

**Universidade de São Paulo  
Escola Superior de Agricultura “Luiz de Queiroz”**

**Estimação dos parâmetros da distribuição beta bivariada:  
aplicações em severidade de doenças em plantas**

**Otávio Akira de Barros**

Dissertação apresentada para obtenção do título  
de Mestre em Ciências. Área de concentração:  
Estatística e Experimentação Agronômica

**Piracicaba  
2015**

Otávio Akira de Barros  
Bacharel em Estatística

**Estimação dos parâmetros da distribuição beta bivariada: aplicações em severidade de doenças em plantas**

Orientador:  
Prof. Dr. **SILVIO SANDOVAL ZOCCHI**

Dissertação apresentado para obtenção do título de Mestre em Ciências. Área de concentração: Estatística e Experimentação Agronômica

**Piracicaba**  
**2015**

**Dados Internacionais de Catalogação na Publicação  
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Barros, Otávio Akira de

Estimação dos parâmetros da distribuição beta bivariada: aplicações em severidade de doenças em plantas / Otávio Akira de Barros. - - Piracicaba, 2015.  
89 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura “Luiz de Queiroz”.

I. Estimação 2. Distribuição beta bivariada 3. Severidade de doença em plantas  
I. Título

CDD 632.3  
B277e

**“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”**

## AGRADECIMENTOS

Meus agradecimentos vão a todos os colegas, professores, familiares e amigos que acompanharam essa trajetória, contribuindo para concretização deste trabalho. Em especial:

Ao meu orientador Prof. Dr. Silvio Sandoval Zocchi, pela sugestão deste tema, pela disposição e competência na condução deste trabalho;

Aos professores do departamento de Matemática e Estatística da ESALQ-USP, pelos ensinamentos;

Aos meus pais Laécio Carvalho de Barros, Maria Cristina Shimako Shimabukuro e minha irmã Luiza Yoko de Barros, pelo amor, dedicação, incentivo, apoio financeiro e presença constante em todas as etapas da minha vida, em especial neste período do mestrado;

À minha namorada Mayra Gobbi Silva, pelo apoio emocional e incentivo para a conclusão deste trabalho;

À minha turma da pós graduação, em especial Bruna, Daniel, Douglas, Erasnilson, Fernando, Rick, Simone, Valiana e Patrícia;

Aos meus amigos da graduação e das bandas Instrumentália e Sábado Sombrio pelos momentos de companheirismo e divertimento durante este período;

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de estudos;

Aos funcionários do departamento de Matemática e Estatística da ESALQ-USP, pela prontidão e atenção dispensada;

Finalmente, a todos que de alguma forma contribuíram direta ou indiretamente para a realização deste trabalho.



## SUMÁRIO

RESUMO . . . . .	7
ABSTRACT . . . . .	9
LISTA DE FIGURAS . . . . .	11
LISTA DE TABELAS . . . . .	13
1 INTRODUÇÃO . . . . .	15
2 REVISÃO BIBLIOGRÁFICA . . . . .	23
3 MATERIAL . . . . .	33
4 MÉTODOS . . . . .	39
4.1 Estimação dos parâmetros . . . . .	39
4.1.1 Método 1 . . . . .	46
4.1.2 Método 2.1 . . . . .	49
4.1.3 Método 2.2 . . . . .	50
4.2 Avaliação dos estimadores . . . . .	52
5 RESULTADOS E DISCUSSÃO . . . . .	55
5.1 Simulações . . . . .	55
5.2 Análise dos dados . . . . .	66
6 CONCLUSÃO . . . . .	71
REFERÊNCIAS . . . . .	73
APÊNDICE . . . . .	75



## RESUMO

### **Estimação dos parâmetros da distribuição beta bivariada: aplicações em severidade de doenças em plantas**

A distribuição beta é apropriada para analisar dados de variáveis medidas no intervalo  $(0, 1)$ , como taxas e proporções, como por exemplo a proporção de severidade de doenças em plantas. Portanto, dados que são pares observações de taxas e proporções, naturalmente pensa-se numa distribuição beta bivariada com suporte  $(0, 1)^2$ . O objetivo deste trabalho constitui-se em encontrar a melhor distribuição beta bivariada na literatura para este caso e, além disso, tentar encontrar estimadores para seus parâmetros, a fim de verificar se esta distribuição escolhida se ajusta bem aos dados. Foi criada uma metodologia para a estimação dos parâmetros, utilizando aquela distribuição que consideramos a mais adequada. Posteriormente foram feitas simulações para avaliar a qualidade desses estimadores e, por fim, foram utilizados três bancos de dados com a finalidade de exemplificar esta metodologia.

Palavras-chave: Estimação; Distribuição beta bivariada; Severidade de doença em plantas.



## ABSTRACT

### Parameters estimation of beta bivariate distribution: applications in disease severity in plants

Beta distribution is suitable for analyzing variable data measured in the range  $(0, 1)$ , as rates and proportions, such as the proportion of disease severity in plants. Therefore, data that are paired observations rates and proportions naturally thinks in a bivariate distribution beta supported  $(0, 1)^2$ . The objective of this work is on finding the best beta bivariate distribution in the literature for this case and, furthermore, try to find estimators for its parameters in order to verify that this chosen distribution fits the data well. A methodology was created for the estimation of parameters using that distribution we consider the most appropriate. Later simulations were performed to evaluate the quality of these estimators and, finally, we use three databases in order to illustrate this methodology.

Keywords: Estimation; Beta bivariate distribution; Disease severity in plants.



## LISTA DE FIGURAS

<p>Figura 1 - Imagens de folíolos de folhas de amendoinzeiros com as seguintes doenças: (a) ferrugem (<i>Puccinia arachidis</i>); (b) mancha castanha (<i>Cercospora arachidicola</i>) e (c) mancha preta (<i>Cercosporidium personatum</i>) (BROOKS, 2014) . . . . .</p> <p>Figura 2 - Escalas diagramáticas para avaliação da severidade em amendoinzeiros das doenças: (a) mancha castanha (<i>Cercospora arachidicola</i>) e (b) mancha preta (<i>Cercosporidium personatum</i>). Valores em porcentagem da área foliar com sintomas (MORAES, 2007) . . . . .</p> <p>Figura 3 - Severidades reais e estimadas visualmente, em porcentagem, por cinco avaliadores com experiência em avaliação de doenças: avaliador 1 (A,K); 2 (B,L); 3 (C,M); 4 (D,N); 5 (E,O), sem auxílio (A a E) e com auxílio (K a O) da escala diagramática e linhas de regressão obtidas entre as severidades reais e as estimadas visualmente (linhas cheias). A linha pontilhada representa a situação ideal, com as estimativas idênticas à real. UTFPR - Campus Pato Branco, 2007 (CAMOCHENA; SANTOS; MAZARO, 2008)</p> <p>Figura 4 - Gráficos da função de densidade de probabilidades da distribuição beta com parâmetros <math>a</math> e <math>b</math> tais que <math>a = b</math>, isto é, com formato simétrico em torno da média 1/2: (a) côncavos para baixo quando <math>a &gt; 1</math> e <math>b &gt; 1</math> e (b) côncavos para cima quando <math>0 &lt; a &lt; 1</math> e <math>0 &lt; b &lt; 1</math> . . . . .</p> <p>Figura 5 - Gráficos da função de densidade de probabilidades da distribuição beta com parâmetros <math>a</math> e <math>b</math> tais que: (a) <math>a &lt; b</math> e (b) <math>a &gt; b</math> . . . . .</p> <p>Figura 6 - Amostras simuladas de tamanho <math>n = 100</math> de distribuições de NADARA-JAH e KOTZ (2005) segundo os cenários de simulação 1 a 4 apresentados na Tabela 4 . . . . .</p> <p>Figura 7 - Gráficos de dispersão: (i) dos pares de observações <math>(x_i, y_i)</math>, <math>(i = 1, \dots, 30)</math> em que <math>X</math> é a severidade real de ferrugem e <math>Y</math> é a severidade estimada visualmente por um avaliador; (ii) de uma amostra de 200 observações de <math>(X, Y)</math> com distribuição BBNK com parâmetros iguais às estimativas obtidas pelo método 2.2 . . . . .</p>	<p>16</p> <p>18</p> <p>20</p> <p>25</p> <p>25</p> <p>36</p> <p>67</p>
--	---

Figura 8 - Gráficos de dispersão: (i) dos pares de observações $(x_i, y_i)$ , ( $i = 1, \dots, 30$ ) em que $X$ é a severidade real de mancha castanha e $Y$ é a severidade estimada visualmente por um avaliador; (ii) de uma amostra de 200 observações de $(X, Y)$ com distribuição BBNK com parâmetros iguais às estimativas obtidas pelo método 1 . . . . .	68
Figura 9 - Gráficos de dispersão: (i) dos pares de observações $(x_i, y_i)$ , ( $i = 1, \dots, 30$ ) em que $X$ é a severidade real de mancha preta e $Y$ é a severidade estimada visualmente por um avaliador; (ii) de uma amostra de 200 observações de $(X, Y)$ com distribuição BBNK com parâmetros iguais às estimativas obtidas pelo método 2.1 . . . . .	69

