

# PRODUTO 6

## PROPOSTA DE METODOLOGIA DE PRECIFICAÇÃO

### Contratada:

Escola Superior de Agricultura Luiz de Queiroz – ESALQ  
Fundação de Estudos Agrários Luiz de Queiroz – FEALQ

### Execução:



### Contratante:



Inversión Pública & Adaptación al  
Cambio Climático en América Latina

## ÍNDICE

INTRODUÇÃO	1
1 CONTEXTUALIZAÇÃO	2
2 CARACTERIZAÇÃO DOS PRODUTOS DE SEGURO AGRÍCOLA E PROAGRO	6
3 METODOLOGIA ATUARIAL DE PRECIFICAÇÃO DE RISCO	8
3.1 ASPECTOS ATUARIAIS DO SEGURO AGRÍCOLA	10
3.2 ESTUDOS ATUARIAIS APLICADOS AO BRASIL	14
3.3 A DISTRIBUICAO NORMAL ASSIMÉTRICA	17
3.4 A DISTRIBUICAO SKEW-T	18
3.5 A DISTRIBUICAO ODD LOG-LOGÍSTICA	20
3.6 A DISTRIBUICAO NÃO-PARAMÉTRICA	22
3.7 CRITÉRIOS PARA ESCOLHA DE DISTRIBUIÇÕES	24
4 CÁLCULO DA TAXA DE PRÊMIO	25
4.1 PRECIFICAÇÃO DO SEGURO DE FATURAMENTO: MODELANDO PRODUTIVIDADE E PREÇO CONJUNTAMENTE	25
5 CÓPULAS	27
5.1 INFERÊNCIA DAS CÓPULAS	28
5.2 MEDIDAS DE AJUSTE DOS MODELOS DE CÓPULAS	29
6 APLICAÇÃO DA METODOLOGIA PARA O CASO DA SOJA NO ESTADO DO PARANÁ: COMPARATIVO COM O PROAGRO	29
7 CONSIDERAÇÕES FINAIS	34
ANEXO	37
REFERÊNCIAS	46
CÓDIGOS R	50

## INTRODUÇÃO

De acordo com os Termos de Referência do Projeto intitulado “Investimento Público e Adaptação à Mudança do Clima na América Latina (IPACC II)”, PN: 2014.9049.9-002: “No marco de implementação do projeto IPACC II, as contrapartes principais do Brasil, o Ministério de Planejamento, Desenvolvimento e Gestão (MP), o Ministério da Fazenda (MF) e o Ministério do Meio Ambiente (MMA), identificaram o setor agrícola como sendo um dos setores que aloca altos níveis de investimento público e privado, e ao mesmo tempo sendo um dos setores de maior vulnerabilidade aos efeitos da mudança do clima. Tais condições motivaram a execução de uma iniciativa piloto, conjuntamente com o Ministério de Agricultura, Pecuária e Abastecimento (MAPA), com a finalidade de inserir a gestão do risco climático nos instrumentos de análise de investimentos no setor e, desta forma, complementar o instrumento atualmente utilizado Zoneamento Agrícola de Risco Climático (ZARC) buscando aprimorá-lo a partir da criação de recomendações de novas metodologias científicas, como suporte à política de gestão agrícola.”

O objetivo geral deste projeto é desenvolver um conjunto de recomendações visando reduzir as perdas agrícolas associadas aos efeitos da mudança do clima e que contribuam para uma alocação eficiente dos investimentos públicos e privados na agricultura. Como objetivo específico encontra-se o produto 6: “*Desenvolver proposta metodológica de precificação do risco: apresentar proposta de metodologia de atuarial de tarificação do risco para o Proagro/Seguro Rural.*”

## 1. CONTEXTUALIZAÇÃO

O seguro agrícola teve início no país em meados do início do século XX no Estado de São Paulo e operacionalizado pelo Instituto Agronômico do Estado. Sem a participação de seguradoras privadas, o prêmio do seguro contra granizo foi incorporado ao preço de venda da semente de algodão e formado uma espécie de fundo contra o granizo.

O Decreto-lei n. 1.186, de 3 de abril de 1939, criou o Instituto de Resseguros do Brasil (IRB) com o objetivo de regular as operações de resseguro, cosseguro e retrocessão e desenvolver a atividade seguradora de modo geral. A partir dessa data todas as operações de seguros eram obrigadas a ressegurar no IRB, inclusive no ramo rural.

Com o aumento de capacidade proporcionada pelo IRB abriu-se a oportunidade para se pensar na criação de uma empresa estatal que pudesse massificar o seguro agrícola em todo o território nacional. Já nessa época havia a preocupação de se elaborar um mecanismo de seguro agrícola que pudesse proteger não apenas o produtor agrícola e sua produção, mas também outras instituições importantes no processo produtivo, por exemplo, instituições financeiras que forneciam o crédito para o setor.

Apesar da importância da necessidade da criação do seguro agrícola, não havia um arcabouço legal para que pudesse nortear o mercado e o próprio Governo. Deste modo, no ano de 1954, teve início a regulamentação e o estabelecimento de normas para o ramo de seguro rural no país. A Lei n. 2.168, de janeiro de 1954, instituiu que o seguro rural destinava-se à preservação das colheitas e dos rebanhos contra a eventualidade de risco que lhes são peculiares.

Permitiu que o IRB realizasse os estudos, levantamentos e planejamentos para a início do seguro rural em todo o país em colaboração com os serviços técnicos das repartições federais, estaduais, municipais, de autarquias e dos estabelecimentos bancários oficiais de financiamento à agricultura e pecuária, e também, elaborasse as condições das apólices e taxas de prêmios de seguros, e que, depois de provadas pelo Departamento Nacional de Seguros Privados e Capitalização, fossem colocadas em vigor, mediante decretos.

A mesma lei ainda autorizava o Poder Executivo a organizar uma sociedade de economia mista, com sede e foro na cidade do Rio de Janeiro, com o objetivo de desenvolver progressivamente operações de seguros rurais, sob a denominação de Companhia Nacional de

Seguro Agrícola (CNSA). O Decreto n. 35.370, de 12 de abril de 1954, regulamentou as operações de seguro rural no país e conferiu a CNSA à exploração dessas operações.

Todas as responsabilidades não assumidas pelas companhias seguradoras ou pela CNSA eram resseguradas no IRB. Se as companhias não encontrassem resseguro das responsabilidades assumidas poderiam, através do IRB, realizar tal operação no exterior. As seguradoras poderiam, caso o IRB rejeitasse ou cancelasse tais responsabilidades, ressegurar em outras firmas no país ou na própria CNSA.

Ao longo de treze anos a CNSA encontrou uma série de dificuldades que culminaram na sua dissolução pelo Decreto-lei n. 73, de 21 de novembro de 1966. Um dos principais problemas encontrados pela Companhia foi a falta de conhecimento da atividade a ser segurada.

Características importantes da produção agrícola e da pecuária não foram levadas em conta, por exemplo, as peculiaridades regionais tanto do ponto de vista da cultura e do manejo do produtor quanto das diferenças climáticas e edáficas, sem as quais não há como, de modo razoável, estabelecer prêmios acessíveis aos produtores e manter a carteira atuarialmente equilibrada. Na época não havia uma base de dados para se realizar os cálculos atuariais nem mesmo metodologias adequadas para realizar a precificação do risco.

Durante quase uma década os produtores ficaram expostos às oscilações e adversidades climáticas, sem qualquer mecanismo de seguros. Apenas em 1975, quase uma década depois, o Governo criou o Programa de Garantia da Atividade Agropecuária (PROAGRO) como um instrumento para garantir o crédito fornecido ao setor.

Pouco se aproveitou da malsucedida experiência da CNSA. Os anos iniciais do PROAGRO foram catastróficos no que se refere aos resultados atuariais, ou seja, a carteira foi altamente deficitária. Basicamente, os mesmos problemas da CNSA se repetiram no contexto do PROAGRO, principalmente na ausência de uma base de dados e de metodologias atuariais de precificação.

Até meados de 2000, o PROAGRO foi o único instrumento utilizado pelo setor para se proteger contra eventos climáticos adversos. Nesse período, destaca-se a atuação de algumas poucas seguradoras estatais pertencentes a governos estaduais, mas com uma abrangência bastante limitada.

No final de 1990, algumas seguradoras privadas começaram a iniciar suas operações no seguro rural, tendo em vista a possível privatização do IRB e a possibilidade de abertura do mercado ressegurador, até então monopólio do IRB.

A abertura do mercado foi importante na atração de resseguradoras internacionais aumentando a capacidade das seguradoras e o interesse por ramos não tradicionais, por exemplo, o ramo rural. Porém, a abertura ocorreria apenas em 2007. Apesar disso, o mercado privado operava a modalidade agrícola de forma cautelosa, devido a falta de experiência, ausência de dados estatísticos de risco e de metodologias atuariais específicas ao setor.

A sanção da Lei 10.823 de 2003, que criou o Programa de Subvenção ao Prêmio do Seguro Rural (PSR), foi um marco legal importante para o florescimento de um mercado privado de seguro rural. Novas seguradoras iniciaram suas operações após a sanção da lei supramencionada.

De forma geral, o Governo forneceu ao setor privado o apoio necessário para operar o seguro rural, mesmo sem uma base histórica necessária para a quantificação e precificação de risco, o que viria a ter início após anos mais tarde mediante o acúmulo de dados pelas companhias seguradoras. No entanto, essa base de dados é confidencial a cada seguradora e restrita apenas as operações realizadas por cada companhia.

A única fonte de dados pública consultada pelo mercado pertence ao Instituto Brasileiro de Geografia e Estatística (IBGE), por meio da “produção agrícola municipal – PAM”. Apesar de possuir uma série de variáveis coletadas por um período relativamente longo, por exemplo, produção agrícola, área e produtividade, os dados estão agregados em nível municipal o que dificulta sua utilização pelo mercado segurador, na medida em que agrega produtores de diferentes perfis de risco em uma única informação.

O mesmo ocorre com os relatórios circunstanciados do PROAGRO publicados pelo BACEN. Os dados presentes nesse relatório são agregados e com uma nomenclatura não usual pelo mercado segurador, dificultando sua comparação com as operações do mercado.

Ademais, variáveis importantes não são coletadas pelos operadores do programa, por exemplo, dados de produtividade observada que são essenciais para os cálculos atuariais. Mesmo assim, os dados individualizados poderiam ser utilizados como “proxy” dos dados de seguros, na medida em que as operações se assemelham em alguns aspectos. No entanto, os microdados não são públicos por razões de confidencialidade.

Por fim, pode-se perceber que o processo histórico de evolução do seguro rural e PROAGRO revelam problemas que até hoje não foram superadas. A ausência de dados individualizados das unidades seguradas, de caráter público, e a uso de metodologias adaptadas de outros ramos, inadequadas às peculiaridades do setor agrícola, são dois dos principais problemas enfrentados para o desenvolvimento do seguro agrícola no país.

No que concerne a ausência de dados individualizados e públicos, o desenvolvimento de um banco de dados com informações úteis ao seguro rural é um esforço que precisa ser respaldado tanto pelo Governo quanto pela iniciativa privada e entidades representativas dos produtores.

Um grande passo foi dado com a criação do “Atlas do Seguro Rural” pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) e disponibilizado no endereço (<http://indicadores.agricultura.gov.br/atlasdoseguro/index.htm>). Ele representa um grande avanço na publicidade dos dados das operações subvencionadas e esforço em fornecer maior transparência à política pública.

Nessa plataforma é possível encontrar uma diversidade de informações relacionadas às operações de seguro rural enquadradas no Programa de Subvenção ao Prêmio, por exemplo, número de produtores atendidos, número de apólices, área segurada, importância segurada, valor de prêmio, valor de subvenção, produtividade segurada, taxas de prêmio, agregadas por município, por seguradora, desde 2006, na forma tabular ou mapas. Falta ainda incluir as variáveis de sinistros, por exemplo, número total de sinistros e valor das indenizações. Os microdados podem ser acessados por meio de solicitação ao MAPA.

O segundo problema apontado e objetivo principal deste relatório é descrever uma proposta de metodologia para a precificação de risco. Para isso, o relatório apresenta a seguinte estrutura. A próxima seção caracteriza os produtores de seguro agrícola e Proagro. A seção 3 apresenta a metodologia atuarial de precificação de risco detalhando em suas subseções as distribuições de probabilidade que podem ser utilizadas para se calcular as taxas. A seção 4 revela a formulação para se determinar as taxas de prêmio, no contexto univariado e bivariado. E, por fim, a seção 5 revela a forma de se modelar no caso bivariado (seguro renda).

## 2. CARACTERIZAÇÃO DOS PRODUTOS DE SEGURO AGRÍCOLA E PROAGRO

Os produtos mais conhecidos e comercializados pelo mercado segurador são o seguro agrícola de custeio e o seguro agrícola de produtividade. Ambos possuem muitas características semelhantes, mas também se diferenciam em outros aspectos.

As principais semelhanças referem-se aos riscos cobertos, níveis de cobertura, produtividade garantida e na forma de avaliação dos sinistros. Os principais riscos cobertos pelas apólices de seguros são:

- chuva excessiva;
- granizo;
- geada;
- incêndio;
- raio;
- seca;
- tromba d'água;
- ventos fortes;
- ventos frios; e
- variação excessiva de temperatura.

A gama de coberturas ofertadas varia conforme a seguradora. A outra variável, nível de cobertura (NC), é outro importante parâmetro do seguro agrícola. Ele é um percentual que varia entre 60 a 80%, em múltiplos de 5, e ser escolhido pelo produtor. A multiplicação do nível de cobertura pela produtividade esperada, em quilogramas por hectare ou sacas por hectare, resulta na produtividade garantida na apólice de seguros. Quanto maior o NC maior será a responsabilidade assumida pela seguradora e maior será a taxa cobrada pelo seguro. Por outro lado, o produtor terá uma proteção maior proporcionada pelo seguro.

A produtividade garantida é o gatilho que dispara o processo indenizatório. Se a produtividade observada ao longo da safra for menor do que a produtividade garantida o produtor terá o direito de receber a indenização. Existem duas situações nesse caso: perda total e perda parcial.



A primeira compreende as perdas decorrentes de um ou mais riscos cobertos que inviabilizem a continuidade da atividade. De outra forma, ocorre perda total quando ocorre morte em 100% das plantas na unidade segurada ou comprometimento total da produção. Por outro lado, as perdas parciais são caracterizadas quando os prejuízos decorrentes dos riscos cobertos não comprometem a continuidade da atividade.

Dessa forma, percebe-se que a variável que caracteriza o prejuízo é a produtividade agrícola. Apesar do produto cobrir perdas causadas por eventos climáticos adversos, o prejuízo é mensurado pela produtividade agrícola que é consequência não apenas dos fatores climáticos mas também das condições de solo, relevo, tipo de cultivar e manejo do produtor.

Por outro lado, os dois tipos de seguro diferenciam-se quanto ao cálculo do limite máximo de indenização (LMI). Enquanto no seguro de custeio o LMI está associado ao custeio por hectare gasto na lavoura, o seguro de produtividade cobre a produtividade esperada multiplicada pela área total a ser segurada e por um preço pré-estabelecido em contrato que permanece fixo até o final da vigência do contrato. Em geral a responsabilidade assumida pela seguradora no seguro de produtividade é maior do que no seguro de custeio.

Por outro lado, o PROAGRO assemelha-se ao seguro de custeio, mas com algumas diferenças. As mais importantes para os objetivos deste relatório referem-se ao LMI total ou também chamado de “valor enquadrado”, no PROAGRO, e ausência de níveis de cobertura.

