4.2.2 Modelos de Gompertz e Langmuir

Analisando o poder relacionado aos modelos de Gompertz (Tabela 4.8) e Langmuir (Tabela 4.9), mediante as situações analisadas no controle do erro tipo I (Tabelas 4.5 e 4.6), verificou-se que, de um modo geral, o teste de identidade de modelos mostrou-se altamente poderoso, com valores iguais de

poder específicos para esses modelos. No entanto, em consonância com os resultados propiciados pelo controle efetivo do erro tipo I, consideraram-se recomendáveis as seguintes situações.

No caso do modelo Gompertz, independentemente da distribuição do resíduo avaliada nesse trabalho e indiferentemente do número de equações, o teste de identidade e modelos mostrou valores de poder coerentes com o controle do erro tipo I para amostras maiores, $n = 50$ e $n = 100$. No caso das amostras menores, $n = 15$, os valores de poder foram considerados adequados a partir de 3 equações.

Tabela 4.8. Probabilidades obtidas via simulação Monte Carlo referentes ao poder de teste para o modelo de Gompertz com desigualdade no parâmetro alfa.

Nº Equações	Tamanho Amostra	$\varepsilon \sim N(0, \delta^2)$	$\varepsilon \sim u(0,1)$
		Prob.	Prob.
2	15	1	1
2	50	1	1
2	100	1	1
3	15	1	1
3	50	1	1
3	100	1	1
4	15	1	1
4	50	1	1
4	100	1	1

Em relação ao modelo Langmuir (Tabela 4.9), verificou-se que no caso do modelo de Gompertz, independentemente da distribuição do resíduo avaliada

nesse trabalho e indiferentemente do número de equações, o teste de identidade e modelos mostrou valores de poder coerentes com o controle do erro tipo I para amostras maiores, $n = 50$ e $n = 100$. No caso das amostras menores, $n = 15$, os valores de poder foram considerados adequados a partir de 3 equações.

Tabela 4.9. Probabilidades obtidas via simulação Monte Carlo referentes ao poder de teste para o modelo de Langmuir, com desigualdade no parâmetro alfa.

Nº Equações	Tamanho	$\varepsilon \sim N(0, \delta^2)$	$\varepsilon \sim u(0,1)$
	Amostra	Prob.	Prob.
2	15	1	1
2	50	1	1
2	100	1	1
3	15	1	1
3	50	1	1
3	100	1	1
4	15	1	1
4	50	1	1
4	100	1	1

5 CONCLUSÕES

O método de identidade aplicado a modelos não lineares apresentou sensibilidade em relação ao modelo especificado.

Diante da normalidade dos resíduos, concluiu-se que o aumento no número de equações produziu um efeito conservativo no controle do Erro Tipo I para os modelos de Gompertz e Langmuir, para todos os tamanhos amostrais avaliados.

O modelo de Gompertz apresentou melhor desempenho quando submetido a situações nas quais o resíduo não tinha distribuição normal.

REFERÊNCIAS BIBLIOGRÁFICAS

- Azzalini , Adelchi , Statistical Inference Base on Likelihood , , London , U.K , Chapman & Hall , 1996.
- Bates, D.; Watts, D.. Nonlinear regression analysis and its applications. New York: J. Willey, 1988.
- Battisti, Iara Denise Endruweit , Comparação entre modelos de regressão com aplicação em biometria florestal – Lavras : UFLA, 2001.
- Bearzoti ,E. , Introdução à Teoria de Probabilidade e à Inferência Estatística; Lavras, Editora da UFLA 1998.
- Bolfarini, H.; Rodrigues, J. , Cordani, L,; O modelo de regressão com erros na variáveis. 10º Simpósio Nacional de Probabilidade e Estatística de 3 a 7 de agosto de 1992. Rio de Janeiro : ABE , 1992.
- Bussab, W. O. ; Morettin , P. A.; Estatística Básica,5ª ed. São Paulo, S.P. ,Saraiva 2004.
- Carmer, S.G. “Optimal significance levels for application of the least significant difference in crop performance trials.” Crop Science, 16: 95-99 ,1976.
- Chow. G. C. Tests of equality between sets of coefficients in two linear regressions. Econometrica, Chicago, v 28, p 591-605, 1960.
- Draper, N.R.; Smith, H. Applied regression analysis, 3 ed. New York: Jonh Wiley and Sons. 1998.
- Farias, Alfredo.A. ; Soares, José. F.; Cesar, Cibele. C.. Introdução a Estatística. 2.ed. Rio de Janeiro: LTC, 2003 .
- Gallant, R., 1987, Nonlinear Statistical Models, John Wiley and Sons, New York .
- Graybill, F. A. Theory and application of linear model . Belmont : Duxbury Press, 1976 .
- Gujarati. D. Use of dummy variables in testing for equality between sets of coeficcients in linear regressions a note. The American Statiscian. Washington , v 24, n. 1 , p. 50-52 . Feb. 1970b.
- Hoffmann, R; Vieira , S.; Análise de Regressão , Uma Introdução à Econometria ,2ª ed., São Paulo , Hucitec , 1977.