## LISTA DE TABELAS

<p>Tabela 1 - Porcentagens reais de tecido doente (severidades) e estimadas visualmente por um avaliador, a partir de imagens de 30 folhas de amendoinzeiros com ferrugem (<i>Puccinia arachidis</i>), geradas pelo pacote computacional Distrain (TOMERLIN; HOWELL, 1988) . . . . .</p> <p>Tabela 2 - Porcentagens reais de tecido doente (severidades) e estimadas visualmente por um avaliador, a partir de imagens de 30 folhas de amendoinzeiros com mancha castanha (<i>Cercospora arachidicola</i>), geradas pelo pacote computacional Distrain (TOMERLIN; HOWELL, 1988) . . . . .</p> <p>Tabela 3 - Porcentagens reais de tecido doente (severidades) e estimadas visualmente por um avaliador, a partir de imagens de 30 folhas de amendoinzeiros com mancha preta (<i>Cercosporidium personatum</i>), geradas pelo pacote computacional Distrain (TOMERLIN; HOWELL, 1988) . . . . .</p> <p>Tabela 4 - Valores dos parâmetros <math>a_1</math>, <math>a_2</math>, <math>b_1</math>, <math>b_2</math>, <math>b</math> e <math>c</math> da distribuição de NADARA-JAH e KOTZ (2005) utilizados para simulação de 1000 amostras . . . . .</p> <p>Tabela 5 - Esperanças e variâncias das distribuições de <math>X</math> e <math>Y</math>, marginais da distribuição de <math>(X, Y)</math> e coeficiente de correlação entre <math>X</math> e <math>Y</math> para cada cenário considerado na Tabela 4 . . . . .</p> <p>Tabela 6 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros <math>a_1</math>, <math>a_2</math>, <math>\beta_1</math> e <math>\beta_2</math> no cenário 1, via método dos momentos e da máxima verossimilhança, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300 . . . . .</p> <p>Tabela 7 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros <math>b_1</math>, <math>b_2</math>, <math>b</math> e <math>c</math> no cenário 1, via método 1, método 2.1 e método 2.2, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300 . . . . .</p> <p>Tabela 8 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros <math>a_1</math>, <math>a_2</math>, <math>\beta_1</math> e <math>\beta_2</math> no cenário 2, via método dos momentos e da máxima verossimilhança, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300 . . . . .</p>	<p>33</p> <p>34</p> <p>35</p> <p>35</p> <p>36</p> <p>56</p> <p>57</p> <p>58</p>
--	---

Tabela 9 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros $b_1$ , $b_2$ , $b$ e $c$ no cenário 2, via método 1, método 2.1 e método 2.2, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300 . . . . .	59
Tabela 10 -Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros $a_1$ , $a_2$ , $\beta_1$ e $\beta_2$ no cenário 3, via método dos momentos e da máxima verossimilhança, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300 . . . . .	61
Tabela 11 -Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros $b_1$ , $b_2$ , $b$ e $c$ no cenário 3, via método 1, método 2.1 e método 2.2, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300 . . . . .	62
Tabela 12 -Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros $a_1$ , $a_2$ , $\beta_1$ e $\beta_2$ no cenário 4, via método dos momentos e da máxima verossimilhança, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300 . . . . .	64
Tabela 13 -Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros $b_1$ , $b_2$ , $b$ e $c$ no cenário 4, via método 1, método 2.1 e método 2.2, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300 . . . . .	65
Tabela 14 -Estimativas dos parâmetros, via método dos momentos para $a_1$ , $a_2$ , $\beta_1$ e $\beta_2$ , e via método 2.2 para $b_1$ , $b_2$ , $b$ e $c$ , dos dados de severidade da ferrugem no amendoim . . . . .	66
Tabela 15 -Estimativas dos parâmetros, via método dos momentos para $a_1$ , $a_2$ , $\beta_1$ e $\beta_2$ , e via método 1 para $b_1$ , $b_2$ , $b$ e $c$ , dos dados de severidade da mancha castanha no amendoim . . . . .	67
Tabela 16 -Estimativas dos parâmetros, via método dos momentos para $a_1$ , $a_2$ , $\beta_1$ e $\beta_2$ , e via método 2.1 para $b_1$ , $b_2$ , $b$ e $c$ , dos dados de severidade da mancha castanha no amendoim . . . . .	69

## 1 INTRODUÇÃO

O amendoim (*Arachis hypogaea*) é uma leguminosa originária da América do Sul, cultivado nas mais variadas regiões tropicais do mundo pela sua ampla adaptabilidade a uma grande diversidade de ambientes. No Brasil, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE, 2013) a área plantada de amendoim foi de 121102 hectares e a quantidade produzida foi de 389783 toneladas. Em particular o estado de São Paulo é o principal produtor com 100675 hectares de área plantada e 349755 toneladas produzidas.

Segundo a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA, 2005), as principais doenças que ocorrem na cultura podem, no entanto, causar uma redução de 10% a mais de 50% na produção de vagens, quando medidas de controle não são utilizadas. Os principais problemas podem ocorrer tanto na fase de plantio (doenças de sementes e plântulas), como durante o desenvolvimento da cultura (doenças causadas por fungo do solo ou da parte aérea). Entre as várias doenças descritas nas diferentes regiões do mundo onde se cultiva o amendoim, algumas são de ocorrência esporádica ou podem aparecer sem contudo causar danos significativos. Outras, como é o caso das cercosporoses e da ferrugem, exigem o uso de medidas de controle para que o amendoim possa ser produzido comercialmente.

A Figura 1(a) apresenta a foto de um folíolo de amendoineiro contaminado pela ferrugem, causada pelo fungo *Puccinia arachidis*. Conforme o Instituto Agronômico (IAC, 1999), essa doença aparece inicialmente como pequenos pontos amarelados que posteriormente passam a apresentar a coloração característica marrom-avermelhada, dando o aspecto ferruginoso às folhas severamente infectadas. Entretanto, mesmo nas infecções severas não ocorre desfolha e as folhas permanecem atadas às plantas com aspecto de queimadas. O ciclo da doença, período entre a infecção e a formação de novos esporos, é rápido 5 a 7 dias, e por isso, o aumento da severidade, sob condições favoráveis, ocorre rapidamente, exigindo maior atenção no controle e a diminuição dos intervalos de aplicação de fungicidas. Nos últimos anos a ferrugem vem se manifestando de forma severa em algumas regiões, podendo tornar-se mais problemática que as demais doenças foliares. Outra doença que merece destaque, a mancha castanha, é causada pelo fungo *Cercospora arachidicola* e se caracteriza por manchas necróticas circulares ou irregulares, como pode ser visto na Figura 1(b). Sua coloração, entretanto, é mais clara, com um halo amarelado mais nítido e de diâ-

metro maior que a mancha preta, outra doença de destaque. Tem seu ciclo de 10 a 12 dias sendo, portanto, mais longo do que o da ferrugem. Além disso, a incidência dessa doença geralmente é observada mais cedo que a da mancha preta. Finalmente, a mancha preta, causada pelo fungo *Cercosporidium personatum* tem como sintomas, conforme ilustra a Figura 1(c), lesões de coloração castanho-escura arredondadas de diâmetros variáveis. Estas são, às vezes, circundadas por um pequeno halo amarelado, como pequenos pontos escuros distribuídos circularmente nos centros das lesões, que apresentam coloração escura a preta, característica que a diferencia da mancha castanha. Embora a mancha preta apareça mais tarde que a mancha castanha, sua curva de aumento de intensidade é mais rápida e severa, causando a desfolha mais rápida das plantas.



(a) Ferrugem



(b) Mancha castanha



(c) Mancha preta

Figura 1 - Imagens de folíolos de folhas de amendoinzeiros com as seguintes doenças:  
 (a) ferrugem (*Puccinia arachidis*); (b) mancha castanha (*Cercospora arachidicola*) e (c) mancha preta (*Cercosporidium personatum*) (BROOKS, 2014)

Apesar de as duas manchas estarem presentes em quase todos os campos de cultivo, a intensidade de cada doença varia com a localidade e as épocas de plantio. No estado de São Paulo a mancha preta tem se mostrado predominante e a mais severa entre as doenças foliares do amendoim, contribuindo para o baixo rendimento da cultura. As manchas castanha e preta do amendoim podem, entretanto, ser controladas por diversas medidas, em conjunto ou isoladamente, destacando-se a rotação de culturas, o controle químico e utilizando-se métodos de monitoramento e avisos climáticos. A quantificação da severidade destas doenças, torna-se, assim, de importância fundamental para predizer a necessidade de pulverização com fungicidas e controle da doença com eficiência.

Segundo BERGAMIN, KIMATI e AMORIM (1995) e KIMATI et al. (1997) a severidade é definida como a proporção, ou porcentagem, da área do tecido da planta coberto por sintomas de uma doença, sendo comumente utilizada na quantificação de doenças foliares, como ferrugens, oídios, mísrios e manchase. O método para sua quantificação, entretanto, é subjetivo e tenta expressar o dano real causado com exatidão, o que nem sempre é uma tarefa fácil.

