

Carlos Sérgio Araújo dos Santos

**MODELOS SIMÉTRICOS TRANSFORMADOS NÃO
LINEARES COM APLICAÇÃO NA ESTIMATIVA
VOLUMÉTRICA EM HÍBRIDO DE *Eucalyptus tereticornis* NO
PÓLO GESSEIRO DO ARARIPE-PE**

RECIFE-PE - JAN/2010



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**MODELOS SIMÉTRICOS TRANSFORMADOS NÃO
LINEARES COM APLICAÇÃO NA ESTIMATIVA
VOLUMÉTRICA EM HÍBRIDO DE *Eucalyptus tereticornis* NO
PÓLO GESSEIRO DO ARARIPE-PE**

Dissertação apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Mestre.

Área de Concentração: Modelagem Estatística e Computacional

Orientador: Prof. PhD José Antonio Aleixo da Silva

Co-orientador: Prof. PhD Gauss Moutinho Cordeiro

Co-orientador: Prof. Dr. Rinaldo Luiz Caraciolo Ferreira

RECIFE-PE - JAN/2010

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

**MODELOS SIMÉTRICOS TRANSFORMADOS NÃO-LINEARES COM APLICAÇÃO NA
ESTIMATIVA VOLUMÉTRICA EM HÍBRIDO DE *Eucalyptus tereticornis* NO PÓLO
GESSEIRO DO ARARIPE-PE**

CARLOS SÉRGIO ARAÚJO DOS SANTOS

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 15/01/2010 pela Comissão Examinadora.

Orientador:

Prof. PhD. José Antonio Aleixo da Silva
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Prof. Dra. Tatijana Stosic
Universidade Federal Rural de Pernambuco

Prof. Dr. Cláudio Tadeu Cristino
Universidade Federal Rural de Pernambuco

Prof. Dr. Paulo de Paula Mendes
Universidade Federal Rural de Pernambuco

Dedico este trabalho, com carinho e atenção,
a meus pais, Rita e Damião.

Agradecimentos

Em primeiro lugar agradeço ao Deus Pai todo poderoso, por me fortalecer e me abençoar. Sem ele não sou e nem serei ninguém.

Ao meu orientador Professor José Antônio Aleixo da Silva, pelos ensinamentos, paciência e pela amizade.

Ao grande Professor Gauss Moutinho Cordeiro, pelas preciosas dicas e sugestões, pela paciência durante todo esse tempo de trabalho, pelos ensinamentos e por ser uma figura de grande importância na minha vida acadêmica.

Ao Professor Rinaldo Luiz Caraciolo Ferreira, pela co-orientação e amizade.

A todos os professores do Programa de Pós-graduação em Biometria e Estatística Aplicada/UFRPE pelos ensinamentos e pelos momentos de aprendizagem e conhecimentos que proporcionaram durante todo mestrado.

Aos meus pais Rita e Damião, aos meus irmãos Solange e Saulo e minha sobrinha Luíza, por me apoiarem e me alicerçarem durante toda minha trajetória acadêmica.

Aos meus familiares por estarem sempre na torcida pela minha vitória.

Ao amigo Joseilme Fernandes Gouveia, por ter sido um grande companheiro durante todo Mestrado e pelas parcerias em diversos trabalhos realizados.

Ao colega Luiz Medeiros de Araújo Lima Filho, pela contribuição dada neste trabalho

Aos amigos Macio Albuquerque, Ulisses, Kleber, Patrícia, Andréa, Cícero Carlos, Rosilda, Jáder, Amanda, David e Rodrigo pela colaboração dada, pelo incentivo, força e amizade

Ao colega Fernando Gadelha do Programa de pós-graduação em Ciências Florestais pelas dicas e ensinamentos.

À funcionária D. Zuleide França por ter sido uma grande amiga durante todo o mestrado.

Ao secretário Marco Antônio dos Santos pelo seu profissionalismo e competência.

À agência de fomento para a formação de acadêmicos do país, o CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo apoio financeiro.

Aos demais colegas de curso, que se tornaram grandes amigos e futuros companheiros de trabalho.

Para não correr o risco da injustiça, agradeço de antemão a todos que de alguma forma passaram pela minha vida e contribuíram para a construção de quem sou hoje.

"O sucesso está onde você enxerga... não pode ser objetivo e sim consequência, faça algo por amor e o sucesso virá. Cuide de seus olhos para enxergar além da imagem"

Victor Chaves

Resumo

Box e Cox (1964) desenvolveram um procedimento numérico para escolher uma transformação da resposta tal que a distribuição da variável transformada esteja o mais próximo possível da distribuição normal. A introdução de uma nova classe de modelos simétricos transformados não lineares visa estender os modelos de Box e Cox para uma classe geral dos modelos simétricos. Esta nova classe de modelos inclui todas as distribuições contínuas simétricas com uma possível estrutura não linear para a média e capacitando o ajustamento de uma larga extensão de modelos para vários tipos de dados. Para ilustrar a utilidade dessa nova classe de modelos de regressão foi realizada uma aplicação na estimativa dos volumes de clones de *Eucalyptus tereticornis* com 7,5 anos oriundos de um experimento que está sendo realizado no Campo Experimental do Araripe do Instituto Agrônomo de Pernambuco (IPA), localizado no Município de Araripina, no semi-árido Pernambucano. O modelo não-linear utilizado para explicar os dados foi o modelo Schumacher-Hall. Diante dos resultados obtidos se concluí que o modelo transformado com erros *t*-Student com dois graus de liberdade foi o que melhor se ajustou os dados.

Palavras-chave: Clones de Eucaliptos, Distribuição Simétrica, Modelo de Box-Cox, Modelo Schumacher-Hall.

Abstract

Box and Cox (1964) developed a numerical procedure to transform the response variable such that the transformed variable should be as closed as possible to the normal distribution. The introduction of a new class of non linear symmetric transformed models aims to extended the Box and Cox models to a general class of symmetric models. The new class of models included all the continuous symmetric distributions with a possible non linear structure to the mean, making possible the use of the new class of regression models. It was applied in the estimate of volumes of the *Eucalyptus tereticornis* clones, with 7,5 years, planted in the Experimental Station of Ararape of the Agronomic Institute of Pernambuco (IPA), in the municipality of Araripina, in the semiarid of Pernambuco. The non linear model used as pattern was the Schumacher and Hall model. The results indicates that the transformed model with t -Student errors with two degrees of freedom adjusted better to the data set.

Keywords: *Eucalyptus* clones, Symmetric distribution, Box-Cox Model, Schumacher-Hall Model.

Lista de Figuras

1	Densidade da distribuição normal segundo diferentes parametrizações. . .	p. 34
2	Densidades da distribuição Cauchy segundo diferentes parâmetros e distribuição normal padrão.	p. 36
3	Densidades da distribuição de Laplace padrão e distribuição normal padrão.	p. 40
4	Densidades da distribuição de t segundo diferentes graus de liberdade e distribuição normal padrão.	p. 43
5	Densidades da distribuição t de Student generalizada com diferentes valores dos parâmetros s e r e distribuição normal padrão.	p. 45
6	Densidades da distribuição de logística I padrão e distribuição normal padrão.	p. 47
7	Densidades da distribuição de logística II padrão e distribuição normal padrão.	p. 49
8	Densidades da distribuição Kotz com diferentes valores dos parâmetros N e r e distribuição normal padrão.	p. 51
9	Densidades da distribuição exponencial potência segundo diferentes valores do parâmetro l e distribuição normal padrão.	p. 54
10	Planta baixa da parcela e área útil.	p. 63
11	Área experimental, de híbridos de <i>Eucalyptus spp.</i> , localizada na Estação Experimental do Araripe (IPA), em Araripina - PE.	p. 65
12	Gráfico de dispersão dos valores observados versus valores ajustados modelo t_2 transformado.	p. 71
13	Gráfico de dispersão dos Resíduos de Pearson versus valores ajustados modelo t_2 transformado.	p. 72
14	Gráfico normal de probabilidades com limite de confiança (envelope de simulação) modelo t_2 transformado.	p. 73

- 15 Erros percentuais para o modelo t_2 transformado. p.74
- 16 Erros percentuais para o modelo t_2 transformado na escala original. . . . p.74

Lista de Tabelas

1	Expressões para $D^{(\lambda)}$, a e b para algumas distribuições simétricas	p. 59
2	Dados sobre Volume, DAP e Altura total Ht do clone Híbrido de <i>Eucalyptus tereticornis</i> com 7,5 anos oriundos do Pólo Gesseiro do Araripe-PE	p. 67
3	Estimativas dos parâmetros e (erros padrões) dos modelos simétricos não transformados.	p. 68
4	Estimativas dos parâmetros e (erros padrões) dos modelos simétricos transformados.	p. 69
5	Estatísticas para seleção dos modelos.	p. 70
6	Comparação dos erros para os modelos ajustados.	p. 70

Sumário

1	Introdução	p. 14
2	Revisão de Literatura	p. 17
2.1	Pólo Gesseiro do Araripe	p. 17
2.2	Os Eucaliptos	p. 18
2.3	Planejamento Florestal	p. 19
2.4	Volume da árvore	p. 20
2.5	Modelos volumétricos	p. 21
2.6	Modelos não lineares	p. 24
2.6.1	Estimação dos Parâmetros	p. 25
2.6.2	Resíduos	p. 26
2.7	Distribuições Simétricas	p. 26
2.7.1	Definições	p. 27
2.8	Distribuição Normal	p. 29
2.8.1	Caracterização	p. 30
2.8.2	Estimação dos parâmetros	p. 32
2.9	Distribuição de Cauchy	p. 34
2.10	Distribuição de Laplace ou Exponencial Dupla	p. 37
2.10.1	Caracterização	p. 37
2.10.2	Distribuições relacionadas	p. 38
2.11	Distribuição t de Student	p. 39
2.11.1	t de Student	p. 41

2.11.2 <i>t</i> de Student Generalizada	p. 43
2.12 Distribuição Logística	p. 46
2.12.1 Logística I	p. 46
2.12.2 Logística II	p. 48
2.13 Distribuição de Kotz	p. 50
2.14 Distribuição Exponencial Potência	p. 51
3 Modelos Simétricos Transformados não lineares	p. 55
3.1 Análise gráfica dos resíduos	p. 62
4 Material e Métodos	p. 63
4.1 Cubagem das árvores amostras	p. 64
4.2 Estimação do Volume dos eucaliptos	p. 64
5 Resultados e Discussão	p. 68
6 Conclusões	p. 77
REFERÊNCIAS BIBLIOGRÁFICAS	p. 78
Apêndice A - Algoritmos e programas	p. 89
Programa do Modelo de Schumacher-Hall para os dados de volume	p. 89

1 Introdução

A vegetação natural da Caatinga vem sofrendo grande pressão antrópica, sendo explorada de forma desordenada, caracterizando-se em grande parte pela ausência de critérios técnicos no corte da vegetação e conseqüentemente, pondo em risco de extinção de várias espécies vegetais e animais (RIBEIRO *et al.*, 2001).

No Pólo Gesseiro do Araripe este fato se deve, principalmente, a crescente demanda por biomassa renovável para energia, aumentando gradativamente a sua degradação, tanto a nível domiciliar, com a utilização da lenha e o carvão vegetal para cocção de alimentos, como à nível industrial e comercial, com a utilização da madeira na calcinação da gipsita para produção do gesso. Portanto, medidas mitigadoras devem ser tomadas para evitar esse cenário.

O aumento da eficiência energética da lenha contribui de forma significativa na diminuição da demanda desse recurso exigido para calcinação da gipsita e, conseqüentemente, aumenta a produtividade dos fornos.

As calcinadoras de gipsita são as principais usuárias de energéticos florestais da região do Araripe, consumindo 56% da produção, seguida da siderurgia, com 33% e do consumo doméstico, com 6%. Os demais ramos juntos representam 5% do consumo total (FUPEF, 2007).

Uma alternativa econômica e ambientalmente viável é o manejo sustentado de povoamentos florestais nativos ou o reflorestamento com florestas de rápido crescimento, com destaque para os *Eucalyptus* por sua versatilidade. Considerado como gênero de rápido crescimento, o *Eucalyptus* pode contribuir no sentido de reduzir a devastação das florestas nativas remanescentes (BAENA, 2005).

Baseado neste contexto, realizar uma modelagem do volume de eucaliptos é essencial para predizer a viabilidade da implantação destas árvores para fins energéticos na região do Pólo Gesseiro do Araripe.

A pesquisa "Módulo de Experimentação Florestal para o Pólo Gesseiro do Araripe", instalada na Estação Experimental do Instituto Agrônomo de Pernambuco (IPA) na Chapada do Araripe-PE, foi implantada em 2002, com 15 clones do gênero *Eucalyptus*. O objetivo dessa pesquisa é atender a carência de informações, servindo de base para o aprofundamento na seleção de clones de *Eucalyptus* na região, por meio de estudos de modelagem de crescimento, mortalidade, sobrevivência e volumetria.

O uso de equações de volume em inventário florestal é uma operação rotineira para cálculo de volume de madeira em pé e estimativa da altura das árvores através da relação DAP (diâmetro à altura do peito) e altura (H). Conhecendo-se a altura total e o DAP de cada árvore, pode-se estimar o volume de cada uma, usando equações de volume, que devem ser restritas para as condições em que se obtiveram dados para suas construções. Qualquer extrapolação além desses limites pode ser perigosa sob o ponto de vista de previsão dos resultados finais (COUTO e BASTOS, 1987). Entre os mais variados modelos, destacam-se os de Spurr (1952), Schumacher-Hall em 1933, Compreensiva (MEYER, 1953), Naslund e Silva (SILVA *et al.*, 1993).

Baseados nas distribuições simétricas Cordeiro e Andrade (2009) introduziram uma nova classe de Modelos Simétricos Transformados (MST) com distribuição simétrica para a variável resposta e uma possível função de ligação não linear para a resposta média. Essa classe de modelos se estende os modelos clássicos de Box e Cox (1964) abrangendo com várias outras distribuições contínuas simétricas com caldas menores ou mais longas que normais.

Essa nova classe de modelos inclui distribuições tais como normal, t-Student, exponencial potência, logísticas I e II e os modelos normais contaminados, e permite ajustar uma ampla variedade de modelos para vários tipos de dados.

O uso de transformações em análise de regressão é muito comum e pode ser útil quando o modelo original não satisfaz as suposições usuais de normalidade. A família de transformações proposta por Box-Cox representa um ferramental útil para estatísticas aplicadas tratando dos aspectos de homogeneidade de discrepância, aditividade e normalidade.

A proposta principal deste estudo foi deduzir um algoritmo iterativo para ajustar os modelos simétricos transformados não lineares pelo método de máxima verossimilhança e discutir a inferência dos parâmetros nesses modelos. A inferência foi deduzida de forma análoga àquela proposta para os modelos de Box-Cox, tentando generalizar resultados desses modelos. A utilidade dessa nova classe de modelos foi testada na estimativa

volumétrica em Híbrido de *Eucalyptus tereticornis* (cruzamento natural), em ponto de corte para produção de lenha.

Considerando a grande utilidade no uso de modelos não-lineares na análise do volume e produção de florestas e a real necessidade de considerar para os erros, distribuições com caudas mais leves e mais pesadas do que a normal, os modelos simétricos transformados tendem a dar uma grande contribuição para trabalhos futuros no manejo sustentado de povoamentos florestais, para promover o melhor uso do recurso florestal resultando na redução de custos no inventário florestal.

Os modelos simétricos transformados não lineares foram estudados do ponto de vista inferencial como no contexto dos modelos lineares generalizados transformados definidos, recentemente, por Cordeiro e Andrade (2007a,b) e de forma mais direta estendendo a inferência tradicional dos modelos de Box e Cox (1964), principalmente, tentando generalizar resultados desses últimos modelos.

Como produto adicional desta pesquisa será desenvolvido um procedimento iterativo para a estimação paramétrica através do uso de algoritmos de máxima verossimilhança, além de alguns resultados inferenciais. Será também desenvolvido um procedimento para quantificação do volume de árvores em plantios clonais de eucalipto oriundos do Módulo de Experimentação Florestal para o Pólo Gesseiro do Araripe.

2 Revisão de Literatura

2.1 Pólo Gesseiro do Araripe

O Pólo Gesseiro do Araripe em Pernambuco engloba os municípios de Araripina, Ipubi, Trindade, Bodocó e Ouricuri, e um total de outros 10 municípios adjacentes dentro do Estado de Pernambuco (Cedro, Dormentes, Exu, Granito, Moreilândia, Parnamirim, Santa Cruz, Santa Filomena, Serrita e Terra Nova). O Pólo Gesseiro do Araripe está localizado na região do semiárido pernambucano, com área de 11.792 km². A Região é caracterizada pelo predomínio de condições ecológicas desfavoráveis, temperaturas elevadas, escassez de chuvas e mal distribuídas, rios temporários e vegetação xerófitas. A principal atividade econômica é a indústria do gesso, além das culturas de subsistência e a pecuária extensiva. A produção agrícola diversificada é impulsionada pelas temperaturas amenas e elevados níveis pluviométricos (ALBUQUERQUE, 2002).

O grande destaque do Pólo Gesseiro do Araripe, é a sua produção de gipsita de onde se produz 95% de todo gesso consumido no Brasil (ALBUQUERQUE, 2002).

O Brasil tem destaque privilegiado em reserva de gipsita que corresponde a cerca de 1.668.570.905 t. A partir de dados do Anuário Mineral Brasileiro 2000, dos Relatórios Anuais de Lavra 2001 e de Relatórios dos Trabalhos de Pesquisa aprovados no ano 2000, do total dessa reserva, 96% localizam-se nos Estados da Bahia (43%), do Pará (35%) e de Pernambuco (18%), vindo em seguida os Estados do Maranhão (2,7%), de Tocantins (0,6%), do Ceará (0,3%), do Piauí (0,2%) e do Amazonas (0,1%) (BRASIL, 2001).

O setor industrial do Pólo gessereiro do Araripe é formado por um total de 26 mineradoras, 72 calcinadoras e 234 fábricas de pré-moldados, responsável pelo maior consumo de biomassa vegetal, cerca de 65% do total da vegetação da região. Com a formação do consórcio formado, a perspectiva é manter os empregos já existentes, mas com maior valor agregado (GEOPHOTO, 2005).

Segundo Albuquerque (2002), o consumo de gesso per capita no Brasil, de 7 Kg/ano,

é relativamente muito inferior ao de países como Chile (41), os Estados Unidos (86) e a Europa (60). Considerando-se que o incentivo ao consumo por meio da diversificação dos produtos ofertados pelo Pólo Gesso do Araripe, mantida a tecnologia de produção atual, no que se refere ao uso de combustíveis, deverá ocorrer, significativamente, a redução da cobertura florestal do Sertão do Araripe e proximidades. Pode-se vislumbrar quão grave é uma política de expansão da capacidade produtiva sem atentar para a atual tecnologia de produção e seus impactos sobre o meio ambiente e sobre a saúde humana.

Um dos aspectos positivos da produção de gesso é a manutenção de um baixo nível de desemprego, pois em Araripina esta cifra foi de 2,6%, enquanto na região metropolitana do Recife atingiu 9,2% e nas seis regiões pesquisadas pelo IBGE, ficou em 7,7%, em julho de 1999 (ALBUQUERQUE, 2002). Atualmente, o Pólo Gesso se apresenta como um conjunto de empresas de micro, pequeno e médio porte que oferecem cerca de 13.000 empregos diretos e aproximadamente 66.000 indiretos (SINDUSGESSO, 2009).

A produção nacional de gesso, atualmente, provém dos Estados de Pernambuco (1.393.113 t - 95% da produção nacional), Ceará (65.303 t - 2%), Maranhão (42.162 t - 2%), Tocantins (11.674 t) e Amazonas (2.33 t - 1%) (SINDUSGESSO, 2009).

2.2 Os Eucaliptos

As sementes de *Eucalyptus* começaram a ser difundida no mundo no início do século XIX na América do Sul. No Brasil, as primeiras mudas chegaram em 1824. Historicamente, a madeira do *Eucalyptus* é utilizada para os mais variados fins, tais como: lenha, carvão vegetal, celulose, dormentes ferroviários, postes para eletrificação, casca para curtir couro, óleos essenciais e construção civil (ANDRADE, 1961; LIMA, 1993). Além dos dormentes usados nas estradas de ferro, o *Eucalyptus* também era usado como combustível (lenha) para movimentar trens (ANDRADE, 1961). Hoje, o Brasil se configura como o maior produtor de *Eucalyptus* do mundo, com uma área plantada que corresponde a 21% das plantações mundiais (FAO, 2009).

Considerado pai da eucaliptocultura no Brasil, Navarro de Andrade desenvolveu experiências de 1904 a 1909, no Horto Florestal de Jundiáí, comparando várias espécies nativas com o eucalipto, tendo esse se destacado das demais espécies quanto à produtividade (MORA e GARCIA, 2000).

Na atualidade a produtividade média dos plantios de eucaliptos é em torno de 30 a 40 m^3 /ha/ano segundo (LIMA, 1993; SBS, 2006) devido a grandes avanços científicos e

tecnológicos nessa cultura.

As plantações têm sido formadas nas mais variadas condições ambientais, desde regiões tropicais equatoriais até condições subtropicais e semi-áridas, em condições climáticas que variam de mediterrânea a temperada, e em elevações que variam desde o nível do mar até mais ou menos 4000 metros de altitude nas regiões dos Andes (LIMA, 1993).