Levine, D.; Berenson, M. L.; Stephan, D. . Estatística: teoria e aplicações usando Microsoft excel em português. Rio de Janeiro: Livros Técnicos e Científicos, 2000.

Meyer, L. Paul ; Probabilidade, Aplicações à Estatística. 2 ed. Rio de Janeiro: LTC-Livros Técnicos e Científicos Editora S.A., 1983.

Metrópolis N. and Ulam , S. , The Monte Carlo Method , J. Am. Stat. Assoc. ; 44 , n 247 , 335- 341 , 1949.

Mood, A. M.; Graybill, F. A. ; Boes, D.C. Introduction to Theory of Statistics. 3. ed. New York: Wiley & Sons , 1974.

Neyman, J. ; Pearson, E. S. On the use and Interpretation of certain test criteria for purposes of statistical inference. Biometrika. London, v. 20 A, p. 175-240 e p.263-294 , 1928.

Rao, C. R. Linear statistical inference and its applications. 2 ed New York: J. Willey, 1988.

Ratkowsky, D. A., Handbook of nonlinear regression models. Marcel Decker Inc, New York, 1990.

Ratkowsky , D. A. Nonlinear Regression Modeling , a unified practical approach. Marcel Decker , New York 1983.

Regazzi, A. J. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear. Rev. Ceres, Viçosa, v.50,n.287,p.9-26,2003.

Regazzi, A. J.; Silva C. H. O. Teste para verificar a igualdade de parâmetros e a identidade de modelos de regressão não-linear.I. Dados no delineamento inteiramente casualizado. Rev. Mat. Estat., Sao Paulo, v.22,n.3,p.33-45,2004.

Ross , Sheldon , A Course in a Simulation , Macmillan Publishing Company , New York ,USA, 1990.

SAS Institute. SAS Language and Procedures: Usage. Version 6, 1 ed. Cary NC. SAS Institute Inc., 1995.373p.

Sobol , Ilya M. ; A Primer for The Monte Carlo Method , CRC Press ,London , UK, 1994.

ANEXOS

ANEXO

ANEXO.....PÁGINA

SIMULAÇÃO	Programa utilizado para gerar os dados e simular a estatística para o teste de identidade de modelos não lineares, por meio de recursos computacionais do sistema estatístico SAS [®] 53
------------------	--