No PROAGRO TRADICIONAL o LMI varia entre 70 a 100%, dependendo da adesão ou não ao PROAGRO nos 36 meses anteriores à data de enquadramento, ou deferimento de cobertura (indenização) referente ao último enquadramento do mesmo empreendimento.

Outro aspecto relevante é a ausência de níveis de cobertura. Em geral, o seguro agrícola não possui franquia ou dedutível associado ao seguro de custeio e o seguro de produtividade. No entanto, a existência do NC nesses produtos funciona de forma semelhante à franquia ou dedutível. Ou seja, é a parte em que o segurado é o responsável pelo seu próprio risco.

Na ausência de franquia, um NC de 100% sobre a produtividade esperada desincentivaria as boas práticas agrícolas, tais como, o uso correto de defensivos e fertilizantes, aumentando a probabilidade de perda da produtividade.

Por sua vez, o PROAGRO possui um sistema de bônus/málus que premia ou penaliza dependendo da comunicação ou não da perda. Nesse caso, a taxa de prêmio do PROAGRO,

denominada alíquota do adicional, é reduzida em 0,25 ponto percentual para cada ano safra sem comunicação de perda e acrescida em 0,50 ponto percentual quando ocorre a comunicação de perda. Na prática é o mesmo princípio do NC no caso do seguro agrícola no sentido de induzir o seguro a boas práticas de manejo e redução do risco moral.

### **3. METODOLOGIA ATUARIAL DE PRECIFICAÇÃO DE RISCO**

Desde as primeiras iniciativas em operar o seguro agrícola, o desenvolvimento de métodos específicos que resultassem em taxas atuarialmente equilibradas foi um grande desafio. A atividade agrícola possui uma série de peculiaridades que dificulta sobremaneira a quantificação e precificação do risco.

As unidades seguradas podem apresentar correlação espacial e dependência temporal e raramente são homogêneas. Ademais é comum o setor apresentar prejuízos de natureza catastrófica ou generalizada. Esses fatores dificultam a precificação, e em última instância o próprio equilíbrio atuarial da carteira, pois violam alguns princípios de segurabilidade.

Geralmente, os ramos tradicionais do seguro vida e não vida possuem métodos atuariais já conhecidos e testados pelo mercado segurador (Booth et al., 1999; Hart et al., 1996; ). Porém, quando novos produtos são desenvolvidos ou novos nichos de atuação são encontrados, depara-se com problemas de ausência de dados estatísticos e também dificuldades de natureza metodológica dificultando o trabalho do atuário. Notoriamente, o seguro agrícola é uma das modalidades mais complexas de precificar o risco em virtude das peculiaridades supramencionadas.

Ao redor do mundo onde o seguro agrícola é operacionalizado o cálculo preciso e equilibrado da taxa de prêmio é fundamental para o sucesso do seguro. Isso porque o seguro está fundamentalmente baseado no conceito de mutualismo, em que os prêmios são proporcionais ao risco individual. Isso garante o equilíbrio atuarial da carteira. Outro ponto importante refere-se a severidade e a frequência dos sinistros. Todos esses fatores influenciam o processo de precificação.

Independente do modelo de governança - públicas, privadas ou mistas - uma das preocupações centrais no desenvolvimento de um mercado de seguro agrícola refere-se a precificação adequada do risco. Nos EUA, por exemplo, que possuem um dos mais antigos e

desenvolvidos mercados de seguro agrícola do mundo a discussão acerca das melhorias e aprimoramentos metodológicos é realizada anualmente em eventos acadêmicos em que participam representantes do mercado (seguradoras e resseguradoras), governo (USDA) e pesquisadores nas áreas de economia, atuária, estatística e agronomia. Durante várias décadas os métodos atuariais aprimoraram-se conjugadas a evolução do próprio ramo de seguro.

Quando os riscos não são adequadamente precificados surgem problemas de desequilíbrio atuarial da carteira e seleção adversa, ou seja, quando há anti-seletividade por problemas de assimetria de informação. Na prática, quando a taxa de prêmio é fixada conforme o risco médio em uma carteira composta por produtores de alto e baixo risco, aqueles com risco maior terão uma probabilidade maior de demandar pelo seguro do que os produtores de baixo risco.

Assim, apenas aqueles de alto risco terão incentivos em demandar pelo seguro. Como consequência a concentração da carteira com indivíduos de alto risco aumenta consideravelmente a chance de tornar a carteira deficitária. Na safra seguinte, o mercado tende a elevar a taxa a fim de compensar as perdas. O problema torna-se ainda pior, e assim, sucessivamente. Esse fenômeno é conhecido como “espiral da morte” na literatura acadêmica.

Percebe-se, dessa forma, que definir uma taxa única para todo o território nacional baseado em uma média global, para cada cultura agrícola, apenas fortalece o problema da seleção adversa, com severas implicações para o equilíbrio da carteira e para a continuidade do seguro agrícola. O histórico do PROAGRO revela tais dificuldades.

O procedimento correto nessas circunstâncias é a separação das taxas por tipo de unidade segurada (propriedade rural) e tipo de cultura, levando-se em conta suas peculiaridades edafoclimáticas, genética e de manejo. Porém, dadas as limitações de dados estatísticos que reflitam a estrutura de risco em nível de propriedade exaurindo a multiplicidade dos riscos envolvidos recorre-se ao uso de dados agregados em nível municipal como uma proxy dos dados individualizados.

O interessante nesse caso é que a metodologia utilizada com os dados agregados pode ser aplicada também aos dados individualizados por propriedade rural, assim que uma massa considerável de dados for acumulada ao longo dos anos. É importante destacar que o método ou metodologia aqui relatado refere-se a um conjunto de procedimentos sistematizados e organizados que baseiam o processo de obtenção do conhecimento científico. A metodologia

apresentada está fundamentada na estatística, ou seja, a ciência que se baseia nas teorias probabilísticas para explicar fenômenos observacionais ou experimentais.

### 3.1 ASPECTOS ATUARIAIS DO SEGURO AGRÍCOLA

Na literatura especializada um dos primeiros trabalhos que consideraram a análise probabilística aplicada a precificação do risco agrícola foi realizado por Botts & Boles (1958), que utilizaram de forma bastante simples a distribuição Normal para parametrizar e quantificar o risco agrícola para uso no sistema de seguro agrícola norte-americano.

Alguns anos mais tarde Day (1965) estudou dados de produtividade resultantes de experimentos com diferentes níveis de nitrogênio para algodão e milho. A hipótese era de que a produtividade tinha assimetria positiva, na medida em que pequenas variações climáticas durante as fases mais sensíveis da cultura poderiam reduzir a produtividade. A conclusão apontou que as distribuições apresentaram, em geral, assimetrias e curtoses não-Normais.

Outro estudo conduzido por Gallagher (1987) procurou evidências de assimetrias nas distribuições de probabilidade em dados de produtividade potencial em nível nacional para a soja nos Estados Unidos. Os resultados mostraram uma assimetria negativa moderada e aumento da variabilidade correspondente a um patamar 20% maior na metade da década de 80 do que em 1972.

Na mesma linha de raciocínio Moss and Shonkwiler (1993) construíram um modelo de tendência estocástica para as séries de milho, agregadas em nível nacional, juntamente com a possibilidade de se transformar os dados para reduzir o efeito de outliers e induzir normalidade. Os resultados sugeriram que a produtividade do milho não segue uma distribuição normal apresentando, particularmente, assimetria negativa.

Taylor (1990) elabora um estudo para as séries de produtividade de milho, soja e trigo corrigidos para a tendência, por meio da construção de dois modelos distintos: método da transformação normal e método da distribuição condicional. Observou-se nos resultados indicativos de assimetria e curtose em todas as culturas.

Ao longo dos anos diversos especialistas analisaram o tema sem uma conclusão definitiva. Diversas abordagens foram apresentadas pelos atuários e econometristas na busca de distribuição de probabilidade que melhor se ajustasse aos dados de produtividade agrícola.

Por exemplo, a modelagem paramétrica, semi-paramétrica (Ker and Coble, 2003), não-paramétrica (Goodwin and Ker, 1998; Turvey and Zhao, 1999, Ozaki et al., 2008b), a abordagem Bayes empírico não-paramétrico (Ker and Goodwin, 2000) e a abordagem bayesiana espaço temporal (Ozaki et al., 2008a).

Dentro da abordagem paramétrica, diversos autores concluem que a produtividade agrícola segue uma distribuição normal (Just and Weninger, 1999). Entretanto, outros pesquisadores como Day (1965), Taylor (1990), Ramirez (1997), Ramirez et. al. (2003), encontraram evidências contra a normalidade na distribuição dos dados.

A adequação do uso da distribuição Normal facilitaria os cálculos atuariais, mas não há consenso quanto ao seu uso. Outras sugestões incluem o uso da distribuição Beta (Nelson and Preckel, 1989), Transformações Seno Hiperbólico Inversa (Moss and Shonkwiler, 1993) e Gama (Gallagher, 1987). Sherrick et al. (2004) ajustaram e compararam diversas distribuições paramétricas, como por exemplo, a Normal, Lognormal, Beta, Weibull e Logística utilizando a dados de produtividades por propriedades rurais nos EUA. Goodwin and Mahul (2004), corretamente apontaram que a distribuição Lognormal impõe assimetria positiva à distribuição, característica que pode ou não estar presentes nos dados.

Além dos testes realizados com diferentes distribuições de probabilidade, diversos estudos apontaram o pequeno número de observações de produtividade, em nível agregado e menor ainda em nível individual, como um empecilho para a modelagem atuarial.

A escassez de uma série temporal relativamente longa impossibilita a detecção de variações estruturais na produtividade e, conseqüentemente, verificar a ocorrência de eventos catastróficos. Quedas bruscas nas series de produtividade causadas por eventos climáticos adversos revelam situações de alto risco relativo.

Apesar de alguns métodos terem sido sugeridos para tentar superar tal problema, tais como, combinar dados de produtividade individuais e regionais ou utilizar dados agregados como proxies para os dados em nível de fazenda, ainda assim, isto dificulta sobremaneira o trabalho de modelar a estrutura de risco do produtor de maneira acurada. Por essa razão deve-se escolher métodos atuariais que reduzam tais problemas, ao mesmo tempo, que os dados sejam coletados e organizados para formar uma base consistida e relativamente longa.

Outra questão relacionada ao nível de agregação dos dados é com relação a alteração na variabilidade das séries. Os métodos de agregação dos dados geralmente induzem a um alisamento da série de dados reduzindo a variabilidade originalmente presente. Em outras

palavras, séries municipais de produtividade revelam um perfil de oscilação menor do que os dados em nível de propriedade rural levando a erros na estimação das taxas de prêmio.

A tendência é outra característica importante que deve ser levado em conta na modelagem dos dados de produtividade agrícola. Em muitos casos, quando existem amostras relativamente grandes, percebe-se que o processo gerador dos dados de produtividade não é constante ou estável, mas varia de acordo com o tempo.

A incorporação de novas tecnologias permitiu ao produtor elevar sua produtividade com o passar do tempo. Assim, a produtividade observada em meados de 1960, em geral é menor do que a produtividade observada em 2018, por exemplo. Para que tais produtividades sejam comparáveis é preciso que a componente de tendência seja captada na modelagem. Para isso ajusta-se um modelo polinomial de ordem 1, utilizando o tempo como covariável.

Ao ajustar um modelo de regressão linear polinomial de ordem 1, com apenas uma variável (tempo) da forma  $y_t = \alpha + \beta T + e_t$ ,  $e_t \sim N(\mu, \sigma^2)$ , capta-se a tendência que as séries possivelmente possam apresentar, em que  $y_t$  é o vetor de produtividade,  $T$  é o vetor de tempo,  $\alpha$  e  $\beta$  os parâmetros de regressão. Utilizando os resíduos  $\hat{e}_t$  da regressão, a estimativa da última observação através do modelo ajustado  $\hat{y}_{2017}$ , a tendência é retirada de acordo com a equação:  $\tilde{y}_t = \hat{y}_{2017} \left(1 + \frac{\hat{e}_t}{\hat{y}_t}\right)$  (Ozaki et al, 2008, Tejada et al., 2008, Duarte et al., 2017). A estatística t foi utilizada para definir a significância dos parâmetros.

Para testar a autocorrelação nos resíduos utiliza-se o teste de Ljung and Box (1978) para tratar a autocorrelação serial da produtividade. Ljung e Box propuseram uma pequena alteração no teste de Box - Pierce, na qual torna o teste mais generalizado. Foi observado na literatura que o teste de Ljung - Box apresenta melhores resultados do que o teste de Box - Pierce. Na prática, ao se captar a tendência em uma série de produtividade, assume-se que a série resultante torna-se estacionária.

Outro ponto a se destacar refere-se a possível presença de correlação espacial entre dados de municípios vizinhos. Como os dados são tratados individualmente, ou seja, os modelos são para séries temporais univariadas e não modelos para dados em painel, a presença de correlação espacial não afeta os resultados. Exceto, na situação abordada por Ozaki et al. (2008), quando a amostra por município era relativamente pequena e optou-se pela expansão da amostra utilizando dados de municípios vizinhos ao município central e sua posterior estimação via abordagem não-paramétrica por meio do estimador Kernel.

Embora essa abordagem seja apropriada para a estimação da probabilidade de perda e, posteriormente, o cálculo da taxa, ela apresenta algumas limitações. A estabilidade do estimador Kernel em pequenas amostras é um fator que merece atenção, além disso, a taxa de convergência para a verdadeira densidade é lenta, embora o estimador produza estimativas consistentes de qualquer densidade. Ademais, percebe-se a grande importância na escolha da distância a ser escolhida para selecionar os municípios vizinhos que normalmente é realizada de forma ad hoc e, portanto, dependente da escolha do usuário.

Para o teste da heterogeneidade de variâncias utiliza-se o teste de Breusch and Pagan (1979). Baseado no teste multiplicador de Lagrange, o teste de Breusch-Pagan é bastante utilizado para testar a hipótese nula de que as variâncias dos erros são iguais (homoscedasticidade) versus a hipótese alternativa de que as variâncias dos erros são uma função multiplicativa de uma ou mais variáveis, sendo que esta(s) variável(eis) pode(m) pertencer ou não ao modelo em questão.

Inicialmente, ajusta-se um modelo de regressão linear (simples ou múltiplo) e obtêm-se os resíduos  $e = (e_1, \dots, e_n)$  e os valores ajustados  $\hat{y} = (\hat{y}_1, \dots, \hat{y}_n)$ . Em seguida, consideram-se os resíduos ao quadrado e estes são padronizados de modo que a média do vetor de resíduos padronizados  $u$  seja igual a 1. Esta padronização é feita dividindo cada resíduo ao quadrado pela  $SQE/n$  em que  $SQE$  é a Soma de Quadrados dos Resíduos do modelo ajustado e  $n$  é o número de observações. Desta forma, temos que cada resíduo padronizado é dado por

$$u_i = \frac{e_i^2}{SQE/n}, i = 1, \dots, n, \text{ em que } SQE = \sum_{i=1}^n e_i^2.$$

Por fim, faz-se a regressão entre  $u = (u_1, \dots, u_n)$  (variável resposta) e o vetor  $\hat{y}$  (variável explicativa) e obtêm-se a estatística do teste  $\chi_{BP}^2$  calculando-se a Soma de Quadrados da Regressão de  $u$  sobre  $\hat{y}$  e dividindo-se o valor encontrado por 2. Sob a hipótese nula, esta estatística tem distribuição qui-quadrada com 1 grau de liberdade. Para a modelagem da produtividade foram testados modelos paramétricos simétricos e assimétricos, que incorporam a unimodalidade e bimodalidade dos dados. Todas estas análises foram realizadas no software R utilizando os pacotes “Optim” e “AdequacyModel”.