Uma das estratégias para avaliação da severidade é a utilização de escalas diagramáticas, que são representações ilustradas de uma série de plantas ou partes de plantas com sintomas em diferentes níveis de severidade. Conforme ilustram as Figuras 2(a) e 2(b), para a mancha castanha e mancha preta, respectivamente, na elaboração de escalas diagramáticas, é necessário que o limite superior da escala corresponda à quantidade máxima de doença observada no campo, a quantidade real de doença e sua representação na escala devem ter alta precisão e as subdivisões da escala devem respeitar as limitações da acuidade visual humana. Mesmo assim, este método pode apresentar problemas quando são necessárias avaliações muito precisas, pois frequentemente a visão humana superestima a doença. Para a validação da escala, costumam-se selecionar alguns avaliadores, experientes ou sem experiência, e pedir para que eles avaliem a severidade por duas vezes, sem e com a escala, em diversas folhas ou tecidos com diferentes porcentagens lesionadas.

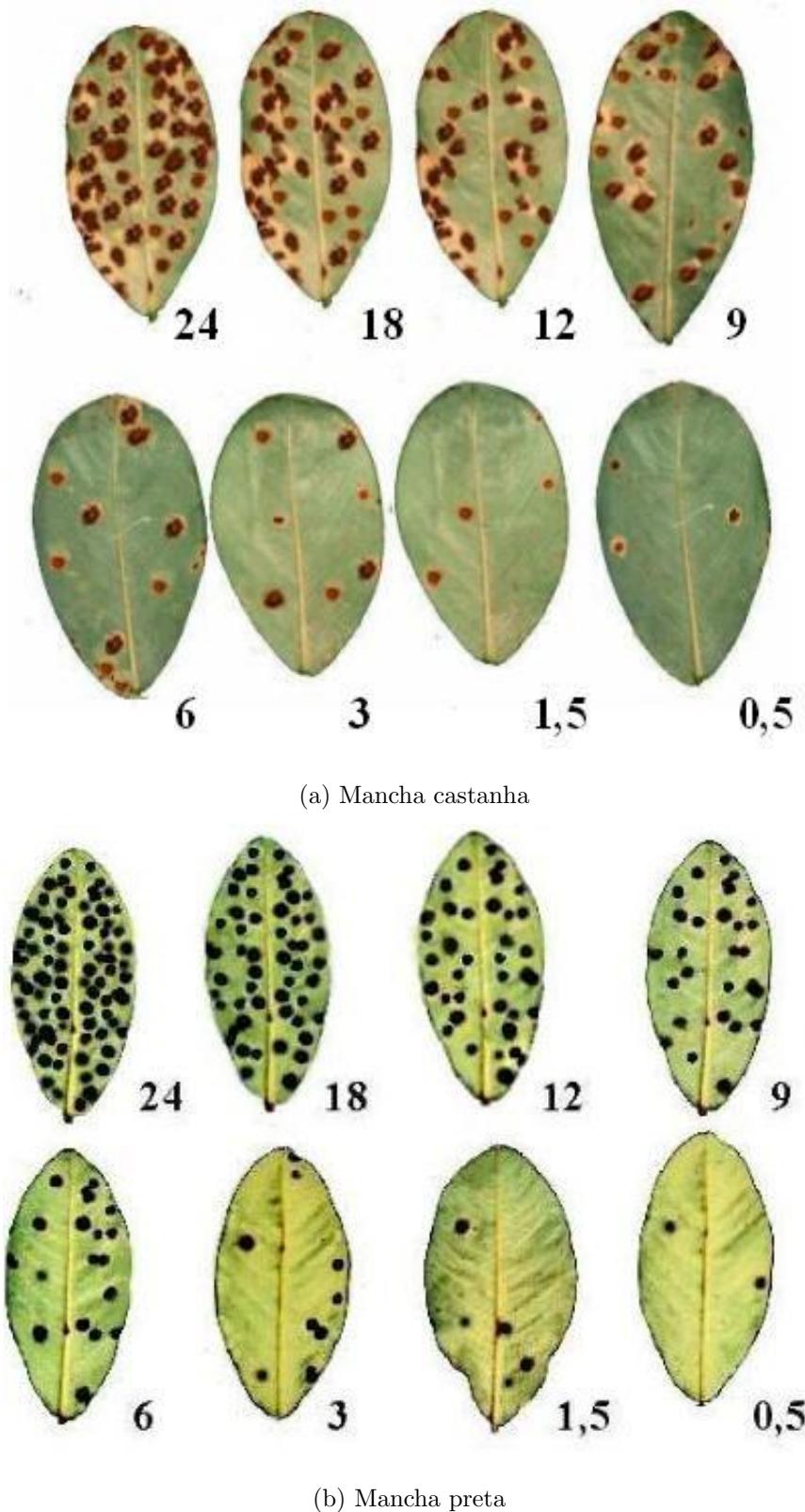


Figura 2 - Escalas diagramáticas para avaliação da severidade em amendoinzeiros das doenças: (a) mancha castanha (*Cercospora arachidicola*) e (b) mancha preta (*Cercosporidium personatum*). Valores em porcentagem da área foliar com sintomas (MORAES, 2007)

Outro método de quantificação da severidade é a análise de imagens obtidas por meio de câmeras fotográficas digitais. As imagens são, então, transferidas para um computador e pacotes computacionais como o “Quant” de VALE, FERNANDES FILHO e LIBERATO (2003) avaliam precisamente a severidade. Este método não está sujeito aos problemas inerentes à vista humana, sendo, portanto, muito preciso, permitindo obter estimativas não subjetivas da severidade de doenças. No entanto, também possui suas limitações, pois as imagens necessitam de uma intensidade luminosa bem controlada para que o contraste entre tecido sadio e doente seja o maior possível, podendo ser aplicado apenas para aquelas doenças em que o tecido lesionado é bem contrastado do sadio. Além disso, algumas vezes o material a ser analisado precisa ser transportado do campo para o laboratório, mas algumas plantas não resistem ao transporte, deformando-se rapidamente, o que impede sua posterior avaliação. A acurácia deste tipo de avaliação é inversamente proporcional ao tamanho da amostra, uma vez que torna quase inviável para amostras muito grandes.

Após mensurada a severidade da doença, costuma-se fazer, para cada avaliador, uma análise de regressão linear simples entre as severidades reais, obtidas por meio de análises de imagens digitalizadas, e as estimadas visualmente pelo avaliador. São analisados então, o intercepto da reta, seu coeficiente angular e o coeficiente de determinação, apresentando os resultados em gráficos como ilustra a Figura 3. Os gráficos de A a E e de K a O apresentam, respectivamente, as severidades estimadas pelos avaliadores sem e com o auxílio da escala diagramática, bem como os coeficientes de determinação (quadrados dos coeficientes de correlação linear de Pearson) da regressão linear tendo como pressuposto que os pares de valores provém de uma distribuição normal bivariada. Entretanto, os valores de severidade, em porcentagem, pertencem ao intervalo (0, 100) ou, quando expressos em proporções, ao intervalo (0, 1). Naturalmente, pode-se pensar na distribuição beta, que apresenta esse intervalo como suporte. Portanto, os pares de valores de severidades não provém de uma distribuição normal bivariada mas sim de uma distribuição beta bivariada com suporte (0, 1)<sup>2</sup>.