SIMULAÇÃO

Programa utilizado para gerar os dados e simular a estatística para o teste de identidade de modelos não lineares, por meio de recursos computacionais do sistema estatístico SAS®

```
/* ***** Programa 1: Identidade de Modelos ***** */
/* ***** H0:a1=...=a5=a ; b1=...=b5=b ; g1=...=g5=g ***** */

options ps=100 ls=70 nodate nonumber;
data teste;

proc iml;

/* ***** ALTERAR PARAMETROS NA SIMULACAO ***** */
***** */

npares=50; /* ***** Numero de pares ***** */
***** */
nexp=1300; /* ***** Numero de experimento ***** */
***** */

/* ***** Parametros do Modelo Completo ***** */

alfac={1.507,
       1.507,
       1.507};

betac={1.025,
       1.025,
       1.025};

gamac={0.090,
       0.090,
       0.090};
```

```

npc=nrow(alfac)+nrow(betac)+nrow(gamac); /* ***** N. Par. Modelo
Compl. ***** */
neq=nrow(alfac); /* ***** N. de Equações
***** */

*print alfac;
*print betac;
*print gamac;

/* ***** FIM ALTERACAO
***** */
obsc=j(npares,4,0);

/* ***** INICIO DA SIMULACAO
***** */
/* ***** FORMAÇÃO DO CONJUNTO DE DADOS
***** */
do i1=1 to nexp;
do i3=1 to neq;

    alfaaux=alfac[i3,1];
    betaaux=betac[i3,1];
    gamaaux=gamac[i3,1];
    do i2=1 to npares;
        u0=ranuni(0)*alfaaux;
        *print u0;

/* ***** GERANDO A VARIÁVEL
ALEATÓRIA ***** */
/* ***** Modelo de
Langmuir ***** */
*x=exp((log(u0/(alfaaux*betaaux-u0*betaaux)))/(1-gamaaux));
/* ***** Modelo de
Gompertz ***** */
*x=log((log(u0)-alfaaux)/(betaaux))/log(gamaaux);
/* ***** Modelo
Logístico ***** */
x=-((log((alfaaux/u0)-1)-betaaux)/gamaaux);
/* *** Armazena Experimento *** */
obsc[i2,1]=i1;

```

```

/* *** Armazena obsc ***** */
obsc[i2,2]=i3;
/* *** Armazena X ***** */
obsc[i2,3]=x;
/* *** Armazena Y ***** */
/* ***** Modelo Langmuir ***** */
*obsc[i2,4]=(alfac[i3,1]*betac[i3,1]*x**(1-
gamac[i3,1]))/(1+betac[i3,1]*x**(1-gamac[i3,1]));
/* ***** Modelo Gompertz ***** */
*obsc[i2,4]=exp(alfac[i3,1]-(betac[i3,1]*(gamac[i3,1]**x)));
/* *** Modelo Logistico ***** */
obsc[i2,4]=alfac[i3,1]/(1+exp(-betac[i3,1]-gamac[i3,1]*x));

end;

media=sum(obsc[,4:4])/npares;
medy=repeat(media,npares);
s=obscllmedy;
v=v//s;
end;
end;

create dados from v;
append from v;
quit;
run;

proc iml;
nexp=2500; r2=0.90; npares=15; neq=3;

use dados;

do exp=1 to nexp*neq;
read next 15 into ms;
yajus=j(nrow(ms),1,0);

dif=ms[,4] - ms[,5];
sqm=t(dif)*dif;
sqres=((1-r2)*sqm)/r2;
sigma=sqres/(npares-3);

```



```

        do a2=1 to nrow(ms);
        e=rannor(0)*sqrt(sigma);
        *e=ranuni(0)*sqrt(12);
        *e=(ranexp(0)/0.5)*sqrt(sigma);
        yajus[a2,1]=ms[a2,4] + e;
        end;

        m=mslyajus;
        mvalajus=mvalajus//m;

end;

        create simula from mvalajus;
        append from mvalajus;

run; quit;

data novo; set simula;

if col2=1 then pop1=1 ; else pop1=0;
if col2=2 then pop2=1 ; else pop2=0;
if col2=3 then pop3=1 ; else pop3=0;
*if col2=4 then pop4=1 ; *else pop4=0;
*if col2=5 then pop5=1 ; *else pop5=0;
run;quit;

/* ***** Estimativas - Modelo Completo ***** */

proc nlin data=novo maxiter=1000 noprint outest=recompc;

parms alfa1=1.507 beta1=1.025 gama1=0.090
      alfa2=1.507 beta2=1.025 gama2=0.090
      alfa3=1.507 beta3=1.025 gama3=0.090;
* alfa4=1,507 beta4=1,025 gama4=0,090;
  *alfa5=0.9323 beta5=0.0505 gama5=2.720;

/* ***** Modelo Gompertz ***** */
/*