Testa-se também se os dados seguem uma distribuição Normal, por meio do teste de Kolmogorov-Smirnov (KS). O teste de Kolmogorov - Smirnov pode ser utilizado para avaliar as hipóteses:  $H_0$ : os dados seguem uma distribuição Normal, contra  $H_1$ : os dados não seguem uma distribuição Normal.

‘Este teste observa a máxima diferença absoluta entre a função de distribuição acumulada dos dados, no caso a Normal, e a função de distribuição empírica dos dados. Como critério, comparamos esta diferença com um valor crítico, para um dado nível de significância. O valor-p resultante do no programa R pode ser interpretado como a medida do grau de concordância entre os dados e a hipótese nula. Quanto menor for o valor-p, menor é a consistência entre os dados e a hipótese nula. Então, a regra de decisão é rejeita-se  $H_0$ , se o valor-p for menor ou igual a um certo valor arbitrado pelo analista.

Em seguida é preciso escolher entre o uso de distribuições paramétricas ou não-paramétricas para ajustar aos dados. Recomenda-se as seguintes distribuições paramétricas: Normal Assimétrica, Skew-T e a distribuição Odd Log-logística, por apresentarem robustez e flexibilidade na detecção de assimetrias e bimodalidades. É importante destacar que ambas as abordagens apresentam desvantagens em amostras relativamente pequenas.

### 3.2 ESTUDOS ATUARIAIS APLICADOS AO BRASIL

No Brasil, alguns trabalhos abrangeram a problemática da precificação utilizando diversas abordagens, que variaram em função do tipo de produto de seguros, ou seja, se o seguro agrícola é de produtividade ou de faturamento. Um dos primeiros trabalhos teve como objetivo de sugerir formas de calcular as taxas de prêmio de um hipotético seguro de renda, no formato dos seguros americanos Income Protection ou Revenue Assurance, para milho nos municípios do Paraná/BR, por meio do uso de cópulas na distribuição conjunta entre preços e produtividade (Miqueleto, 2011).

Outro trabalho inovador, proposto por Brisolara (2013), considerou a metodologia para o cálculo da taxa de prêmio utilizando séries univariadas de receita agrícola e séries multivariadas de preço e produtividade, supondo independência entre as duas variáveis. A modelagem foi realizada por meio da Normal Multivariada relacionando a correlação entre preço e produtividade na matriz de variância e covariância da própria distribuição. A aplicação teve como foco o seguro agrícola de receita para soja no estado do Paraná.



Santos, Scalon e Ozaki (2014) sugeriram a distribuição normal-assimétrica como uma alternativa à distribuição normal para modelar a distribuição da produtividade agrícola no Brasil. Os pesquisadores destacam que a distribuição normal-assimétrica foi escolhida por apresentar a possibilidade de ajuste de uma série que contenha assimetria em sua distribuição, característica comum nos dados de produtividade. Porém, destacam que existem outras distribuições como a log-normal e beta que seriam possíveis também em dadas condições.

Um ponto a destacar refere-se ao fato de que as séries de produtividade utilizadas no trabalho apresentaram tendência, o que exigiu métodos de correção das mesmas. Sabe-se que as séries corrigidas podem apresentar valores negativos, fato que impossibilita o uso de algumas distribuições paramétricas, como é o caso da densidade log-normal.

Uma desvantagem da distribuição normal-assimétrica em relação à normal é a possibilidade de não ocorrer convergência no algoritmo que estima os parâmetros do modelo. Os pesquisadores ressaltaram que ambos os modelos apresentam uma limitação que é a impossibilidade de bimodalidade ou multimodalidade da distribuição de produtividade.

Os autores observaram também que a distribuição normal-assimétrica apresentou alguma assimetria positiva em relação à normal e, portanto, sugeriram que a normal estaria subestimando o pagamento esperado em alguns municípios. O prêmio levou em conta o valor do pagamento esperado, como consequência, pode-se subestimá-lo e cobrar menos do que o necessário para compensar as perdas e ainda, em caso de sinistro, a seguradora deverá fazer um pagamento maior do que o previsto.

Duarte (2015), em sua tese de doutorado, apresentou métodos alternativos para precificação dos riscos de produtividade e faturamento agrícola. Na modelagem da distribuição de produtividade, utilizou-se distribuições paramétricas que capturam a simetria, a assimetria e a bimodalidade dos dados, características estas geralmente encontradas em produtividades brasileiras. Além disso, a metodologia de cópulas foi utilizada na análise multivariada entre produtividade e preço.

Esta metodologia apresenta um ganho significativo para estudos de problemas multivariados comparados à distribuição normal multivariada, pois pode-se levar em consideração estruturas de dependências linear, não linear e dependência somente nas caudas da distribuição multivariada. Além disso especifica qualquer tipo de distribuição para as distribuições marginais.

Polo (2016), em sua dissertação de mestrado, teve por objetivo aplicar redes Bayesianas (grafo acíclico direcionado, ou modelo hierárquico Bayesiano) para a estimação da distribuição de probabilidade de rendimento da soja em alguns municípios do Paraná, com foco na análise comparativa de riscos.

O pesquisador descreve que a modelagem baseada em redes bayesianas apresenta vantagens de flexibilidade de interpretação da estrutura do modelo. Em especial, a modulação do modelo a partir de um grafo da rede bayesiana permite dar uma interpretação física e agrônômica a um modelo empírico. Se a abordagem de ajuste é aceitável, ganha-se grande simplicidade no ajuste de um modelo complexo.

Por fim o autor relata que o modelo ainda permite a sua aplicação em momentos diferentes do ciclo de produção agrícola para obter informações diferentes (úteis para cada momento), e associar dados meteorológicos aos dados de sensoriamento remoto sem que haja necessidade de observações de todas as variáveis estarem disponíveis.

Adami e Ozaki (2016) tiveram como objetivo de analisar a viabilidade do seguro receita, considerando-se a receita bruta da soja da região oeste do Paraná, em nível de propriedade rural. Os autores relatam que tradicionalmente, o método usado pelo mercado para a precificação do risco, tanto de custeio quanto de receita, é baseado na relação entre as perdas esperadas e o valor garantido pela empresa seguradora, conhecido como "taxas empíricas" (ER). Uma das principais desvantagens do método ER é a dependência de um grande número de informações para refletir acuradamente a distribuição de probabilidade da produtividade ou da receita.

De modo a resolver esse problema, os autores assumiram que a receita da propriedade agrícola segue uma distribuição normal assimétrica, que incorpora um parâmetro flexível para captar, caso exista, a assimetria da distribuição de probabilidade. Optaram pela distribuição normal assimétrica em função de sua flexibilidade e devido a evidências de assimetria na distribuição da produtividade agrícola e da receita agrícola em estudos anteriores (SHERRICK et al., 2004; OZAKI e SILVA, 2009; SANTOS, SCALON e OZAKI, 2014).

### 3.3 A DISTRIBUICAO NORMAL ASSIMÉTRICA

Considerou-se a distribuição Normal Assimétrica (DNA) pela sua flexibilidade em acomodar assimetrias presentes nas distribuições dos dados. A DNA foi proposta por Azzalini (1985) e é uma generalização da distribuição Normal. A vantagem está no fato de se captar assimetrias tanto negativas quanto positivas. Tais assimetrias tem impacto direto no calculo da probabilidade de perda e, conseqüentemente, no cálculo da taxa de prêmio.

Nesse contexto, considere  $f$  a função densidade de probabilidade (f.d.p.) simétrica ao redor de zero e  $G$  função distribuição cumulativa (f.d.c.) contínua e absoluta, tal que  $g = G'$  é simétrica ao redor de zero (Ozaki e Silva, 2009). Então,

$$2 f(y) G(\gamma y), y \in R \quad (1)$$

É uma f.d.p. para qualquer  $\gamma \in R$ . Da equação (1), a DNA, com parâmetro de locação  $\mu$ , escala  $\sigma$  e forma  $\gamma$ , é definida como:

$$\phi(y|\mu, \sigma, \gamma) = \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\frac{\gamma(y-\mu)}{\sigma}\right) \quad (2)$$

Onde  $\phi$  e  $\Phi$  são respectivamente a f.d.p. e a f.d.c. de uma variável aleatória Normal padrão. O grau de assimetria da DNA é dada pela equação

$$\varepsilon = \delta^3 \left[ \frac{4}{\pi} - 1 \right] \left[ 1 - \frac{2\delta^2}{\pi} \right]^{-3/2} \sqrt{2/\pi} \quad (3)$$

Onde  $\delta = \gamma(1 + \gamma^2)^{-1/2}$  e  $-0.99527 < \varepsilon < 0.99527$  com casos limites ( $\gamma \rightarrow \pm\infty$ ) sendo uma distribuição Meio Normal. De fato, utilizou-se a abordagem de Henze (1986), representando a DNA como uma mistura de média variância de uma Normal e uma Meia Normal.

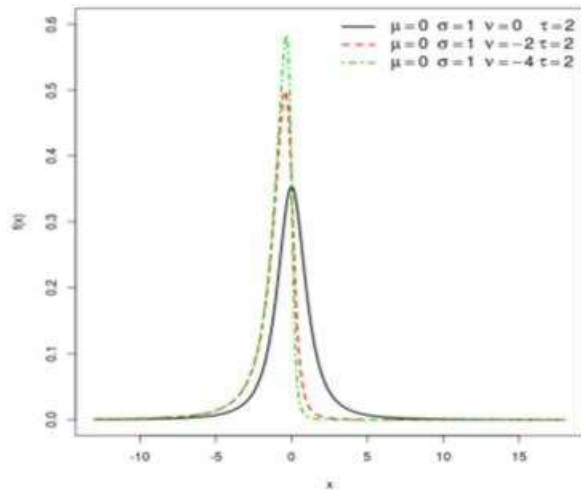
### 3.4 A DISTRIBUICAO SKEW-T

Quando se tem dados com um comportamento assimétrico e presença de valores extremos ou discrepantes, a distribuição Skew-Normal pode não se ajustar bem. Em tais situações, uma alternativa mais robusta ao modelo Skew-Normal é proposta por Azzalini e Capitanio (2003), a versão assimétrica da distribuição t-Student (ST), denominada t-Student Assimétrica.

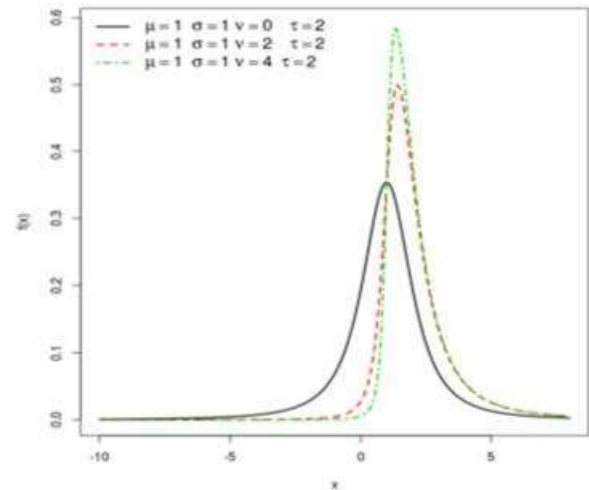
Uma variável  $X$  é chamada “Skew-t Student”, com parâmetros de locação  $\mu$ ,  $\mu \in \mathbb{R}$ , escala  $\sigma$ ,  $\sigma \in \mathbb{R}_+^*$ , assimetria  $v$ ,  $v \in \mathbb{R}$ , e curtose  $\tau$ ,  $\tau \in \mathbb{R}_+^*$ , se sua f.d.p é dada por:

$$f(x; \mu, \sigma, \lambda) = \frac{2}{\sigma} t\left(\frac{x-\mu}{\sigma}\right) T\left[v\sqrt{\lambda}\left(\frac{x-\mu}{\sigma}\right)\right], \quad (4)$$

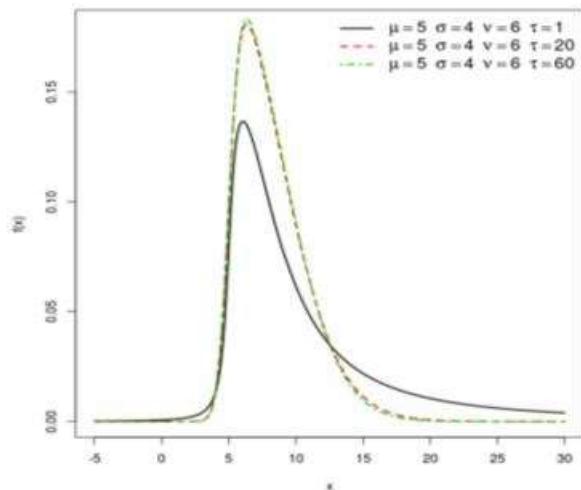
em que  $z = \left(\frac{x-\mu}{\sigma}\right)$ ,  $w = v\lambda^{1/2}z$ ,  $\lambda = \frac{\tau+1}{\tau+z^2}$ ,  $tZ1$  e  $TZ1$  são respectivamente, a f.d.p da t-Student padrão com  $\tau$  graus de liberdade e a fda da t-Student padrão com  $\tau + 1$  graus de liberdade. Como notação, adota-se  $Z \sim ST(\mu, \sigma, v, \tau)$ . Note que, quando  $v = 0$ , a distribuição é equivalente a t-Student simétrica.



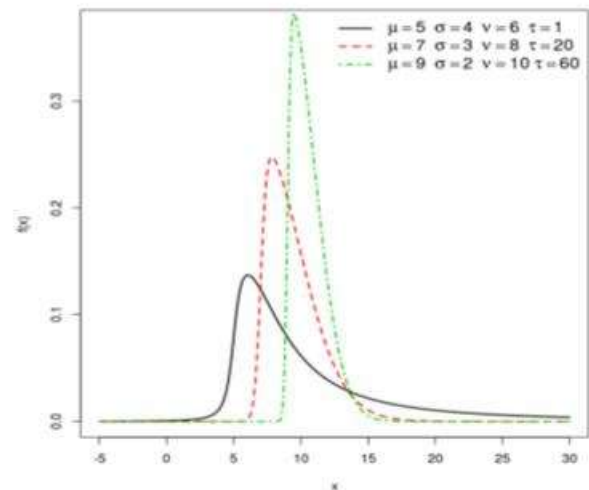
(a) Variando  $\nu < 0$



(b) Variando  $\nu > 0$



(c) Variando  $\tau > 0$



(d) Variando  $\mu$  e  $\tau$

Figura 1 : Gráfico da f.d.p da Skew-t para diferentes valores dos parâmetros.

A figura 1 ilustra o comportamento da f.d.p da ST. Nota-se nas figuras 1(a,b) a contribuição do parâmetro  $\nu$  na forma da distribuição, com  $\mu$ ,  $\sigma$ ,  $\tau$  fixos. Os valores positivos de  $\nu$  indicam assimetria positiva e os valores negativos de  $\nu$  indicam a assimetria negativa.

Na figura 1(c), observa-se o efeito do parâmetro de curtose na distribuição quando os outros parâmetros são fixos. Na figura 1(d) observa-se o efeito de deslocamento da distribuição pela variação do parâmetro  $\mu$  e caudas mais “pesadas” com a variação do parâmetro  $\tau$ .