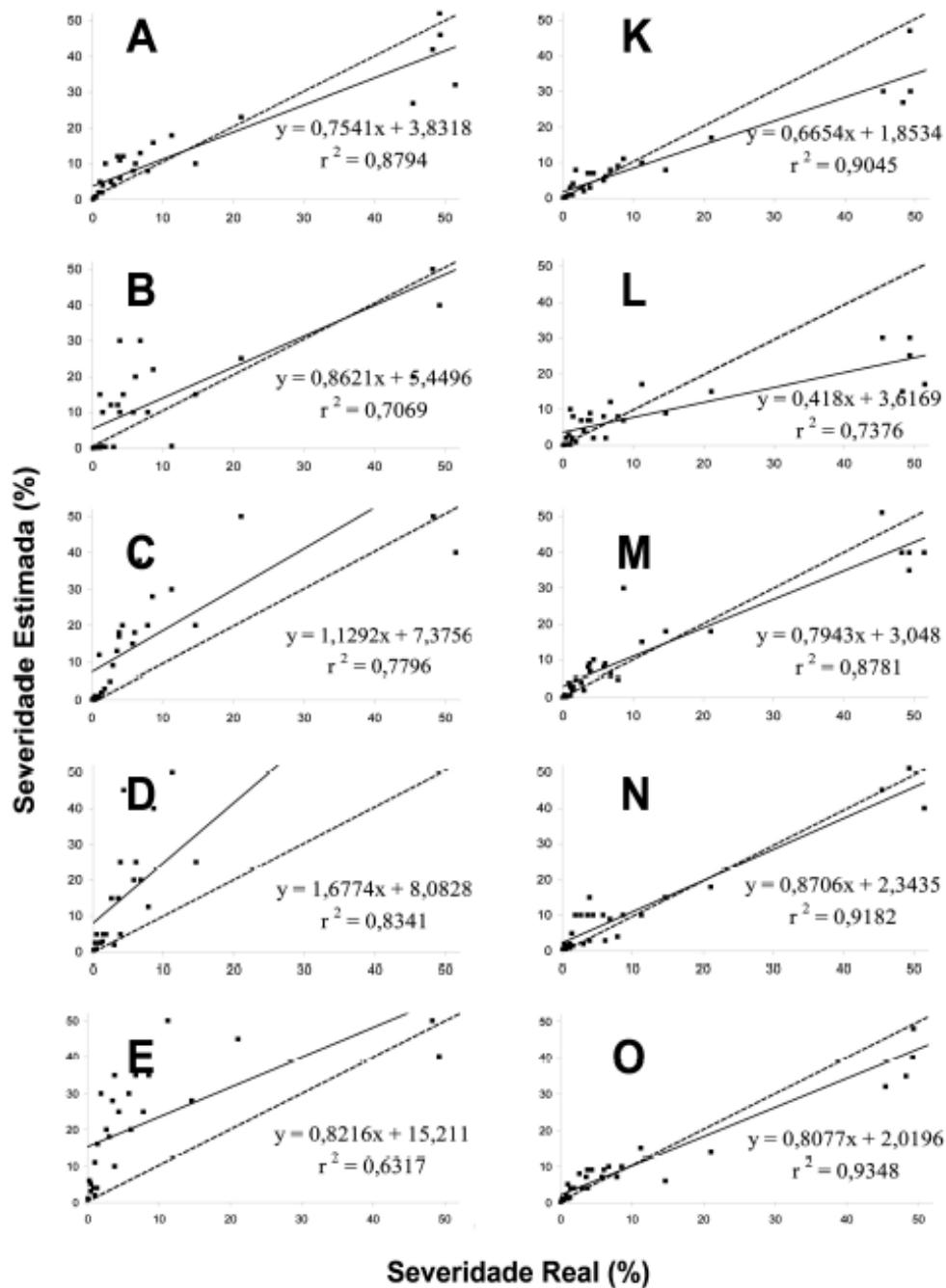


Figura 3 - Severidades reais e estimadas visualmente, em porcentagem, por cinco avaliadores com experiência em avaliação de doenças: avaliador 1 (A,K); 2 (B,L); 3 (C,M); 4 (D,N); 5 (E,O), sem auxílio (A a E) e com auxílio (K a O) da escala diagramática e linhas de regressão obtidas entre as severidades reais e as estimadas visualmente (linhas cheias). A linha pontilhada representa a situação ideal, com as estimativas idênticas à real. UTFPR - Campus Pato Branco, 2007 (CAMOCHEA; SANTOS; MAZARO, 2008)

Com esse suporte, há na literatura, algumas distribuições beta bivariadas, como as propostas por GUPTA e WONG (1985), JONES (2002), OLKIN e LIU (2003), NADARAJAH e KOTZ (2005), NADARAJAH (2006), NADARAJAH (2008), ARNOLD e NG (2011) e OLKIN e TRIKALINOS (2015).

Neste trabalho, propõe-se métodos de estimação dos parâmetros da distribuição beta bivariada de NADARAJAH e KOTZ (2005). Em seguida, os estimadores são avaliados por meio de simulações. Finalmente, tal distribuição é ajustada a dados originais de porcentagens de severidade reais e estimadas visualmente de doenças em amendoinzeiros.



## 2 REVISÃO BIBLIOGRÁFICA

Segundo GUPTA e NADARAJAH (2004), a distribuição beta é apropriada nas situações em que o pesquisador precisa analisar dados de variáveis medidas no intervalo  $(0, 1)$ , como taxas e proporções, ou até mesmo como uma versão mais flexível da distribuição uniforme, assumindo diferentes formas dependendo do valor dos seus parâmetros. Tal distribuição é chamada de beta por possuir em sua função de densidade de probabilidades a função beta, que aparece como uma constante normalizadora para garantir que a função seja de densidade, ou seja, que a integral em todo o espaço definido resulte em 1. A função beta é definida por:

$$B(a, b) = \int_0^1 z^{a-1}(1-z)^{b-1} dz, \quad a > 0, \quad b > 0.$$

Esta função, por sua vez, possui a seguinte relação com a função gama:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

em que  $\Gamma(t) = \int_0^\infty z^{t-1}e^{-z} dz$ , podendo ser representada pela série:

$$B(a, b) = \frac{1}{b} \sum_{k=0}^{\infty} (-1)^k b \frac{(b-1) \cdots (b-k)}{k!(a+k)}.$$

De forma análoga à função beta, a função beta incompleta é definida por:

$$B_x(a, b) = \int_0^x w^{a-1}(1-w)^{b-1} dw,$$

ou por meio de sua relação com a função hipergeométrica de Gauss, ou seja,

$$B_x(a, b) = \frac{x^a}{a} {}_2F_1(a, 1-b; a+1; x), \quad (1)$$

em que  ${}_2F_1$  é a função hipergeométrica de Gauss definida por:

$${}_2F_1(\alpha, \beta; \gamma; x) = \sum_{i=0}^{\infty} \frac{(\alpha)_i (\beta)_i}{(\gamma)_i} \frac{x^i}{i!}$$

e  $(z)_i = z(z+1) \cdots (z+i-1)$  denota um fatorial ascendente.

Seja  $X$  uma variável aleatória contínua com suporte  $(0, 1)$ . Conforme JOHN-SON, KOTZ e BALAKRISHNAN (1995), essa variável tem distribuição beta com parâmetros  $a$  e  $b$ , ou seja,  $X \sim Beta(a, b)$  se e somente se sua função de densidade de probabilidades é:

$$f(x; a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \mathbb{I}_{(0,1)}(x), \quad (2)$$

em que:  $\mathbb{I}_{(0,1)}(x)$  é uma função indicadora que assume o valor 1 para  $0 < x < 1$  e 0, caso contrário e  $B(a, b)$  é a função beta, sendo  $a > 0$  e  $b > 0$ .

A função de distribuição acumulada de  $X$ , por sua vez, é:

$$F(x) = \frac{1}{B(a, b)} \int_0^x w^{a-1} (1-w)^{b-1} dw = \frac{B_x(a, b)}{B(a, b)},$$

para  $0 < x < 1$ , em que  $B_x(a, b)$  é a função beta incompleta, definida pela expressão (1).

Além disso, esperança e variância de  $X$  são dadas, respectivamente, por:

$$E(X) = \frac{a}{a+b}$$

e

$$Var(X) = \frac{ab}{(a+b+1)(a+b)^2} = E(X)[1 - E(X)] \frac{1}{a+b+1}.$$

Note que se  $a = b = 1$ ,  $f(x) = 1$  para  $0 < x < 1$ , ou seja,  $X$  tem distribuição uniforme em  $(0, 1)$ . De forma geral, se  $a = b$ , têm-se que  $E(X) = 1/2$  e  $f(1/2 - x) = f(1/2 + x)$ , ou seja, a distribuição  $X$  é simétrica em torno da média, como ilustram as Figuras 4(a) e 4(b).

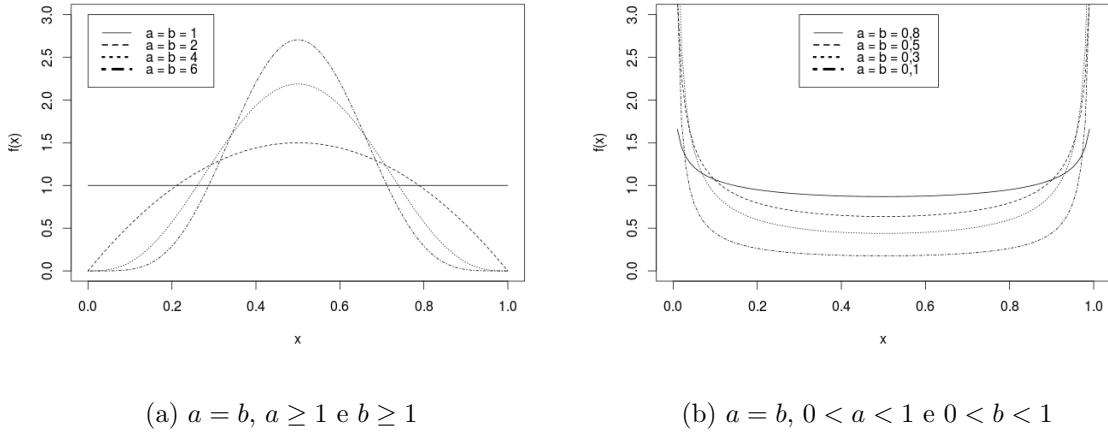


Figura 4 - Gráficos da função de densidade de probabilidades da distribuição beta com parâmetros  $a$  e  $b$  tais que  $a = b$ , isto é, com formato simétrico em torno da média 1/2: (a) côncavos para baixo quando  $a > 1$  e  $b > 1$  e (b) côncavos para cima quando  $0 < a < 1$  e  $0 < b < 1$

Por outro lado, quando  $a < b$ , o gráfico da função de densidade de probabilidades é assimétrico à direita e quando  $a > b$ , é assimétrico à esquerda como ilustram as Figuras 5(a) e 5(b), respectivamente.