A principal razão da grande aceitação dessa essência florestal é a elevada taxa de crescimento, a facilidade de reprodução, a rusticidade, o altíssimo nível de melhoramento genético em produtividade e qualidade da madeira e as técnicas aprimoradas de formação de mudas, plantio, manejo e exploração. Devido a esses grandes avanços científicos e tecnológicos nessa cultura, a silvicultura brasileira se destaca como uma das mais avançadas do mundo (BAENA, 2005).

A necessidade de dispor de uma ferramenta simplificada que permita a obtenção de valores precisos da existência do volume de madeira de uma espécie é vital para a planificação da gestão das massas florestais e para sua ordenação sustentável. Em áreas de reflorestamentos industriais qualquer sistema de manejo florestal requerirá da estimação precisa dos estoques de madeira correntes e futuros, seja para produção de celulose, carvão ou lenha (PARENT, 2003).

No Pólo Gesseiro do Araripe não existiam florestas plantadas de *Eucalyptus*. Em março de 2002, foi implantado na Estação Experimental do Instituto Agrônomo de Pernambuco (IPA) na Chapada do Araripe - PE, o Módulo de Experimentação Florestal que utilizou 15 clones do gênero *Eucalyptus* em uma área de 2,352 ha, entre híbridos e espécies do gênero *Eucalyptus*. Visando atender a carência de informações, este experimento servirá de base para o aprofundamento na seleção de clones de *Eucalyptus* na região, por meio de estudos de modelagem de crescimento, mortalidade, sobrevivência, volumetria, calorimetria, entre outros.

2.3 Planejamento Florestal

O reflorestamento em pequenas, médias e grandes propriedades rurais é de interesse público, por ser uma fonte de renda e contribuir para evitar o êxodo rural, o desemprego e, simultaneamente, possibilitar inúmeros e imprescindíveis benefícios ambientais (GALVÃO, 2000). A predição da produção de madeira do reflorestamento é fundamental para o planejamento das atividades florestais, constituindo-se em um dos elementos principais do manejo florestal (DAVIS e JOHNSON, 1987).

A previsão do crescimento de uma floresta, através de curvas de projeção do desenvolvimento em diâmetro, altura e volume, é a condição básica em qualquer processo de planejamento racional, uma vez que essas curvas são importantes para o planejamento da produção, através da determinação da idade ótima e econômica de corte; bem como para o planejamento e a pesquisa silvicultural, que permitem agrupar espécies e procedências, e para analisar eficientemente os experimentos florestais, entre outros (BERGER *et al.*, 2000).

O cuidado na obtenção das informações, no caso do volume, deve ser tomado, porque os erros podem ocasionar sérios problemas, uma vez que as estimativas são, geralmente, utilizadas quando se trata de negociações de serviços, ou venda de madeira em pé, com base no inventário florestal (TONINI, 2004).

2.4 Volume da árvore

Segundo Pellico Netto (2004), a volumetria baseada na metodologia matemática trata, fundamentalmente, de uma avaliação de subconjuntos (volume comercial e total) para se obter o volume de madeira em uma unidade de área, geralmente, um hectare. Primeiramente, consideram-se as árvores de uma espécie, depois se pode formar um subconjunto com espécies que apresentam similaridades de forma e tamanho, e assim por diante.

A forma de cada árvore, geralmente, é semelhante a uma forma geométrica que na literatura florestal recebe o nome de protótipo dendrométrico. Na literatura, os mais comuns são: cilindro, neilóide, parabolóide e cone (GOMES, 1957; SILVA e PAULA NETO, 1979).

Segundo Clutter *et al.* (1983), o volume do tronco (V) é, considerado uma função do diâmetro a 1,30 m do nível do terreno (DAP), da altura (H) e a expressão de forma (F), estimada pela razão entre o volume do tronco e o volume de um cilindro, que tem diâmetro igual ao DAP e altura igual à altura total da árvore, sendo a relação envolvida representada por: $V = f(DAP, H, F)$.

A atividade de cubagem rigorosa do volume é feita em separado das medições das unidades amostrais, o que gera aumento no custo do inventário. Esse é o método direto mais utilizado na rotina de inventários florestais e consiste na medição sucessiva de diâmetros ao longo do tronco, dividindo-o em seções que variam de comprimento de acordo com o medidor (LEITE e ANDRADE, 2002).

Existem várias fórmulas para o cálculo do volume da árvore, mas no entanto se desta-

cou entre elas as de Huber, Smalian e Newton expressas por:

i) Formula de Huber:

$$V = g_{\frac{1}{2}} \cdot l$$

ii) Formula de Smalian:

$$V = \left(\frac{g_1 + g_2}{2} \right) \cdot l$$

iii) Formula de Newton:

$$V = \frac{1}{6} l (g_1 + 4g_{\frac{1}{2}} + g_2)$$

em que V é volume da seção, em m^3 , g_1 a área seccional da base de tora, $g_{1/2}$ a área seccional no meio da tora, g_2 a área seccional no final da tora, em m^2 e l o comprimento da seção, em m . A somatória de todas as secções da árvore resulta no volume total da mesma.

2.5 Modelos volumétricos

Na totalização de parcelas de inventário florestal é comum o uso de equações volumétricas, correlacionando o volume (V) com o diâmetro (DAP) e a altura total (H), cabendo mencionar o modelo de Schumacher & Hall em 1933. Como H é uma variável de difícil medição dentro do povoamento florestal, em alguns casos, equações hipsométricas são obtidas individualmente por parcela, ao correlacionar H com o DAP , conforme alguns exemplos de modelos estatísticos apresentados em Husch *et al.* (1972), Avery e Burkhart (1983), Guimarães (1994) e Huang *et al.* (2000), mas no geral essas equações hipsométricas são evitadas, por aumentarem os erros de estimação.

O uso de equações de volume em inventário florestal se constitui em uma operação rotineira para cálculo de volume de madeira em pé por meio da relação DAP e altura (COUTO e BASTOS, 1987).

A necessidade de dispor de uma ferramenta simplificada que permita a obtenção de valores precisos do volume de madeira de uma espécie é vital para a planificação da gestão das massas florestais e para sua ordenação sustentável. Em áreas de reflorestamentos

industriais qualquer sistema de manejo florestal requerirá de estimação precisa dos estoques de madeira corrente e futuros, seja para produção de celulose, carvão ou lenha. Uma destas ferramentas utilizadas nas estimativas e prognoses de produção madeireira é o emprego de equações volumétricas, constituindo um procedimento eficiente para quantificação da produção em volume de um povoamento florestal (PARENT, 2003).

Existem três tipos de equações volumétricas: simples entrada ou local que usa, unicamente, o diâmetro a altura do peito (*DAP*) ou a circunferência a altura do peito (*CAP*) como variável independente; as de dupla entrada ou regionais, que usam o (*DAP*) ou (*CAP*) com a altura da árvore e são as mais utilizadas; e as formais, que além do *DAP* ou *CAP* e altura, incluem também a forma da árvore como variável independente (SILVA, 1977).

A maioria das metodologias desenvolvidas para estimativa do volume de árvores considera que, se o volume de uma árvore foi determinado corretamente, o valor encontrado é válido para outra árvore de igual diâmetro, altura e forma (THIERSCH *et. al.*, 2006).

Seguem a seguir alguns modelos, comentados utilizados para estimativa de volumes de *Eucalyptus*:

Modelo de Schumacher e Hall

$$vol_i = \beta_0 \cdot DAP_i^{\beta_1} \cdot H_i^{\beta_2} \cdot \epsilon_i$$

Modelo de Spurr

$$vol_i = \beta_0 \cdot (DAP_i^2 \cdot H_i)^{\beta_1} \cdot \epsilon_i$$

Modelo da Variável Combinada de Spurr

$$vol_i = \beta_0 + \beta_1(DAP_i^2 \cdot H_i) + \epsilon_i$$

Modelo de Honner

$$vol_i = \frac{DAP_i^2}{\beta_0 + \beta_1 \cdot \frac{1}{H_i}} + \epsilon_i$$

Modelo de Naslund

$$vol_i = \beta_1 DAP_i^2 + \beta_2 DAP_i^2 H_i + \beta_3 DAP_i H_i^2 + \beta_4 H_i^2 + \epsilon_i$$

Veiga (1984) afirmou que apesar dos bons resultados que comprovadamente são obtidos na estimativa do volume de madeira pelo uso de equações, o seu emprego era pouco freqüente, por causa das dificuldades de cálculo e desconhecimento sobre o assunto por parte de técnicos ligados à área florestal.

O primeiro trabalho abordando o assunto profundamente sob o ponto de vista estatístico foi realizado por Veiga (1972), com *Eucalyptus saligna*, em que comparou várias equações de volume aritméticas não formais visando a escolha das mais adequadas para *Eucalyptus saligna Smith* em maciços em ocasião de primeiro corte.

Em geral, as equações de volume publicadas se restringem a uma determinada área, idade, espaçamento, notação ou espécie. Paula Neto *et al.* (1979). apresentam modelos de equações de volume comercial de *Eucalyptus microcorys* em regime de altofuste na região de Coronel Fabriciano, MG. Já Paula Neto *et al.* (1975) produziram uma equação de volume para *Eucalyptus grandis* plantado na região de Viçosa, MG.

Veiga e Brasil (1981) testaram vários modelos de equações de volume para *Eucalyptus propinqua* com 5 anos de idade plantado na região de Itupeva e Mogi Guaçu, SP.

Silva (1977) comparou vários modelos volumétricos em espécies de eucalíptos em regime de altofuste e talhadia e o modelo de Schumacher e Hall foi que gerou melhores ajustes.

Na literatura florestal, existem vários modelos volumétricos que têm se tornado uma ferramenta cada vez mais importante dentro dos empreendimentos de base florestal. Esses modelos resultam em uma ferramenta especial, as equações são capazes de prognosticar o crescimento e a produção sob certas condições (SPATHELF e NUTTO, 2000; SOARES *et al.*, 2001).

Dado aos bons ajustes proporcionados os modelos mais clássicos são de Schumacher e Hall de 1933 e o de Spurr (1952). Eles descrevem os processos relacionados ao crescimento das árvores até a estimação volumétrica de povoamentos florestais, baseando-se em equações ou sistemas de equações volumétricas que têm como variáveis independentes o (DAP) e a altura da árvore (H), entre outras (MACHADO *et al.*, 2002; ASSIS *et al.*, 2004; SANTANA *et al.*, 2005; SILVA, 1977).

2.6 Modelos não lineares

Os modelos não lineares têm uma base teórica em que os parâmetros desses modelos fornecem um maior conhecimento sobre o fenômeno em estudo do que os modelos lineares, o que indica ser mais apropriado para modelagem de fenômenos biológicos, econômicos, ecológicos, agrícolas, etc. A principal característica dos modelos não lineares é que eles são deduzidos a partir de suposições teóricas e os parâmetros resultantes são interpretáveis. Assim, aproximá-los por modelos lineares, mesmo que sejam alcançados ajustes satisfatórios, prejudicaria bastante a obtenção de estimativas mais realistas dos parâmetros de interesse. Além disso, existem modelos que são intrinsecamente não lineares, isto é, não podem ser linearizados por transformação (SILVA e SILVA, 1995).

A aplicação dos modelos não lineares requer a suposição de homocedasticidade dos erros, ou seja, variabilidade constante dos erros. Os modelos não lineares são aplicados em diversas áreas, tais como Ecologia, Biologia, Ciência Florestal, Agricultura, Farmacologia, Econometria, Engenharia, etc.

Até o início da década de 70, as principais técnicas desenvolvidas para modelos de regressão não lineares se restringiam à suposição de normalidade para a variável resposta. Nelder e Wedderburn (1972) ampliaram a distribuição da variável resposta para a família exponencial de distribuições definindo os Modelos Lineares Generalizados. Mesmo assim, os modelos não lineares continuaram recebendo um tratamento especial, surgindo diversos artigos na década de 70 e nas décadas posteriores. Particularmente, destaca-se Ratkowsky (1983), em que vários modelos não lineares são discutidos segundo vários aspectos (CORDEIRO e LIMA NETO, 2004).

A forma clássica do modelo normal não-linear é dada por

$$y_i = h(\beta; x_i) + \varepsilon_i = \mu(\beta) + \varepsilon_i, \quad (2.1)$$

em que os ε_i 's são distribuídos normalmente com média zero e variância constante σ^2 , $h(\beta; x_i)$ é uma função diferenciável, $\beta = (\beta_1, \dots, \beta_p)^T$ contém os parâmetros desconhecidos a serem estimados e $x = (x_1, \dots, x_q)$ representa os valores de q variáveis explicativas.

Uma característica importante do modelo não linear é que o número de parâmetros regressores não são necessariamente relacionados com o número de variáveis explicativas

no modelo. No modelo linear, se existem $p - 1$ variáveis explicativas então existem p parâmetros regressores no modelo, o que pode não acontecer no modelo não linear.

Na literatura, Draper e Smith (1981) e Montgomery e Peck (1982) apresentaram de forma introdutória o modelo de regressão não-linear. Por outro lado, duas referências que abrangem de forma completa o tema são Bates e Watts (1988) e Seber e Wild (1989).

2.6.1 Estimação dos Parâmetros

Sejam y_1, \dots, y_n variáveis aleatórias independentes com estrutura dada na equação (2.1). O estimador do vetor de parâmetros β é obtido pelo método dos mínimos quadrados que, nesse caso, coincide com a máxima verossimilhança, pois o modelo tem respostas normais independentes com mesma variância. O estimador de β é, então, obtido, minimizando-se a função quadrática

$$S(\beta) = \sum_{i=1}^n [y_i - \mu_i(\beta)]^2$$

Derivando-se $S(\beta)$ em relação a β_r obtém-se

$$\frac{\partial S(\beta)}{\partial \beta_r} = 2 \sum_{i=1}^n [y_i - \mu_i(\beta)] \frac{\partial \mu_i}{\partial \beta_r}$$

A estimativa $\hat{\beta}$ do vetor de parâmetros β é obtida igualando-se $\frac{\partial S(\beta)}{\partial \beta_r}$ a zero para $r = 1, \dots, p$. Em geral, as equações $\frac{\partial S(\beta)}{\partial \beta_r} = 0$, $r = 1, \dots, p$ não são lineares e têm que ser resolvidas numericamente por processos iterativos do tipo Newton-Raphson (CORDEIRO e LIMA NETO, 2004).

Expandindo $S(\beta)$ em série de Taylor em torno de um valor β^0 até a segunda ordem, chega-se ao seguinte processo iterativo para obter $\hat{\beta}$:

$$\beta^{(m+1)} = \beta^{(m)} + \{\tilde{X}^{(m)T} \tilde{X}^{(m)}\}^{-1} \tilde{X}^{(m)T} \{y - \mu(\beta^{(m)})\}$$

em que \tilde{X} é a matriz Jacobiana da transformação de $\mu(\beta)$ em β .

2.6.2 Resíduos

Uma etapa importante na análise de regressão é a verificação de possíveis afastamentos das suposições feitas para o modelo, bem como a existência de observações extremas com alguma interferência desproporcional nos resultados do ajuste. Tal etapa, conhecida como análise de diagnóstico, iniciou-se com a análise de resíduos para detectar a presença de pontos extremos e avaliar a adequação da distribuição proposta para a variável resposta. Uma referência importante é o artigo de Cox e Snell (1968) em que é apresentada uma forma bastante geral de definir resíduos, usada até os dias atuais.

No modelo de regressão linear, como os erros apresentam distribuição normal e são não correlacionados, testes estatísticos e intervalos de confiança baseados nas distribuições t de Student e F podem ser usados para realizar inferência sobre as estimativas dos parâmetros que apresentam ainda propriedades interessantes como: não viés, mínima variância e distribuição Normal.

O resíduos ordinários não lineares r_i são definidos por

$$r_i = y_i - \mu_i(\hat{\beta}), \quad i = 1, \dots, n.$$

Em geral, mesmo assintoticamente, os resíduos ordinários são viesados e possuem distribuição não normal dificultando a verificação dos modelos pelos métodos tradicionais.

2.7 Distribuições Simétricas

É usual em análise de regressão supor normalidade para os erros aleatórios, porém, nem sempre a modelagem se ajusta adequadamente aos dados. Sendo assim, a busca de outras alternativas à suposição de erros normalmente distribuídos tem sido proposta na literatura. Uma solução é considerar modelos com distribuições mais adequadas do que o modelo normal para este tipo de problema.

A família de distribuições simétricas que será definida a seguir, retém a estrutura da distribuição normal, mas elimina a forma específica da densidade normal. Esta família

inclui densidades simétricas que têm caudas menores ou mais longas que normais.

Na literatura, destacam-se diversos estudos direcionados a classe das distribuições simétricas. Entre eles destacam-se Chmielewski (1981), Fang *et al.* (1990), Fang e Zhang (1990), Fang e Anderson (1990) e Gupta e Varga (1993), nas quais são realizadas aplicações das distribuições simétricas. Lange *et al.* (1989) propuseram o modelo baseado na suposição de erros t de Student. Little (1988) e Yamaguchi (1990) utilizaram o modelo baseado na suposição de erros com distribuição normal contaminada. Nesses modelos, adicionam-se parâmetros os quais permitem ajustar a curtose da distribuição dos dados. Taylor (1992) propôs o ajuste de um modelo de regressão linear supondo erros com distribuição exponencial potência com um parâmetro extra de forma.

Arellano-Valle (1994) apresentou vários resultados sobre propriedades, inferência e aplicações a modelos de regressão em distribuições elípticas. Ferrari e Uribe-Opazo (2001) estendem esses resultados para modelos de regressão linear simétricos. Cordeiro *et al.* (2000) obtiveram a correção do viés do estimador de máxima verossimilhança na classe de modelos não-lineares simétricos. Cordeiro (2004) desenvolveu uma correção de Bartlett para os modelos de regressão não-lineares simétricos generalizando os resultados de Ferrari e Uribe-Opazo (2001). Galea *et al.* (2005) apresentaram alguns resultados sobre modelagem, em particular sobre o desenvolvimento da análise inferencial e de diagnóstico na classe não lineares com erros simétricos independentes. Cordeiro e Andrade (2009) introduziram uma nova classe de modelos simétricos transformados que estendem o modelo de Box e Cox (1964).

2.7.1 Definições

A seguir será definida a classe simétrica univariada e alguns resultados teóricos.

Uma variável aleatória Y com suporte em \mathbb{R} , com parâmetro de locação $\mu \in \mathbb{R}$ e de escala $\phi > 0$ com função de densidade de probabilidade especificada por

$$f(y; \mu, \phi) = \frac{1}{\sqrt{\phi}} g \left\{ \frac{(y - \mu)^2}{\phi} \right\}, \quad y \in \mathbb{R},$$

para alguma função $g(\cdot)$ denominada função geradora de densidades com $g(u) > 0$, para $u > 0$ e $\int_0^\infty u^{-1/2} g(u) du = 1$. Esta condição é necessária para que $f(y; \mu, \phi)$ seja uma função de densidade de probabilidade. Denota-se por $y \sim S(\mu, \phi)$ e se denomina variável

aleatória simétrica (CYSNEIROS *et al.*, 2005).

Pode-se citar como distribuições pertencentes a esta classe a normal, Cauchy, t -Student, logística tipos I e II, Laplace, exponencial potência, Kotz, normal contaminada, t -Student generalizada, logística generalizada, Kotz generalizada, dentre outras (CYSNEIROS *et al.*, 2005).

Algumas propriedades da distribuição normal podem ser estendidas para a classe simétrica de distribuições, tais como:

- i) Se $y \sim S(\mu, \phi)$ então a função característica de y , $\psi_y(t) = E(e^{ity})$ é dada por $e^{it\mu}\varphi(t^2\phi)$, $t \in \mathbb{R}$ para alguma função φ , com $\varphi(u) \in \mathbb{R}$ para $u > 0$. Quando existem, $E(y) = \mu$ e $Var(y) = \xi\phi$, em que $\xi > 0$ é uma constante dada por $\xi = -2\varphi'(0)$, com $\varphi'(0) = d\varphi(u)/du|_{u=0}$ e que não depende dos parâmetros μ e ϕ (FANG *et al.*, 1990). Se $u^{-\frac{1}{2}(k+1)}g(u)$ for integrável então o k -ésimo momento de Y existe.
- ii) Se $Y \sim S(\mu, \phi)$ então $a + bY \sim S(a + b\mu, b^2\phi)$, em que $a, b \in \mathbb{R}$ com $b \neq 0$, isto é, a distribuição de qualquer combinação linear de uma variável aleatória com distribuição simétrica é também simétrica.
- iii) Berkane e Bentler (1986), considerando uma distribuição simétrica padrão e que seus momentos existem, mostram que a função característica de Z pode ser expandida como

$$\psi_Z(t) = \sum_{k=0}^{\infty} i^k \mu'_k \frac{t^k}{k!}$$

em que $\mu'_k = E(y^k) = i^k \psi_Z^{(k)}(0)$, com $\psi_Z^{(k)}(0)$ denotando a k -ésima derivada de $\psi_Z(t)$ avaliada em $t = 0$. Então

$$\mu'_k = \begin{cases} 0, & k \text{ ímpar} \\ \frac{(2m)!}{2^m m!} (\mu'_2)^m [k(m) + 1], & k = 2m, \quad m = 1, 2, \dots \end{cases}$$

sendo

$$k(m) = \frac{\varphi^{(m)}(0)}{\{\varphi^{(1)}(0)\}^m} - 1$$

em que $\varphi^{(r)}(0)$ é a r -ésima derivada de função φ , avaliada em zero. Os coeficientes $k(m)$, $m = 1, 2, \dots$, são conhecidos como parâmetros de momentos e generalizam o coeficiente de curtose $\gamma_2 = 3\{k(2) + 1\}$ de uma distribuição $S(\mu, \phi)$ (MUIRHEAD, 1982).