### 3.5 A DISTRIBUICAO ODD LOG-LOGÍSTICA

Uma nova classe de família de distribuições contínuas recentemente introduzidas por Cruz (2016) é chamada odd log-logistic-F (OLLF). A inferência e estimação pelo método da máxima verossimilhança para grandes amostras são também abordados para dados de sobrevivência censurados. Esta nova classe de distribuições é interessante, pois através de suas simulações pode-se capturar a bimodalidade dos dados.

Dado uma função base de probabilidade acumulada qualquer com  $F(x; \xi)$  um vetor de parâmetros  $\xi$ , a função distribuição cumulativa (f.d.c.) da distribuição odd log-logistic-F (OLL-F) com um parâmetro de forma  $\alpha > 0$  é definido por:

$$G(x; \xi) = \int_0^{\frac{F(x;\xi)}{\bar{F}(x;\xi)}} \frac{\alpha x^{\alpha-1}}{(1+x^\alpha)^2} dx = \frac{F(x;\xi)^\alpha}{F(x;\xi)^\alpha + \bar{F}(x;\xi)^\alpha} \quad (5)$$

em que  $\bar{F}(x, \xi) = 1 - F(x, \xi)$  e  $\alpha = \frac{\log \frac{G(x, \xi)}{G(x, \xi)}}{\log \frac{F(x, \xi)}{\bar{F}(x, \xi)}}$ .

O parâmetro  $\alpha$  representa o quociente do log da razão de chances para a distribuição base. Se  $F(x, \xi) = \frac{x}{1+x}$ , a equação (5) então se reduz a função de probabilidade da distribuição log-logística. Várias distribuições podem ser geradas da equação acima, como as apresentados nos trabalhos de Alizadeh et al. (2015), Cruz et al. (2015) e Braga et al. (2016).

A nova família de distribuição OLLF permite uma maior flexibilidade nas caudas da distribuição. Se considerarmos a função base  $G(x; \xi)$  como a distribuição acumulada da normal  $N(\mu, \sigma)$  Assim, para a nova classe de família odd log logistic normal (OLL-N) a equação (5) se reduz a:

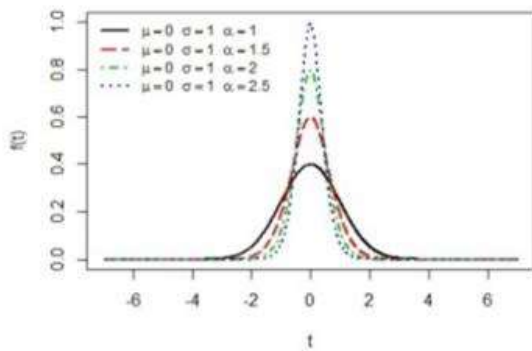
$$G(x; \mu, \sigma, \alpha) = \int_0^{\frac{\Phi(x;\mu,\sigma,\alpha)}{\bar{\Phi}(x;\mu,\sigma,\alpha)}} \frac{\alpha x^{\alpha-1}}{(1+x^\alpha)^2} dx = \frac{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right)}{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha} \quad (6)$$

em que  $\bar{\Phi}(x; \mu, \sigma) = 1 - \Phi(x; \mu, \sigma)$ .

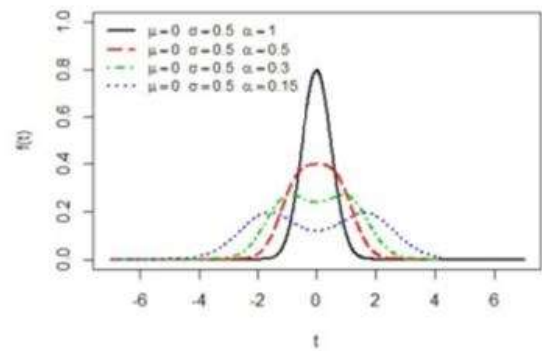
A função densidade de probabilidade (f.d.p) é dada por:

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right) \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}}{\sigma \left\{ \Phi^{\alpha}\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha} \right\}^2} \quad (7)$$

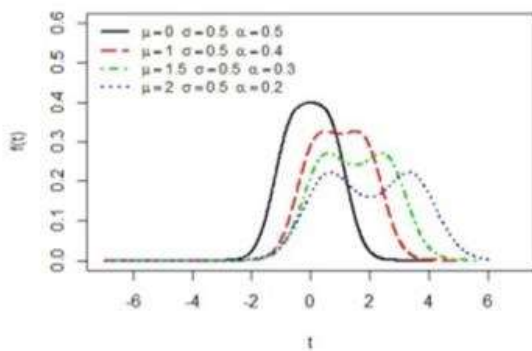
Note que  $\alpha > 0$  é um parâmetro de forma. A partir de agora, assumimos que uma variável aleatória com função de densidade como da equação (7) é denotado por  $X \sim \text{OLLN}(\alpha, \mu, \sigma)$ .



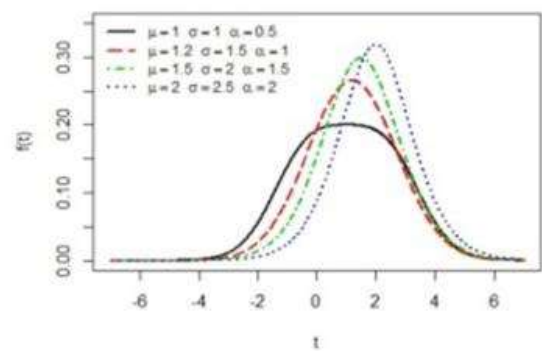
(a) Variando  $\alpha > 1$



(b) Variando  $\alpha < 1$



(c) Variando  $\mu > 0$  e  $0.2 < \alpha < 0.5$



(d) Variando todos os parâmetros

Figura 2: Gráfico da f.d.p da OLLN para diferentes valores dos parâmetros.

A figura 2 apresenta algumas formas que a função densidade do modelo OLLN pode assumir para diferentes valores dos parâmetros  $\mu$ ,  $\sigma$  e  $\alpha$ . Em simulações, pode-se observar a contribuição do parâmetro  $\alpha$  sobre a unimodalidade e bimodalidade da distribuição, quando  $\mu$ ,  $\sigma$  são fixos, figura 2(a,b).

Quando o parâmetro  $\alpha$  aproxima-se de zero, a f.d.p apresenta bimodalidade. Por outro lado, quando o valor de  $\alpha$  aumenta, a função apresenta unimodalidade. Além disso, quando  $\mu$  varia, os gráficos são transladados no eixo das abcissas, independentemente da forma como mostra a Figura 1c. A figura 2d apresenta a f.d.p da OLLN com variação de todos os parâmetros.

### 3.6 A DISTRIBUICAO NÃO-PARAMÉTRICA

Diferentemente da abordagem paramétrica, a análise não-paramétrica não especifica à priori a distribuição dos dados. A sua vantagem é deixar que os dados revelem a forma da densidade. Ademais, certas características da distribuição podem ser captadas, tais como, assimetrias positivas e negativas e a bi ou multimodalidade.

Dentre os diversos tipos de estimadores da densidade, o histograma é o mais utilizado e conhecido nas análises exploratórias de dados. Dois aspectos merecem destaque na estimação dos histogramas, que são a origem e a largura das classes  $h$  ou “janelas”.

Em função da escolha da origem, diferentes histogramas podem ser construídos e, conseqüentemente, diferentes interpretações podem surgir. O outro aspecto, a largura das classes determina o alisamento da série. Apesar de sua simplicidade, este estimador apresenta desvantagens em relação a outros estimadores, como por exemplo, o estimador Kernel. Recomenda-se trabalhar com histogramas para realizar uma primeira análise dos dados.

No contexto de precificação de contratos de seguro agrícola, Goodwin e Ker (1998), Turvey e Zhao (1999) e Ozaki et al. (2008b) utilizaram o estimador Kernel para determinar a forma da densidade da produtividade condicional. O estimador kernel da densidade  $\hat{f}(y)$  pode ser representada como uma convolução da distribuição amostral, utilizando-se uma função kernel  $K$ , tal que:



$$\hat{f}(y) = \int K_h(y-v) dF_n(v) \quad (8)$$

onde  $K_h(v) = 1/hK(v/h)$  e  $F_n(v)$  é a função de distribuição amostral (Goodwin e Ker, 2002).

O estimador kernel é a soma de “saltos” localizados em cada observação, de modo que a função kernel determina a forma destes saltos e a janela sua largura. Quanto maior o valor da janela, a série terá um alisamento maior e os detalhes tendem a desaparecer. Por outro lado, os saltos terão uma forma de pico quanto mais próxima de zero estiver a janela, tornando mais pronunciado os detalhes na densidade.

Algumas suposições devem ser feitas à respeito de  $K$ . A função kernel será não negativa, simétrica e constituirá uma função de densidade, tal que  $\int K(v)dv = 1$ . Além disso,  $\int vK(v)dv = 0$  e  $\int v^2K(v)dv = \theta_2 \neq 0$ . Caso a função kernel seja representada pela densidade Normal, então  $\theta_2$  será a variância da distribuição.

Como uma medida de discrepância do estimador da densidade  $\hat{f}$  em relação a verdadeira densidade  $f$ , Silverman (1986) adota o erro quadrático integrado médio (MISE) dado por  $\int E[\hat{f}(y) - f(y)]^2 dx$ , que pode ser decomposta em dois componentes, o viés quadrático integrado e a variância integrada. Sob certas pressuposições, o viés pode ser aproximado por  $\frac{1}{4}h^4\theta_2 \int f''(y)^2 dy$  e a variância por  $(nh)^{-1} \int k(v)^2 dv$ .

Ao se minimizar o MISE escolhe-se um alto valor para o parâmetro de alisamento, reduz-se a variação aleatória (variância) levando ao aumento do erro sistemático (viés). Caso contrário, um baixo valor do parâmetro resulta em aumento da variância integrada e diminuição do viés quadrático integrado.

Na busca pelo parâmetro de alisamento ótimo e o kernel que minimize o MISE, o autor mostra que, se o  $h$  for ótimo,  $h_{opt} = \theta_2^{-2/5} (\int K(v)^2 dv)^{1/5} (\int f''(y)^2 dy)^{-1/5} n^{-1/5}$ , então o valor

aproximado para o erro quadrático integrado médio será dado por  $\frac{5}{4}W(K)(\int f''(y)^2 dy)^{1/5}n^{-4/5}$ , onde  $W(K) = \theta_2^{2/5}(\int K(v)^2 dv)^{4/5}$ .

O problema de minimizar o MISE torna-se um problema de escolher o mínimo  $W(K)$ , dado o parâmetro de alisamento. Seguindo certas restrições, se a função kernel escolhida for o kernel de *Epanechnikov*  $W(K_e)$ , então o MISE será minimizado.

Para comparar outros kernel simétricos, procedeu-se na elaboração de um índice de eficiência igual a  $[W(K_e) / W(K)]^{5/4}$ . Utilizando as densidades biweight, triangular, gaussiana e retangular, observou-se que todos os kernel apresentaram eficiência próximas de um. Desta forma, a escolha do kernel resulta em pouca influência sobre o critério escolhido.

A escolha do parâmetro de alisamento ótimo, no caso da distribuição Gaussiana, será igual a  $1.06\sigma n^{-1/5}$ , onde  $\sigma$  é o desvio padrão da produtividade. Recomenda-se  $\sigma = \min(\text{desvio padrão}, \text{interquartil}/1.34)$  para levar em conta desvios da normalidade. Além disso, se o fator 1,06 for reduzido para 0,9, obtêm-se melhores resultados empíricos. Desta forma, recomenda-se o uso desta estimativa para determinação do parâmetro de alisamento.

### 3.7 CRITÉRIOS PARA ESCOLHA DE DISTRIBUIÇÕES

Para a seleção da distribuição de probabilidade a ser utilizada na precificação de risco dois critérios são utilizados na literatura: o critério de informação Akaike (AIC), proposto por Akaike (1974) e o critério Bayesiano (BIC), proposto por Schwarz (1978).

Estes critérios são baseados no logaritmo da função de verossimilhança estimada. Assim, o critério proposto por Akaike e o critério Bayesiano são, respectivamente:

$$\text{AIC} = -2l(\hat{\theta}) + 2k,$$

$$\text{BIC} = -2l(\hat{\theta}) + \ln(n) k$$

em que  $l(\cdot)$  é o logaritmo da função de verossimilhança maximizada sob o modelo considerado,  $k$  o número de parâmetros e  $\hat{\theta}$  o vetor de parâmetros estimado. O modelo com o menor valor utilizando um dos critérios deve ser escolhido.

#### 4. CÁLCULO DA TAXA DE PRÊMIO

Uma vez escolhida a abordagem probabilística por meio do critério de informação de Akaike (AIC) ou critério de informação bayesiano (BIC), a ser descrita posteriormente, gera-se a estimativa da probabilidade de perda que é um dos parâmetros da taxa de prêmio e que pode ser calculada através da área sob a curva menor do que a produtividade garantida. Desta forma, se denominarmos o nível de cobertura por  $\lambda$ , tal que  $0 < \lambda < 1$ , e a produtividade esperada por  $y_e$ , então a probabilidade da produtividade ser menor do que  $\lambda y_e$  será calculada através da área sob a densidade, numericamente estimada através da regra do trapézio. A taxa de prêmio será dada por (Goodwin e Ker, 1998):

$$\text{Taxa de prêmio} = \frac{F_Y(\lambda y^e) E_Y[\lambda y^e - (Y | y < \lambda y^e)]}{\lambda y^e} \quad (9)$$

onde  $E$  é o operador de esperança e  $F$  a distribuição cumulativa da produtividade.

##### 4.1 PRECIFICAÇÃO DO SEGURO DE FATURAMENTO: MODELANDO A PRODUTIVIDADE E O PREÇO CONJUNTAMENTE

Nota-se que, no caso anterior, a taxa de prêmio é calculada de forma univariada considerando apenas a produtividade agrícola. Outro produto que apresenta rápido crescimento e aceitação por parte dos produtores é o seguro agrícola de faturamento ou receita. Esse tipo de seguro é uma reivindicação relativamente antiga dos produtores rurais, mas atendida apenas em meados de 2010, quando o mercado ofertou o seguro de forma limitada em apenas 8 operações e consumindo apenas R\$ 63 mil de subvenção.

De 2010 até 2014 houve um aumento significativo do seguro de faturamento. Em 2014 foram aproximadamente 3.200 apólices e pouco mais de R\$ 35 milhões gastos em

subvenção, para uma importância segurada de quase R\$ 1 bilhão. Em 2015, devido ao forte contingenciamento orçamentário sofrido pelo MAPA, o seguro de faturamento recuou drasticamente em razão de sua elevada dependência com o orçamento do PSR.

A partir de 2017, o MAPA priorizou o seguro de faturamento atrelando um maior percentual de subvenção nas operações do seguro de faturamento. Como resultado, em 2018, o seguro de faturamento alcançou quase 2.900 apólices, R\$ 34,4 milhões de subvenção gastos e uma importância segurada de quase R\$ 2 bilhões, praticamente o dobro de 2014 com praticamente o mesmo total gasto em subvenção. Atualmente, o seguro de faturamento consome aproximadamente 10% dos recursos do PSR, equivalente a 16% de toda a importância segurada em 2018 e 4% das apólices realizadas, segundo o Atlas do Seguro Rural do MAPA.

No que tange a sua precificação, o seguro de faturamento é formado por duas ou três variáveis conforme as características do produto. A primeira delas é a produtividade e a segunda é o preço futuro, em reais. Caso o preço seja fixado em dólar, existe ainda uma terceira variável que é a taxa de câmbio.