```

```

model col6=pop1*exp(alfa1+(beta1*gama1**col3))+
pop2*exp(alfa2+(beta2*gama2**col3))+
pop3*exp(alfa3+(beta3*gama3**col3));
      *pop4*exp(alfa4+(beta4*gama4**col3));
      *pop5*exp(alfa5+(beta5*gama5**col3));

      der.alfa1=pop1*exp(alfa1+(beta1*gama1**col3));
      der.alfa2=pop2*exp(alfa2+(beta2*gama2**col3));
      der.alfa3=pop3*exp(alfa3+(beta3*gama3**col3));
      *der.alfa4=pop4*exp(alfa4+(beta4*gama4**col3));
      *der.alfa5=pop5*exp(alfa5+(beta5*gama5**col3));

der.beta1=pop1*(gama1**col3*exp(alfa1+beta1*gama1**col3));
der.beta2=pop2*(gama2**col3*exp(alfa2+beta2*gama2**col3));
der.beta3=pop3*(gama3**col3*exp(alfa3+beta3*gama3**col3));
*der.beta4=pop4*(gama4**col3*exp(alfa4+beta4*gama4**col3));
*der.beta5=pop5*(gama5**col3*exp(alfa5+beta5*gama5**col3));

der.gama1=(pop1*beta1*gama1**col3*col3*exp(alfa1+beta1*gama1**col3))/g
ama1;

der.gama2=(pop2*beta2*gama2**col3*col3*exp(alfa2+beta2*gama2**col3))/g
ama2;

der.gama3=(pop3*beta3*gama3**col3*col3*exp(alfa3+beta3*gama3**col3))/g
ama3;;

*der.gama4=(pop4*beta4*gama4**col3*col3*exp(alfa4+beta4*gama4**col3))/
gama4;

*der.gama5=(pop5*beta5*gama5**col3*col3*exp(alfa5+beta5*gama5**col3))/
gama5;

*/
/* ***** Modelo Langmuir ***** */
/*
  model col6=pop1*((alfa1*beta1*col3**(1-gama1))/(1+beta1*col3**(1-
gama1)))+
      pop2*((alfa2*beta2*col3**(1-gama2))/(1+beta2*col3**(1-
gama2)))+

```

```

pop3*((alfa3*beta3*col3**(1-
gama3))/(1+beta3*col3**(1-gama3)))+
pop4*((alfa4*beta4*col3**(1-
gama4))/(1+beta4*col3**(1-gama4)));
*pop5*((alfa5*beta5*col3**(1-
gama5))/(1+beta5*col3**(1-gama5)));

*der.alfa1=(pop1*(beta1*col3**(1-
gama1)))/(1+beta1*col3**(1-gama1))+
*der.alfa2=(pop2*(beta2*col3**(1-
gama2)))/(1+beta2*col3**(1-gama2))+
*der.alfa3=(pop3*(beta3*col3**(1-
gama3)))/(1+beta3*col3**(1-gama3));
*der.alfa4=(pop4*(beta4*col3**(1-
gama4)))/(1+beta4*col3**(1-gama4));
*der.alfa5=(pop5*(beta5*col3**(1-
gama5)))/(1+beta5*col3**(1-gama5));

*der.beta1=pop1*((alfa1*col3**(1-
gama1))/(1+beta1*col3)-(alfa1*beta1*(col3**(1-
gama1)**2))/((1+beta1*col3**(1-gama1))**2))+
*der.beta2=pop2*((alfa2*col3**(1-
gama2))/(1+beta2*col3)-(alfa2*beta2*(col3**(1-
gama2)**2))/((1+beta2*col3**(1-gama2))**2))+
*der.beta3=pop3*((alfa3*col3**(1-
gama3))/(1+beta3*col3)-(alfa3*beta3*(col3**(1-
gama3)**2))/((1+beta3*col3**(1-gama3))**2));
*der.beta4=pop4*((alfa4*col3**(1-
gama4))/(1+beta4*col3)-(alfa4*beta4*(col3**(1-
gama4)**2))/((1+beta4*col3**(1-gama4))**2));
*der.beta5=pop5*((alfa5*col3**(1-
gama5))/(1+beta5*col3)-(alfa5*beta5*(col3**(1-
gama5)**2))/((1+beta5*col3**(1-gama5))**2));

*der.gama1=pop1*(-(alfa1*beta1*col3**(1-
gama1)*log(col3))/(1+beta1*col3**(1-gama1))+(alfa1*beta1**2*(col3**(1-
gama1)**2))/((1+beta1*col3**(1-gama1))**2))+
*der.gama2=pop2*(-(alfa2*beta2*col3**(1-
gama2)*log(col3))/(1+beta2*col3**(1-gama2))+(alfa2*beta2**2*(col3**(1-
gama2)**2))/((1+beta2*col3**(1-gama2))**2))+