Cambanis *et al.* (1981) observaram que a família de distribuições simétricas coincide com a classe de distribuições elípticas univariadas. Na última década surgiram contribuições importantes a partir dos trabalhos de Keller (1970) para as distribuições elípticas univariadas e multivariadas. Pode-se citar alguns trabalhos que discutem alguns aspectos dessas distribuições, tais como Berkane e Bentler (1986), Muirhead (1980 e 1982), Rao (1990), Cambanis *et al.* (1981), Anderson e Fang (1987) e Cysneiros *et al.* (2005).

A seguir são apresentadas algumas distribuições simétricas com suporte na reta real para $u = \frac{(y-\mu)^2}{\phi}$, em que $y \sim S(\mu, \phi)$.

2.8 Distribuição Normal

A distribuição Normal, também conhecida por distribuição Gaussiana, segunda lei de Laplace, Laplace, Laplace-Gauss, de Moivre, é uma família importante das distribuições contínuas de probabilidade, aplicável em muitas áreas (JOHNSON e KOTZ, 1970). Suas propriedades, além de fundamentar decisões, medir e prevenir riscos e até explicar curiosidades, descrevem bem variáveis como comprimento de pinos e diâmetros de discos, altura, peso, inteligência e tempo de gestação de seres vivos, rendas e despesas de famílias ou categorias profissionais, rendimentos de máquinas e campos de trigo, qualidade do ar, velocidade de molécula, distribuições diamétricas e volumétricas de árvores, etc. Cada membro dessa família pode ser definido por dois parâmetros, locação e escala: a média μ e a variância σ^2 , respectivamente. A distribuição normal padrão possui média zero e variância um (JOHNSON e KOTZ, 1970).

Uma variável normal, de modo geral, retrata bem fenômenos cujo efeito final corresponde à soma de múltiplas causas ou é afetado por diversas variáveis independentes (típico de variáveis físico químicas, socioeconômicas, psicossociais, etc).

Carl Friedrich Gauss em 1809 se tornou associado com essa distribuição quando ele analisou dados astronômicos, e definiu a equação desta densidade de probabilidade. Ela é frequentemente chamada de curva de sino porque o gráfico da sua densidade de probabilidade se assemelha um sino (JOHNSON e KOTZ, 1970).

A importância da distribuição normal como um modelo de fenômeno quantitativo nas

ciências naturais e comportamentais é devido em parte ao Teorema do Limite Central. A distribuição normal é amplamente usada entre as famílias de distribuições na Estatística e muitos testes estatísticos são baseados na suposição de normalidade (JOHNSON e KOTZ, 1970).

A distribuição normal foi primeiramente introduzida por Abraham de Moivre em um artigo em 1733, no contexto da aproximação exata da distribuição binomial para n grande. Seu resultado foi estendido por Laplace em seu livro Teoria das Probabilidades Analíticas em 1812, e é hoje conhecido por "O Teorema de Moivre".

Laplace usou a distribuição normal nas análises dos erros de experimentos. O nome "distribuição normal" foi estabelecido independentemente por Charles S. Peirce, Francis Galton e Wilhelm Lexis por volta de 1875. Em 1774 Laplace obteve a distribuição normal com uma aproximação pela distribuição hipergeométrica (JOHNSON e KOTZ, 1970).

2.8.1 Caracterização

Para indicar que uma variável aleatória Y é normalmente distribuída com média μ e variância $\sigma^2 \geq 0$, tem-se:

$$Y \sim N(\mu, \sigma^2)$$

A função densidade de probabilidade da distribuição normal é a função Gaussiana expressa por:

$$f(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, \quad y \in \mathbb{R},$$

em que $\sigma > 0$ é o desvio-padrão, o parâmetro real μ é o valor esperado e

$$f(y; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}, \quad y \in \mathbb{R},$$

é a função densidade da distribuição normal "padrão": com $\mu = 0$ e $\sigma^2 = 1$.

A função densidade de probabilidade possui propriedades importantes incluindo:

- i) simetria em relação a média μ .
- ii) a moda e mediana são iguais a média μ .
- iii) o ponto de inflexão da curva ocorre em um desvio-padrão da média, isto é, em $\mu - \sigma$ e $\mu + \sigma$.

A função de distribuição acumulada da distribuição normal é expressada em termos da função densidade como a seguir:

$$\Phi_{\mu, \sigma^2}(y) = \int_{-\infty}^y f(u, \mu, \sigma^2) du = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^y \exp\left[-\frac{(u - \mu)^2}{2\sigma^2}\right] du, \quad y \in \mathbb{R},$$

A função geratriz de momentos é expressada por

$$M_y(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

A função geratriz de cumulantes que é o logaritmo da função geratriz de momentos: $g(t) = \mu t + \frac{\sigma^2 t^2}{2}$. Desde então, este é um polinômio quadrático em t e somente são diferentes de zero nos dois primeiros cumulantes.

A função característica é dada por

$$M_y(it) = \exp\left(i\mu t + \frac{\sigma^2 t^2}{2}\right)$$

Algumas propriedades da distribuição normal (JOHNSON e KOTZ, 1970):

- i) Se $Y \sim N(\mu, \sigma^2)$ e a e b são números reais, então $aY + b \sim N(a\mu + b, (a\sigma)^2)$.

- ii) Se $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$ são variáveis aleatórias normais e independentes, então, a soma dela é normalmente distribuídas com $U = X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$ e a diferença é normalmente distribuída com $V = X - Y \sim N(\mu_x - \mu_y, \sigma_x^2 + \sigma_y^2)$.
- iii) Se Y_1, \dots, Y_n são variáveis aleatórias normais padrão independentes, então Y_1^2, \dots, Y_n^2 tem uma distribuição qui-quadrado com n graus de liberdade.

Seguem, a seguir, algumas distribuições relacionadas com a distribuição normal

- 1) $R \sim \text{Rayleigh}(\sigma)$ é uma distribuição Rayleigh se $R = \sqrt{X^2 + Y^2}$ em que $X \sim N(0, \sigma^2)$ e $Y \sim N(0, \sigma^2)$ são distribuições normais independentes.
- 2) $Y \sim \chi_k^2$ é uma distribuição qui-quadrado com v graus de liberdade se $Y = \sum_{i=1}^v X_k^2$ em que $X_k \sim N(0, 1)$ para $k = 1, \dots, v$ e são independentes.
- 3) $Y \sim \text{Cauchy}(\mu = 0, \lambda = 1)$ é uma distribuição Cauchy se $Y = \frac{X_1}{X_2}$ para $X_1 \sim N(0, 1)$ e $X_2 \sim N(0, 1)$ são duas distribuições normais independentes.
- 4) $Y \sim \text{Log} - N(\mu, \sigma^2)$ é uma distribuição log-normal se $Y = e^X$ e $X \sim N(\mu, \sigma^2)$

2.8.2 Estimação dos parâmetros

Suponha X_1, \dots, X_n são independentes e cada uma é normalmente distribuída com esperança μ e variância $\sigma^2 > 0$.

Os estimadores de máxima verossimilhança dos parâmetros da distribuição normal são dados por

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

e

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

em que \bar{x} é o estimador da verdadeira média μ e s^2 é o estimador não viesado da verdadeira variância σ^2 . Estas estimativas são conjuntamente suficientes para μ e σ^2 . Para a maioria dos estudos \bar{x} é o melhor estimador para μ , para σ conhecido ou não.

A distribuição normal pertencente à classe simétrica é a mais utilizada devido a todo desenvolvimento teórico e aplicado estabelecido no decorrer dos anos (CYSNEIROS *et al.*, 2005).

Seja $Y \sim S(\mu, \phi)$ e a função geradora de densidades $g(\cdot)$ é especificada por

$$g(u) = \frac{1}{\sqrt{2\pi}} \exp\{-u/2\}, \quad u > 0,$$

então Y tem uma distribuição normal denotada por $Y \sim N(\mu, \phi)$ e os momentos de ordem r são

$$\mu_r = E\{(y - \mu)^r\} = \begin{cases} 0, & r \text{ ímpar} \\ \phi^r r! / \{2^{r/2} (r/2)!\}, & r \text{ par} \end{cases}$$

e, portanto o coeficiente de curtose é $\gamma_2 = 3$.

Como ilustração, na Figura 1 é apresentado o gráfico da função normal representada com diferentes parametrizações.

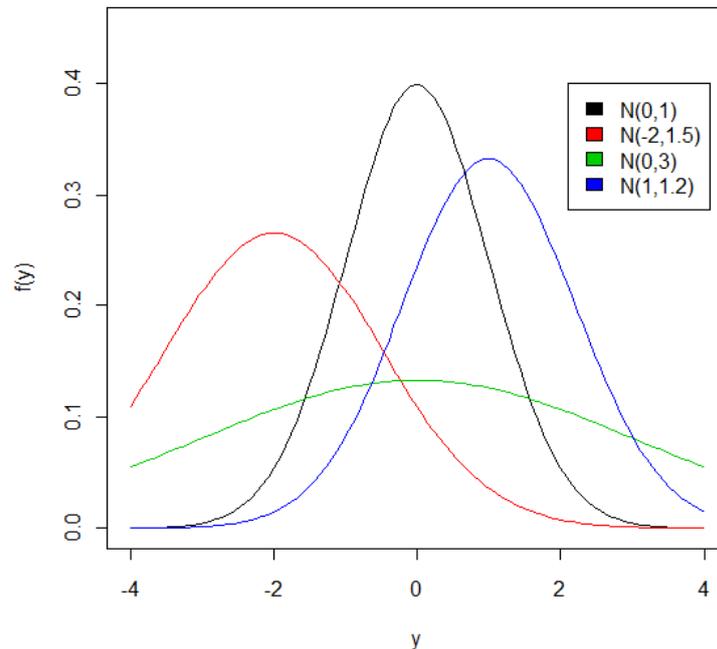


Figura 1: Densidade da distribuição normal segundo diferentes parametrizações.

2.9 Distribuição de Cauchy

A distribuição, conhecida como "distribuição de Cauchy", tornou-se associada com Augustin Cauchy em 1853, um dos matemáticos mais produtivos que já existiu, professor na Escola Politécnica de Paris em 1816 (JOHNSON e KOTZ, 1970).

A distribuição Cauchy-Lorentz, nomeada conforme Augustin Cauchy e Hendrick Lorentz, é uma distribuição de probabilidade contínua e simétrica em relação a média. Como uma distribuição de probabilidade, ela é conhecida como distribuição de Cauchy, enquanto que, entre os físicos ela é conhecida como uma distribuição de Lorentz ou distribuição Breit-Wigner. A distribuição de Cauchy é usada na teoria elétrica e mecânica e na física.

Sua função densidade de probabilidade é definida por

$$f(y; \mu, \lambda) = \frac{1}{\pi\lambda \left[1 + \left(\frac{y-\mu}{\lambda} \right)^2 \right]}, \quad \lambda > 0 \quad e \quad -\infty < y < \infty,$$

em que μ é o parâmetro de locação, especificando a locação do pico da distribuição, e λ é o parâmetro de escala.

O caso especial com $\mu = 0$ e $\lambda = 1$ é chamada a distribuição Cauchy padrão com função densidade de probabilidade dada por

$$f(y; 0, 1) = \frac{1}{\pi(1 + y^2)}, \quad -\infty < y < \infty.$$

A função de distribuição acumulada é:

$$F(y; \mu, \lambda) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left(\frac{y - \mu}{\lambda} \right)$$

A distribuição de Cauchy é um exemplo das distribuições que não possuem média, variância e outros momentos definidos, o que provoca grande interesse dos matemáticos.

Se Y_1, \dots, Y_n são variáveis aleatórias independentes e identicamente distribuídas, cada uma com distribuição Cauchy padrão, a média amostral $(Y_1 + \dots + Y_n)/n$ também é uma distribuição Cauchy padrão.

A função característica da distribuição de Cauchy é dada por

$$E[e^{itY}] = \exp(it\mu - \lambda|t|)$$

Seguem, a seguir, algumas distribuições relacionadas com a distribuição de Cauchy (JOHNSON e KOTZ, 1970):

- 1) A razão entre duas variáveis aleatórias normais padrão independentes é uma variável Cauchy padrão, ou seja, uma $Cauchy(0, 1)$. Assim, a distribuição Cauchy é uma distribuição razão.
- 2) A distribuição Cauchy padrão $(0, 1)$ segue como um caso especial da distribuição t -Student com um grau de liberdade.
- 3) A recíproca de uma variável Cauchy $C(\mu, \lambda)$ é uma variável Cauchy $C(\mu', \lambda')$, ou seja, $1/C(\mu, \lambda) \sim C(\mu', \lambda')$, em que $\mu' = \mu/(\mu^2 + \lambda^2)$ e $\lambda' = \lambda/(\mu^2 + \lambda^2)$

A variável aleatória $Y \sim S(\mu, \phi)$ tem distribuição de Cauchy se sua função geradora de densidades $g(\cdot)$ tem a seguinte forma

$$g(u) = \frac{1}{\pi}(1 + u)^{-1}, \quad u > 0.$$

Então se denota por $Y \sim C(\mu, \phi)$. Uma característica importante da distribuição de Cauchy é que ela tem caudas mais pesadas do que a normal. Da relação $Y = \mu + \phi N_1/N_2$ em que $N_i \sim N(0, 1)$ para $i = 1, 2$ independentes, é possível definir um gerador de números aleatórios para a distribuição Cauchy.

Na Figura 2 é apresentada uma comparação entre as densidades da distribuição Cauchy com diferentes parametrizações e da distribuição normal padrão.

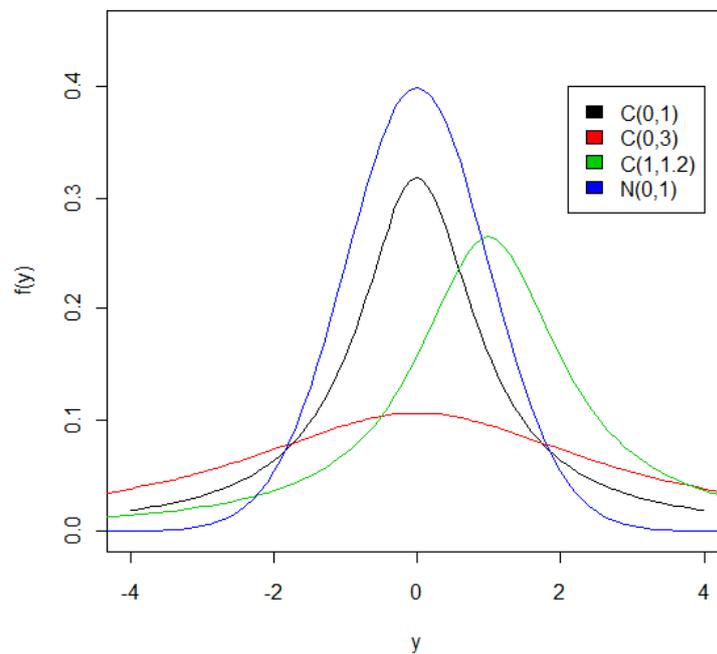


Figura 2: Densidades da distribuição Cauchy segundo diferentes parâmetros e distribuição normal padrão.

2.10 Distribuição de Laplace ou Exponencial Dupla

A distribuição de Laplace foi desenvolvida por Pierre-Simon Laplace em 1774, como a distribuição para qual a função de verossimilhança é maximizada pelo conjunto de parâmetros de locação iguais a mediana dos valores observados de um número de variáveis aleatórias independentes e identicamente distribuídas. Este resultado surgiu no artigo fundamental de Laplace sobre distribuições simétricas para descrever erros de medidas, e ele é apropriadamente conhecido como a **primeira lei de Laplace** (JOHNSON e KOTZ, 1970).

A distribuição de Laplace tem sido encontrada em uma variedade de muitos usos específicos. Ela tem se tornado recentemente totalmente popular na modelagem de variáveis financeiras (movimento Browniano de Laplace) (JOHNSON e KOTZ, 1970).

2.10.1 Caracterização

A variável aleatória Y tem uma distribuição de Laplace(μ, λ) se sua função densidade de probabilidade é

$$f(y; \mu, \lambda) = \frac{1}{2\lambda} \exp\left(-\frac{|y - \mu|}{\lambda}\right), \quad -\infty < y < \infty, \quad \lambda > 0$$

em que μ é o parâmetro de locação e λ é o parâmetro de escala. A função densidade de probabilidade da distribuição Laplace é remanescente da distribuição normal: Entretanto, considerando que a distribuição normal é expressada em termos da diferença quadrática da média μ e a densidade de Laplace é expressada em termos da diferença absoluta da média. Consequentemente a distribuição Laplace tem caldas mais pesadas do que a distribuição normal.

A função de distribuição acumulada é dada por

$$F(y) = \begin{cases} \frac{1}{2} \exp\left(-\frac{\mu - y}{\lambda}\right) & \text{se } y < \mu \\ 1 + \frac{1}{2} \exp\left(-\frac{\mu - y}{\lambda}\right) & \text{se } y \geq \mu. \end{cases}$$

A variável Laplace(0, λ) pode também ser gerada como a diferença entre duas variáveis Exponenciais($1/\lambda$) independentes e identicamente distribuídas. É desta relação que

o nome Exponencial dupla está associado a distribuição de Laplace.

Dada N amostras independentes e identicamente distribuídas Y_1, \dots, Y_N , um estimador $\hat{\mu}$ de μ é a mediana amostral, e o estimador de λ é:

$$\hat{\lambda} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{\mu}|,$$

usando o estimador de máxima verossimilhança.

O r -ésimo momento central é

$$\mu'_r = \begin{cases} 0 & r \text{ ímpar} \\ r! \lambda^r & r \text{ par} \end{cases}$$

e a função geratriz de momentos é dada por

$$E[e^{ty}] = \frac{\exp(\mu t)}{1 - \lambda^2 t^2}, \quad |t| < 1/\lambda.$$

A média, a moda e a mediana são iguais a μ , a variância é igual a $2\lambda^2$ e o coeficiente de curtose $\gamma_2 = 6$.

O r -ésimo cumulante é

$$\kappa_r(y) = \begin{cases} 0, & r \text{ ímpar} \\ 2[(r-1)!] \lambda^r, & r \text{ par} \end{cases}$$

2.10.2 Distribuições relacionadas

- 1) Se $Y \sim \text{Laplace}(0, \lambda)$ então $|Y| \sim \text{Exponencial}(\lambda^{-1})$ é uma distribuição exponencial.
- 2) Se $X \sim \text{Exponencial}(\lambda)$ e $Y \sim \text{Bernoulli}(0.5)$ independente de X então $X(2Y - 1) \sim \text{Laplace}(0, \lambda^{-1})$

- 3) Se $X_1 \sim \text{Exponencial}(\lambda_1)$ e $X_2 \sim \text{Exponencial}(\lambda_2)$ independente de X_1 , então $\lambda_1 X_1 - \lambda_2 X_2 \sim \text{Laplace}(0, 1)$.
- 4) Duas variáveis Laplace independentes, com parâmetro $\mu = 0$, estão relacionadas com a variável F com parâmetros $v = w = 2$, $F(2, 2)$ por $F(2, 2) \sim |L(0, \lambda)_1 / L(0, \lambda)_2|$.

Uma variável aleatória $Y \sim S(\mu, \phi)$ tem distribuição de Laplace se sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = \frac{1}{2} \exp\{-\sqrt{u}\}, \quad u > 0$$

e se denota por $Y \sim \text{Laplace}(\mu, \phi)$. A função característica é dada por $\psi_y(t) = \frac{e^{ity}}{1+t^2\phi}$, $t \in \mathbb{R}$. Se $Z \sim \text{Laplace}(0, 1)$ temos os momentos μ'_r dados por

$$\mu'_r = E(z^r) = \begin{cases} 0 & r \text{ ímpar} \\ r! & r \text{ par} \end{cases}$$

Os quantis da distribuição são $(\mu \pm 0, 534\sqrt{\phi})$.

Na Figura 3 observa-se os gráficos das densidades da distribuição de Laplace padrão e da distribuição normal padrão.

2.11 Distribuição t de Student

A distribuição t de Student é uma distribuição de probabilidade que aparece naturalmente no problema de se determinar a média de uma população (que segue a distribuição normal) a partir de uma amostra.

Student é um pseudônimo de Willian Sealy Gosset. Quando se formou, em 1899, foi trabalhar para a destilaria de Dublin de Arthur Guinness & Son (JOHNSON e KOTZ, 1970).

A Guinness era uma empresa de Agro-Química progressista e Gosset iria aplicar os seus conhecimentos de estatística tanto na cervejaria como nas quintas, para seleção das melhores espécimes de cervada. Para prevenir fugas de informação e futuras revelações dos “segredos” da marca, a Guinness proibiu que os seus empregados pudessem publicar

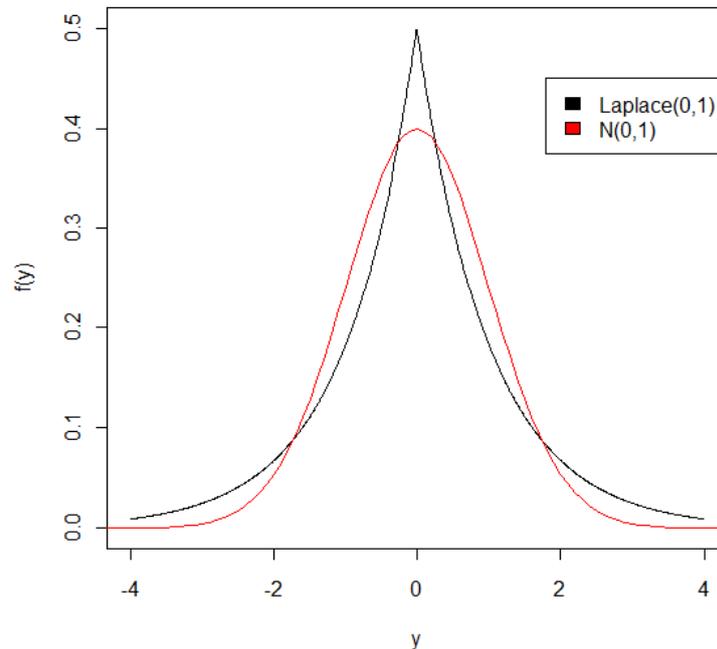


Figura 3: Densidades da distribuição de Laplace padrão e distribuição normal padrão.

quaisquer trabalhos independentemente do conteúdo. Então, Gosset usou o pseudônimo Student para as suas publicações evitando ser detectado pela entidade empregadora. Desta forma, o seu feito mais famoso, é hoje conhecido como a distribuição t -Student, publicada em 1908 que em outras circunstâncias seria conhecida com a distribuição t -Gosset (JOHNSON e KOTZ, 1970).