Considerando o caso em que o preço seja fixado em reais, a taxa de prêmio será fixada levando-se em conta a relação entre as duas variáveis, de acordo com a seguinte equação:

$$\text{Taxa de Prêmio} = \frac{F_y(X < x; Y < y | x, y < x_g, y_g) [x_g \cdot y_g - E(XY | x, y < x_g, y_g)]}{x_g \cdot y_g} \quad (10)$$

onde E é o operador de esperança, F a distribuição cumulativa da produtividade,  $x_g$  representa o preço garantido, que em geral baseia-se em uma proporção dos preços futuros negociados em bolsas de mercadorias e futuros;  $y_g$  diz respeito à produtividade garantida, obtida através de uma proporção da produtividade média histórica do produtor; x representa o preço ao final da safra do produtor enquanto que y representa a produtividade obtida ao final da safra. Esse seguro paga, de acordo com  $I = \max[(x_g \cdot y_g - x \cdot y); 0]$  caso haja perdas inesperadas na receita do produtor rural.

Nesse caso, a modelagem torna-se um pouco mais complexa. Em virtude da inclusão de mais uma variável aleatória, há necessidade de obter uma relação matemática entre distribuições multivariadas e suas distribuições marginais univariadas.

## 5. CÓPULAS

A dependência entre variáveis aleatórias  $X_1, X_2, \dots, X_n$ , é completamente descrita pela sua distribuição conjunta  $H(X_1, X_2, \dots, X_n)$ . Na maioria das situações aplicadas, apenas os marginais são conhecidas (estimadas ou fixadas a priori) e a distribuição conjunta pode ser desconhecida ou de difícil estimação. Na prática, a distribuição multivariada normal é muito utilizada em problemas multivariados. No entanto, essa suposição de normalidade restringe a tipo de associação entre as marginais, tornando-a linear e simétrica.

Por isso, a teoria de cópulas é uma ferramenta de modelagem multivariada muito utilizada em diferentes ramos da ciência, pois é possível identificar a estrutura de dependência linear, não linear, dependência em uma ou ambas as caudas da distribuição. A ideia de separar  $H$  em uma parte que descreve a estrutura de dependência e as partes que descrevem apenas o comportamento marginal levou ao conceito de cópulas, introduzido na literatura estatística por Sklar (Sklar, 1959).

De acordo com Joe (2014), a cópula é uma distribuição multivariada em que todas as distribuições unidimensionais marginais são uniformes  $U(0, 1)$ . Consequentemente, se  $C$  é uma cópula, então a distribuição de um vetor de variáveis aleatórias dependentes é  $U(0, 1)$ .

**Definição de cópulas:** Seja  $X_1, X_2, \dots, X_n$  variáveis aleatórias contínuas com função de distribuição conjunta  $H(X_1, X_2, \dots, X_n)$  e distribuições marginais  $F(X_1), F(X_2), \dots, F(X_n)$ , respectivamente. Para todo  $(x_1, \dots, x_n) \in [-\infty, \infty]^n$  considere o ponto em  $[0, 1]^{n+1}$  com coordenadas  $(F_{x_1}(x_1), \dots, F_{x_n}(x_n))$ ,  $H(x_1, x_2, \dots, x_n)$ . Este mapeamento de  $[0, 1]^n$  a  $[0, 1]$  é uma cópula  $n$ -dimensional.

A definição de cópula é conectada pelo teorema de Sklar (Sklar, 1959) que é o principal resultado na teoria da cópula.

**Teorema de Sklar.** Para uma distribuição d-dimensional  $F \in \mathcal{F}(F_1, F_2, \dots, F_d)$  em que  $F_j$  é a j-ésima distribuição univariada marginal, a cópula associada a  $F$  é uma função de distribuição  $C : [0,1]^d \rightarrow [0, 1]$ ,  $F(y) = C(F_1(y_1), F_2(y_2), \dots, F_d(y_d))$ ,  $y \in \mathbb{R}^d$ , com marginais  $U(0,1)$  que satisfazem:

1. Se  $F$  é uma distribuição d-dimensional contínua com marginais uniformes  $F_1, \dots, F_d$  e funções quantílicas  $F_{11}, \dots, F_{d1}$ , então:  $C(u) = F_{11}(u_1), \dots, F_{d1}(u_d)$ ,  $u \in [0,1]^d$  é única; e,
2. Se  $F$  é uma distribuição d-dimensional de variáveis aleatórias discretas, então a cópula é única somente no conjunto  $\text{Dom}(F_1) \times \dots \times \text{Dom}(F_d)$ .

A cópula  $C$  é parametrizada por um vetor de parâmetros  $\delta$ , chamado de parâmetro de dependência multivariado. As famílias de cópulas são geralmente dadas como função de distribuição acumulada (para a inferência via verossimilhança) e a cópula densidade é obtida pela diferenciação. Se  $C(u)$  é uma distribuição acumulada absolutamente contínua então sua função densidade é dada por:  $c(u) = c(u_1, \dots, u_d) = \frac{\partial^d C(u)}{\partial u_1 \dots \partial u_d}$ ,  $u \in [0,1]^d$

## 5.1 INFERÊNCIA DAS CÓPULAS

Suponha uma amostra independente e identicamente distribuída  $y_{i1}, \dots, y_{id}$ ,  $i = 1, \dots, d$ , então a distribuição conjunta é  $F(y; \eta_1, \dots, \eta_d, \delta) = C(F_1(y, \eta_1), \dots, F_d(y, \eta_d), \delta)$ . Se todas as variáveis forem contínuas e  $F_1, \dots, F_d, C$  forem absolutamente contínuas com relação as densidades  $f_1, \dots, f_d, c$ , então a log-verossimilhança é dada por:

$$l(\theta) = \sum_{i=1}^n [\log c(F_1(y_{i1}; \eta_1), \dots, F_d(y_{id}; \eta_d); \delta) + \sum_{j=1}^d \log f_j(y_{ij}; \eta_j)], \quad (11)$$

em que  $\theta = (\eta_1, \dots, \eta_d, \delta)t$  é o vetor de todos os parâmetros das marginais e da cópula. Assim, o estimador de máxima verossimilhança (EMV) é obtido maximizando a expressão acima,  $\hat{\theta}_{EMV} = \max_{\theta} l(\theta)$ .

Sob condições usuais de regularidade, o estimador de máxima verossimilhança existe e é consistente e assintoticamente eficiente. Além disso, verifica-se a propriedade de assintoticamente normal:  $\sqrt{d}(\hat{\theta}_{EMV} - \theta_0) \rightarrow N(0, \mathfrak{F}^{-1}(\theta_0))$ , em que  $\mathfrak{F}$  é a matriz de informação de Fisher e  $\theta_0$  o valor verdadeiro usual.

Neste método é necessário estimar conjuntamente os parâmetros das distribuições marginais e os parâmetros da cópula, por isso, o método de máxima verossimilhança pode ser computacionalmente dispendioso. Assim, a literatura sugere a estimação dos parâmetros em dois passos (Joe e Xu, 1996):

No primeiro passo, estima-se os parâmetros das marginais  $\theta_1 = (\eta_1, \dots, \eta_d)$  ao realizar as estimativas das marginais  $\hat{\theta}_1 = \max \sum_{i=1}^n \sum_{j=1}^d \log f_j(y_{i,j}; \eta_j)$ .

No segundo passo, dado  $\hat{\theta}_1$  estima-se os parâmetros da cópula  $\theta_2$ :

$$\hat{\theta}_2 = \max \sum_{i=1}^n [\log c(F_1(y_{i1}), \dots, F_d(y_{id}); \theta_2; \hat{\theta}_1)] \quad (12)$$

Este método é chamado de inferência por marginais (IFM). O estimador IFM é definido como o vetor  $\hat{\theta}_{IFM} = (\hat{\theta}_1, \hat{\theta}_2)t$ . Joe e Xu (1996) mostra que o método IFM é altamente eficiente quando comparado com o método de MV.

## 5.2 MEDIDAS DE AJUSTE DOS MODELOS DE CÓPULAS

Para a seleção do modelo de cópulas os critérios mais utilizados são o critério de informação Akaike (AIC), proposto por Akaike (1974) e o critério Bayesiano (BIC), proposto por Schwarz (1978), conforme descrito na seção 3.6.

## **6. APLICAÇÃO DA METODOLOGIA PARA O CASO DA SOJA NO ESTADO DO PARANÁ: COMPARATIVO COM O PROAGRO**

O problema da precificação mal especificada é conhecido no mercado segurador. Notoriamente, a causa está relacionada não apenas a falta de dados estatísticos adequados aos cálculos das taxas de prêmio, mas também ao desconhecimento de metodologias que acomodem satisfatoriamente as peculiaridades do setor agrícola.

A implicação direta deste fato é a definição de taxas que não refletem adequadamente o risco de cada unidade segurada, sejam eles, estados, municípios, propriedades agrícolas ou mesmo talhões dentro de uma propriedade. Em países em que o seguro agrícola está altamente desenvolvido é comum existirem produtos que cobrem talhões ao invés de propriedades. Nesse caso, a taxa deve ser fixada em nível de talhão ao invés de propriedade.

No Brasil, até 2016, o PROAGRO definia uma alíquota única para todas as culturas agrícolas em todo o território nacional repetindo o modelo de tarifação da CNSA, que juntamente com outros fatores, levaram ao fracasso da companhia e a sua dissolução em 1966.

A partir de 2017, houve uma mudança na tarifação do PROAGRO. As alíquotas passaram a se diferenciar por culturas, mas ainda assim, permaneceram as mesmas para todo o território nacional. O Manual do Crédito Rural define as seguintes alíquotas em empreendimentos em regime de sequeiro:

I - milho (verão) e soja: 4,5%;

II - milho safrinha (2ª safra): 6%;

III - ameixa, maçã, nectarina e pêsego, sem estrutura de proteção contra granizo: 6,5%;

IV - ameixa, maçã, nectarina e pêsego, com estrutura de proteção contra granizo: 3,5%;

V - cevada e trigo: 6,5%; e,

VI - demais culturas zoneadas: 4%.

Apesar da diferenciação por tipo de cultura, o problema da má especificação ainda persiste. O fato é que o PROAGRO não distingue entre os produtores de soja, por exemplo, aqueles de risco maior ou risco menor. A consequência desse problema é o aumento da



antiseletividade, ou seja, a concentração de indivíduos de maior risco na carteira, o que pode levar a elevados prejuízos e déficit nos resultados da carteira.

Por definição, indivíduos de maior risco devem possuir taxas maiores, e vice-versa. Quando o PROAGRO define uma única taxa (alíquota) para todos os produtores de soja, dois graves problemas surgem: i) perda de receita, pois os produtores de alto risco deveriam pagar uma taxa bem maior do que aquela fixada no PROAGRO; e, ii) os produtores de baixo risco serão penalizados ao pagar uma taxa maior do que deveriam arcar, comprometendo sua margem operacional. Como a contratação do PROAGRO ou do seguro rural privado é compulsório para produtores que contratam crédito de custeio até R\$ 300 mil, a escolha muitas vezes é definida pelo preço. Possivelmente os produtores optarão pelo seguro privado, em função da subvenção ao prêmio.

Para que o exercício de cálculo das taxas de prêmio tenha efetividade e se aproxime à realidade do mercado usamos as mesmas fontes de dados das companhias credenciadas no PSR, ou seja, os dados de produtividade municipais do IBGE. A base de dados do IBGE é reconhecida pelo mercado segurador e, principalmente, pelos resseguradores como a principal fonte de dados para a precificação do risco agrícola, por pertencer a uma instituição governamental idônea no processo. Por esse motivo, a taxa de prêmio foi calculada utilizando os dados da produtividade média municipal do IBGE/SIDRA que pode ser acessada no link <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>

Com raras exceções, a taxa de prêmio é calculada com base nos dados das propriedades rurais pertencentes a determinadas cooperativas, principalmente do estado do Paraná. Nesses casos, é preciso um trabalho minucioso feito pelas seguradoras e aceito pelos resseguradores, no sentido de mostrar a confiabilidade na fonte supracitada.

Nesse contexto, para fins de comparação, calculamos as taxas de prêmio por municípios do estado do Paraná, para a cultura da soja, conforme metodologia descrita neste produto. Ao todo, 332 municípios foram selecionados para compor a amostra. Os dados de produtividade municipal utilizados correspondem ao período de 1974 a 2017 e são provenientes do IBGE no link <https://sidra.ibge.gov.br/pesquisa/pam/tabelas>.

Nesse exercício é importante verificar o impacto quando se define as taxas por municípios ao invés de uma única taxa para todo o país. Observa-se uma grande variabilidade de taxas entre os municípios e que são completamente distintos da linha vermelha, que representa a taxa do PROAGRO (Fig. 3)

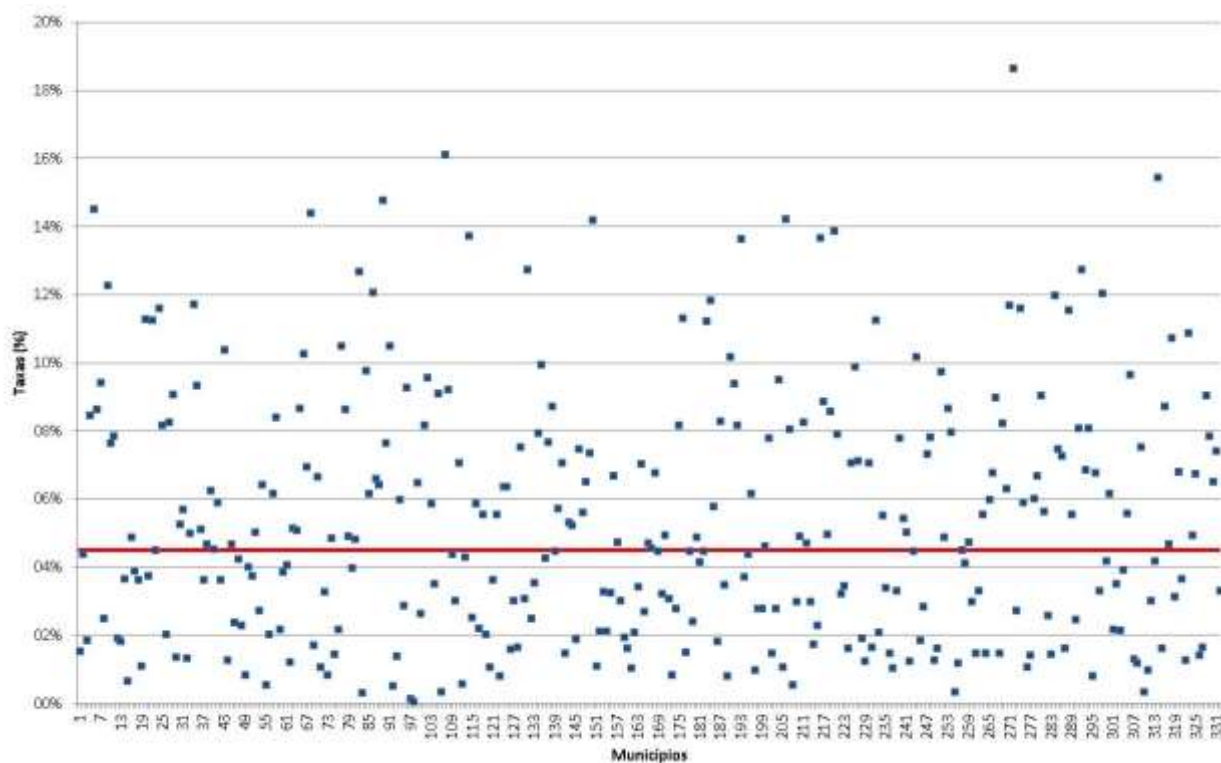


Figura 3. Comparativo entre a taxa do PROAGRO (4,5%) e as taxas municipalizadas calculadas pela consultoria.