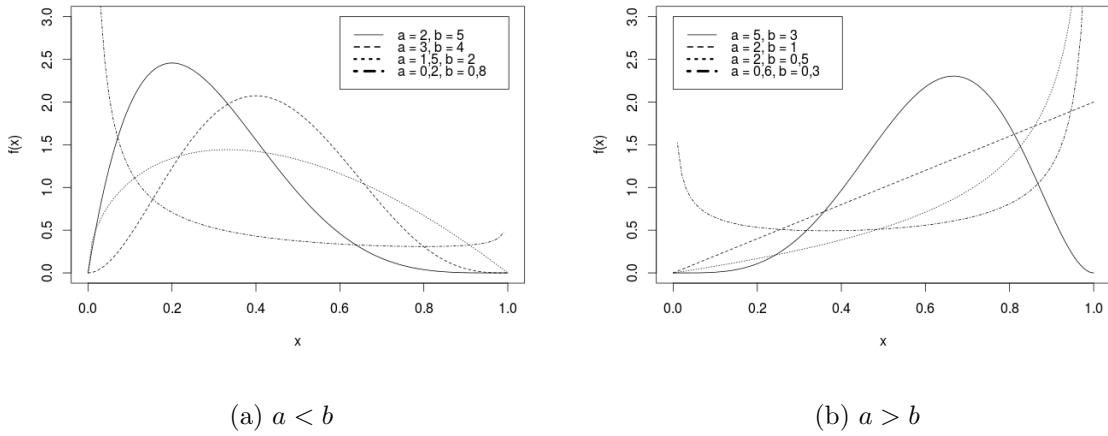


Figura 5 - Gráficos da função de densidade de probabilidades da distribuição beta com parâmetros  $a$  e  $b$  tais que: (a)  $a < b$  e (b)  $a > b$

Em modelos de regressão em que a variável resposta tem distribuição beta, entretanto, é comum utilizar a reparametrização da função de densidade de probabilidades (2) proposta por FERRARI e CRIBARI-NETO (2004). Neste caso, são definidos dois novos

parâmetros  $\mu = \frac{a}{a+b}$  e  $\phi = a + b$  e a função de densidade, neste caso, fica:

$$f(x; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} x^{\mu\phi-1} (1-x)^{(1-\mu)\phi-1} \mathbb{I}_{(0,1)}(x)$$

com  $0 < \mu < 1$  e  $\phi > 0$ . Assim, se uma variável aleatória  $X$  tem distribuição beta com parâmetros  $\mu$  e  $\phi$ , ou seja,  $X \sim Beta(\mu, \phi)$ , sua esperança e variância são dadas respectivamente por:

$$E(X) = \mu$$

e

$$Var(X) = \frac{\mu(1-\mu)}{1+\phi}.$$

Note que, nesta reparametrização, o parâmetro  $\phi$  está relacionado com a precisão, pois para um valor fixo de  $\mu$ , quanto maior for o valor de  $\phi$ , menor será a variância.

Até agora, foram consideradas, apenas, variáveis aleatórias unidimensionais, no entanto, quando há duas variáveis e estas estão relacionadas entre si, a análise bivariada torna-se mais apropriada por considerar as associações entre essas variáveis.

Para estender o caso univariado para o caso bivariado, precisamos conhecer a distribuição conjunta das variáveis de interesse ou introduzir, na distribuição, estruturas que induzem a dependência entre essas variáveis, devendo ser flexível o suficiente para acomodar uma gama grande de possíveis tipos de dependência. Métodos para a construção de distribuições conjuntas para variáveis aleatórias bivariadas ou multivariadas foram apresentados por BALAKRISHNAN e LAI (2009) e JOHNSON, KOTZ e BALAKRISHNAN (2002), respectivamente. No que diz respeito a distribuições bivariadas envolvendo duas distribuições beta, no entanto, há na literatura as denominadas beta bivariadas, a serem vistas a seguir.

Inicialmente, GUPTA e WONG (1985) sugeriram duas distribuições beta bivariadas. A primeira, com cinco parâmetros é obtida a partir do sistema de Morgenstern (MORGENSTERN, 1956) dado por:

$$f(x_1, x_2) = g_1(x_1)g_2(x_2)[1 + \lambda(2G_1(x_1) - 1)(2G_2(x_2) - 1)]\mathbb{I}_{(0,1)}(x_1)\mathbb{I}_{(0,1)}(x_2),$$

em que  $g(\cdot)$  e  $G(\cdot)$  são as funções de densidade e acumuladas de duas distribuições beta independentes e  $-1 < \lambda < 1$  é um parâmetro diretamente proporcional ao coeficiente

de correlação. A segunda distribuição por eles proposta é uma beta bivariada com três parâmetros  $\alpha_1$ ,  $\alpha_2$  e  $\alpha_3$  positivos, que se trata de uma Dirichlet bivariada com função de densidade de probabilidades dada por:

$$f(x_1, x_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} x_1^{\alpha_1-1} x_2^{\alpha_2-1} (1 - x_1 - x_2)^{\alpha_3-1}$$

com  $x_1, x_2 \geq 0$ ,  $x_1 + x_2 \leq 1$  e distribuições marginais  $X_1 \sim Beta(\alpha_1, \alpha_2 + \alpha_3)$  e  $X_2 \sim Beta(\alpha_2, \alpha_1 + \alpha_3)$ . Esta última distribuição, entretanto, não é de interesse para o caso em questão pois não apresenta suporte  $(0, 1)^2$ .

JONES (2002), posteriormente, definiu uma distribuição beta multivariada partindo da distribuição  $F$  multivariada, definida em  $(0, \infty)^m$ , com  $m + 1$  parâmetros  $2\nu_j$ ,  $j = 0, \dots, m$  e marginais com distribuição  $F(2\nu_i, 2\nu_0)$ ,  $i = 1, \dots, m$  dada por:

$$f_F(f_1, \dots, f_m) = G_v \prod_{j=0}^m \nu_j^{\nu_j} \frac{\prod_{i=1}^m f_i^{\nu_i-1}}{(\nu_0 + \sum_{k=1}^m \nu_k f_k)^n},$$

em que  $n = \nu_0 + \dots + \nu_m$  e  $G_v = \frac{\Gamma(n)}{\Gamma(\nu_0) \dots \Gamma(\nu_m)}$ . Considere que  $X_i \sim Beta(\nu_i, \nu_0)$  e  $F_i \sim F(2\nu_i, 2\nu_0)$ , então a relação entre  $X_i$  e  $F_i$  é dada da seguinte forma:

$$F_i = \frac{\nu_0}{\nu_i} \frac{X_i}{1 - X_i} \quad \Leftrightarrow \quad X_i = \frac{\nu_i F_i}{\nu_0 + \nu_i F_i}.$$

Com base nisso e no fato de que  $dF_i/dX_i = (\nu_0/\nu_i)(1 - X_i)^{-2}$ , a densidade conjunta de  $X_1, \dots, X_m$  é dada por:

$$f_X(x_1, \dots, x_m) = \frac{G_v \prod_{j=1}^m \frac{x_j^{\nu_j-1}}{(1 - x_j)^{\nu_j+1}}}{\left(1 + \sum_{k=1}^m \frac{x_k}{1 - x_k}\right)^n},$$

em que  $0 < x_1 < 1, \dots, 0 < x_m < 1$  e as marginais possuem distribuição  $Beta(\nu_i, \nu_0)$ . Em particular, o caso bivariado possui a seguinte função de densidade conjunta:

$$f(x_1, x_2) = B(\nu_0, \nu_1, \nu_2) \frac{x_1^{\nu_1-1} (1 - x_1)^{\nu_0+\nu_2-1} x_2^{\nu_2-1} (1 - x_2)^{\nu_0+\nu_1-1}}{(1 - x_1 x_2)^{\nu_0+\nu_1+\nu_2}} \mathbb{I}_{(0,1)}(x_1) \mathbb{I}_{(0,1)}(x_2), \quad (3)$$

em que  $B(\nu_0, \nu_1, \nu_2) = \frac{\Gamma(\nu_0)\Gamma(\nu_1)\Gamma(\nu_2)}{\Gamma(\nu_0 + \nu_1 + \nu_2)}$ . Além disso, o autor deduziu que:

$$\begin{aligned} E(X_1 X_2) &= \frac{\nu_1 \nu_2}{\nu_0 + \nu_1 + \nu_2} \frac{\Gamma(\nu_0 + \nu_1)\Gamma(\nu_0 + \nu_2)}{\Gamma(\nu_0)\Gamma(\nu_0 + \nu_1 + \nu_2 + 1)} \\ &\times {}_3F_2(\nu_0 + \nu_1 + \nu_2, \nu_2 + 1, \nu_1 + 1; \nu_0 + \nu_1 + \nu_2 + 1, \nu_0 + \nu_1 + \nu_2 + 1; 1). \end{aligned}$$

OLKIN e LIU (2003), por sua vez, propuseram uma distribuição beta bivariada partindo de três variáveis aleatórias independentes  $U$ ,  $V$  e  $W$  com distribuições gama, parâmetros de forma  $a$ ,  $b$  e  $c$  respectivamente e parâmetros de escala unitários. Neste caso, as variáveis aleatórias  $X$  e  $Y$ , definidas por:

$$X = \frac{U}{U + W} \quad \text{e} \quad Y = \frac{V}{V + W},$$

possuem distribuições  $Beta(a, c)$  e  $Beta(b, c)$ , respectivamente, sendo que  $X$  e  $Y$  são correlacionadas. Segundo os autores,  $(X, Y)$  possui distribuição beta bivariada definida para  $0 < x < 1$  e  $0 < y < 1$  com a seguinte função de densidade conjunta:

$$f(x, y) = \frac{x^{a-1} y^{b-1} (1-x)^{c-1} (1-y)^{a+b+c-1}}{B(a, b, c) (1-xy)^{a+b+c}} \mathbb{I}_{(0,1)}(x) \mathbb{I}_{(0,1)}(y), \quad (4)$$

em que  $B(a, b, c) = \frac{\Gamma(a)\Gamma(b)\Gamma(c)}{\Gamma(a+b+c)}$ , sendo  $a$ ,  $b$  e  $c$  parâmetros positivos. Adicionalmente, demonstraram que:

$$E(XY) = \frac{ab}{a+b+c} \frac{\Gamma(a+c)\Gamma(b+c)}{\Gamma(c)\Gamma(a+b+c+1)} {}_3F_2(a+1, b+1, a+b+c; a+b+c+1, a+b+c+1; 1).$$

Note que, embora tenham sido propostas independentemente, as distribuições bivariadas de  $(X_1, X_2)$  e  $(X, Y)$ , dadas pelas expressões (3) e (4) são equivalentes.