A distribuição t é a base do famoso teste t -Student de significância estatística da diferença entre duas médias amostrais, e para intervalos de confiança para a diferença entre duas médias populacionais. A distribuição t de Student também é utilizada para modelar o comportamento de dados que provém de uma distribuição com caudas mais pesadas que a distribuição normal, permitindo reduzir a influência de observações aberrantes. Essa distribuição permite ajustar a curtose dos dados através dos graus de liberdade " v ". Para valores de n maiores do que 30 as diferenças entre as distribuições normal e t de Student são pequenas, ou seja, com o aumento de v a distribuição t de Student se aproxima da distribuição normal (JOHNSON e KOTZ, 1970).

Algumas referências relevantes são Praetz (1972), Praetz e Wilson (1978), Blattberg e Gonedes (1974), McLeay (1986) e Taylor e Kingsman (1979). Esta distribuição também pode ser aplicada com uma distribuição a priori para valores esperados de variáveis multinormais.

2.11.1 t de Student

Sejam Z e S variáveis aleatórias independentes tal que

$$Z \sim N(0, 1) \quad e \quad nS^2 \sim \chi_n^2$$

A distribuição de $t = Z/S$ é chamada t de Student com grau de liberdade $g.l$ v . A variável aleatória t de Student com $g.l = v$ é comumente denotada por t_v , e sua função de densidade de probabilidade é

$$f(y; v) = \frac{\Gamma[(v+1)/2]}{\Gamma(v/2)\sqrt{v\pi}} \cdot \frac{1}{(1+y^2/v)^{(v+1)/2}} \quad -\infty < y < \infty, \quad v \geq 1$$

em que Γ é a função Gamma.

Após algumas álgebras, a função acumulada da distribuição t de Student é definida por:

$$F(y) = \begin{cases} \frac{1}{2} - \frac{1}{2} I_{\frac{y^2}{v+y^2}} \left(\frac{1}{2}, \frac{v}{2} \right) & \text{para } -\infty < y < 0 \\ \frac{1}{2} + \frac{1}{2} I_{\frac{y^2}{v+y^2}} \left(\frac{1}{2}, \frac{v}{2} \right) & \text{para } 0 < y < \infty \end{cases}$$

Tem-se que a média é $E(y) = 0$ para $v > 1$ e $Var(y) = \frac{v}{v-2}$ para $v > 2$. O desvio médio é dado por:

$$E(|u|) = \frac{\sqrt{v}\Gamma(\frac{v-1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{v}{2})}$$

O coeficiente de curtose é $\gamma_2 = 3 + \frac{6}{v-4}$, para $v > 4$. Observa-se que este coeficiente é maior do que o coeficiente da distribuição normal.

Se $v > r$, seus momentos de ordem r existem e são definidos por:

$$F(y) = \begin{cases} 0, & r \text{ ímpar} \\ \frac{v^{r/2} \Gamma(\frac{r+1}{2}) \Gamma(\frac{v-r}{2})}{\Gamma(\frac{1}{2}) \Gamma(\frac{v}{2})}, & r \text{ par} \end{cases}$$

O momento de ordem r é indefinido se r for par e $r \geq v$.

Seguem, a seguir, algumas distribuições relacionadas com a distribuição t de Student (JOHNSON e KOTZ, 1970):

- 1) $Y \sim t_v$ tem uma distribuição t se $\sigma^2 \sim Inv - \chi^2(v, 1)$ tem uma distribuição χ^2 inversa e $Y \sim N(0, \sigma^2)$ tem uma distribuição normal.
- 2) $X \sim F(v_1 = 1, v_2 = v)$ tem uma distribuição F se $X = Y^2$ e $Y \sim t_v$ tem uma distribuição t de Student.
- 3) $X \sim N(0, 1)$ tem uma distribuição normal, como $X = \lim_{v \rightarrow \infty} Y$ em que $Y \sim t_v$.
- 4) $Y \sim \text{Cauchy}(0, 1)$ tem uma distribuição de Cauchy se $Y \sim t_1$.

A variável aleatória y tem distribuição t de Student com v graus de liberdade se $y \sim S(\mu, \phi)$ e se sua função geradora de densidades for da forma:

$$g(u) = \frac{v^{v/2}}{B(\frac{1}{2}, \frac{v}{2})} (v + u)^{-\frac{v+1}{2}}, \quad v > 0, \quad u > 0$$

em que $B(., .)$ é a função beta e $y \sim t(\mu, \phi, v)$. Assim, a função de densidade de y é obtida aplicando a função $g(.)$ acima.

Como ilustração, na Figura 4 é apresentada uma comparação entre a função de densidade da distribuição t -Student com diferentes graus de liberdade e da distribuição normal padrão.

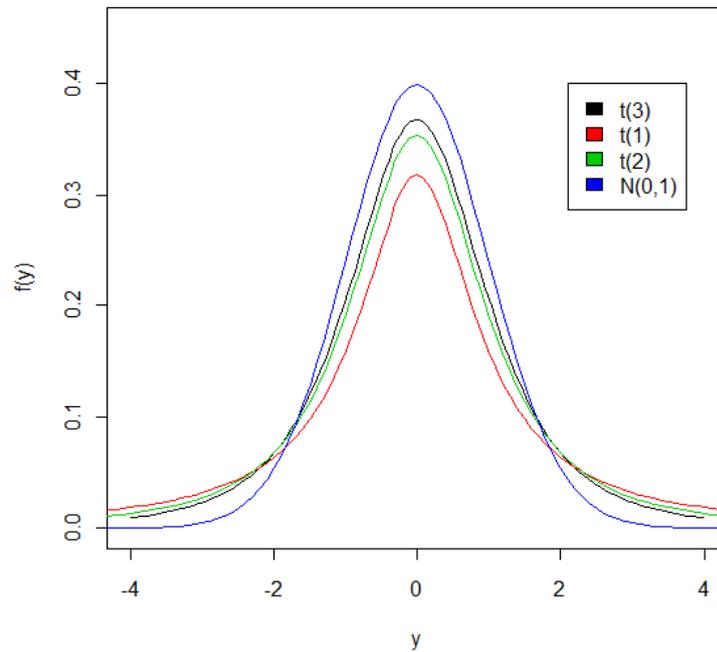


Figura 4: Densidades da distribuição de t segundo diferentes graus de liberdade e distribuição normal padrão.

2.11.2 t de Student Generalizada

McDonald e Newey (1988) introduziram a distribuição t de Student generalizada com função densidade dada por:

$$f(y, \sigma, p, q) = \frac{p}{2q^{1/p} B(p^{-1}, q) (1 + |y|^p / q)^{p+q-1}} \quad -\infty < y < \infty, \quad p, q > 0,$$

em que $B(.,.)$ é a função beta, σ é o parâmetro de escala, enquanto que p e q são os parâmetros de forma da densidade. Quanto maior os valores de p e q mais suaves serão as caudas da distribuição. Uma caso especial dessa distribuição é quando $p = 2$ e $q = 2v$. Neste caso a distribuição se reduz a distribuição t de Student. A densidade generalizada definida acima também inclui, como um caso especial, a densidade exponencial potência,

$$f(y) = \frac{pe^{|y|^p}}{2\Gamma(p^{-1})} \quad -\infty < y < \infty, \quad p > 0,$$

quando $q \rightarrow \infty$.

McDonald (1984) mostrou que a t de Student generalizada, de fato, é uma mistura da distribuição gamma generalizada e da distribuição exponencial potência de Box e Tiao. McDonald (1991) apresentou valores do coeficiente de curtose, para vários valores de p e q .

McDonald e Newey (1988) usaram a t de Student generalizada para desenvolver estimação parcialmente adaptada de modelos de regressão. Butler *et al.* (1990) discutiram a estimação robusta de modelos de regressão usando a t de Student generalizada. Similarmente, estimação parcialmente adaptada de modelos ARMA de séries temporais através da distribuição t de Student generalizada foi desenvolvida por McDonald (1989).

Uma variável aleatória $Y \sim S(\mu, \phi)$ com função geradora de densidades definida por

$$g(u) = \frac{s^{r/2}}{B(1/2, r/2)} (s + u)^{-\frac{r+1}{2}}, \quad s, r > 0, \quad u > 0,$$

é chamada t de Student generalizada com parâmetros (μ, ϕ, s, r) (DICKY, 1967). Como membro desta família de distribuições temos a t -Student ($s = r = v$) e Cauchy ($s = r = 1$).

Suponha que $Y/V = v \sim N(\mu, v\phi)$ em que $V \sim GI(r/2, s/2)$, independentes com $s, r > 0$ podendo não ser inteiros. Pode-se relacionar algumas propriedades:

- i) $Y \sim tG(\mu, \phi, s, r)$;
- ii) $E(Y) = \mu$ para $r > 1$, $Var(Y) = \{s/(r - 2)\}\phi$ para $r > 2$ e o coeficiente de curtose $\gamma_2 = 3 + 6/(r - 4)$ para $r > 4$. Vale salientar que o coeficiente de curtose não depende do parâmetro s . O coeficiente de curtose da t de Student generalizada é maior do que o coeficiente de curtose da distribuição normal;
- iii) $V/Y \sim GI((r + 1)/2, (s + z^2)/2)$, em que $z^2 = (y - \mu)^2/\phi$;
- iv) $U^2 = rZ^2/s \sim F_{(1,r)}$;
- v) Se $W = (r + 1)/(s + Z^2)$ então

$$E(Z^{2k}W^\ell) = \frac{(-\frac{r+1}{2})^\ell}{s^{\ell-k}} \frac{B[(2k + 1)/2, \{r + 2(\ell - k)\}/2]}{B(1/2, r/2)},$$

para $\ell = 0, 1, 2$ e $k = 1, 2, \dots$;

vi) Os parâmetros s e r , apresentados a seguir, têm uma relação com o parâmetro de curtose e o segundo momento central (JOHNSON e KOTZ, 1970).

$$r = \frac{2(2\gamma_2 - 3)}{\gamma_2 - 3} \quad e \quad s = \frac{2\mu_2\gamma_2}{\gamma_2 - 3};$$

vii) O ℓ -ésimo momento existe se e somente se $r > \ell$;

viii) para a variável aleatória $Y = v^{-1/2}Z$, Z e V variáveis aleatórias independentes, em que $Z \sim N(0, 1)$ e $V \sim GI(r/2, s/2)$ então $Y \sim tG(0, 1, s, r)$.

Na Figura 5, é possível observar os gráficos das funções de densidades das distribuições t de Student generalizada com diferentes valores dos parâmetros s e r e da distribuição normal padrão.

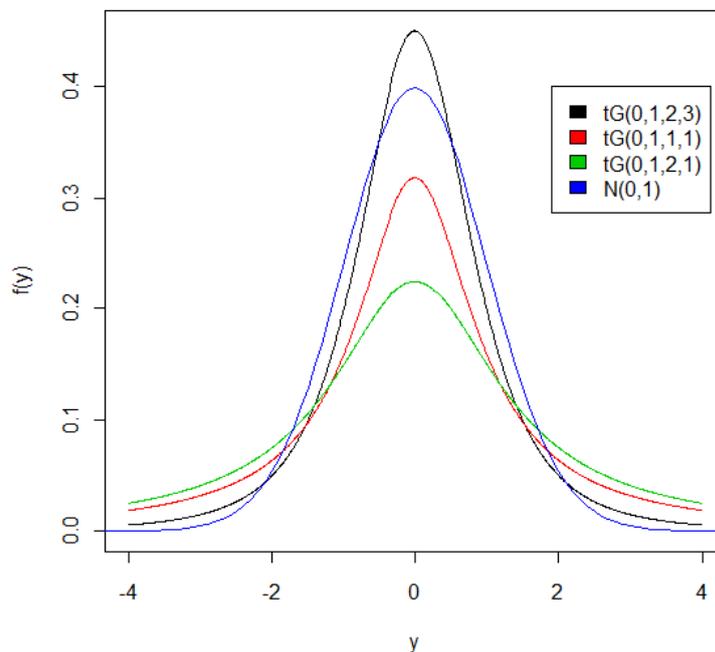


Figura 5: Densidades da distribuição t de Student generalizada com diferentes valores dos parâmetros s e r e distribuição normal padrão.

2.12 Distribuição Logística

A distribuição logística foi primeiramente referenciada como uma curva de crescimento por Verhulst em 1838 e 1845. É uma distribuição bastante popular entre demógrafos e economistas desde o início do século XIX. Pearl e Reed em 1920 e 1924, Pearl em 1940 e Schultz em 1930 aplicaram o modelo logístico como modelo de crescimento em população humana, assim como em alguns organismos biológicos. Schultz em 1930 usou a função logística para dados de produção agrícola. Outras aplicações bastantes interessantes são nas áreas de sobrevivência (PLACKETT, 1961) e em modelagem de distribuição de renda.

De acordo com Gumbel em 1944, a distribuição logística surgiu de uma maneira puramente estatística como distribuição limite ($n \rightarrow \infty$) de medidas padronizadas (média de maiores e menores valores) de amostras aleatórias de tamanho n .

O fato adicional de que a distribuição logística tem forma similar a distribuição normal torna ela proveitosa em certas ocasiões. Substituindo a normal pela logística se obtém uma simplificação das análises sem trazer grandes discrepâncias na teoria. Assim, tais substituições devem ser feitas com cuidado e entendimento da similaridade entre as duas distribuições.

Entre as aplicações da distribuição logística incluem: introdução de um novo produto no mercado, crescimento populacional, expansão da produção agrícola e peso ganho em animais.

2.12.1 Logística I

Diz-se que a variável aleatória $y \sim S(\mu, \phi)$ tem distribuição logística I (CYSNEIROS *et al.*, 2005) se sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = c \cdot \frac{e^{-u}}{(1 + e^{-u})^2}$$

Denota-se por $y \sim LI(\mu, \phi)$ e sua função de densidade de probabilidade é dada por

$$f(y; \mu, \phi) = \frac{c}{\sqrt{\phi}} \cdot \frac{e^{-\left\{\frac{(y-\mu)^2}{\phi}\right\}}}{(1 + e^{-\left\{\frac{(y-\mu)^2}{\phi}\right\}})^2}, \quad -\infty < y < \infty$$

em que c é a constante normalizadora obtida da relação $\int_0^\infty u^{-1/2}g(u) = 1$, logo $c \approx 1,484300029$.

Tem-se que $E(y) = \mu$, $Var(y) \approx 0,79569\phi$ e $\gamma_2 \approx 2,385165$. Pode-se observar que o coeficiente de curtose da distribuição logística- I é menor do que o coeficiente de curtose da distribuição normal.

Na Figura 6, é possível observar o gráfico da função de densidade da distribuição logística- I padrão comparada com a função de densidade da distribuição normal padrão.

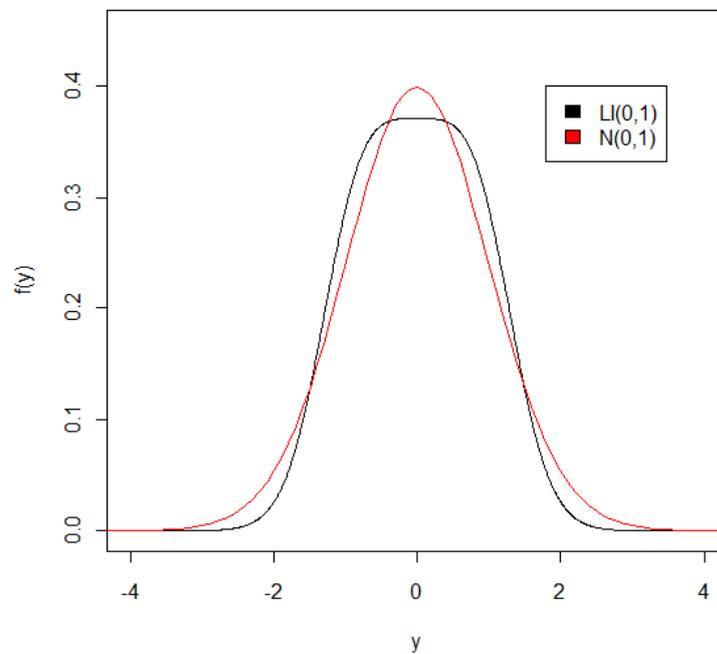


Figura 6: Densidades da distribuição de logística I padrão e distribuição normal padrão.

2.12.2 Logística II

A função de densidade de probabilidade da distribuição logística-*II* com parâmetro de locação μ e de escala β é dada por:

$$f(y; \mu, \beta) = \frac{1}{\beta} \cdot \frac{e^{-\left\{\frac{y-\mu}{\beta}\right\}}}{\left(1 + e^{-\left\{\frac{y-\mu}{\beta}\right\}}\right)^2}, \quad -\infty < y < \infty, \quad -\infty < \mu < \infty, \quad \beta > 0.$$

A função de distribuição acumulada que é usada como modelo para crescimento é dada por

$$F(y; \mu, \beta) = \left[1 + e^{-\left(\frac{y-\mu}{\beta}\right)}\right]^{-1}$$

Tem-se que $E(y) = \mu$, $Var(y) = \frac{\pi^2\beta^2}{3}$ e $\gamma_2 = 4,2$, que é maior que o coeficiente de curtose das distribuições normal e logística I.

A distribuição logística II tem sido extensivamente utilizada em diferentes áreas tais como: Biologia, Epidemiologia, Psicologia, Tecnologia e Market (JOHNSON e KOTZ, 1970).

Diz-se que a variável aleatória $y \sim S(\mu, \phi)$ tem distribuição logística II se sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = \frac{e^{-u^{1/2}}}{(1 + e^{-u^{1/2}})^2} \quad u > 0,$$

denotada por $y \sim LII(\mu, \phi)$. A função característica é dada por

$$\psi_y(t) = \frac{2(e^{it\mu}\pi\phi^{1/2}t)}{(e^{\pi\phi^{1/2}t} - e^{-\pi\phi^{1/2}t})}, \quad t \in \mathbb{R}.$$

Uma relação bastante útil para gerar amostras aleatórias é dada por Hastings e Pea-

cock (1975). Seja $u \sim U(0, 1)$ e $y = \mu + \sqrt{\phi} \log\{u/(1-u)\}$, então $y \sim LII(\mu, \phi)$.

Seguem, algumas distribuições relacionadas com a distribuição logística II.

- 1) Se Y é a variável aleatória Logística II(μ, ϕ), então $\frac{(y-\mu)}{\beta} \sim LII(0, 1)$.
- 2) Se Y é a variável aleatória exponencial padrão, então

$$-\ln\left[\frac{e^{-y}}{1-e^{-y}}\right] \sim LII(0, 1).$$

- 2) Se Y_1 e Y_2 são variáveis aleatória exponenciais padrão independentes, então

$$-\ln\left(\frac{Y_1}{Y_2}\right) \sim LII(0, 1).$$

Na Figura 7, é apresentado o gráfico da função de densidade da distribuição logística II padrão comparando com a função de densidade da distribuição normal padrão.

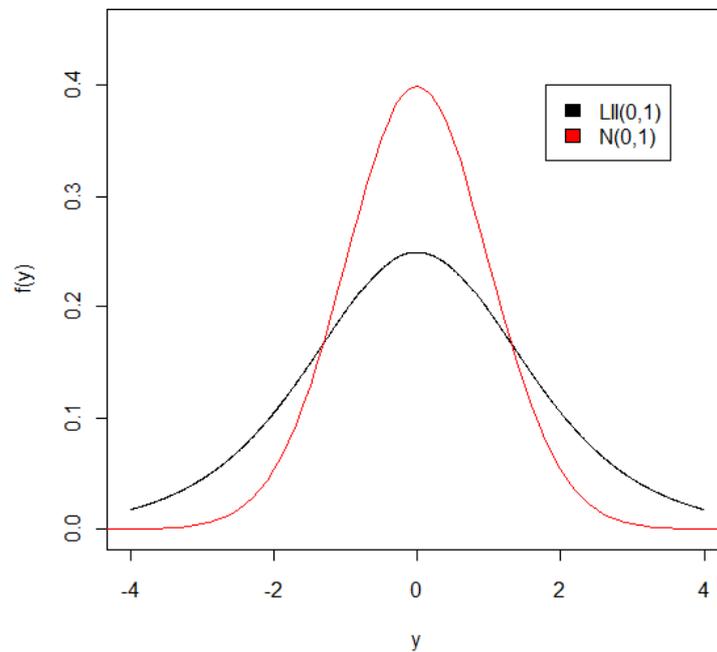


Figura 7: Densidades da distribuição de logística II padrão e distribuição normal padrão.