Na figura, observa-se duas taxas próximas a 0% refletindo o risco muito baixo desses municípios. Aqui vale uma ressalva: as taxas calculadas neste exercício refletem apenas o risco puro da atividade. Normalmente, sobre o risco puro é acrescido uma taxa de carregamento, que corresponde ao custo tributário, administrativo, operacional e o lucro das seguradoras.

É importante ressaltar que os pontos acima da linha vermelha representam os municípios que possuem taxas maiores do que a atual taxa cobrada pelo programa (4,5%). Nesse contexto é possível ver claramente a receita perdida. Por outro lado, existe um conjunto de produtores que deveriam pagar menos por estarem em municípios com riscos menores. Ao todo, 53% dos municípios apresentaram taxas maiores do que o PROAGRO, e 47%, taxas menores.

Quando se avalia a diferença entre a taxa calculada e a taxa do PROAGRO (4,5%), nota-se uma variabilidade bastante significativa. As diferenças podem chegar a quase a um

valor mínimo de -4,4% a um valor máximo de 14,2% (Fig. 4). Ademais, percebe-se a assimetria em direção as diferenças positivas.

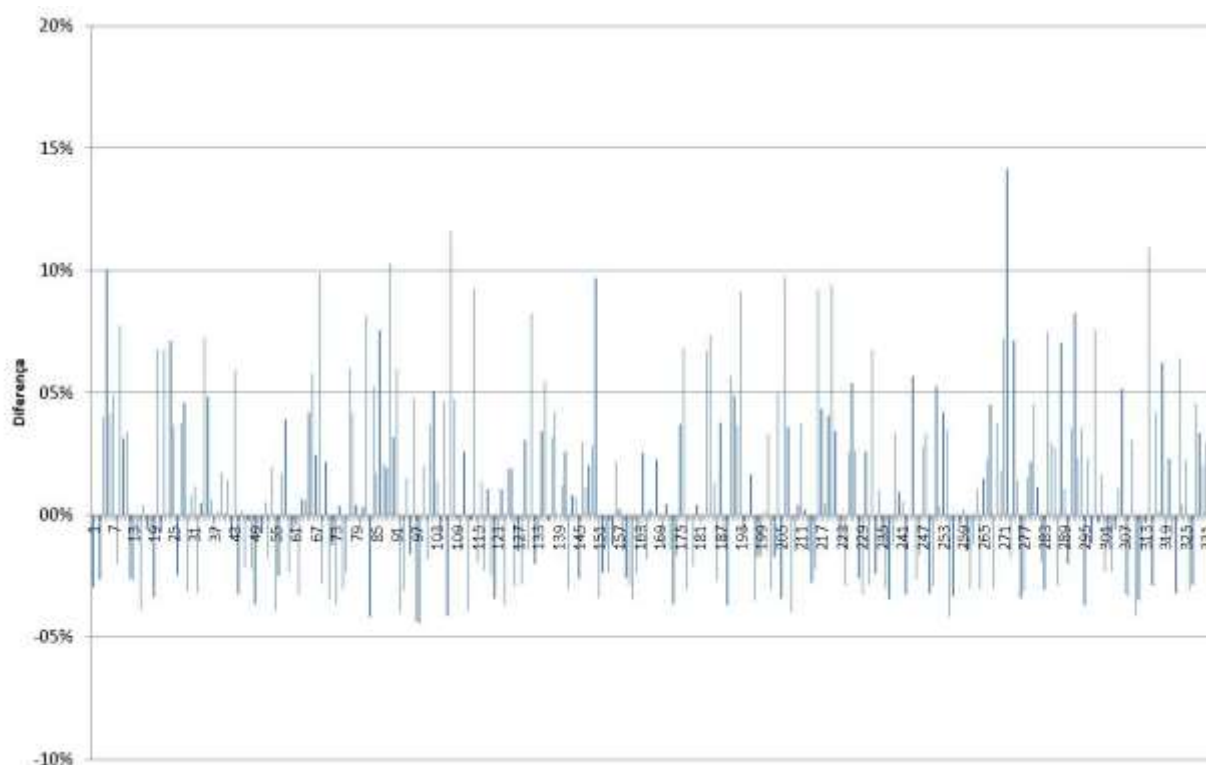


Figura 4. Diferença entre a taxa calculada usando a metodologia proposta neste projeto e a taxa de 4,5% do PROAGRO.

Para se ter uma ideia do ganho financeiro em precificar de acordo com a recomendação metodológica é preciso quantificar as perdas impostas pelo atual modelo de tarifação. Para isso é necessário levantar os dados do volume de crédito rural de custeio contratado no Paraná no ano safra 2017/18, em cada um dos municípios selecionados.

Utilizando a base do Sicor, levantamos esses dados. Para a soja e em todos os municípios selecionados, o valor total do custeio contratado foi de R\$ 4,7 bilhões na safra 2017/18. Assim, comparando a receita total utilizando a taxa única de 4,5% versus a receita total com a nova proposta, teremos uma boa noção do ganho em se aprimorar o modelo atual de precificação. Utilizando a taxa calculada para cada um dos municípios e multiplicando-os pelos respectivos valores de crédito de custeio, obtemos a PERDA DE RECEITA DO PROAGRO, da ordem de **R\$ 88,3 milhões**.

Ademais, ressalta-se que além da perda de receita pela má especificação da taxa, há o fenômeno da migração dos produtores de baixo risco para o seguro privado, já que em função da subvenção ao prêmio, a taxa é muitas vezes menor do que no PROAGRO. Nesse caso, a PERDA DE RECEITA EM FUNÇÃO DA MIGRAÇÃO para o seguro privado é da ordem de **R\$ 53,8 milhões**. Essa é uma fonte de ineficiência no PROAGRO, na medida em que deixa na carteira apenas os produtores de risco maior, com maior probabilidade de perdas na lavoura, e conseqüentemente, maior a chance do PROAGRO ter que pagar as indenizações. O aumento das indenizações reflete-se em aumento da chance de déficit na carteira e possíveis dispêndios não programados pelo Governo.

Se considerarmos as duas perdas, atinge-se um valor máximo (teto) de perda da ordem de **R\$ 142,1 milhões**. Esse valor representa a perda de receita apenas para a soja em um único estado. Cabe ressaltar que o PROAGRO atua com muitas outras culturas em praticamente todo o território nacional, portanto a perda de receita é muito maior do que o valor supramencionado.

## **7. CONSIDERAÇÕES FINAIS**

O estudo mostra uma ampla revisão de literatura sobre a precificação de risco agrícola e detalha a sua metodologia atuarial. A formulação apresentada segue os padrões internacionais utilizados, principalmente, pelo programa de seguro agrícola norte-americano que é o maior programa em nível mundial.

A metodologia presente neste estudo pode ser utilizada tanto para as apólices do mercado privado de seguros quanto para o PROAGRO. Uma vez definido o produto de seguros a ser precificado, depura-se os dados com o objetivo de verificar a presença de tendência e retirá-la da série de produtividade. Na sequência é preciso ajustar as distribuições aos dados, dentre as apresentadas neste estudo, ou seja, a Normal Assimétrica, t-Assimétrica, Odd Log-Logística e a modelagem não paramétrica.

As distribuições paramétricas supramencionadas são suficientes para captar certas idiosincrasias presentes nos dados, por exemplo, assimetrias e bimodalidades. Outras distribuições, por exemplo, Normal, Gama e Beta podem ser consideradas casos particulares das distribuições apresentadas.

Após o ajuste das distribuições é preciso escolher a distribuição que melhor representa os dados. Os critérios AIC e BIC podem ser utilizados para a escolha de modelos. Recomenda-se utilizar apenas um dos critérios apresentados e não os dois concomitantemente. Uma vez definida a distribuição de probabilidade o próximo passo consiste no cálculo das taxas de prêmio. Caso o produto escolhido seja o seguro de faturamento, há ainda um passo complementar que é a escolha da cópula que melhor se ajusta aos dados. Após a escolha da cópula o atuário deverá seguir os passos anteriores para realizar a precificação de risco.

Como exemplo prático, para mostrar os ganhos ao alterar a forma de precificação, vamos comparar as receitas obtidas com a forma atual e a receita com a nova proposta de precificação. Para isso, calculamos as taxas por municípios do Paraná, para a cultura da soja. Antes, porém, é importante destacar um ponto. As perdas financeiras no PROAGRO podem ocorrer de duas formas distintas: i) perdas de receita pela má especificação da taxa, ou seja, não se distingue entre municípios de alto, médio e baixo risco, associando a todos uma única taxa. Dessa forma, perde-se receita, pois a taxa atual de 4,5% é menor do que aquela que deveria ser cobrada; e, ii) perdas de receita devido a migração de produtores ao seguro privado, pois não estão dispostos a pagar pelo PROAGRO e enxergam no seguro privado uma solução mais barata, devido a subvenção ao prêmio.

Utilizando a base do Sicor, levantamos esses dados. Para a soja e em todos os municípios selecionados, o valor total do custeio contratado foi de R\$ 4,7 bilhões na safra 2017/18. Assim, comparando a receita total utilizando a taxa única de 4,5% versus a receita total com a nova proposta, teremos uma boa noção do ganho em se aprimorar o modelo atual de precificação. Utilizando a taxa calculada para cada um dos municípios e multiplicando-os pelos respectivos valores de crédito de custeio, obtemos a PERDA DE RECEITA DO PROAGRO, da ordem de **R\$ 88,3 milhões**.

Ademais, ressalta-se que além da perda de receita pela má especificação da taxa, há o fenômeno da migração dos produtores de baixo risco para o seguro privado, já que em função da subvenção ao prêmio, a taxa é muitas vezes menor do que no PROAGRO. Nesse caso, a PERDA DE RECEITA EM FUNÇÃO DA MIGRAÇÃO para o seguro privado é da ordem de **R\$ 53,8 milhões**. Essa é uma fonte de ineficiência no PROAGRO, na medida em que deixa na carteira apenas os produtores de risco maior, com maior probabilidade de perdas na lavoura, e conseqüentemente, maior a chance de o PROAGRO ter que pagar as indenizações.

O aumento das indenizações reflete-se em aumento da chance de déficit na carteira e possíveis dispêndios não programados pelo Governo.

Se considerarmos as duas perdas, atinge-se um valor máximo (teto) de perda da ordem de **R\$ 142,1 milhões**. Esse valor representa a perda de receita apenas para a soja em um único estado. Cabe ressaltar que o PROAGRO atua com muitas outras culturas em praticamente todo o território nacional, portanto a perda de receita é muito maior do que o valor supramencionado.

## ANEXO: ETAPAS DA METODOLOGIA DE PRECIFICAÇÃO

I) Tratamento da série de produtividade quanto a tendência e autocorrelação.

II) Seleção da distribuição de probabilidade (Normal, Skew-Normal, OLLN e Skew-t) que melhor se ajusta aos dados, através do procedimento descrito em 3.

III) Cálculo da distribuição acumulada de probabilidade, ou também chamada de probabilidade de perda, escolhida para o valor de  $Y_g$  (produtividade garantida pelo seguro).

IV) Cálculo da Expectativa de Perda, através da fórmula:

$$E(perda) = E(\max(y_g - Y); 0) = (\lambda y_e - E(Y|Y < \lambda y_e)).$$

Observe que é necessário calcular a probabilidade condicional acumulativa da produtividade dado que ela é menor do que a produtividade garantida.

V) Cálculo da Taxa de Prêmio, através da fórmula:

$$TP = \frac{F_y(\lambda y^e) E_y[\lambda y^e - (Y|y < \lambda y^e)]}{\lambda y^e}$$

Em que  $E$  é o operador de esperança e  $F$  a distribuição cumulativa da produtividade (probabilidade de perda),  $\lambda$  representa o nível de cobertura, tal que  $0 < \lambda < 1$ , e  $y^e$  é a produtividade esperada. O numerador é formado por dois termos, a probabilidade de perda e a perda média dada pelo operador de esperança, ou seja, a média das perdas considerando toda a série temporal. A multiplicação da probabilidade de perda pela perda esperada, dividida pela produtividade garantida resulta na taxa de prêmio para cada nível de cobertura.

## 2. TRATAMENTO PRELIMINAR DOS DADOS

Para a precificação do risco de produtividade municipal é necessário fazer um tratamento preliminar nos dados com relação à tendência crescente da produtividade ao longo dos anos, de acordo com a metodologia proposta por Gallagher (1987).

Uma regressão linear é ajustada aos dados de produtividade tendo o tempo como a variável explicativa, então o último rendimento observado é corrigido utilizando-se os resíduos do modelo para cada ano. Os rendimentos corrigidos pela tendência são definidos pela seguinte equação:

$$\tilde{y}_t = \widehat{y}_{2016} \left( 1 + \frac{\hat{e}_t}{\hat{y}_t} \right)$$

onde  $\tilde{y}_t$ ,  $\hat{y}_t$  e  $\hat{e}_t$  são, respectivamente, o rendimento corrigido, o rendimento ajustado e o residual do ano  $t$ ,  $y_{2016}$  é o rendimento ajustado para 2016.

No próximo passo foi utilizado o teste de Ljung and Box(1978) para tratar a autocorrelação serial da produtividade. Ljung e Box propuseram uma pequena alteração no teste de Box - Pierce, na qual torna o teste mais generalizado. Foi observado na literatura que o teste de Ljung - Box apresenta melhores resultados do que o teste de Box - Pierce.

As hipóteses do teste são as mesmas:

$$\begin{cases} H_0: Os\ resíduos\ são\ i.\ i.\ d.; \\ H_1: Os\ resíduos\ não\ são\ i.\ i.\ d.. \end{cases}$$

Calculamos as estimativas de autocorrelações por



$$\widehat{r}_k = \frac{\sum_{t=k+1}^n \widehat{a}_t \widehat{a}_{t-k}}{\sum_{t=1}^n \widehat{a}_t^2}$$

É possível mostrar que  $\widehat{r}_k \sim N\left(0, \frac{1}{n}\right)$ . Se o modelo for apropriado, a estatística do teste

$$Q(k) = n(n-2) \sum_{j=1}^K \frac{\widehat{r}_j^2}{(n-j)}$$

terá aproximadamente uma distribuição  $\chi^2$  com  $(K - p - q)$  graus de liberdade, onde  $K$  é o número de defasagens tomada na função de autocorrelação,  $p$  e  $q$  são as ordens do modelo ajustado. Portanto, rejeitamos a hipótese nula se  $Q > \chi_{1-\alpha, k-p-q}^2$  com um nível de significância  $\alpha$ .

Para a modelagem da produtividade testa-se modelos paramétricos simétricos e assimétricos, que incorporam a unimodalidade e bimodalidade dos dados. Todas estas análises foram realizadas no software R utilizando os pacotes “Optim” e “AdequacyModel”.

Num terceiro momento foi utilizado o teste de Breusch and Pagan (1979) para tratar a heterogeneidade de variâncias. Baseado no teste multiplicador de Lagrange, o teste de Breusch-Pagan é bastante utilizado para testar a hipótese nula de que as variâncias dos erros são iguais (homoscedasticidade) versus a hipótese alternativa de que as variâncias dos erros são uma função multiplicativa de uma ou mais variáveis, sendo que esta(s) variável(eis) pode(m) pertencer ou não ao modelo em questão.