```

```

*der.gama3=pop3*(-(alfa3*beta3*col3**(1-
gama3)*log(col3))/(1+beta3*col3**(1-gama3))+(alfa3*beta3**2*(col3**(1-
gama3)**2)/((1+beta3*col3**(1-gama3)**2)));
*der.gama4=pop4*(-(alfa4*beta4*col3**(1-
gama4)*log(col3))/(1+beta4*col3**(1-gama4))+(alfa4*beta4**2*(col3**(1-
gama4)**2)/((1+beta4*col3**(1-gama4)**2)));
*der.gama5=pop5*(-(alfa5*beta5*col3**(1-
gama5)*log(col3))/(1+beta5*col3**(1-gama5))+(alfa5*beta5**2*(col3**(1-
gama5)**2)/((1+beta5*col3**(1-gama5)**2)));
*/
/* ***** Modelo Logistico ***** */

model col6=pop1*(alfa1/(1+exp(-beta1-gama1*col3)))+
pop2*(alfa2/(1+exp(-beta2-gama2*col3)))+
pop3*(alfa3/(1+exp(-beta3-gama3*col3)));
*pop4*(alfa4/(1+exp(beta4-gama4*col3)));
*pop5*(alfa5/(1+exp(beta5-gama5*col3)));
/*

der.alfa1=pop1/(1+exp(-beta1-gama1*col3));
der.alfa2=pop2/(1+exp(-beta2-gama2*col3));
*der.alfa3=pop3/(1+exp(-beta3-gama3*col3));
*der.alfa4=pop4/(1+exp(-beta4-gama4*col3));
*der.alfa5=pop5/(1+exp(-beta5-gama5*col3));

der.beta1=pop1*(alfa1*exp(-beta1-gama1*col3))/(1+exp(-beta1-
gama1*col3)**2);
der.beta2=pop2*(alfa2*exp(-beta2-gama2*col3))/(1+exp(-beta2-
gama2*col3)**2);
*der.beta3=pop3*(alfa3*exp(-beta3-gama3*col3))/(1+exp(-beta3-
gama3*col3)**2);
*der.beta4=pop4*(alfa4*exp(-beta4-gama4*col3))/(1+exp(-beta4-
gama4*col3)**2);
*der.beta5=pop5*(alfa5*exp(-beta5-gama5*col3))/(1+exp(-beta5-
gama5*col3)**2);

der.gama1=pop1*(alfa1*col3*exp(-beta1-gama1*col3))/(1+
exp(-beta1-gama1*col3)**2);
der.gama2=pop2*(alfa2*col3*exp(-beta2-gama2*col3))/(1+
exp(-beta2-gama2*col3)**2);
*der.gama3=pop3*(alfa3*col3*exp(-beta3-gama3*col3))/(1+
exp(-beta3-gama3*col3)**2);