2.13 Distribuição de Kotz

A distribuição de Kotz, na classe de distribuições elípticas univariadas, pode ser interpretada como sendo gerada por uma distribuição de Weibull ou uma distribuição valor extremo tipo III. Essa distribuição tem caudas mais pesadas do que a distribuição normal e um fato interessante é que os estimadores de máxima verossimilhança são as medianas espaciais generalizadas (JOHNSON e KOTZ, 1970).

A distribuição de Kotz tem função de densidade de probabilidade dada por

$$f(y) = \frac{sr^{(2q-1)/2s}}{\Gamma((2q-1)/2s)} y^{2(q-1)} e^{-ry^2s}, \quad r, s > 0 \quad q > 1/2.$$

Uma variável aleatória $Y \sim S(\mu, \phi)$ tem distribuição de Kotz se sua função geradora de densidade $g(\cdot)$ é dada por

$$g(u) = \frac{r^{(2N-1)/2}}{\Gamma\left(\frac{2N-1}{2}\right)}, \quad r > 0, \quad N \geq 1, \quad u > 0,$$

e assim denotada por $y \sim K(\mu, \phi, N, r)$. Tem-se que $E(y) = \mu$ e $Var(y) = \{(2N-1)/2r\}\phi$ e $\gamma_2 = (2N+1)/(2N-1)$ e os momentos de ordem central $2m$ são dados por

$$\mu_{2m} = \frac{\Gamma\{(2N+2m-1)/2\}}{r^m \Gamma\{(2N-1)/2\}} \phi^m, \quad m > 0.$$

Se $N = 1$ a distribuição de Kotz se reduz a $N(\mu, \phi/2r)$. Se $N > 1$ a distribuição é bimodal com modas $Y = \mu \pm \sqrt{(N-1)/(r\phi)}$. Além disso, tem-se que, se $Z^2 = (y-\mu)^2/\phi$ então $Z^2 \sim G(\{2N-1\}/2r)$. Em um caso particular, quando $N = 1$ e $r = 1/2$, tem-se que $Z^2 \sim \chi_1^2$.

Na Figura 8, é possível observar os gráficos das densidades da distribuição de Kotz com diferentes valores dos parâmetros N r e da distribuição normal padrão.

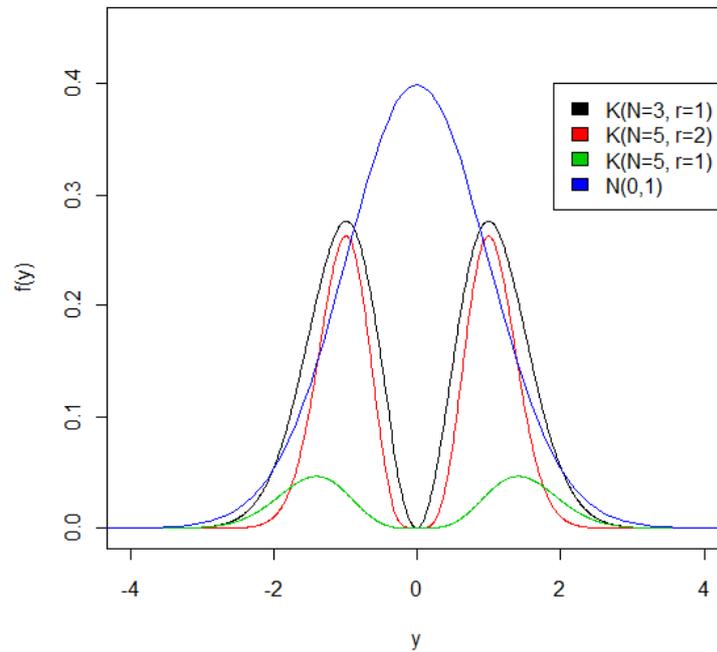


Figura 8: Densidades da distribuição Kotz com diferentes valores dos parâmetros N r e distribuição normal padrão.

2.14 Distribuição Exponencial Potência

A distribuição exponencial potência pode ser considerada como uma distribuição geral para erros aleatórios. A primeira formulação dessa distribuição pode ser atribuída para Subbotin em 1923. Subbotin obteve a função densidade de probabilidade dada por

$$f(\varepsilon) = \frac{mh}{2\Gamma(1/m)} e^{-h^m|\varepsilon|^m},$$

em que $-\infty < \varepsilon < \infty$, $h > 0$ e $m > 1$.

Seguindo o procedimento introduzido por Pearson em 1895, Lunetta (1963) derivou uma diferente parametrização da distribuição exponencial potência resolvendo a seguinte equação diferencial:

$$\frac{\partial \log f}{\partial y} = p \left(\frac{\partial f - \log a}{y - c} \right),$$

em que a e c são constantes

Assim, a função de densidade da distribuição exponencial potência é dada por:

$$f(y) = \frac{1}{2\sigma p^{1/2}\Gamma(1 + 1/p)} \exp\left(-\frac{|y - \mu|^p}{p\sigma^p}\right),$$

em que $-\infty < y < \infty$, $-\infty < \mu < \infty$, $\sigma > 0$ e $p > 0$.

É importante notar que, embora com uma diferente parametrização, a distribuição exponencial potência pode ser tratada também por métodos Bayesianos quando existe o problema para especificar uma distribuição a priori adequada (BOX e TIAO 1992, CHOY e SMITH 1997 e ACHAR e PEREIRA 1999). Uma distribuição exponencial potência bivariada foi introduzida por De Simone (1968) e Taguchi (1978), ao passo que uma formulação multivariada dessa distribuição pode ser encontrada em Fang *et al.* (1990) e Krzanowski e Marriott (1994).

Na estimação dos parâmetros da distribuição exponencial potência, assumindo que o parâmetro de forma p é conhecido, os parâmetros de locação e escala devem ser facilmente obtidos pelo uso do método de estimação de máxima verossimilhança. Entretanto, a estimação do parâmetro de forma é, em geral, um problema. Vários procedimentos interessantes têm sido propostos na literatura, sendo um deles baseado no método de máxima verossimilhança. Em geral, a derivação de estimadores de máxima verossimilhança não traz, formalmente, muitos problemas e os estimadores têm propriedades adequadas, pelo menos assintoticamente.

Uma variável aleatória $y \sim S(\mu, \phi)$ tem distribuição exponencial potência (BOX e TIAO, 1973) se sua função geradora de densidades $g(\cdot)$ é da forma

$$g(u) = C(l)\exp\left\{-\frac{1}{2}u^{1/(1+l)}\right\}, \quad -1 < l \leq 1, \quad u > 0,$$

em que $C(l)^{-1} = \Gamma(1 + \frac{1+l}{2})2^{1+(1+l)/2}$ e denota-se por $y \sim EP(\mu, \phi, k)$.

Tem-se ainda que:

$$E(y) = \mu, \quad Var(y) = 2^{(1+l)} \left[\frac{\Gamma(\frac{3(1+l)}{2})}{\Gamma(\frac{1+l}{2})} \right]$$

e

$$\gamma_2 = \frac{\Gamma[\frac{5}{2}(1+l)]\Gamma(\frac{1+l}{2})}{\Gamma^2[\frac{3}{2}(1+l)]}.$$

Observa-se que para $l > 0$, tem-se que $\gamma_2 > 3$, ou seja, a distribuição é leptocúrtica e para $l < 0$, tem-se $\gamma_2 < 3$, ou seja, a distribuição é platicúrtica. Pode-se observar o parâmetro l como uma medida de curtose, ou mesmo, uma medida de não normalidade pois quando $l = 0$ tem-se distribuição normal. Em particular, quando $l = 1$ tem-se a distribuição exponencial dupla. Se l tende a -1 , a distribuição tende a uma distribuição uniforme no intervalo $(\mu - \sqrt{3\phi}, \mu + \sqrt{3\phi})$

Se $Y = (2W)^{1/r}$ em que $V \sim U(-1, 1)$, $W \sim G(1 + 1/r, 1)$ e $r = 2/(1 + l)$ independentes (DEVROYE, 1986), então $Y \sim EP(0, 1, l)$. Essa relação é suficiente para gerar amostras de uma distribuição $EP(0, 1, l)$.

Na Figura 9 a seguir é apresentado o gráfico da função de densidade da distribuição exponencial potência segundo diferentes valores do parâmetro l comparado com a função de densidade da distribuição normal padrão.

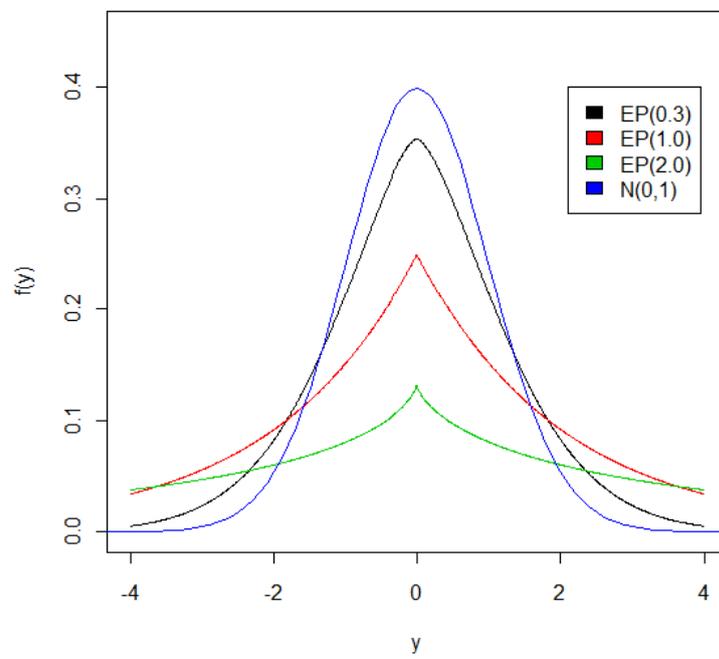


Figura 9: Densidades da distribuição exponencial potência segundo diferentes valores do parâmetro l e distribuição normal padrão.

3 Modelos Simétricos Transformados não lineares

Neste estudo se considera uma família paramétrica geral de transformações da variável resposta Y especificada por:

$$Y^{(\lambda)} = \Lambda(Y, \lambda), \quad (3.1)$$

em que λ é um parâmetro escalar definido para uma transformação particular. Admite-se que para cada λ , $Y^{(\lambda)}$ é uma função monotônica de Y . Usualmente, considera-se a transformação de potência de Box e Cox (1964), $Y^{(\lambda)} = (Y^\lambda - 1)/\lambda$ para tal $\lambda \neq 0$ e $Y^{(\lambda)} = \log(Y)$ para $\lambda = 0$. Geralmente, é assumido que existe um valor de λ para a variável resposta tal que $Y^{(\lambda)}$ segue um modelo de regressão não-linear $\mu = h(x_i, \beta)$ com erro normal e variância constante.

A transformação potência de Box–Cox tem gerado muito interesse em pesquisas teóricas e em aplicadas. Procedimentos de estimação para os coeficientes de regressão e para o parâmetro da transformação de Box–Cox foram estudados extensivamente. Claramente nem todos os dados podem ser transformados com uma transformação potência para produzir uma distribuição normal. Silva *et al.* (1994) utilizaram a família de transformações de Box-Cox comparando com os modelos volumétricos em florestas plantadas.

Assumindo-se que as variáveis aleatórias transformadas $Y_1^{(\lambda)}, \dots, Y_n^{(\lambda)}$ em $Y^{(\lambda)}$ são independentes e cada $Y_i^{(\lambda)}$ tem uma distribuição contínua simétrica com parâmetro de locação $\mu_i \in \mathbb{R}$ e parâmetro de dispersão $\phi > 0$ dada por

$$\pi(y_i^{(\lambda)}; \mu, \phi) = \frac{1}{\sqrt{\phi}} h \left\{ \left(\frac{y_i^{(\lambda)} - \mu_i}{\phi} \right)^2 \right\}, \quad y_i^{(\lambda)} \in \mathbb{R}, \quad (3.2)$$

em que a função $h(\cdot)$ (conhecida como a geradora de densidades) tal que $h(u) > 0$, para $u > 0$ e $\int_0^\infty u^{-1/2} f(u) du = 1$, em que $u = \frac{(y-\mu)^2}{\phi}$. Esta condição é necessária para que $\pi(y_i^{(\lambda)}; \mu, \phi)$ seja uma função de densidade de probabilidade da variável aleatória padronizada $Z^{(\lambda)} = \frac{y_i^{(\lambda)} - \mu_i}{\sqrt{\phi}}$ é $\pi(z, 0, 1) = f(z^2)$ $v \in \mathbb{R}$, isto é, $Z^{(\lambda)} \sim S(0, 1)$.

A função característica de $Y_i^{(\lambda)}$ é $\psi(t) = e^{it\mu_i} \varphi(t^2\phi)$, $t \in C$, para alguma função $\varphi(x) \in \mathbb{R}$ e $x > 0$. Mostrando que eles existem, $E(Y_i^{(\lambda)}) = \mu_i$ e $Var(Y_i^{(\lambda)}) = \kappa_2\phi$ em que $\kappa_2 = -2\varphi'(0)$ é uma constante, e $\varphi' = \varphi(x)/dx |_{x=0}$.

Para introduzir uma estrutura de regressão na classe de modelos (3.2), assume-se um componente sistemático para o vetor da média $\mu = E(Y^{(\lambda)})$ dado por

$$g(\mu) = \eta_i(\beta) = h(x_i, \beta) \quad (3.3)$$

em que $g(\cdot)$ é conhecida e duplamente diferenciável, $\eta_i(\beta)$ é o preditor não-linear, X é uma matriz $n \times p$ de posto completo e $\beta = (\beta_1, \dots, \beta_p)^T$ é um conjunto de parâmetros não-lineares desconhecidos a serem estimados.

Os modelos simétricos transformados não lineares assumem a existência de alguns valores de λ em (3.1) tal que as variáveis aleatórias transformadas $Y_1^{(\lambda)}, \dots, Y_n^{(\lambda)}$ podem ser tratadas como independentemente distribuídas seguindo a componente aleatória (3.2) e a componente sistemática (3.3).

A forma generalizada (3.1) é usada para determinar a forma específica dentre uma particular classe de funções de transformações que é otimizada em referência ao critério de máxima verossimilhança. O valor de λ é obtido para produzir um modelo simétrico ajustado mais aproximado para os dados transformados.

A família de densidades simétrica de locação-escala (3.2) retém a estrutura da distribuição normal, mas elimina a forma específica da densidade normal. Esta família inclui densidades simétricas que têm caldas menores ou mais longas que normais.

Sejam X a matriz do modelo, os dados originais $y = (y_1, \dots, y_n)^T$ e assumindo que as observações transformadas $y = (y_1^{(\lambda)}, \dots, y_n^{(\lambda)})^T$, para alguma transformação desconhecida do parâmetro λ em (3.1), satisfazem as suposições (3.2) e (3.3). O principal objetivo na análise de modelos transformados simétricos não lineares é fazer inferências no vetor de parâmetros β e nos escalares ϕ e λ .

Seja $J(\lambda, y)$ o Jacobiano da transformação de y para $y^{(\lambda)}$. A log-verossimilhança para

os parâmetros do modelo pode ser expressa como

$$\ell(\beta, \phi, \lambda) = -\frac{n}{2} \log \phi + \sum_{i=1}^n \log[f\{\phi^{-1}(y_i^{(\lambda)} - \mu_i)^2\}] + \sum_{i=1}^n \log\{J(\lambda, y_i)\} \quad (3.4)$$

em que

$$J(\lambda, y_i) = \left| \frac{d\Lambda(y_i, \lambda)}{dy_i} \right| \quad (3.5)$$

Os modelos simétricos potência com transformação de Box e Cox produzem $J(\lambda, y_i) = |y_i|^{(\lambda-1)}$. A log-verossimilhança apresentada pode ser maximizada incondicionalmente usando alguns software como o SAS, Matlab, R ou a linguagem de programação Ox.

Alternativamente, pode-se assumir λ fixo e assim obter as equações de log-verossimilhança condicionadas para estimar β e ϕ . Sejam $\hat{\beta}^{(\lambda)}$, $\hat{\eta}^{(\lambda)} = h(x_i, \hat{\beta}^{(\lambda)})$, $\hat{\mu}^{(\lambda)} = g^{-1}(\hat{\eta}^{(\lambda)})$ e $\hat{\phi}^{(\lambda)}$ as estimativas de máxima verossimilhança de β , η , μ e ϕ , respectivamente, para dado λ . As estimativas condicionais de $\hat{\beta}^{(\lambda)}$ e $\hat{\phi}^{(\lambda)}$ são obtidas facilmente por iteração reponderando o quadrado médio.

O modelo simétrico não-linear é definido por

$$y_i = \mu_i(\beta, x_i) + \varepsilon_i$$

em que $\mu_i = \mu_i(\beta, x_i)$ é uma função não-linear contínua diferenciável de $\beta = (\beta_1, \dots, \beta_p)^T$ tal que a matriz de derivadas $D_\beta = \frac{\partial \mu}{\partial \beta}$ tenha posto p ($p < n$) para todo β com $\mu = (\mu_1, \dots, \mu_n)^T$, $y = (y_1, \dots, y_n)^T$ é o vetor de respostas observadas, $x_i = (x_{i1}, \dots, x_{in})^T$ contém valores de p variáveis explanatórias e $\varepsilon_i \sim S(0, \phi)$. Seja $t(x) = \log\{g(x)\}$ e, sua primeira derivada $t'(x)$ e $D^{(\lambda)} = \text{diag}\{v_1, \dots, v_n\}$, $d_i^{(\lambda)} = -2t'(u_i^{(\lambda)}, u_i^{(\lambda)}) = \phi^{-1}(y_i^{(\lambda)} - \mu_i)^2$ é a raiz da i -ésima observação transformada padronizada $y_i^{(\lambda)}$, $Q(\mu^{(\lambda)}, y^{(\lambda)}) = (y^{(\lambda)} - \mu)^T D^{(\lambda)} (y^{(\lambda)} - \mu)$.

Para obter a função escore para o vetor de parâmetro β e parâmetro de dispersão ϕ para dado λ , precisamos derivar $\ell(\beta, \phi, \lambda)$ com respeito aos parâmetros desconhecidos. Supõe-se aqui que tais derivadas existem. As funções escore para β e ϕ tomam, respecti-

vamente, as formas

$$U_{\beta}^{(\lambda)} = \phi^{-1} D_{\beta}^T D^{(\lambda)} (y^{(\lambda)} - \mu) \quad (3.6)$$

e

$$U_{\phi}^{(\lambda)} = (2\phi)^{-1} \{ \phi^{-1} Q(\mu^{(\lambda)}, y^{(\lambda)}) - n \} \quad (3.7)$$

Condicionando em λ , pode-se observar facilmente que os parâmetros β e ϕ são globalmente ortogonais e então as estimativas de $\hat{\beta}^{(\lambda)}$ e $\hat{\phi}^{(\lambda)}$ são assintoticamente independentes devido a sua normalidade assintótica e a estrutura bloco diagonal da matriz de informação conjunta, isto é, $K = \text{diag}(K_{\beta}, \kappa_{\phi})$ em que $K_{\beta} = 4a\phi^{-1} D_{\beta}^T D_{\beta}$ e $\kappa_{\phi} = n(4b - 1)/(4\phi^2)$ são, respectivamente, as matrizes de informação para β e ϕ . As constantes a e b , independentes em μ e ϕ , podem ser obtidas da variável aleatória $Z^{(\lambda)} \sim S(0, 1)$, através de $a = E\{t'(Z^{(\lambda)^2})^2 Z^{(\lambda)^2}\}$ e $b = E\{t'(Z^{(\lambda)^2}) Z^{(\lambda)^4}\}$.

Os estimadores de máxima verossimilhança de β e ϕ são obtidos por meio da solução do sistema de equações dado por

$$U_{\beta}^{(\lambda)} = 0 \quad (3.8)$$

e

$$U_{\phi}^{(\lambda)} = 0 \quad (3.9)$$

As equações acima são não lineares e não podem ser resolvidas explicitamente. No entanto, podem ser resolvidas por meio de um método iterativo, como por exemplo o método escore de Fisher. Assim, o processo iterativo reduz para

$$\beta^{(\lambda)(m+1)} = \beta^{(m)} + (4a)^{-1} \{D_\beta^{T(m)} D_\beta^{(m)}\}^{-1} D_\beta^{T(m)} D^{(\lambda)(m)} \{y^{(\lambda)} - \mu^{(\lambda)(m)}\} \quad (3.10)$$

e

$$\phi^{(\lambda)(m+1)} = \frac{1}{n} Q(\mu^{(\lambda)(m+1)}, y^{(\lambda)}) \quad (m = 0, 1, 2, \dots) \quad (3.11)$$

Note-se que em (3.11) se tem sempre uma solução positiva para $\phi^{(\lambda)(m+1)}$. As expressões de $D^{(\lambda)}$, a e b são facilmente derivadas para muitas distribuições simétricas e podem ser vistas na tabela 1 adaptada de Cysneiros *et al.* (2005).