Inicialmente, ajusta-se um modelo de regressão linear (simples ou múltiplo) e obtém-se os resíduos  $e = (e_1, \dots, e_n)$  e os valores ajustados  $\widehat{y} = (\widehat{y}_1, \dots, \widehat{y}_n)$ . Em seguida, consideram-se os resíduos ao quadrado e estes são padronizados de modo que a média do vetor de resíduos padronizados seja igual a 1. Esta padronização é feita dividindo cada resíduo ao quadrado

pela  $SQE/n$  em que  $SQE$  é a Soma de Quadrados dos Resíduos do modelo ajustado e  $n$  é o número de observações. Desta forma, temos que cada resíduo padronizado é dado por

$$u_i = \frac{e_i^2}{SQE/n}, i = 1, \dots, n, \text{ em que } SQE = \sum_{i=1}^n e_i^2.$$

Por fim, faz-se a regressão entre  $u = (u_1, \dots, u_n)$  (variável resposta) e o vetor  $\hat{y}$  (variável explicativa) e obtem-se a estatística do teste  $\chi_{BP}^2$  calculando-se a Soma de Quadrados da Regressão de  $u$  sobre  $\hat{y}$  e dividindo-se o valor encontrado por 2. Sob a hipótese nula, esta estatística tem distribuição qui-quadrada com 1 grau de liberdade. Para a modelagem da produtividade foram testados modelos paramétricos simétricos e assimétricos, que incorporam a unimodalidade e bimodalidade dos dados. Todas estas análises foram realizadas no software R utilizando os pacotes “Optim” e “AdequacyModel”.

Testa-se também se os dados seguem uma distribuição Normal, por meio do teste de Kolmogorov-Smirnov (KS). O teste de Kolmogorov - Smirnov pode ser utilizado para avaliar as hipóteses:  $H_0$ : os dados seguem uma distribuição Normal, contra  $H_1$ : os dados não seguem uma distribuição Normal.

Este teste observa a máxima diferença absoluta entre a função de distribuição acumulada dos dados, no caso a Normal, e a função de distribuição empírica dos dados. Como critério, comparamos esta diferença com um valor crítico, para um dado nível de significância. O valor-p resultante do no programa R pode ser interpretado como a medida do grau de concordância entre os dados e a hipótese nula. Quanto menor for o valor-p, menor é a consistência entre os dados e a hipótese nula. Então, a regra de decisão é rejeita-se  $H_0$ , se o valor-p for menor ou igual a um certo valor arbitrado pelo analista.

Questões relativas à melhoria do ponto 6 ("Aplicação da metodologia para o caso da soja no Estado do Paraná: comparativo do Proagro") do Produto 6

### 3. AJUSTE E TESTES DAS DISTRIBUIÇÕES DE PROBABILIDADE

As distribuições de probabilidade consideradas para a análise foram:

Skew-Normal

$$f(x) = \frac{2}{\omega} \phi\left(\frac{x-\xi}{\omega}\right) \Phi\left(\alpha\left(\frac{x-\xi}{\omega}\right)\right)$$

Onde  $\phi$  representa a densidade de probabilidade da distribuição normal e  $\Phi$  a sua densidade acumulada.

Odd log-logística normal (OLLN)

$$F(x; \mu, \sigma, \alpha) = \int_0^{\frac{\Phi(x; \mu, \sigma)}{\bar{\Phi}(x; \mu, \sigma)}} \frac{\alpha t^{\alpha-1}}{(1+t^\alpha)^2} dt = \frac{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right)}{\Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha}$$

em que  $\bar{\Phi}(x; \mu, \sigma) = 1 - \Phi(x; \mu, \sigma)$ . A fdp é dada por:

$$f(x; \mu, \sigma, \alpha) = \frac{\alpha \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{x-\mu}{\sigma}\right) \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^{\alpha-1}}{\sigma \left\{ \Phi^\alpha\left(\frac{x-\mu}{\sigma}\right) + \left[1 - \Phi\left(\frac{x-\mu}{\sigma}\right)\right]^\alpha \right\}^2}$$

respectivamente. Note que  $\alpha > 0$  é um parâmetro de forma.

Skew-t

$$f_{ST}(x; \mu, \sigma, \lambda, p, q) = \frac{p}{2\nu\sigma q^{1/p} B\left(\frac{1}{p}, q\right) \left( \frac{|x-\mu+m|^p}{q(\nu\sigma)^p (\lambda \text{sign}(x-\mu+m) + 1)^p} + 1 \right)^{\frac{1}{p}+q}}$$

onde  $B$  é a função beta,  $\mu$  é o parâmetro de localização,  $\sigma > 0$  é o parâmetro de escala,  $-1 < \lambda < 1$  é o parâmetro de assimetria, e  $p > 0$  e  $q > 0$  são os parâmetros que controlam a curtose. Note que  $m$  e  $v$  não são parâmetros, mas funções dos outros parâmetros que são usados aqui para mudar a escala ou deslocar a distribuição. Na parametrização original,

$$m = \frac{2\nu\sigma\lambda q^{\frac{1}{p}} B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)}$$

e

$$v = \frac{q^{\frac{1}{p}}}{\sqrt{(3\lambda^2 + 1) \frac{B\left(\frac{3}{p}, q - \frac{2}{p}\right)}{B\left(\frac{1}{p}, q\right)} - 4\lambda^2 \frac{B\left(\frac{2}{p}, q - \frac{1}{p}\right)^2}{B\left(\frac{1}{p}, q\right)^2}}}$$

Para escolha da melhor distribuição paramétrica foram utilizados os critérios AIC de Akaike (1974) e BIC, Schwarz et al (1978). O Critério de Informação de Akaike (AIC) é definido como

$$AIC_p = -2\log(L_p) + 2[(p + 1) + 1]$$

em que  $L_p$  é a função de máxima verossimilhança do modelo e  $p$  é o número de variáveis explicativas consideradas no modelo.

O Critério de Informação Bayesiano (BIC) é definido como

$$BIC_p = -2\log(L_p) + [(p + 1) + 1]\log(n).$$

Tanto o AIC quanto o BIC aumentam conforme a  $SQE$  aumenta. Além disso, ambos critérios penalizam modelos com muitas variáveis sendo que valores menores de AIC e BIC são preferíveis.

Como modelos com mais variáveis tendem a produzir menor  $SQE$  mas usam mais parâmetros, a melhor escolha é balancear o ajuste com a quantidade de variáveis.

Além disso, pode-se utilizar estatísticas modificadas de Anderson-Darling ( $W^*$ ) e Cramer-von Mises ( $A^*$ ), proposto por Lin et al. (2008) e Pakyari and Balakrishnan (2012), respectivamente.

Seja  $x_1, x_2, \dots, x_n$  uma amostra aleatória e suponha que um provável candidato para a f.d.a. dos dados seja  $F(x)$ , o teste de hipóteses para verificar a adequação da distribuição  $F(x)$  aos dados é

$$\begin{cases} H_0: a \text{ amostra tem distribuição } F(x) \\ H_1: a \text{ amostra não tem distribuição } F(x) \end{cases} \quad (1)$$

Anderson e Darling (1952, 1954) propuseram a seguinte estatística para testar (1):

$$A^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)(1 - F(x))} dF(x)$$

sendo  $F_n(x)$  a função de distribuição acumulada empírica definida por

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i \leq x\}} = \begin{cases} 0, & \text{se } x < x_{(1)}, \\ \frac{k}{n}, & \text{se } x_{(k)} \leq x < x_{(k+1)}, \\ 1, & \text{se } x > x_{(n)}, \end{cases}$$

sendo  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ , as estatísticas de ordem da amostra aleatória e  $1_{\{x_i \leq x\}}$  a função indicadora que vale 1 se  $x_i \leq x$  e 0 se  $x_i > x, i = 1, \dots, n$ .

A estatística  $A^2$  pode ser representada numa forma equivalente como

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n \left[ (2i - 1) \ln(F(x_{(i)})) + (2(n - i) + 1) \ln(1 - F(x_{(i)})) \right]$$

A transformação  $F(x_{(i)})$  leva  $x_{(i)}$  em  $U_{(i)}$ , sendo  $U_{(1)}, \dots, U_{(n)}$  uma amostra de tamanho  $n$  com distribuição uniforme em  $(0,1)$ . Logo,

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n \left[ (2i - 1) \ln(U_{(i)}) + (2(n - i) + 1) \ln(1 - U_{(i)}) \right]$$

Para calcular o valor da estatística  $A^2$ , devemos seguir os passos abaixo:

1. Ordene os valores da amostra:  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ ;
2. Quando necessário, estime os parâmetros da distribuição de interesse;
3. Calcule  $U_i = F(x_{(i)})$  e calcule o valor da estatística de Anderson Darling:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n \left[ (2i - 1) (\ln(U_i) + \ln(1 - U_{n+1-i})) \right]$$

4. Para cada uma das distribuições calcule, se for o caso, o valor da estatística modificada de acordo com as tabelas dadas para cada uma delas.

Similarmente se obtém a estatística de Cramér-von-Mises:

$$W^2 = \sum_{i=1}^n \left( U_i - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n}$$

#### 4. CÁLCULO DA PROBABILIDADE DE PERDA

Para o cálculo da probabilidade de perda é feito o cálculo da distribuição acumulada no valor de  $Y_g$  (produtividade garantida).

Depois de selecionada a função de distribuição que melhor se ajustou os dados de acordo com o foi calculada densidade acumulada de cada distribuição utilizando-se o software R. Esta fase da modelagem tem como input os níveis de cobertura desejados e como output os valores de probabilidade de perda.

#### 5. CÁLCULO DA TAXA DE PRÊMIO

O cálculo da taxa de prêmio é realizada pela equação

$$TP = \frac{F_y(\lambda y^e) E_y[\lambda y^e - (Y|y < \lambda y^e)]}{\lambda y^e}$$

Em que  $E$  é o operador de esperança e  $F$  a distribuição cumulativa da produtividade (probabilidade de perda),  $\lambda$  representa o nível de cobertura, tal que  $0 < \lambda < 1$ , e  $y^e$  é a produtividade esperada. O numerador é formado por dois termos, a probabilidade de perda e a perda média dada pelo operador de esperança, ou seja, a média das perdas considerando toda a série temporal. A multiplicação da probabilidade de perda pela perda esperada, dividida pela produtividade garantida resulta na taxa de prêmio para cada nível de cobertura.

## REFERÊNCIAS

Adami, A. C. O., & Ozaki, V. A. (2016). Estudo Sobre a Viabilidade do Seguro Receita: o caso da região oeste do Paraná. *Revista de Economia e Sociologia Rural*, 54(2), 305-318.

Akaike, H. (1974) A new look at the statistical model identification, *Automatic Control, IEEE Transactions on* 19, 716–723.

Alizadeh, M; Emadi, M.; Doostparast, M.; Cordeiro, G.M.; Ortega, E.M.; Pescim, R.R. (2015) A new family of distributions: the kumaraswamy odd log-logistic, properties and applications, *Hacettepe Journal of Mathematics and Statistics*, 44.

Azzalini, A. A class of distributions wich includes the normal ones, *Scand. J. Statist.* 12 (1985), pp. 171-178.

Azzalini, A.; Capitanio, A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 367–389.

Booth, P.; Chadburn, R.; Cooper, D.; Haberman, S.; James, D. (1999) *Modern actuarial theory and practice*. London: Chapman & Hall/CRC, 716p.

Botts, R.R.; Boles, J.N. (1958) Use of normal-curve theory in crop insurance ratemaking, *Journal of Farm Economics* 40, 733-740.

Braga, A. S.; Cordeiro, G.M.; Ortega, E.M.; Cruz, J.N. (2016) The odd log–logistic normal distribution: Theory and applications in analysis of experiments, *Journal of Statistical Theory and Practice*, 1–25.

Brisolara, C. S. (2013). *Proposições para o desenvolvimento do seguro de receita agrícola no Brasil: do modelo teórico ao cálculo das taxas de prêmio*. Tese de Doutorado, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. doi:10.11606/T.11.2013.tde-02102013-141823. Recuperado em 2019-03-20, de [www.teses.usp.br](http://www.teses.usp.br).

Cruz, J.N. (2016) *A nova família de distribuições odd log-logística: teoria e aplicações*, Ph.D. diss., Escola Superior de Agricultura “Luiz de Queiroz”.



Cruz, J.N.; Ortega, E.M.; Cordeiro, G.M. (2015) The log-odd log-logistic weibull regression model: modelling, estimation, influence diagnostics and residual analysis, *Journal of Statistical Computation and Simulation*, 1–23.

Day, R.H. (1965) Probability distributions of field crop yields, *Journal of Farm Economics*, 47, 713-741.

Duarte, G. V. (2018). Precificação do seguro agrícola: novas abordagens utilizando distribuições de probabilidades alternativas e o uso de cópulas bidimensionais e tridimensionais. Tese de Doutorado, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. Recuperado em 2019-03-20, de <http://www.teses.usp.br/teses/disponiveis/11/11134/tde-17012019-152335/>.

Gallagher, P. (1987) U.S. Soybean yields: estimation and forecasting with nonsymmetric disturbances, *American Journal of Agricultural Economics*, 69, 796-803.

Goodwin, B.K.; Ker, A.P. (1998) Nonparametric estimation of crop yield distributions: implications for rating group-risk crop insurance contracts, *American Journal of Agricultural Economics*, 80, 139-153.

Goodwin, B.K.; Mahul, O. (2004) Risk modeling concepts relating to the design and rating of agricultural insurance contracts, World Bank.

Hart, D. G.; Buchanan, R. A.; Howe, B. A. (1996) The actuarial practice of general insurance. 5 ed. Australia: The Institute of Actuaries of Australia, 592p.

Joe, H. (2014) Dependence modeling with copulas. CRC Press.

Joe, H.; Xu, J. J. (1996) The estimation method of inference functions for margins for multivariate models. Technical report.

Just, R.E.; Weninger, Q. (1999) Are crop yields normally distributed?, *American Journal of Agricultural Economics*, 81, 287-304.

Ker, A.P.; Goodwin, B.K. (2000) Nonparametric estimation of crop insurance rates revisited, *American Journal of Agricultural Economics*, 83, 463-478.

Ker, A.P.; Coble, K. (2003) Modeling conditional yield densities, *American Journal of Agricultural Economics*, 85, 291-304.

Moss, C.B.; Shonkwiler, J.S. (1993) Estimating yield distributions with a stochastic trend and nonNormal errors, *American Journal of Agricultural Economics*, 75, 1056-1062.

Miqueleto, G. J. (2011). Contribuições para o desenvolvimento do seguro agrícola de renda para o Brasil: evidências teóricas e empíricas. Tese de Doutorado, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. doi:10.11606/T.11.2011.tde-12092011-163544. Recuperado em 2019-03-20, de [www.teses.usp.br](http://www.teses.usp.br).

Nelson, C.H.; Preckel, P.V. (1989) The conditional beta distribution as a stochastic production function, *American Journal of Agricultural Economics*, 71, 370-378.

Ozaki, V.A.; Goodwin, B. K. ; Ghosh, S. ; Shirota, R. (2008a) Spatio-temporal modeling of agricultural yield data applied to pricing crop insurance contracts. *American Journal of Agricultural Economics*, 90, 951-961.

Ozaki, V.A.; Goodwin, B. K. ; Shirota, R. (2008b) Parametric and nonparametric statistical modeling of crop yield: implications for pricing crop insurance contracts. *Applied Economics*, 40, 1151-1164.

Ozaki, V.A.; Silva, R. S. (2009) Bayesian ratemaking procedure of crop insurance contracts with skewed distribution. *Journal of Applied Statistics*, 36, 443-452.