Outra distribuição beta bivariada foi proposta por NADARAJAH e KOTZ (2005), obtida a partir de três variáveis aleatórias independentes  $U, V$  e  $W$  com as seguintes distribuições beta dadas por:  $U \sim Beta(a_1, b_1)$ ,  $V \sim Beta(a_2, b_2)$  e  $W \sim Beta(c, b)$  em que  $c = a_1 + b_1 = a_2 + b_2$ . Utilizando então as seguintes transformações:

$$X = UW \quad \text{e} \quad Y = VW,$$

chegaram à distribuição conjunta de  $(X, Y)$  em que  $X \sim Beta(a_1, b_1+b)$  e  $Y \sim Beta(a_2, b_2+b)$ .

$b$ ), cuja função densidade de probabilidades é:

$$f(x, y) = Cx^{a_1-1}y^{a_2-1} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{(-1)^{i+j}(b_1-i)_i(b_2-j)_j x^i y^j B_{1-\max(x,y)}(b, b_2 - a_1 - i - j)}{i!j!}$$

para  $0 < x < 1$  e  $0 < y < 1$ , em que  $C = \frac{1}{B(a_1, b_1)B(a_2, b_2)B(b, c)}$ . Deduziram, ainda, que:

$$\begin{aligned} E(XY) &= \frac{\beta(a_1 + 1, b_1)\beta(a_2 + 1, b_2)\beta(b, c + 2)}{\beta(a_1, b_1)\beta(a_2, b_2)\beta(c, b)} \\ &= \frac{\Gamma(a_1 + 1)\Gamma(b_1)\Gamma(a_2 + 1)\Gamma(b_2)\Gamma(b)\Gamma(c + 2)}{\Gamma(a_1 + b_1 + 1)\Gamma(a_2 + b_2 + 1)\Gamma(b + c + 2)} \\ &= \frac{\Gamma(a_1)\Gamma(b_1)\Gamma(a_2)\Gamma(b_2)\Gamma(c)\Gamma(b)}{\Gamma(a_1 + b_1)\Gamma(a_2 + b_2)\Gamma(c + b)} \\ &= \frac{\Gamma(a_1 + 1)\Gamma(a_2 + 1)\Gamma(c + 2)\Gamma(a_1 + b_1)\Gamma(a_2 + b_2)\Gamma(c + b)}{\Gamma(a_1 + b_1 + 1)\Gamma(a_2 + b_2 + 1)\Gamma(b + c + 2)\Gamma(a_1)\Gamma(a_2)\Gamma(c)}, \end{aligned}$$

mas como  $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$  e  $\Gamma(\alpha + 2) = (\alpha + 1)\Gamma(\alpha + 1)$ , tem-se, então, que:

$$\begin{aligned} E(XY) &= \frac{a_1\Gamma(a_1)a_2\Gamma(a_2)(c + 1)c\Gamma(c)\Gamma(a_1 + b_1)\Gamma(a_2 + b_2)\Gamma(c + b)}{(a_1 + b_1)\Gamma(a_1 + b_1)(a_2 + b_2)\Gamma(a_2 + b_2)(b + c + 1)(b + c)\Gamma(b + c)\Gamma(a_1)\Gamma(a_2)\Gamma(c)} \\ &= \frac{a_1a_2(c + 1)c}{(a_1 + b_1)(a_2 + b_2)(b + c + 1)(b + c)} \\ &= \frac{a_1a_2(c + 1)}{c(b + c)(b + c + 1)}. \end{aligned}$$

Logo:

$$\begin{aligned} Cov(X, Y) &= \frac{a_1a_2(c + 1)}{c(b + c)(b + c + 1)} - \frac{a_1}{a_1 + b_1 + b} \frac{a_2}{a_2 + b_2 + b} \\ &= \frac{a_1a_2(c + 1)}{c(b + c)(b + c + 1)} - \frac{a_1a_2}{(b + c)^2} \\ &= \frac{(b + c)^2a_1a_2(c + 1) - c(b + c)(b + c + 1)a_1a_2}{c(b + c)^3(b + c + 1)} \\ &= \frac{a_1a_2(c + 1)(b + c) - c(b + c + 1)a_1a_2}{c(b + c)^2(b + c + 1)} \\ &= \frac{a_1a_2[(c + 1)(b + c) - c(b + c + 1)]}{c(b + c)^2(b + c + 1)} \\ &= \frac{a_1a_2[cb - c^2 + b + c - cb - c^2 - c]}{c(b + c)^2(b + c + 1)} \\ &= \frac{a_1a_2b}{c(b + c)^2(b + c + 1)}. \end{aligned}$$

A correlação sai, portanto, como resultado imediato:

$$\begin{aligned}
Cor(X, Y) &= \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \\
&= \frac{\frac{a_1 a_2 b}{c(b+c)^2(b+c+1)}}{\sqrt{\frac{a_1(b_1+b)}{(a_1+b_1+b)^2(a_1+b_1+b+1)} \frac{a_2(b_2+b)}{(a_2+b_2+b)^2(a_2+b_2+b+1)}}} \\
&= \frac{\frac{a_1 a_2 b}{c(b+c)^2(b+c+1)}}{\sqrt{\frac{a_1 a_2 (b_1+b)(b_2+b)}{(c+b)^2(c+b+1)(c+b)^2(c+b+1)}}} \\
&= \frac{\frac{a_1 a_2 b}{c(b+c)^2(b+c+1)}}{\frac{1}{(b+c)^2(b+c+1)} \sqrt{a_1 a_2 (b_1+b)(b_2+b)}} \\
&= \frac{a_1 a_2 b}{c \sqrt{a_1 a_2 (b_1+b)(b_2+b)}} \\
&= \frac{b}{c} \sqrt{\frac{(a_1 a_2)^2}{a_1 a_2 (b_1+b)(b_2+b)}} \\
&= \frac{b}{c} \sqrt{\frac{a_1 a_2}{(b_1+b)(b_2+b)}}.
\end{aligned}$$

É importante notar que como  $a_1 > 0$ ,  $a_2 > 0$ ,  $b_1 > 0$ ,  $b_2 > 0$ ,  $b > 0$  e  $c > 0$ ,

então:

$$Cor(X, Y) = \frac{b}{c} \sqrt{\frac{a_1 a_2}{(b_1+b)(b_2+b)}} > 0.$$

Serão apresentados, a seguir, alguns casos particulares em que a correlação tende a 1 ou tende a 0. Convém ressaltar, entretanto, que este estudo não consta no artigo original dos autores. O coeficiente de correlação entre  $X$  e  $Y$  tenderá a 1 se:

caso i):  $a_1 = a_2 = \alpha$ ;  $b_1 = b_2 = \delta$ ;  $b = \beta$ ;  $c = \alpha + \delta$  e  $\delta \rightarrow 0^+$ . Neste caso:

$$Cor(X, Y) = \frac{\beta}{\alpha + \delta} \sqrt{\frac{\alpha \alpha}{(\beta + \delta)(\beta + \delta)}} = \frac{\beta}{\alpha + \delta} \frac{\alpha}{\beta + \delta}$$

e

$$\lim_{\delta \rightarrow 0^+} Cor(X, Y) = \lim_{\delta \rightarrow 0^+} \frac{\beta}{\alpha + \delta} \frac{\alpha}{\beta + \delta} = 1.$$

caso ii):  $a_1 = a_2 = b = \gamma$ ;  $b_1 = b_2 = \delta$ ;  $c = \gamma + \delta$  e  $\delta \rightarrow 0^+$ . Neste caso:

$$\text{Cor}(X, Y) = \frac{\gamma}{\gamma + \delta} \sqrt{\frac{\gamma\gamma}{(\gamma + \delta)(\gamma + \delta)}} = \frac{\gamma}{\gamma + \delta} \frac{\gamma}{\gamma + \delta} = \left[ \frac{\gamma}{\gamma + \delta} \right]^2$$

e

$$\lim_{\delta \rightarrow 0^+} \text{Cor}(X, Y) = \lim_{\delta \rightarrow 0^+} \left[ \frac{\gamma}{\gamma + \delta} \right]^2 = 1$$