```

```

                *der.gama4=pop4*(alfa4*col3*exp(-beta4-gama4*col3))/(1+
exp(-beta4-gama4*col3)**2);
                *der.gama5=pop5*(alfa5*col3*exp(-beta5-gama5*col3))/(1+
exp(-beta5-gama5*col3)**2);

*/
    by col1;
run;
quit;

/* ***** Fim do Modelo Completo ***** */
/* ***** limpa arquivo ***** */

data compc ; set recompc;
    if _type_='ITER' then delete;
    if _type_='COVB' then delete;
run; quit;

/* ***** Estimativas - Modelo Reduzido ***** */
*proc print data=novo;
*run;*quit;

proc nlin data=novo maxiter=5000 noprint outest=recomw1;

    parms alfa=1.507 beta=1.025 gama=0.090;

/* ***** Modelo Gompertz ***** */
/*
model col6=exp(alfa+(beta*gama**col3));
                der.alfa=exp(alfa+(beta*gama**col3));
                der.beta=gama**col3*exp(alfa+beta*gama**col3);

der.gama=beta*gama**col3*col3*exp(alfa+beta*gama**col3)/gama;
*/
/* ***** Modelo Langmuir ***** */
/*
    model col6=((alfa*beta*col3**(1-gama))/(1+beta*col3**(1-gama)));

    *der.alfa=(beta*col3**(1-gama))/(1+beta*col3**(1-gama));
    *der.beta=(alfa*col3**(1-gama))/(1+beta*col3)-(alfa*beta*(col3**(1-
gama)**2))/((1+beta*col3**(1-gama))**2);

```

```

*der.gama=-(alfa*beta*col3**(1-gama)*log(col3))/(1+beta*col3**(1-
gama))+((alfa*beta**2*(col3**(1-gama))**2)/((1+beta*col3**(1-gama))**2));
*/
/* ***** Modelo Logistico ***** */

model col6=alfa/(1+exp(-beta-gama*col3));
/*
der.alfa=1/(1+exp(-beta-gama*col3));
der.beta=alfa*exp(-beta-gama*col3)/(1+exp(-beta-gama*col3))**2;
der.gama=alfa*col3*exp(-beta-gama*col3)/(1+exp(-beta-gama*col3))**2;
*/

by col1;

run; quit;

/* ***** Fim Estimativas - Modelo Reduzido ***** */
/* ***** limpa arquivo ***** */

data compw1 ; set recomw1;
if _type_='ITER' then delete;
if _type_='COVB' then delete;
run; quit;

/* ***** Calculo da estatistica qui-quadrado ***** */

proc iml;

/* ***** ALTERAR PARAMETROS NA SIMULAÇÃO ***** */
***** */

neq=3; /* ***** N de equacoes ***** */
npares=15; /* ***** N de pares ***** */
glc=6; /* ***** Gl (glmc-glmr) muda em função de neq ***** */
nPMC=9; /* ***** Numero de parametros do modelo completo ***** */
nPMR=3; /* ***** Numero de parametros do modelo reduzido ***** */
nexp=2500; /* ***** Numero de Simulacoes ***** */

contasim=0; prop5=0;

use compc; read all into matrizc; *print matrizc; close compc;
use compw1; read all into matrizw1; close compw1; *print matrizw1;

```

```

mfiltra=j(nexp,5,0);

do i=1 to nrow(matrizc);
  a=matrizc[i,1];
do ii=1 to nrow(matrizw1);
  if (a=matrizw1[ii,1]) then do;
    mfiltra[ii,1]=matrizc[i,1];
    mfiltra[ii,2]=matrizc[i,3]; /* SQR - Matriz completa */
    mfiltra[ii,3]=matrizw1[ii,3]; /* SQR - Matriz Reduzida */
    difm=mfiltra[ii,3]-mfiltra[ii,2];

    if (difm>1e-8) then do;
      qrmc=mfiltra[ii,2]/(neq*npares-npmc);
      qrmr=mfiltra[ii,3]/(neq*npares-npmr);
      est=(neq*npares)*(log(qrmr/qrmc));
      mfiltra[ii,4]=est;
    end;
  end;
end;
end;

do iii=1 to nrow(mfiltra);
  estaux=mfiltra[iii,4];
  if (mfiltra[iii,4]>1e-6) & (mfiltra[iii,1]>1e-6) then do;
    pr5=1-probchi(estaux,glc);
    mfiltra[iii,5]=pr5;
    contasim=contasim+1;
    if pr5<0.05 then prop5=prop5+1;
    mfiltra[iii,5]=pr5;
    *contasim=contasim+1;
  end;
end;

res5=prop5/contasim;

/* ***** Resultados ***** */

print 'numero de simulações validas',contasim;

```

```
    print 'probabilidade resultante .....5%:', res5;  
run;  
quit;
```