Tabela 1: Expressões para $D^{(\lambda)}$, a e b para algumas distribuições simétricas

Distribuição	$D^{(\lambda)}$	a	b
Normal	I	$\frac{1}{4}$	$\frac{3}{4}$
t -Student	$diag\left\{\frac{v+1}{v+u_i^{(\lambda)}}\right\}$	$\frac{(v+1)}{4(v+3)}$	$\frac{3(v+1)}{4(v+3)}$
t -Student generalizada	$diag\left\{\frac{(r+1)}{(s+u_i^{(\lambda)})}\right\}$	$\frac{r(r+1)}{4s(r+3)}$	$\frac{3(r+1)}{4(r+3)}$
Logística I	$2diag\left\{\tan h\left(\frac{u_i^{(\lambda)}}{2}\right)\right\}$	0,369310044	1,003445984
Logística II	$diag\left\{\frac{\exp(-\sqrt{u_i^{(\lambda)}})-1}{\sqrt{u_i^{(\lambda)}}(1+\exp(-\sqrt{u_i^{(\lambda)}}))}\right\}$	$\frac{1}{12}$	0,60749
Exponencial potência	$diag\left\{\frac{1}{(1+k)u_i^{k/(k+1)(\lambda)}}\right\}$	$\frac{\Gamma\{(3-k)/2\}}{4(2^{k-1})(1+k)^2\Gamma\{(k+1)/2\}}$	$\frac{(k+3)}{4(k+1)}$

Substituindo as estimativas de máxima verossimilhança de $\hat{\beta}^{(\lambda)}$ e $\hat{\phi}^{(\lambda)}$ em (3.4) obtém-se a log-verossimilhança perfilada para λ

$$\ell_p(\lambda) = -\frac{n}{2} \log \hat{\phi}^{(\lambda)} + \sum_{i=1}^n \log[f\{\hat{\phi}^{(\lambda)(-1)}(y_i^{(\lambda)} - \hat{\mu}_i^\lambda)^2\}] + \sum_{i=1}^n \log\{J(\lambda, y_i)\} \quad (3.12)$$

Seja $S_i^{(\lambda)2} = f\{\hat{\phi}^{(\lambda)(-1)}(y_i^{(\lambda)} - \hat{\mu}_i^\lambda)^2\}$ e definindo as médias geométricas

$$\tilde{S}^{(\lambda)2} = \left(\prod_{i=1}^n S_i^{(\lambda)2} \right)^{1/n} \quad (3.13)$$

e

$$\tilde{J}^{(\lambda)} = \left(\prod_{i=1}^n J(\lambda, y_i) \right)^{1/n} \quad (3.14)$$

Então, a log-verossimilhança perfilada para λ reduz-se para

$$\ell_p(\lambda) = n \log \left(\frac{\tilde{S}^{(\lambda)2} \tilde{J}^{(\lambda)}}{\sqrt{\hat{\phi}^{(\lambda)}}} \right) \quad (3.15)$$

A log-verossimilhança perfilada $\ell_p(\lambda)$ plotada versus λ para uma série de ensaios de valores determina numericamente o valor da estimativa de verossimilhança $\hat{\lambda}$. Para maximizar a log-verossimilhança perfilada só é necessário encontrar um valor de λ que minimize a seguinte taxa

$$\hat{\lambda} = \arg \min_{\lambda} \left(\frac{\sqrt{\hat{\phi}^{(\lambda)}}}{\tilde{S}^{(\lambda)2} \tilde{J}^{(\lambda)}} \right) \quad (3.16)$$

Para modelos simétricos potência, $\tilde{J}^\lambda = \tilde{y}^{\lambda-1}$, em que \tilde{y} é a média geométrica dos dados originais.

Uma vez que a estimativa de máxima verossimilhança $\hat{\lambda}$ é obtida, ela pode ser substituída dentro das equações (3.10) e (3.11) para produzir as estimativas irrestritas $\hat{\beta} = \hat{\beta}^{(\hat{\lambda})}$ e $\hat{\phi} = \hat{\phi}^{(\hat{\lambda})}$ e então a log-verossimilhança maximizada $\hat{\ell} = \ell_p(\hat{\lambda})$ segue da equação (3.16) para o modelo considerado.

Essencialmente faz-se inferências sobre β e ϕ condicionado em $\lambda = \hat{\lambda}$. Então, $\hat{\lambda}$ é considerada conhecida, e intervalos de confiança para os parâmetros β , η , μ e ϕ , testes

de hipóteses, resíduos e medidas de diagnóstico podem ser obtidas em um contexto similar de modelos lineares generalizados para valores ajustados $\hat{\beta}$, $\hat{\eta}$, $\hat{\mu}$ e $\hat{\phi}$. A matriz de variância-covariância assintótica de $\hat{\beta}$ e $\hat{\phi}$ dada por $(4a)^{-1}\hat{\phi}(D_{\hat{\beta}}D_{\hat{\beta}}^T)^{-1}$ e $(4\hat{\phi}^2)/n(4b-1)$, respectivamente, pode ser usada para obter intervalos normais assintóticos para esses parâmetros.

Para realizar testes de hipótese sobre o parâmetro de transformação da família definida em (3.1) é possível obter de (3.15) e (3.16) por meio da estatística $w = 2\{\ell_p(\hat{\lambda}) - \hat{\ell}_p(\lambda^{(0)})\}$ para testar $\lambda = \lambda^{(0)}$ a qual tem distribuição assintótica χ_1^2 . Baseado nesse teste, um intervalo de confiança para λ é facilmente obtido para grandes amostras. A precisão desta aproximação segue o fato que $P_r\{w \geq \chi_1^2(\alpha)\} = \alpha + O(n^{-1/2})$.

Procedimentos de identificação e comparação de modelos têm sido propostos na literatura. Uma nova perspectiva do problema de identificação é obtida pela análise do método de máxima verossimilhança. O fato de que sob certas condições de regularidade o estimador de máxima verossimilhança é assintoticamente eficiente mostra que a função de verossimilhança tende a ser um critério mais sensível a pequenos desvios dos parâmetros do modelo de seus valores verdadeiros. Baseado nessa idéia, Akaike (1972) apresentou um método de identificação de modelos. Em 1974, descreveu como o problema de seleção de modelos pode ser manuseado pelo uso do critério de informação introduzido em 1972 (AKAIKE, 1974). Esse critério de informação de Akaike (AIC) é uma estatística bem conhecida e de fácil interpretação para seleção de modelos de regressão. Desta forma, para comparar todos os modelos não transformados e transformados ajustados aos dados, pode-se usar o critério de informação de Akaike definido por:

$$AIC = -2\hat{\ell} + 2r \quad (3.17)$$

em que $\hat{\ell}$ é a log-verossimilhança maximizada, e $r = p + 1$ ou $r = p + 2$ para modelos não transformados e transformados, respectivamente e p é o número de parâmetros estimados. O modelo com menor valor do AIC , entre todos os modelos ajustados, pode ser considerado como o que melhor explica a variabilidade dos dados.

Para avaliar melhor os ajustes obtidos calculou-se o erro médio quadrático (EQM) e o erro percentual absoluto médio ($EPAM$) dados por:

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i^{(\hat{\lambda})} - \hat{\mu}_i)^2 \quad (3.18)$$

e

$$EPAM = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i^{(\hat{\lambda})} - \hat{\mu}_i}{y_i^{(\hat{\lambda})}} \right| \quad (3.19)$$

3.1 Análise gráfica dos resíduos

Conforme Schneider (1998) e Menon (2006) a análise gráfica dos resíduos, embora visual, possibilita a verificação do ajuste do modelo através da distribuição dos dados observados ou estimados ao longo da linha de regressão, mostrando sua tendenciosidade ou não, permitindo assim, decidir se a equação utilizada é realmente apropriada. Na área de Ciências Florestais um dos mais importantes elementos para avaliação de uma equação é a visualização dos seus resíduos percentuais especificado por:

$$E\% = \left(\frac{y_i^{(\hat{\lambda})} - \hat{\mu}_i}{y_i^{(\hat{\lambda})}} \times 100 \right)$$

De forma geral erros percentuais em torno de 20 % são considerados aceitáveis.

4 Material e Métodos

A região na qual o experimento está implantado possui as seguintes coordenadas geográficas de posição $07^{\circ}29'00\text{S}$ e $40^{\circ}36'00\text{W}$ e altitude de 816 metros. O clima regional predominante é do tipo BShw' de Koppen, quente e seco das baixas latitudes, com chuvas de verão. A presença da Chapada do Araripe confere uma diferenciação no clima regional, elevando a pluviometria, registrando cerca de 735 mm anuais (ALVES, 2007).

O experimento (Figura 10) foi implantado no início do mês de março de 2002, em uma área de 2,352 ha, com espaçamento de 3 x 2 m. Foram utilizados 15 tratamentos, sendo utilizados 7 híbridos de *Eucalyptus*, com 4 repetições, estabelecidos em um delineamento multivariado de medidas repetidas ao nível de 1% de probabilidade.

As parcelas são compostas por 49 plantas, sendo que a área útil, isto é, sem bordadura, é composta de 25 plantas (Figura 10).

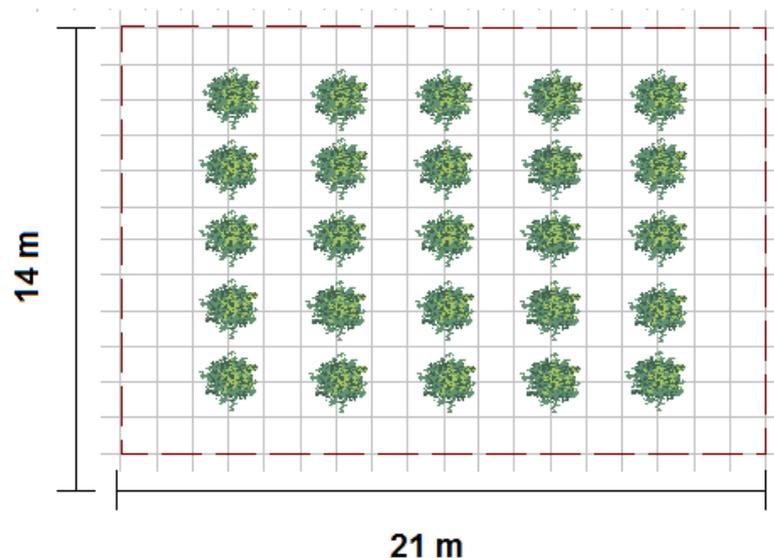


Figura 10: Planta baixa da parcela e área útil.

Foi deduzido um algoritmo iterativo para ajustar os Modelos Simétricos Transformados não lineares pelo método de máxima verossimilhança e discutida a inferência dos parâmetros nesses modelos. A inferência foi deduzida de forma análoga àquela proposta para os

modelos de Box - Cox conforme apresentada no Capítulo 3.

Na aplicação dessa nova classe de modelos foram utilizadas para este trabalho 56 árvores pertencentes ao clone 01 referente ao Híbrido de *Eucalyptus tereticornis* cujo os dados estão apresentados na Tabela 2.

4.1 Cubagem das árvores amostras

As árvores amostras foram derrubadas e seccionadas com motosserra, para se proceder à cubagem rigorosa das mesmas na área do povoamento.

Nas árvores derrubadas foram medidas a altura total, a altura do fuste e as circunferências a 0,30; 0,50; 0,70; 0,90; 1,10; 1,30; 1,50; 1,70; 2,30 m; e daí por diante de 1,00 m em 1,00 m até o final da árvore. As alturas foram medidas com auxílio da trena graduada em centímetros. As circunferências foram tomadas com fita métrica, graduada em centímetro com aproximação em milímetro, conforme o procedimento descrito por SILVA (1977).

Devido à facilidade dos cálculos e rapidez na medição das pequenas seções, a fórmula de Smalian (LOESTCH *et al.*, 1973), foi empregada, para calcular o volume rigoroso total e o volume das seções.

4.2 Estimação do Volume dos eucaliptos

A metodologia dos Modelos Simétricos Transformados Não Lineares foi aplicada na modelagem volumétrica em um clone de eucalipto proveniente do experimento realizado na Estação Experimental do Araripe, pertencente ao Instituto Agrônomo de Pernambuco (IPA) (Figura 11) no semi-árido de Pernambuco.

Os modelos foram ajustados supondo diferentes distribuições simétricas para os erros (normal, *t*-Student, Exponencial Potência, Cauchy, logística *I* e logística *II*).

Para os dados da variável volume em (m^3) a transformação de Box-Cox utilizada mais apropriada devido a natureza dos dados foi

$$y^{(\lambda)} = \begin{cases} y^\lambda & (\lambda \neq 0) \\ \log(y) & (\lambda = 0) \end{cases}$$

Para estimar os volumes das árvores foi utilizado o modelo de Schumacher-Hall de

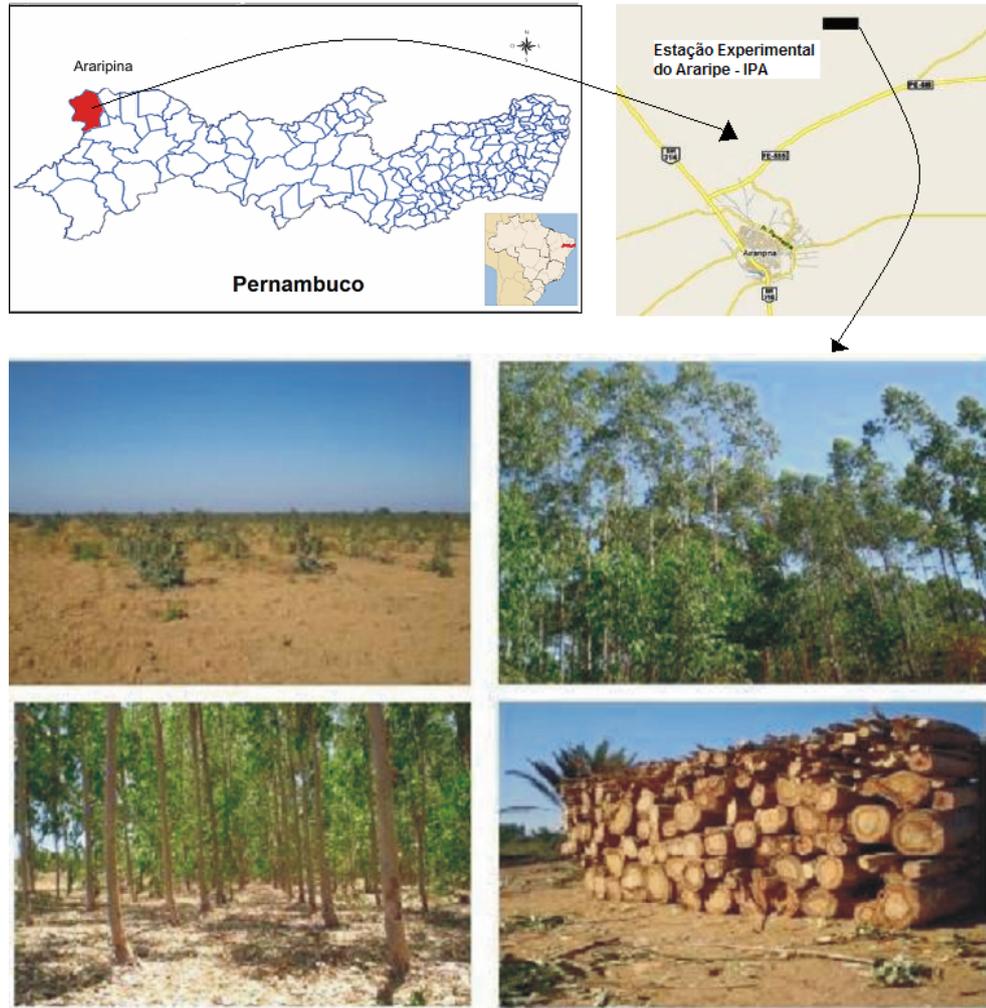


Figura 11: Área experimental, de híbridos de *Eucalyptus spp.* , localizada na Estação Experimental do Araripe (IPA), em Araripina - PE.

1933 especificado por:

$$vol_i^{(\lambda)} = \beta_0 \cdot DAP_i^{\beta_1} \cdot Ht_i^{\beta_2} \cdot \epsilon_i$$

em que λ é o parâmetro da transformação de Box-Cox, β_0 , β_1 e β_2 são os parâmetros a serem estimados, vol representa o volume da árvore em m^3 , DAP representa o diâmetro à altura do peito em cm (medida a 1,30m) e Ht representa a altura total da árvore em m e ϵ_i é o erro aleatório.

Os modelos foram ajustados supondo diferentes distribuições simétricas para os erros (normal, t de Student, Cauchy, exponencial potência, logística I e logística II). Segue abaixo a estrutura dos erros considerando a distribuição t de Student

$$g(u) = \frac{v^{v/2}}{B(1/2, v/2)} (v + u)^{-\frac{v+1}{2}}$$

em que $B(., .)$ é a função Beta. Para demais distribuições da estrutura dos erros ver Capítulo 3.

Em seguida a estrutura dos erros das distribuições simétricas citadas acima foi incluída na log-verossimilhança descrita na equação (3.4) no Capítulo 3. Assim, os parâmetros foram estimados pelo método de maximização da log-verossimilhança

Inicialmente foram calculadas as estimativas dos parâmetros através dos modelos simétricos transformados e não transformados supondo diferentes distribuições simétricas. O parâmetro λ da transformação de Box-Cox foi estimado conforme a verossimilhança perfilada conforme descrita na equação (3.12), Capítulo 3.

Os parâmetros foram estimados utilizando os softwares R e SAS. O software R é interpretado como uma linguagem computacional designada para análise de dados estatísticos que se caracteriza pelo compromisso entre a flexibilidade oferecida pelas linguagens compiladas, tais como C e C++, e a conveniência de softwares estatísticos tradicionais. Inclui uma ampla variedade de métodos estatísticos tradicionais e modernos, com a vantagem de ser gratuito. O SAS é um sistema integrado de aplicações para análise de dados que consiste de: recuperação de dados, gerenciamento de arquivos, análise estatística, acesso a banco de dados e geração de gráficos.

Tabela 2: Dados sobre Volume, *DAP* e Altura total *Ht* do clone Híbrido de *Eucalyptus tereticornis* com 7,5 anos oriundos do Pólo Gesseiro do Araripe-PE .

Volume	<i>DAP</i>	<i>Ht</i>
0.1944	16.01	17.20
0.1625	15.28	16.70
0.1600	14.64	16.46
0.1538	14.01	17.55
0.1526	15.92	13.29
0.1471	13.53	16.80
0.1315	13.69	15.40
0.1298	13.69	14.37
0.1242	13.69	14.98
0.0217	6.68	8.62
0.1168	12.57	15.19
0.0267	7.00	9.79
0.1093	12.57	14.20
0.0330	7.80	10.10
0.1090	12.73	14.91
0.1086	12.73	14.17
0.0338	7.96	11.08
0.0353	7.48	12.33
0.0355	7.48	11.65
0.0357	7.96	11.00
0.0360	8.59	12.50
0.0361	7.64	11.54
0.1055	13.05	12.93
0.0372	8.59	9.85
0.1048	13.05	14.28
0.1046	11.62	13.90
0.0384	7.96	10.85
0.0390	8.12	11.05
0.1013	13.05	13.70
0.0413	8.28	11.10
0.0419	8.12	12.17
0.0423	8.59	9.82
0.0430	8.28	11.16
0.0433	8.12	11.85
0.0439	8.59	11.12
0.0442	8.59	10.88
0.0460	9.23	10.90
0.0465	8.75	12.20
0.0957	12.10	14.55
0.0954	12.25	13.46
0.0471	8.59	11.90
0.0473	8.91	11.15
0.0937	11.30	14.22
0.0491	9.39	11.18
0.0491	9.23	10.46
0.0492	9.55	11.67
0.0496	9.23	10.40
0.0924	11.94	14.75
0.0504	8.75	12.30
0.0916	11.46	14.90
0.0506	9.07	11.53
0.0508	9.23	11.56
0.0513	10.03	11.40
0.0908	11.62	13.57
0.0526	10.03	11.18
0.0528	9.07	11.60

5 Resultados e Discussão

As estimativas dos parâmetros por meio dos modelos simétricos não transformados ($\lambda = 1$) não lineares se encontram na Tabela 3 bem como a estimativa do parâmetro de dispersão ϕ condicionado em λ fixo e seus respectivos erros-padrões. As estimativas não apresentaram parâmetros não significativos na modelagem dos modelos simétricos não transformados.

Tabela 3: Estimativas dos parâmetros e (erros padrões) dos modelos simétricos não transformados.

Distribuição	λ	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}$
Normal	1,000000	0,000109 (0,000014)	1,770904 (0,061631)	0,901799 (0,083267)	0,000019 (0,000004)
Student t_2	1,000000	0,000113 (0,000010)	1,793792 (0,035118)	0,867191 (0,050285)	0,000006 (0,000002)
Exp. Potência (0, 5)	1,000000	0,000114 (0,000010)	1,796848 (0,030194)	0,861188 (0,049069)	0,000007 (0,000002)
Cauchy	1,000000	0,000115 (0,000009)	1,791883 (0,028475)	0,863092 (0,043244)	0,000004 (0,000001)
Logística I	1,000000	0,000104 (0,000016)	1,739356 (0,063115)	0,946841 (0,080446)	0,000027 (0,000004)
Logística II	1,000000	0,000110 (0,000013)	1,783865 (0,051036)	0,886091 (0,070647)	0,000005 (0,000001)

Na Tabela 4, encontram-se as estimativas dos parâmetros dos modelos simétricos transformados não-lineares, bem como do parâmetro de dispersão ϕ condicionado em λ estimado e seus respectivos erros-padrões. O parâmetro λ da transformação de Box-Cox foi estimado usando a verossimilhança perfilada. Assim como nas estimativas dos parâmetros dos modelos não-transformados, não houveram estimativas não-significativas.

Tabela 4: Estimativas dos parâmetros e (erros padrões) dos modelos simétricos transformados.