Polo, L. (2016). Redes Bayesianas aplicadas a estimação da taxa de prêmio de seguro agrícola de produtividade. Dissertação de Mestrado, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. doi:10.11606/D.11.2016.tde-10082016-132524. Recuperado em 2019-03-20, de [www.teses.usp.br](http://www.teses.usp.br).

Ramirez, O.A. (1997) Estimation and use of a multivariate parametric model for simulating heteroskedastic, correlated, nonNormal random variables: the case of corn belt corn, soybean and wheat yields, *American Journal of Agricultural Economics*, 79, 191-205.

Ramirez, O.A.; Misra, S.; Field, J. (2003) Crop-yield distributions revisited, *American Journal of Agricultural Economics*, 85, 108-120.

Santos, C. O., Scalon, J. D., & Ozaki, V. A. (2014). A distribuição normal-assimétrica como modelo para produtividade de milho aplicada ao seguro agrícola. *Revista de Economia e Sociologia Rural*, 52(4), 725-742.

Schwarz, G. (1978) Estimating the dimension of a model, *The annals of statistics*, 6, 461–464.

Sherrick, B.J.; Zanini, F.C.; Schnitkey, G.D.; Irwin, S.H. (2004) Crop insurance valuation under alternative yield distributions, *American Journal of Agricultural Economics*, 86, 406-419.

Silverman, B.W. (1986) *Density estimation for statistics and data analysis*, Chapman and Hall, London, 175p.

Sklar, M. (1959) *Fonctions de repartition a n dimensions et leurs marges*. Universite Paris, 8.

Taylor, C.R. (1990) Two practical procedures for estimating multivariate nonNormal probability density functions, *American Journal of Agricultural Economics*, 72, 210-217.

Turvey, C.; Zhao, C. (1999) Parametric and nonparametric crop yield distributions and their effects on all-risk crop insurance premiums, Working paper, Dep. of Agricultural Economics and Business, University of Guelph.

## CÓDIGOS EM R

```
rm(list=ls(all=TRUE))

setwd("~/Área de Trabalho/PR/produtividade")

require(ggplot2)

dados=read.csv("nortecentralPR_IPARDES.csv",sep=";",header = TRUE)

(ano=dados[,1])

(y=dados[,2]) ## produtividade em cascavel

par(mfrow=c(2,2))

plot(ts(y,start=c(1980,1),end=c(2014,1)),type="o",lwd=1,main="Cascavel",ylim=c(500,3900))

hist(y,main="Cascavel",nclass=15)

#####

##### yt = alfa+ beta*t + epsilon

#####

y1=ts(y,start=c(1980,1),end=c(2014,1))

ano=dados[,1]

cor(y1,ano) # coeficiente de correlacao linear

m1= lm(y1~ano);m1 # MMQ

summary(m1)

par(mfrow=c(1,1))

plot(ts(y1,start=c(1980,1),end=c(2014,1)),type="o",lwd=1,main="Cascavel",ylim=c(500,3900))

abline(m1,col=20)

names(m1)

res= m1$residuals

ypred= m1$fitted.values
```

```

par(mfrow=c(2,2))

plot(m1)

shapiro.test(res)

shapiro.test(rstudent(m1))

m2= lm(y1~ano+l(ano^2));m2

summary(m2)

predict(m2)

lines(ano,predict(m2),col=2)

names(m2)

res= m2$residuals

ypred= m2$fitted.values

ynew= ypred[35]*(1+res/ypred)

ynew2=res+ ypred[35]

ynew=ts(ynew,start=c(1980,1),end=c(2014,1))

ynew2=ts(ynew2,start=c(1980,1),end=c(2014,1))

par(mfrow=c(1,1))

plot(y1,type="o",ylim=c(700,4200),main="Cascavel ",ylab="Productivity (Kg/ha)",xlab="Years")

lines(ynew,col="red",type="o",lwd=1,lty=2)

lines(ynew2,col="green ",type="o",lwd=1,lty=2)

legend('bottomleft', c("Original", "Corrigida Metodo Multiplicativo", "Corrigida Metodo Aditivo"),

      col=c('1', 'red', "green"),lty=c(1,2,2),lwd=c(1,1,1))

```

```

y=ynew
y=y1
plot(y)
hist(y,nclass=15)

####Estatistica DEscritiva

require(moments)

cbind(n=length(y),media=mean(y),mediana=median(y),desviopadiao=sd(y),assimetria=skewness(y),c
urtose= kurtosis(y))

cbind(max=max(y),min=min(y))# OBS A mediana > media entao tem dist. assimetrica a esquerda

# Verificacao da autocorrelaço

library(lmtest)

lmy=lm(y~ano)

#pvalor>0.05 entao nao rejeito H0

Box.test(y,type=c("Ljung-Box")) #pvalor>0.05 nao rejeito H0
Box.test(y*y,type=c("Ljung-Box")) #pvalor>0.05 nao rejeito H0

par(mfrow=c(2,2))

acf(y)

pacf(y)

acf(y*y)

pacf(y*y)

require(tseries)

m2= arima(y,order=c(1,0,0));m2

tsdiag(m2)

```

```
Box.test(m2$residuals,type=c("Ljung-Box"))
```

```
r=m2$residuals
```

```
plot(series)
```

```
#####
```

```
Ye=Produtividade esperada
```

```
Ye2=mean(y[30:35]);Ye2
```

```
xt<- c(0.6, 0.65, 0.7, 0.75, 0.8) # nivel de cobertura 70%, 75%,80,85 e 90%
```

```
Yg =seq(1,5)
```

```
for(i in 1:5) { Yg[i] = Ye2*xt[i]}
```

```
Yg # Produtividade garantida
```

```
#####
```

```
pdfnorm=function(par,x){
```

```
  mu = par[1]
```

```
  sigma = par[2]
```

```
  (1/(sigma*sqrt(2*pi)))*exp(-0.5*((x - mu)/sigma)^2)
```

```
}
```

```
cdfnorm=function(par,x){
```

```
  mu = par[1]
```

```
  sigma = par[2]
```

```
  pnorm(x,mean=mu,sd=sigma)
```

```
}
```

```

require(AdequacyModel)

aj=goodness.fit(pdf=pdfnorm,cdf=cdfnorm, starts=c(mean(y),sd(y)),data=y, method="Nelder-Mead",
domain=c(0,Inf));aj

aj$mle

result=goodness.fit(pdf=pdfnorm,cdf=cdfnorm, starts=c(aj$mle[1],aj$mle[2]),data=y,
method="Nelder-Mead", domain=c(0,Inf));result

#####

x = seq(min(y)-100,max(y)+150,0.1)

fnorm =pdfnorm(result$mle,x)

dev.off()

hist(y,freq=F,main="Cascavel",ylim=c(0,0.0015))

hist(y,freq=F,main="",ylim=c(0,0.0015),xlabel="")

#lines(density(y,bw="nrd0",kernel="gaussian"),col=6,lwd = 2,lty=2) #kernel

lines(x,fnorm , col = '1',lwd =2,lty=1) #Normal

####

plot(ecdf(y),main="Cascavel ")#ylabel="Cumulative Dist. Function",

lines(x,cdfnorm(result$mle,x),col=2,lwd = 1,lty=1) #Dist. Acumulada Normal

##cdfNormal =fcao probal. cumulativa

cbind("Prob. Perda(60%)" = cdfnorm(result$mle,Yg[1]),"Prob.
Perda(65%)"=cdfnorm(result$mle,Yg[2]),"Prob. Perda(70%)"=cdfnorm(result$mle,Yg[3]),

"Prob. Perda(75%)"=cdfnorm(result$mle,Yg[4]),"Prob. Perda(80%)"=cdfnorm(result$mle,Yg[5])) #
Probabilidade de Perda

#####

pdf_sn=function(par,x){

```



```

mu = par[1]
sigma = par[2]
gama=par[3]
z = (x - mu)/sigma
f=dnorm(z,0,1)
F=pnorm(gama*z)
(2/sigma)*f*F
}
require(sn)
cdf_sn=function(par,x){
  mu = par[1]
  sigma = par[2]
  gama=par[3]
  z = (x - mu)/sigma
  pnorm(z)-2*TOwen(z,gama)
}
aj=goodness.fit(pdf=pdf_sn,cdf=cdf_sn, starts=c(mean(y),sd(y),0.2),data=y, method="L",
domain=c(0,Inf));aj
aj$mle
result=aj
result=goodness.fit(pdf=pdf_sn,cdf=cdf_sn, starts=c(aj$mle[1],aj$mle[2],aj$mle[3]),data=y,
method="Nelder-Mead", domain=c(0,Inf));result
cbind(result$mle[1],result$Erro[1],result$mle[2],result$Erro[2],result$mle[3],result$Erro[3])
#####
x = seq(min(y)-100,max(y)+150,0.1)
fsn =pdf_sn(result$mle,x)
lines(x,fsn, col = '3', lwd = 2,lty=3) #Skew-Normal

```

```
#####Distribuicao acumulada

lines(x,cdf_sn(result$mle,x),col=3,lwd = 2,lty=2) #Dist. Acumulada Normal

cbind("Prob. Perda(70%" = cdf_sn(result$mle,Yg[1]),"Prob.
Perda(75%)=cdf_sn(result$mle,Yg[2]),"Prob. Perda(80%)=cdf_sn(result$mle,Yg[3]),

      "Prob. Perda(85%)=cdf_sn(result$mle,Yg[4]),"Prob. Perda(90%)=cdf_sn(result$mle,Yg[5]))
```

```
#####
```

```
pdf_olln=function(par,x){
  mu = par[1]
  sigma = par[2]
  alpha = par[3]
  z = (x - mu)/sigma
  f = (1/(sigma*sqrt(2*pi)))*exp(-0.5*(z)**2)
  F = pnorm(z)
  num= alpha*f*( F*(1-F))^(alpha-1)
  den = (F^(alpha) + (1-F)^(alpha))^2
  num/den
}
```

```
cdf_olln=function(par,x){
  mu = par[1]
  sigma = par[2]
  alpha = par[3]
  z = (x - mu)/sigma
  G = pnorm(z)
  (G^(alpha))/((G^(alpha))+(1-G)^(alpha))
}
```

```
aj=goodness.fit(pdf=pdf_oln,cdf=cdf_oln, start=c(mean(y),sd(y),2),data=y, method="Nelder-Mead",
domain=c(0,Inf))
```

```
aj$mle
```

```
result=goodness.fit(pdf=pdf_oln,cdf=cdf_oln, starts=c(aj$mle[1],aj$mle[2],aj$mle[3]),data=y,
method="L-BFGS-B", domain=c(0,Inf));result
```

```
#cbind(result$mle[1],result$Erro[1],result$mle[2],result$Erro[2],result$mle[3],result$Erro[3])
```

```
#####
```

```
x = seq(min(y)-100,max(y)+150,0.1)
```

```
folln = pdf_oln(result$mle,x)
```

```
lines(x,folln , col = '4', lwd = 2,lty=1)
```

```
lines(x,cdf_oln(result$mle,x),col=4,lwd = 2,lty=1) #Dist. Acumulada Normal
```

```
#####
```

```
cbind("Prob. Perda(70%)" = cdf_oln(result$mle,Yg[1]),"Prob.
```

```
Perda(75%)"=cdf_oln(result$mle,Yg[2]),"Prob. Perda(80%)"=cdf_oln(result$mle,Yg[3]),
```

```
"Prob. Perda(85%)"=cdf_oln(result$mle,Yg[4]),"Prob. Perda(90%)"=cdf_oln(result$mle,Yg[5])) #
Probabilidade de Perda
```

```
#####
```

```
require(ExtDist)
```

```
pdfbeta=function(par,x){
```

```
  par1=par[1]
```

```
  par2=par[2]
```

```
  par3=par[3]
```

```
  par4=par[4]
```

```
  dBeta_ab(x, par1,par2,par3,par4)
```

```
}
```

```

cdfbeta=function(par,x){
  par1=par[1]
  par2=par[2]
  par3=par[3]
  par4=par[4]
  pBeta_ab(x,par1,par2,par3,par4)
}

require(AdequacyModel)

pp=eBeta_ab(y);pp

aj=goodness.fit(pdf=pdfbeta,cdf=cdfbeta, starts = c(pp$shape1,pp$shape2,pp$a,pp$b),data=y,
method="N", domain=c(0,Inf));aj

aj$mle

result=aj

result=goodness.fit(pdf=pdfbeta,cdf=cdfbeta,
starts=c(aj$mle[1],aj$mle[2],aj$mle[3],aj$mle[4]),data=y, method="N", domain=c(0,Inf));result

aj=goodness.fit(pdf=pdfbeta,cdf=cdfbeta, starts=c(1.65,0.78,2362.50,3661.33),data=y, method="S",
domain=c(0,Inf));aj

result=aj

require(EnvStats)

#fb= dBeta_ab(y, pp$shape1,pp$shape2,pp$a,pp$b)#params =
list(pp$shape1,pp$shape2,pp$a,pp$b))

#facumb= pBeta_ab(y, pp$shape1,pp$shape2,pp$a,pp$b);facumb

t = seq(min(y)-100,max(y)+150,0.1)

length(t)

fb =pdfbeta(result$mle,t)

```

```

length(fb)

####

plot(ecdf(y),main="Castro")#ylabel="Cumulative Dist. Function",

lines(x,cdfbeta(result$mle,x),col="Sienna",lwd =2,lty=2) #Dist. Acumulada Normal

legend("topleft",lty=c(1,2,1,2),lwd=c(1,2,2,,2), c("N","SN","OLL-N","Beta"),col=c(2,3,4,"Sienna"))

cbind("Prob. Perda(70%" = cdfbeta(result$mle,Yg[1]),"Prob.
Perda(75%)"=cdfbeta(result$mle,Yg[2]),"Prob. Perda(80%"=cdfbeta(result$mle,Yg[3]),

  "Prob. Perda(85%"=cdfbeta(result$mle,Yg[4]),"Prob. Perda(90%"=cdfbeta(result$mle,Yg[5])) #
Probabilidade de Perda

#####3

require(skewt)

pdf_skewt=function(par,x){

  mu = par[1]

  sigma = par[2]

  gama=par[3]

  df= par[4]

  z = (x - mu)/sigma

  dskt(z,df,gama)

}

require(sn)

cdf_skewt=function(par,x){

  mu = par[1]

  sigma = par[2]

  gama=par[3]

  df= par[4]

```

```

z = (x - mu)/sigma
pskt(z,df,gama)
}
aj=goodness.fit(pdf=pdf_skewt,cdf=cdf_skewt, starts=c(mean(y),sd(y),gama=2,df=2),data=y,
method="Nelder-Mead", domain=c(0,Inf));aj
aj$mle
result=aj
result=goodness.fit(pdf=pdf_sn,cdf=cdf_sn, starts=(aj$mle[1],aj$mle[2],aj$mle[3],aj$mle[4]),data=y,
method="Nelder-Mead", domain=c(0,Inf));result
cbind(result$mle[1],result$Erro[1],result$mle[2],result$Erro[2],result$mle[3],result$Erro[3])

#####
x = seq(min(y)-100,max(y)+150,0.1)
fst =pdf_skewt(result$mle,x)
hist(y,prob=T,ylim=c(0,0.0025))
lines(x,fst,col="2",lty=4,lwd=2)
legend("topleft",lty=c(1,3,1,4),lwd=c(2,2,2,2),cex=0.8, c("N","SN","OLL-N","RS"),col=c(1,3,4,2))
lines(x,cdf_skewt(result$mle,x),col=3,lwd = 2,lty=2) #Dist. Acumulada Normal

```