Por outro lado, o coeficiente de correlação entre  $X$  e  $Y$  tenderá a 0 se:

caso i):  $a_1 \rightarrow 0^+$

$$\lim_{a_1 \rightarrow 0^+} \text{Cor}(X, Y) = \lim_{a_1 \rightarrow 0^+} \frac{b}{c} \sqrt{\frac{a_1 a_2}{(b_1 + b)(b_2 + b)}} = 0$$

caso ii):  $a_2 \rightarrow 0^+$

$$\lim_{a_2 \rightarrow 0^+} \text{Cor}(X, Y) = \lim_{a_2 \rightarrow 0^+} \frac{b}{c} \sqrt{\frac{a_1 a_2}{(b_1 + b)(b_2 + b)}} = 0$$

caso iii):  $b \rightarrow 0^+$

$$\lim_{b \rightarrow 0^+} \text{Cor}(X, Y) = \lim_{b \rightarrow 0^+} \frac{b}{c} \sqrt{\frac{a_1 a_2}{(b_1 + b)(b_2 + b)}} = 0$$

Mais recentemente, ARNOLD e NG (2011) definiram uma distribuição beta bivariada partindo de distribuições gama. Considere  $U_i \sim \text{gama}(\alpha_i, 1)$ ,  $i = 1, \dots, 5$ , ou seja,  $U_i$  tem distribuição gama com parâmetro de forma  $\alpha_i$  e parâmetro de escala unitário. Considere, ainda, que:

$$X = \frac{U_1 + U_3}{U_1 + U_3 + U_4 + U_5} \quad \text{e} \quad Y = \frac{U_2 + U_4}{U_2 + U_3 + U_4 + U_5}.$$

Neste caso, os autores demonstram que  $(X, Y)$  tem distribuição beta bivariada com distribuições marginais dadas por:  $X \sim \text{Beta}(\alpha_1 + \alpha_3, \alpha_4 + \alpha_5)$  e  $Y \sim \text{Beta}(\alpha_2 + \alpha_4, \alpha_3 + \alpha_5)$ . Como casos particulares, têm-se a distribuição bivariada de Dirichlet, proposta por GUPTA e WONG (1985), que ocorre quando  $\alpha_1 = \alpha_2 = 0$ , e a beta bivariada de JONES (2002) e OLKIN e LIU (2003), quando  $\alpha_3 = \alpha_4 = 0$ .

Uma alternativa à construção de distribuições multivariadas quando as marginais são dadas ou conhecidas é por meio de cópulas, como pode-se ver em NELSEN (1999) e TRIVEDI e ZIMMER (2007). Estas possuem flexibilidade para tratamento de relações

não-lineares entre as variáveis envolvidas. OAKES e RITZ (2000), por exemplo, utilizam entre outras, a seguinte função cópula para encontrar a densidade de uma distribuição bivariada:

$$f(y_1, y_2; \mu_1, \mu_2, \phi_1, \phi_2, \psi) = f(y_1; \mu_1, \phi_1)f(y_2; \mu_2, \phi_2)[1 + \psi(y_1 - \mu_1)(y_2 - \mu_2)],$$

em que  $\theta = (\mu_1, \mu_2, \phi_1, \phi_2, \psi)$  é o vetor de parâmetros, com  $y_j \in (0, 1)$ ,  $j = 1, 2$  e  $f(y_j; \mu_j, \phi_j)$  denotam as densidades das distribuições marginais, reparametrizadas em termos das médias  $\mu_j \in (0, 1)$  e dos parâmetros de precisão  $\phi_j > 0$ . As marginais são densidades univariadas, e a correlação é dada por:

$$\text{Cor}(Y_1, Y_2) = \psi [\mu_1(1 - \mu_1)\mu_2(1 - \mu_2)(1 + \phi_1)^{-1}(1 + \phi_2)^{-1}]^{1/2}.$$

O parâmetro  $\psi$  pode ser visto como um parâmetro de dependência:  $\psi = 0$  corresponde a independência entre as variáveis.

Distribuições beta bivariadas estão sendo até hoje construídas, como por exemplo, os trabalhos de NADARAJAH (2006), NADARAJAH (2008) e OLKIN e TRIKALINOS (2015). A diferença é que estas distribuições mais recentes utilizam como marginais distribuições betas generalizadas. Desta forma, podemos notar que algumas distribuições beta bivariadas ainda estão sendo desenvolvidas para tratar dados de taxas e proporções, oferecendo a possibilidade de analisar conjuntamente as variáveis relacionadas, aproveitando os benefícios da abordagem conjunta.

### 3 MATERIAL

A fim de ilustrar os procedimentos de estimação propostos neste trabalho, são considerados os dados obtidos por meio do pacote computacional Distrain (TOMERLIN; HOWELL, 1988), criado com o intuito de treinar avaliadores para a quantificação visual da severidade de uma determinada doença. Este programa exibe imagens de plantas, com percentuais aleatórios de tecido doente, para que o avaliador forneça suas estimativas visuais. No caso em questão optou-se pela exibição de imagens de folhas de amendoim (*Arachis hypogaea*) cobertas por ferrugem (*Puccinia arachidis*), mancha castanha (*Cercospora arachidicola*) e mancha preta (*Cercosporidium personatum*) com diferentes porcentagens de tecido doente devido a importância dessas doenças no Brasil. Estas foram então, avaliadas uma a uma visualmente e os resultados para 30 imagens relativas a cada doença estão nas Tabelas 1 a 3.

Tabela 1 - Porcentagens reais de tecido doente (severidades) e estimadas visualmente por um avaliador, a partir de imagens de 30 folhas de amendoeiros com ferrugem (*Puccinia arachidis*), geradas pelo pacote computacional Distrain (TOMERLIN; HOWELL, 1988)

Imagem	Severidade		Imagem	Severidade		Imagem	Severidade	
	Real	Estimada		Real	Estimada		Real	Estimada
1	31	38	11	15	16	21	21	26
2	42	46	12	35	53	22	13	12
3	14	8	13	32	41	23	11	16
4	45	67	14	25	22	24	34	55
5	36	29	15	52	86	25	32	43
6	23	18	16	35	43	26	9	20
7	38	44	17	33	39	27	25	37
8	45	52	18	8	14	28	23	41
9	10	7	19	34	43	29	25	40
10	34	29	20	27	38	30	40	67

A Tabela 1 mostra os valores de severidade, estimados visualmente e reais, relativos a 30 imagens de folhas de amendoim com ferrugem. A Tabela 2 mostra, por

sua vez, os valores de severidade estimados visualmente e reais, relativos a 30 imagens de folhas de amendoim com mancha castanha. A Tabela 3, finalmente, mostra os valores de severidade, estimados visualmente e reais, relativos a 30 imagens de folhas de amendoim com mancha preta.

Tabela 2 - Porcentagens reais de tecido doente (severidades) e estimadas visualmente por um avaliador, a partir de imagens de 30 folhas de amendoinzeiros com mancha castanha (*Cercospora arachidicola*), geradas pelo pacote computacional Distrain (TOMERLIN; HOWELL, 1988)

Imagen	Severidade		Imagen	Severidade		Imagen	Severidade	
	Real	Estimada		Real	Estimada		Real	Estimada
1	43	45	11	7	9	21	53	54
2	54	63	12	47	37	22	9	12
3	51	46	13	20	21	23	46	27
4	4	8	14	55	47	24	2	6
5	17	14	15	37	32	25	6	11
6	44	34	16	51	41	26	40	28
7	54	55	17	27	22	27	53	45
8	1	6	18	4	9	28	23	26
9	8	11	19	32	26	29	30	30
10	53	63	20	32	31	30	38	37

De modo a avaliar os métodos de estimação propostos neste trabalho, situações comumente encontradas em dados de severidade de doença, reais e estimadas visualmente, serão considerados quatro cenários de simulação diferentes, todos com tamanhos amostrais  $n = 10, 30, 100$  ou  $300$  apresentados na Tabela 4.

Tabela 3 - Porcentagens reais de tecido doente (severidades) e estimadas visualmente por um avaliador, a partir de imagens de 30 folhas de amendoinzeiros com mancha preta (*Cercosporidium personatum*), geradas pelo pacote computacional Distrain (TOMERLIN; HOWELL, 1988)

Imagen	Severidade		Imagen	Severidade		Imagen	Severidade	
	Real	Estimada		Real	Estimada		Real	Estimada
1	4	5	11	33	18	21	3	8
2	51	55	12	27	12	22	46	35
3	47	40	13	15	14	23	31	25
4	7	6	14	29	20	24	23	13
5	14	9	15	51	46	25	23	14
6	23	12	16	29	20	26	22	16
7	51	50	17	5	9	27	23	18
8	42	43	18	27	15	28	22	16
9	32	20	19	53	54	29	11	8
10	33	19	20	38	43	30	40	28

Tabela 4 - Valores dos parâmetros  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ ,  $b$  e  $c$  da distribuição de NADARA-JAH e KOTZ (2005) utilizados para simulação de 1000 amostras

Cenário	$a_1$	$a_2$	$b_1$	$b_2$	$b$	$c$	$\beta_1 = b_1 + b$	$\beta_2 + b_2 + b$
1	2	2	0,5	0,5	2	2,5	2,5	2,5
2	2	2	0,2	0,2	2	2,2	2,2	2,2
3	1,5	2	0,7	0,2	2	2,2	2,7	2,2
4	2	1,5	0,2	0,7	2	2,2	2,2	2,7

Na Tabela 5, por sua vez, são apresentadas, para cada cenário considerado, as esperanças e variâncias das distribuições de  $X$  e de  $Y$  marginais, da distribuição de  $(X, Y)$  e coeficiente de correlação entre  $X$  e  $Y$ .