Distribuição	$\hat{\lambda}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}$
Normal	0,261653 (0,005842)	0,089971 (0,003478)	0,469752 (0,024329)	0,236445 (0,024722)	0,000064 (0,000012)
Student t_2	0,377258 (0,006476)	0,031938 (0,001304)	0,673226 (0,025671)	0,334996 (0,026328)	0,000027 (0,000008)
Exp. Potência (0, 5)	0,653398 (0,011022)	0,002523 (0,000194)	1,171680 (0,048876)	0,582250 (0,048762)	0,000018 (0,000004)
Cauchy	0,403273 (0,007193)	0,025371 (0,001367)	0,720387 (0,021871)	0,355926 (0,021444)	0,000017 (0,000006)
Logística I	0,348162 (0,009353)	0,039636 (0,002730)	0,639566 (0,040220)	0,310258 (0,039824)	0,000096 (0,000016)
Logística II	0,418490 (0,008910)	0,021468 (0,001347)	0,739828 (0,033674)	0,386234 (0,033898)	0,000019 (0,000004)

Para escolher o modelo que melhor se ajustou aos dados, apresenta-se na Tabela 5 a log-verossimilhança maximizada $\hat{\ell}$ e o critério de informação de Akaike (AIC), para os modelos transformados e não-transformados ($\lambda = 1$). O modelo supondo erro com distribuição t -Student transformado produziu maior valor para a log-verossimilhança $\hat{\ell}$ (362,21) e um menor valor do AIC (-714,43) comparado com os demais modelos simétricos, sendo assim, o modelo mais indicado para representar os dados de acordo com esses critérios. É importante notar que os modelos transformados produziram maiores valores para a log-verossimilhança maximizada e menores valores do AIC em relação aos modelos não transformados, confirmando ainda mais a utilidade desta nova classe de modelos segundo esses critérios.

Tabela 5: Estatísticas para seleção dos modelos.

Distribuição	λ	$\hat{\ell}$	AIC
Normal	1,000000	275,86	-543,72
	0,261653	354,80	-699,61
Student t_2	1,000000	304,56	-601,61
	0,377258	362,21	-714,43
Exp. Potência (0, 5)	1,000000	227,28	-446,56
	0,653398	254,12	-498,24
Cauchy	1,000000	288,04	-568,09
	0,403273	341,31	-672,62
Logística I	1,000000	219,95	-431,90
	0,348162	284,12	-558,25
Logística II	1,000000	226,51	-445,03
	0,418490	280,25	-550,51

Na Tabela 6 encontram-se os valores das estatísticas EQM e EPAM. De acordo com os resultados obtidos o modelo t_2 transformado continua sendo o mais adequado, pois apresenta menores valores para os dois critérios considerados. É possível observar que em comparação com os modelos não-transformados, os modelos transformados apresentam menores valores para esses critérios.

Tabela 6: Comparação dos erros para os modelos ajustados.

Distribuição	λ	EQM	EPAM
Normal	1,000000	0,001939	0,773123
	0,261653	0,000674	0,327349
Student t_2	1,000000	0,001871	0,935777
	0,377258	0,000535	0,004994
Exp. Potência (0, 5)	1,000000	0,002012	1,008951
	0,653398	0,000588	0,067221
Cauchy	1,000000	0,002163	1,215905
	0,403273	0,000703	0,282932
Logística I	1,000000	0,002182	0,149070
	0,348162	0,000756	0,522813
Logística II	1,000000	0,002559	0,718893
	0,418490	0,000692	0,075364

De acordo com os critérios analisados, verificou-se que o modelo com erros t_2 transformado foi o modelo mais indicado para representar os dados de volume de clones de eucaliptos.

Para verificar se o modelo t_2 transformado é razoável para ajustar os dados de volume de eucaliptos, apresenta-se na Figura 12 o gráfico de dispersão dos valores observados versus médias ajustadas. Observa-se através da figura que os pontos se apresentam de forma linear, indicando que os dados foram bem ajustados. Além disso, na Figura 13 os resíduos de Pearson apresentaram uma distribuição aleatória quando feita a sua dispersão versus os valores ajustados, isto é, eles não contêm nenhum padrão evidente, indicando assim, que os resíduos são não correlacionados, ou seja, a hipótese de independência e variância constante para os resíduos são aceitas.

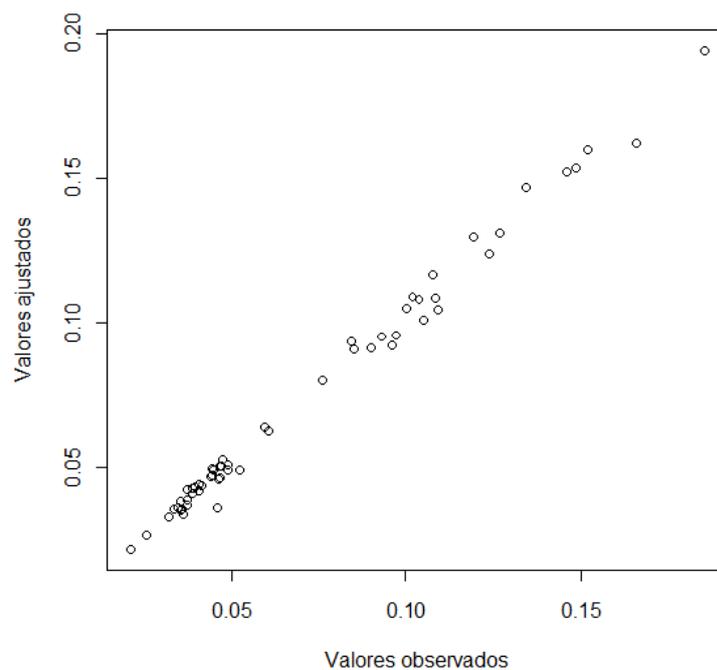


Figura 12: Gráfico de dispersão dos valores observados versus valores ajustados modelo t_2 transformado.

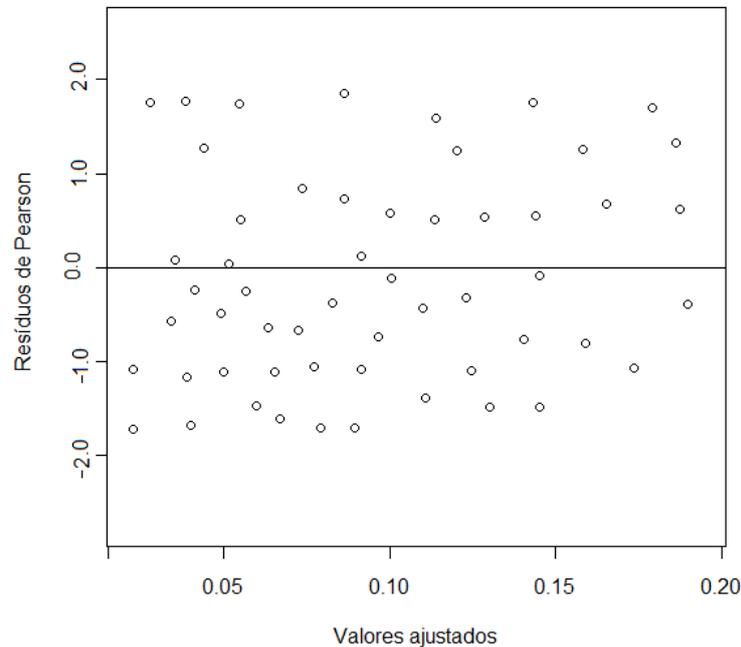


Figura 13: Gráfico de dispersão dos Resíduos de Pearson versus valores ajustados modelo t_2 transformado.

Para construir o gráfico de probabilidade normal dos resíduos plotou-se no eixo horizontal o resíduo ordenado e no eixo vertical plotamos a freqüência cumulativa, em uma escala de probabilidade normal. O gráfico de probabilidade meio-normal com limite de confiança (envelope simulado) é construído baseado nos resíduos padronizados. O envelope simulado é a banda de confiança. A ocorrência de pontos fora da banda de confiança indica que o modelo não está apropriado. Na Figura 14 são amostrados que todos os pontos se encontram dentro do limite de confiança, indicando o bom ajuste dos dados. Não observamos pontos muito fora do alinhamento. Por conseguinte, não há indicação de que a consideração de normalidade dos resíduos deva ser rejeitada.

Na Figura 15 são apresentados os erros percentuais versus o índice das observações para os ajustes através da distribuição t_2 transformada. Verifica-se que os erros percentuais estão mais próximos de zero e o maior valor apresentado foi de aproximadamente -11%.

Na prática, é de interesse que as estimativas estejam na escala original. Na Figura 16 é apresentado os erros percentuais versus o índice das observações para os ajustes através da distribuição t_2 transformada na escala original, ou seja, para a transformação

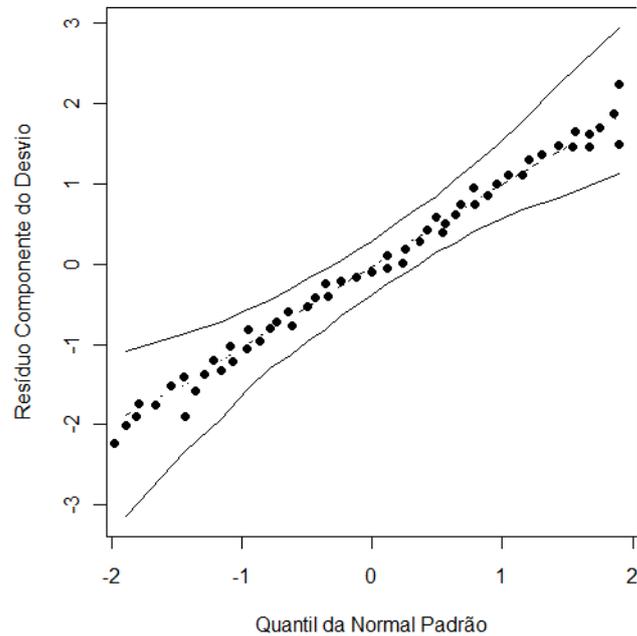


Figura 14: Gráfico normal de probabilidades com limite de confiança (envelope de simulação) modelo t_2 transformado.

utilizada $Y^{(\lambda)} = Y^\lambda$ retorna-se para a escala original $Y = Y^{1/\lambda}$. Verifica-se que os erros percentuais estão próximos de zero porém, encontram-se mais afastados e dispersos se comparando com o gráfico na escala transformada. Nota-se na figura que o maior valor apresentado foi de aproximadamente -33% .

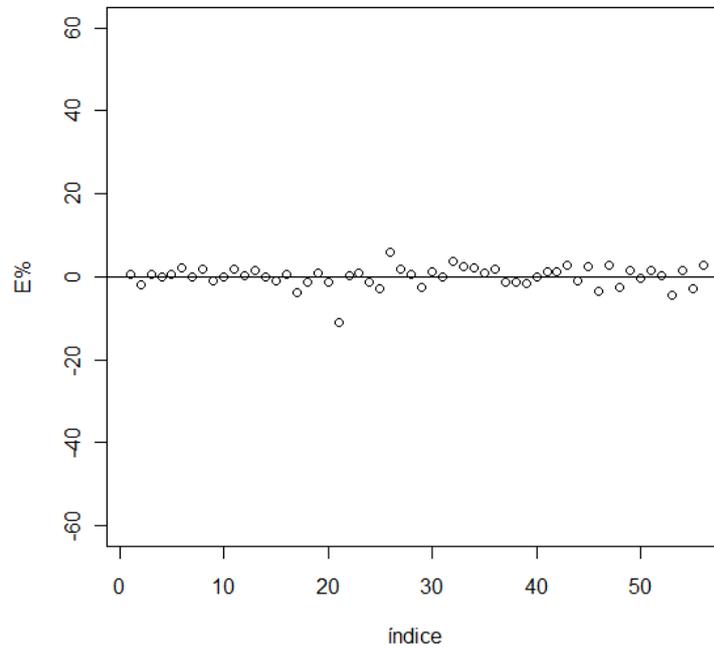


Figura 15: Erros percentuais para o modelo t_2 transformado.

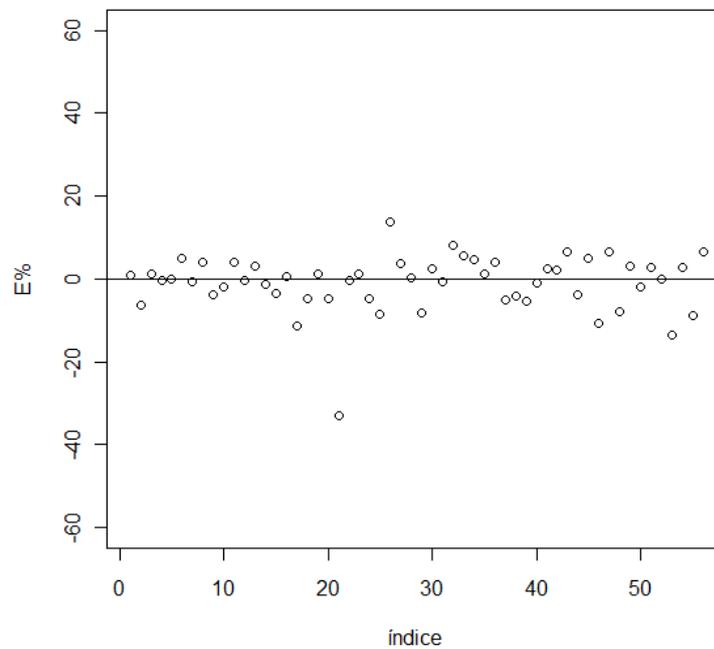


Figura 16: Erros percentuais para o modelo t_2 transformado na escala original.

A partir das estimativas dos parâmetros do modelo com erros t_2 transformado representado pelo modelo não-linear de Schumacher-Hall, pode-se agora, escrever o modelo

ajustado com parâmetro de dispersão $\phi = 0,000027$.

$$\hat{vol}_i^{(0,377258)} = 0,031938 \cdot DAP_i^{0,673226} \cdot Ht_i^{0,334996}.$$

Logo,

$$\hat{vol}_i = 0,000108 \cdot DAP_i^{1,784524} \cdot Ht_i^{0,887986}. \quad (5.1)$$

Dentre os modelos simétricos transformados não lineares testados o modelo com erros *t*-Student com 2 graus de liberdade representado pelo modelo não-linear de Schumacher-Hall mostrado na equação (5.1) foi o que mostrou maior confiabilidade estatística e apresentou melhor ajuste para o volume de Híbrido de *Eucalyptus tereticornis* no Pólo Gesseiro do Araripe. Vale ressaltar que os demais modelos com erros simétricos se ajustaram satisfatoriamente aos dados. O modelo apresentado em (5.1) permite inferir o volume individual do clone utilizado neste estudo a partir dos valores do diâmetro à altura do peito (*DAP*) e altura total (*Ht*).

A utilização de características de fácil mensuração como *DAP* e *Ht* para predição da resposta de uma característica de mensuração complexa como o volume (*vol*), é uma das principais vantagens da utilização das equações volumétricas para prognose da produtividade.

De acordo com os resultados apresentados, observa-se que a utilização dos modelos simétricos transformados não lineares aumentou consideravelmente a precisão comparado aos modelos não transformados, ou seja, um melhor ajuste foi dado por modelos que utilizaram a transformação dos dados. Silva *et al.* (1994) utilizaram a transformação de Box-Cox na estimativa volumétrica de *Pinus elliotti* através da comparação do Modelo de Schumachere e Hall e o da variável Combinada (Spurr) e comprovaram o aumento significativo da precisão do modelo.

Guimarães e Leite (1996) utilizaram o modelo de Schumacher e Hall para determinar o volume de *Eucalyptus Grandis* e concluíram que equações baseadas em 150 árvores resultaria em estimativas similares às obtidas pela equação original.

Tomé *et al.* (2007) testaram várias equações para estimação do volume para *Eucalyptus Globulus Labil* e constataram que o mdelo de Schumacher e Hall apresentou melhores

qualidades de ajustamento.

A vantagem da utilização dessa nova classe de modelos é a maior adequação e a flexibilidade em face ao modelo normal pois permitem ajustar uma ampla variedade de modelos para diversos tipos de dados, além do mais, buscar outras alternativas à suposição de erros normalmente distribuídos torna-se importante pois nem sempre a modelagem se ajusta adequadamente aos dados.

Vale ressaltar que os Modelos Simétricos Transformados Não Lineares possui uma limitação quanto ao ajuste de dados com um número n de observações grande, pois a medida que $n \rightarrow \infty$ a distribuição dos erros converge assintoticamente para a distribuição normal. Pode-se considerar que os Modelos Simétricos Transformados não lineares são mais adequados para ajustar dados com o número de observações relativamente pequeno.

6 Conclusões

A nova classe de Modelos Simétricos Transformados Não Lineares, introduzida neste estudo mostrou-se uma ferramenta bastante eficaz para modelagem do volume de Híbrido de *Eucalyptus Tereticornis* devido a flexibilidade da distribuição dos erros. Para as condições em que se realizou este estudo, pode-se concluir que o modelo t_2 transformado se ajustou aos dados mais adequadamente utilizando o modelo não linear de Schumacher e Hall, quando comparado aos demais modelos simétricos, diante dos critérios estabelecidos para escolha do melhor ajuste. Entretanto, os demais modelos também se ajustaram satisfatoriamente aos dados experimentais.

Assim, as possíveis correções volumétricas deverão ser estudadas para cada caso particular de aplicação, permanecendo, portanto, o problema em aberto para futuras pesquisas. Entretanto, para a finalidade de produção energética, foi apresentado um estudo da volumetria em Híbrido de *Eucalyptus Tereticornis* que servirá de suporte em estudos do volume de eucaliptos na região do Pólo Gesseiro do Araripe.

Espera-se que este estudo possa contribuir para definição de estratégias visando melhor aproveitamento, em termos de produção e qualidade, das fontes de energia utilizadas no processo de fabricação do gesso, em virtude da escassez cada vez maior desses recursos energéticos.

Espera-se também que este estudo sirva de modelo no oferecimento de informações mais precisas para o planejamento energético, economicamente eficiente, que atenda às necessidades da indústria, que preserve o meio ambiente, bem como contribuindo para a promoção do desenvolvimento regional no Pólo Gesseiro do Araripe.

Referências Bibliográficas

ACHAR, J. A.; PEREIRA, G. A. Use of exponential power distribution for mixture models in the presence of covariates. *Journal of Applied Statistics*, v. 26, n. 6, p. 669-679, 1999.

AKAIKE, H. Use of an information theoretic quantity for statistical model identification. *Proceedings of 5th Hawaii International Conference on System Sciences*, p. 249-250, 1972.

AKAIKE, H. A new look at the statistical model identification. *IEEE Trans. Automatic Control*, v. 19, p. 716-723, 1974.

ALBUQUERQUE, J. de L. Diagnóstico ambiental e questões estratégicas: uma análise considerando o Pólo Gesseiro do Sertão do Araripe - Estado de Pernambuco. 185p. Tese (Doutorado em Engenharia Florestal) - Universidade Federal do Paraná, Curitiba, 2002.

ALVES, A. M. C. Quantificação da produção de biomassa e do teor de carbono fixado por clones de eucalipto, no pólo gesseiro do Araripe-PE. 62f. Dissertação (Mestrado em Ciências Florestais) - Universidade Federal Rural de Pernambuco, Recife, 2007.

ANDERSON, T.W.; FANG, K.T. Cochran's theorem for elliptically contoured distributions. *Sankhya*, v. 49, p. 305-315, 1987.

ANDRADE, E. N. O eucalipto. 2 ed., Juniaí, Cia Paulista de Estradas de Ferro, 681p, 1961.

ARELLANO-VALLE, R.B. Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão. 112p. 1994. Tese (Doutorado em Estatística) - Universidade de São Paulo, Brazil.

ASSIS, A.L. *et al.* Avaliação de modelos polinomiais não-segmentados na estimativa e diâmetros e volumes comerciais de *Pinus taeda*. *Ciência Florestal*, Santa Maria, v. 12, n. 1, p. 89-107, 2004.

AVERY, T. E.; BURKHART, H. E. *Forest measurements*. 3.ed. New York: McGraw-Hill, 1983. 331 p.

BAENA, E.S. A rentabilidade econômica da cultura do *Eucalyptus* e sua contribuição ao agronegócio brasileiro. *Conhecimento Interativo*, São José dos Pinhais, v.1, n.1, p. 3-9, jul./dez, 2005.

BATES, D.M. e WATTS, D.G. Relative curvature measures of non-linearity. *J. R. Statist. Soc. B*, 42, 1-25, 1988.

BERGER, R. *et al.* Efeito do espaçamento e da fertilização sobre o crescimento e qualidade da madeira de um clone de *Eucalyptus saligna* Smith. In: SIMPÓSIO LATINO-AMERICANO SOBRE MANEJO FLORESTAL, 2000, Santa Maria. *Anais...Santa Maria: UFSM*, 2000. p. 209 236.

BERKANE, M.; BENTLER, P.M. Moments of elliptical distributed random variates. *Statistics and Probability Letters*, v. 4, p. 333-335, 1986.

BLATTBERG, R.C.; GONEDDES, N.J. A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business*, v. 47, p. 244-280, 1974.

BOX, G. E. P., COX, D.R. An analysis of transformation. *Journal of the Royal Statistical Society B*, 26, 211-252, 1964.

BOX, M.J.; TIAO, G.C. *Bayesian inference in statistical analysis*. London: Addison-Wesley, 1973.

BOX, M.J.; TIAO, G.C. *Bayesian Inference in Statistical Analysis*. New York: John Wiley, 1992.

BRASIL. Ministério de Minas e Energia. Balanço mineral brasileiro, Brasília, (2001). Disponível em <http://www.dnpm.gov.br.br/>. Acesso em 08 jan. 2008.

BUTLER, R.J.; MCDONALD, J.B.; NELSON, R.D.; WHITE, S.B. Robust and partially adaptive estimation of regression models. *Review of Economics and Statistics*, v. 72, p. 321-326, 1990.

CAMBANIS, S.; HUANG, S.; SIMONS, G. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, v. 11, p. 368-385, 1981.

CHMIELEWSKI, M. A. Elliptically symmetric distributions: a review and bibliography. *International Statistical Review*, 49, 67-74, 1981.

CHOY, S.T.B.; SMITH, A.F.M. On Robust Analysis of a Normal Location Parameter. *Journal of the Royal Statistical Society. Series B*, v. 59, p. 463-474, 1997.

CLUTTER, J.L. *et al.* Timber management: a quantitative approach. New York: John Wiley & Sons, 1983. 333 p.

CORDEIRO, G.M.; ANDRADE, M.G. Transformed Generalized Linear Models, Nota Técnica, Série Estatística, ICMC/USP, São Carlos, SP, 2007a.

CORDEIRO, G.M.; ANDRADE, M.G. Transformed Generalized ARMA Models, Nota Técnica, Série Estatística, ICMC/USP, São Carlos/SP, 2007b.

CORDEIRO, G.M.; ANDRADE, M.G. Transformed symmetric models, *Statistical Modelling*, (artigo submetido), 2009.

CORDEIRO, G.M. Corrected LR tests in symmetric nonlinear regression models. *Journal of Statistical Computation and Simulation*, v. 74, n. 5, p. 609-620, 2004.

CORDEIRO, G.M.; *et al.* Corrected maximum likelihood estimation in a class of symmetric nonlinear regression models. *Statistics and Probability Letters*, v. 46, p. 317-328, 2000.

CORDEIRO, G. M. ; LIMA NETO, E. A. Modelos Paramétricos. São Paulo: Associação

ciação Brasileira de Estatística, 2004. 246 p.

COUTO, H. T. Z.; BASTOS, N. L. M. Modelos de equações de volume e relações hipsométricas para plantações de *eucalyptus* no Estado de São Paulo. IPEF, n.37, p.33-44, 1987.

COX, D.R. e SNELL, E.J. A general definition of residual (with discussion). J. R. Statist. Soc. B, 30, 248-275, 1968.

CYSNEIROS, F. J. A.; PAULA, G. A.; GALEA, M. Modelos Simétricos Aplicados. 9 Escola de Modelos de Regressão, Águas de São Pedro, 2005.

DAVIS, L.S.; JOHNSON, K.N. Forest management. 3. ed. New York: McGraw-Hill inc, 1987. 789p.

DEVROYE, L. Non-uniform random variable generator. New York: Springer-Verlag, 1986.

DICKEY, J.M. Multivariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. Annals of Mathematical Statistics, v. 38, p. 511-518, 1967.

DRAPER, N.R.; SMITH, H. Applied regression analysis. 407p. John Wiley, New York, 1981.

FANG, K. T.; ANDERSON, T. W. Statistical Inference in Elliptical Contoured and Related Distributions. New York: Allerton Press, 1990.

FANG, K. T. e ZHANG, Y. T. Generalized Multivariate Analysis. New York: Springer-Verlag, 1990.

FANG, K. T.; *et al.* Symmetric Multivariate and Related Distributions. London: Chapman and Hall, 1990.

FAO. State of the World's Forests. Roma, 2009.

FERRARI, S.L.P.; URIBE-OPAZO, M.A. Corrected likelihood ratio tests in a class of symmetric linear regression models. *Brazilian Journal of Probability and Statistics*, v. 15, p. 49-67, 2001.

Fundação de Pesquisas Florestais do Paraná - FUPEF. Apoio técnico e institucional para o desenvolvimento do programa florestal da chapada do Araripe em Pernambuco, Curitiba/PR, 2007.

GALEA, M.; PAULA, G.A.; CYSNEIROS, F.J.A. On Diagnostic in Symmetrical Nonlinear Models. *Statistics and Probability Letters*. V. 73, n. 4, p. 459-467, 2005.

GALVÃO, A. P. M. Reflorestamento de propriedade rurais para fins produtivos e ambientais: um guia para ações municipais e regionais. Brasília: EMBRAPA-CTT, 2000. 351p.

GEOPHOTO - GEOPROCESSAMENTO E SENSORIAMENTO REMOTO. Pólo Geográfico de Pernambuco: diagnóstico e perspectivas de utilização dos energéticos florestais na região do Araripe. Fortaleza-CE, 2005. 15p.

GOMES, A.M.A. Medição dos arvoredos. Lisboa: Livraria Sá da Costa, 1957. 413p.

GUIMARÃES, D. P. Desenvolvimento de um modelo de distribuição diamétrica de passo invariante para prognose e projeção da estrutura de povoamentos de eucalipto. 1994. 160 f. Tese (Doutorado em Ciência Florestal) Universidade Federal de Viçosa, Viçosa, 1994.

GUIMARÃES, D. P.; LEITE, H. G. Influência do número de árvores na determinação de equação volumétrica para *Eucalyptus Grandis*. *Scientia Forestalis*. n. 50, p. 37-42, dez., 1996.

GUPTA, A. K. e VARGA, T. *Elliptically Contoured Models in Statistics*. Kluwer Academic Publishers, 1993.

HASTINGS, N.A.J.; PEACOCK, J.B. *Statistical distributions*. New York: John Wiley, 1975.

HUANG, S.; PRICE, D.; TITUS, S. J. Development of ecoregion-based height-diameter models for white spruce in boreal forests. *Forest Ecology and Management*, v. 129, p. 125-141, 2000.

HUSCH, B.; MILLER, C. I.; BEERS, T. W. *Forest mensuration*. 2.ed. New York: Ronald Press, 1972. 409 p.

JOHNSON, R.; KOTZ, S. *Continuous univariate distributions*. V.2. Boston: Houghton Mifflin, 1970.

KELLER, D. Distribution theory of spherical distributions and a location scale parameter generalization. *Sankhya*, v. 32, p. 419-430, 1970.

KRZANOWSKI, W.J.; MARRIOTT, F.H.C. *Multivariate Analysis*. London: Edward Arnold, 1994.

LANGE, K.L.; LITTLE, R.J.A.; TAYLOR, J.M.G. Robust statistical modeling using the *t* distribution. *Journal of the American Statistical Association*, v. 84, p. 881-896, 1989.

LEITE, H.G.; ANDRADE, V.C.L. Um método para condução de inventários florestais sem o uso de equações volumétricas. *Revista Árvore*, Viçosa-MG, v.26, n.3, p.321-328, 2002.

LIMA, W.P. *Impacto ambiental do eucalipto*. 2a Edição, São Paulo: EDUSP, p. 301, 1993.

LITTLE, R.J.A. Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics*, v. 37, p. 23-39, 1988.

LOESTCH, F. *et al.* *Forest inventory*. Munchen: BLV Verlagsgesellschaft, 1973. 2v. 469p.

LUNETTA, G. Di una Generalizzazione dello Schema della Curva Normale. *Annali della Facoltà di Economia e Commercio di Palermo*, v. 17, p. 237-244, 1963.

MACHADO, S.A. *et al.* Modelagem do volume individual para diferentes idades e regimes de desbastes em plantações de *Pinus oocarpa*. *Ciências Exatas e Naturais*,

Curitiba, v.4, n.2, p.185-196, 2002.

MCDONALD, J.B. Some generalized functions for the size distribution of income. *Econometrica*, v. 52, p. 647-663, 1984.

MCDONALD, J.B.; NEWHEY, W.K. Partially adaptive estimation of regression models via the generalized t distribution. *Econometric Theory*, v. 4, p. 428-457, 1988.

MCDONALD, J.B. Partially adaptive estimation of ARMA time series models. *International Journal of Forecasting*, v. 5, p. 217-230, 1989.

MCDONALD, J.B. Parametric models for partially adaptive estimation with skewed and leptokurtic residuals. *Economics Letters*, v. 37, p. 237-278, 1991.

MCLEAY, S. Students t and the distribution of financial ratios. *Journal of Business Finance and Accounting*, v. 13, p. 209-222, 1986.

MENON, M.U. Ajuste dos modelos de Kozak e do sistema Murkhart e Cao para plantações de *Pinus oocarpa*. *Revista Ambiência*, Guarapuava, n.1, p. 197-205, 2006.

MEYER, H.A. *Forest mensuration*. Pennsylvania: Penns Valley Publ., 357p, 1953.

MONTGOMERY, D.C. e PECK, E. A. *Introduction to linear regression analysis*. John Wiley, New York, 1982.

MORA, A. L.; GARCIA, C. H. *A cultura do eucalipto no Brasil (Eucalyptus cultivation in Brazil)*. São Paulo: Sociedade Brasileira de Silvicultura, 112p, 2000.

MUIRHEAD, R. The effects of symmetric distributions on some standard procedures involving correlation coefficients. *Multivariate Statistical Analysis North-Holland*, p. 143-159, 1980.

MUIRHEAD, R. *Aspects of multivariate statistical theory*. New York: John Wiley, 1982.

NELDER, J.A. e WEDDERBURN, R.W.M. Generalized linear models. *J. R. Statist. Soc. A*, 135, 370-384, 1972.

PARENT, D. R., A stand volume equation for cruising small-dia-meter material. *Journal of Forestry*, n. 185, p. 5-6, 2003.

PAULA NETO, F. *et al.* Tabela de volume para *Eucalyptus grandis*. *Revista Ceres, Viçosa*, 22: 212-22, mai./jun, 1975.

PAULA NETO, F. *et al.* Estimativa do volume comercial para *Eucalyptus microcorys* em regime de alto fuste na região de Coronel Fabriciano-MG. *Brasil Florestal*, 9(34): 57-61, abr./jul, 1979.

PELLICO NETTO, S. Equivalência volumétrica: uma nova metodologia para estimativa do volume de árvores. *Revista Acadêmica: ciências agrárias e ambientais*, Curitiba, v.2, n.1, p.17-30, jan./mar. 2004.

PLACKETT, R.L. The analysis of life-test data. *Technometrics*, v. 1, p. 9-19, 1961.

PRAETZ, P.D. The distribution of share price changes. *Journal of Business*, v. 45, p. 49-55, 1972.

PRAETZ, P.D.; WILSON, E.J.G. The distribution of stock market returns: 1958-1973. *Australian Journal of Management*, v. 3, p. 79-90, 1978.

RAO, B.L.S.P. Remarks on univariate symmetric distributions. *Statistics and Probability Letters*, v. 10, p. 307-315, 1990.

RATKOWSKY, D.A. *Nonlinear regression modelling*. 267p. Marcel Dekker, New York, 1983.

RIBEIRO, C.A.S. *et al.* Seleção de modelos volumétricos para leucena no Agreste de Pernambuco. *Brasil Florestal*, Brasília, DF, ano 20, n. 72, nov, 2001.

SANTANA, C. *et al.* Sistema de equações para simulação do crescimento e da produção em povoamentos de *Eucalyptus grandis* Hill ex. Maiden. sem desbaste baseado no modelo de Clutter. *Revista Ambiência*, n.1, p.239-256. Guarapuava: UNICENTRO, 2005.

SCHNEIDER, P.R. Análise de regressão aplicada à Engenharia Florestal. Santa Maria: UFSM, 1998. 236p.

SEBER, G. A. F. e WILD, C. J. Nonlinear Regression. John Wiley, New York, 1989.

SILVA, J. A. A., *et al.* Aumento da precisão de modelos volumétricos através do uso a transformação de Box-Cox. CERNE, V.1, N.1, P.013-016, 1994.

SILVA, J.A.A. Análise de equações volumétricas para construção de tabelas de volume comercial para *Eucalyptus* spp., Segundo a Espécie, Região e Método de Regeneração. 1977. 93f. Dissertação (Mestrado em Manejo Florestal) Universidade Federal de Viçosa, Viçosa MG.

SILVA, J.A.A. *et al.* Equação volumétrica para *Eucalyptus camaldulensis*, na região de Barbalha, Ceará, usando o volume da primeira tora como variável independente. Revista Árvore, Viçosa, MG, v.17, n.1, p. 30-37, 1993.

SILVA, J.A.A; SILVA, I.P. Estatística experimental aplicada à ciência florestal, Recife, Pernambuco: Imprensa Universitária da UFRPE, 1995. 292 p.

SILVA, J.A.A; PAULA-NETO, F. Princípios básicos de dendrometria. Recife: Universidade Federal Rural de Pernambuco, 1979. 191p.: il.

SIMONE, S. de. Su una estensione dello schema delle curve normali di ordine r alle variabili doppie. Statistica, v. 37, p. 447-474, 1968.

Sindicato das Indústrias do Gesso - SINDUSGESSO. Pólo gesseiro. Disponível em: <http://www.sindusgesso.org.br/>. Acesso em 14 jan. de 2009.

SOARES, C.P.B. *et al.* Estimativas de múltiplos volumes utilizando modelos tipo povoamento total In: SIMPÓSIO LATINO-AMERICANO SOBRE MANEJO FLORESTAL, 3., 2001, Santa Maria. Anais...Santa Maria: UFSM, 2001. p. 279 -286.

Sociedade Brasileira de Silvicultura - SBS. Fatos e números do Brasil florestal. São Paulo, 2006. 109p.

SPATHELF, P.; NUTTO, L. Modelagem aplicada ao crescimento florestal. Santa Maria: UFSM, 2000. 70 p.

SPURR, S.H. Forest inventory. New York: Ronald Press, 476 p, 1952.

SUBBOTIN, M.T. On the law of frequency of errors. *Matematicheskii Sbornik*, v. 31, p. 296-301, 1923.

VEIGA, R. A. A. Equações volumétricas para *Eucalyptus saligna* Smith em ocasião de primeiro corte. Botucatu. 174p, 1972.(Tese-Livre-Docência-UNESP)

VEIGA, R. A. A. Uso de equações de volume em levantamentos florestais. In: SIMPÓSIO SOBRE INVENTÁRIO FLORESTAL, 2, Piracicaba, 1984. Anais. Piracicaba, IPEF. p.93-102, 1984.

VEIGA, R.A.A.; BRASIL, M.A.M. Peso de matéria seca e volume de *Eucalyptus* propinqua Deane ex Maiden - equações. Boletim Técnico - Instituto Florestal. São Paulo, 35(2): 73-84, dez, 1981.

TAGUCHI, T. On a generalization of gaussian distribution. *Annals of the Institute of Statistical Mathematics*, v. 30, p. 211-242, 1978.

TAYLOR, J.M.G. Properties of modelling the error distribution with an extra shape parameter. *Computational statistics and data analysis*, v. 13, p. 33-46, 1992.

TAYLOR, S.J.; KINGSMAN, B.G. An analysis of the variance and distribution of commodity price-changes. *Australian Journal of Management*, v. 4, p. 135-149, 1979.

THIERSCH, C. R.; *et al.* Acurácia dos métodos para estimativa do volume comercial de clones de *Eucalyptus sp.*. *Cerne*, Lavras, v.12, n.2, p.167-181, 2006.

TOMÉ, M.; TOMÉ, J.; RIBEIRO, F.; FAIAS, S. Equação de Volume Total, Volume Percentual e de Perfil do Tronco para *Eucalyptus globulus Labill.* em Portugal. *Silva Lusitana* 15(1): 25 - 39, 2007.

TONINI, H. Emprego de diferentes metodologias para o cálculo do volume comercial de *Eucalyptus saligna* Smith. In: SIMPÓSIO LATINO-AMERICANO SOBRE MANEJO FLORESTAL, 3., 2004, Santa Maria. Anais... Santa Maria: UFSM, 2004. p. 28-34.

YAMAGUCHI, K. Generalized EM algorithm for model with contaminated error term. Proceedings of the seven Japan and Korea joint conference of statistical, p. 107-114, 1990.

APÊNDICE A - Algoritmos e programas

Programa do Modelo de Schumacher-Hall para os dados de volume

```
data arvore; input vol d ht; datalines;
```

```
0.1944 16.01 17.20
```

```
0.1625 15.28 16.70
```

```
.
```

```
.
```

```
.
```

```
.
```

```
;
```

```
run;
```

```
*/PROC NLP NORMAL NÃO TRANSFORMADO;
```

```
proc nlp data=arvore cov=2 pcov vardef=n covariance=h maxiter=10000;
```

```
max loglikO;
```

```
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1;
```

```
media=beta0*(d**beta1)*ht**beta2;
```

```
u=((vol-media)**2)/phi;
```

```
loglikO=(-0.5)*log(phi)-u/2;
```

```
run;
```

```
*/PROC NLP t STUDENT NÃO TRANSFORMADO;
```

```
proc nlp data=arvore cov=2 pcov vardef=n covariance=h;
```

```
max loglikO;
```

```
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1;
```

```
media=beta0*(d**beta1)*ht**beta2;
```

```

u=(vol-media)**2/phi;
m=2;
den=gamma((m+1)/2)/gamma(1/2)*gamma(m/2);
t1=m**(m/2);
t2= t1*(m+u)**(-(m+1)/2);
ger=t2/den;
loglikO=-0.5*log(phi)+log(ger);
run;

```

```
*/PROC NLP CAUCHY NÃO TRANSFORMADO;
```

```

proc nlp data=arvore cov=2 pcov vardef=n covariance=h;
max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1;
media=beta0*(d**beta1)*ht**beta2;
u=(vol-media)**2/phi;
ger=1/(1+u);
loglikO=-0.5*log(phi)+log(ger);
run;

```

```
*/PROC NLP LOGISTICA I NÃO TRANSFORMADO;
```

```

proc nlp data=arvore cov=2 pcov vardef=n covariance=h;
max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1;
media=beta0*(d**beta1)*ht**beta2;
u=(vol-media)**2/phi;
ger= 1.484300029*exp(-u)/(1+exp(-u))**2;
loglikO=-0.5*log(phi)+log(ger);
run;

```

```
*/PROC NLP LOGISTICA II NÃO TRANSFORMADO;
```

```

proc nlp data=arvore cov=2 pcov vardef=n covariance=h;

```

```

max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1;
media=beta0*(d**beta1)*ht**beta2;
u=(vol-media)**2/phi;
ger= exp(-u**0.5)/(1+exp(-u**0.5))**2;
loglikO=-0.5*log(phi)+log(ger);
run;

*/PROC NLP EXPONENCIAL POTENCIA NÃO TRANSFORMADO;

proc nlp data=arvore cov=2 pcov vardef=n covariance=h;
max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1;
media=beta0*(d**beta1)*ht**beta2;
u=(vol-media)**2/phi;
m=0.5;
t=1+0.5*(1+m);
kon=gamma(t)*2**t;
rr=(1+m)**(-1);
ger= kon**(-1)*exp(-0.5*u**rr);
loglikO=-0.5*log(phi)+log(ger);
run;

*/PROC NLP NORMAL TRANSFORMADO;

proc nlp data=arvore cov=2 pcov vardef=n covariance=h maxiter=10000;
max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1, lambda=1;
media=beta0*(d**beta1)*ht**beta2;
volt =(vol**lambda);
u=((volt-media)**2)/phi;
loglikO=(-0.5)*log(phi)-u/2+(lambda-1)*log(vol);
run;

```

```
*/PROC NLP t STUDENT - TRANSFORMADO;
```

```
proc nlp data=arvore cov=2 pcov vardef=n covariance=h;
max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1, lambda=1;
media=beta0*(d**beta1)*ht**beta2;
volt =(vol**lambda);
u=(volt-media)**2/phi;
m=2;
den=gamma((m+1)/2)/gamma(1/2)*gamma(m/2);
t1=m**(m/2);
t2= t1*(m+u)**(-(m+1)/2);
ger=t2/den;
loglikO=-0.5*log(phi)+log(ger)+(lambda-1)*log(vol);
run;
```

```
*/PROC NLP CAUCHY - TRANSFORMADO;
```

```
proc nlp data=arvore cov=2 pcov vardef=n covariance=h;
max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1, lambda=1;
media=beta0*(d**beta1)*ht**beta2;
volt =(vol**lambda);
u=(volt-media)**2/phi;
ger=1/(1+u);
loglikO=-0.5*log(phi)+log(ger)+(lambda-1)*log(vol);
run;
```

```
*/PROC NLP LOGISTICA I - TRANSFORMADO;
```

```
proc nlp data=arvore cov=2 pcov vardef=n covariance=h;
max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1, lambda=1;
media=beta0*(d**beta1)*ht**beta2;
```

```

volt =(vol**lambda);
u=(volt-media)**2/phi;
ger= 1.484300029*exp(-u)/(1+exp(-u))**2;
loglikO=-0.5*log(phi)+log(ger)+(lambda-1)*log(vol);
run;

```

```
*/PROC NLP LOGISTICA II - TRANSFORMADO;
```

```

proc nlp data=arvore cov=2 pcov vardef=n covariance=h;
max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1, lambda=1;
media=beta0*(d**beta1)*ht**beta2;
volt =(vol**lambda);
u=(volt-media)**2/phi;
ger= exp(-u**0.5)/(1+exp(-u**0.5))**2;
loglikO=-0.5*log(phi)+log(ger)+(lambda-1)*log(vol);
run;

```

```
*/PROC NLP EXPONENCIAL POTENCIA TRANSFORMADO;
```

```

proc nlp data=arvore cov=2 pcov vardef=n covariance=h;
max loglikO;
parms BETA0=0.0001, BETA1=2, BETA2=0.5, phi=1, lambda=1;
media=beta0*(d**beta1)*ht**beta2;
volt =(vol**lambda);
u=(volt-media)**2/phi;
m=0.5;
t=1+0.5*(1+m);
kon=gamma(t)*2**t;
rr=(1+m)**(-1);
ger= kon**(-1)*exp(-0.5*u**rr);
loglikO=-0.5*log(phi)+log(ger)+(lambda-1)*log(vol);
run;

```