Tabela 5 - Esperanças e variâncias das distribuições de  $X$  e  $Y$ , marginais da distribuição de  $(X, Y)$  e coeficiente de correlação entre  $X$  e  $Y$  para cada cenário considerado na Tabela 4

Cenário	$E(X)$	$Var(X)$	$E(Y)$	$Var(Y)$	$Corr(X, Y)$
1	0,444	0,045	0,444	0,045	0,640
2	0,476	0,048	0,476	0,048	0,826
3	0,357	0,044	0,476	0,048	0,646
4	0,476	0,048	0,357	0,044	0,646

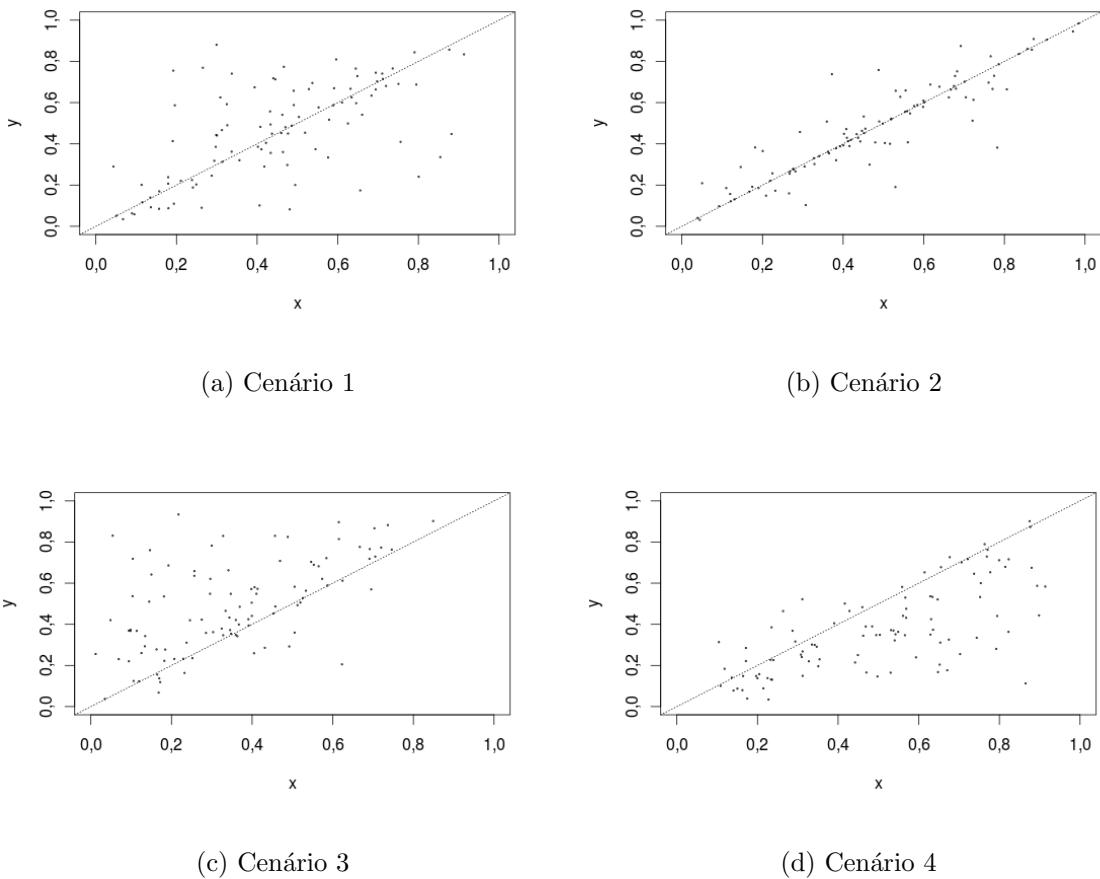


Figura 6 - Amostras simuladas de tamanho  $n = 100$  de distribuições de NADARA-JAH e KOTZ (2005) segundo os cenários de simulação 1 a 4 apresentados na Tabela 4

Da Tabela 4 observa-se que  $a_1 = a_2$  e  $b_1 = b_2$  para os cenários 1 e 2, o que leva ao fato de haver, em média, concordância entre as severidades observadas e as estimadas.

Entretanto, no caso 2, há uma maior correlação entre as observações (ver Tabela 5). Isso pode ser visto comparando-se as Figuras 6(a) a 6(d), que apresentam amostras simuladas de tamanho 100 das distribuições cujos parâmetros estão na Tabela 4.

Considere que  $Y$  seja a variável aleatória severidade da doença estimada visualmente e  $X$  a severidade real. Nesse caso, os cenários de simulação 3 e 4 revelam a tendência de o avaliador superestimar ou subestimar a severidade, respectivamente, como ilustram as Figuras 6(c) e 6(d).



## 4 MÉTODOS

Neste trabalho decidiu-se utilizar a distribuição beta bivariada proposta por NADARAJAH e KOTZ (2005), devido à sua maior flexibilidade na estrutura de dependência entre as variáveis em relação às demais distribuições beta bivariadas. Uma vez que, a função de densidade conjunta desta distribuição é descrita por expansões em séries, torna-se inviável encontrar os estimadores dos seus parâmetros analiticamente via método da máxima verossimilhança. Para estimar seus parâmetros propõe-se, como passo inicial, estimar os parâmetros das distribuições marginais de  $X$  e  $Y$ , ou seja,  $a_1, a_2, \beta_1 = b_1 + b$  e  $\beta_2 = b_2 + b$ . Na segunda etapa é apresentado um método para estimar os parâmetros restantes, isto é,  $b_1, b_2, b$  e  $c$ . Finalmente, é feito um estudo de simulação para avaliar o método de estimação proposto.

### 4.1 Estimação dos parâmetros

Como foi visto na seção 2, as distribuições marginais de  $(X, Y)$  com distribuição beta bivariada de NADARAJAH e KOTZ (2005) são distribuições beta, ou seja,  $X \sim Beta(a_1, \beta_1)$  e  $Y \sim Beta(a_2, \beta_2)$ .

Seja  $X$  uma variável aleatória contínua com distribuição beta com parâmetros  $a_1$  e  $\beta_1$  e  $X_1, X_2, \dots, X_n$  uma amostra aleatória de tamanho  $n$ , ou seja,  $X_i \stackrel{iid}{\sim} Beta(a_1, \beta_1)$ ,  $i = 1, \dots, n$ .

Uma forma de encontrar estimadores de parâmetros é por meio do método dos momentos. Este método é baseado nos momentos teóricos e amostrais das variáveis aleatórias envolvidas. Seja  $\mu_k$  o  $k$ -ésimo momento de  $X$ , dado por:

$$\begin{aligned}\mu_k &= \mathbb{E}(X^k) = \int_0^1 x^k \frac{x^{a_1-1}(1-x)^{\beta_1-1}}{B(a_1, \beta_1)} dx \\ &= \frac{1}{B(a_1, \beta_1)} \int_0^1 x^{a_1+k-1}(1-x)^{\beta_1-1} dx = \frac{B(a_1+k, \beta_1)}{B(a_1, \beta_1)}\end{aligned}$$

para  $k$  inteiro positivo. Em particular, tem-se que os dois primeiros momentos são:  $\mu_1 =$

$\frac{a_1}{a_1+\beta_1}$  e  $\mu_2 = \frac{a_1(a_1+1)}{(a_1+\beta_1)(a_1+\beta_1+1)}$ . Seja, ainda,  $m_k$  o  $k$ -ésimo momento amostral, dado por:

$$m_k = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

Em particular, tem-se que  $m_1 = \bar{X}$  e  $m_2 = \sum_{i=1}^n X_i^2/n$ . Logo o estimador  $\hat{a}_1$  e  $\hat{\beta}_1$  de  $a_1$  e  $\beta_1$ , respectivamente, obtido pelo método dos momentos são os valores  $a_1$  e  $\beta_1$  que solucionam o sistema de equações:

$$m_k = \mu_k, \quad k = 1, 2,$$

dado por  $\hat{a}_1 = \frac{m_1(m_1-m_2)}{m_2-m_1^2}$  e  $\hat{\beta}_1 = \hat{a}_1 \frac{(1-m_1)}{m_1}$ .

Analogamente pode-se obter os estimadores solucionando o sistema de equações obtido igualando-se  $\mu_1 = E(X)$  com  $\bar{X}$  e  $E[(X - \mu_1)^2] = Var(X)$  com  $S_X^2$ , em que:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

e

$$S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

No caso em questão,  $X \sim Beta(a_1, \beta_1)$  e os estimadores de  $a_1$  e  $\beta_1$  são obtidos solucionando-se o sistema de equações:

$$\begin{cases} \bar{X} &= \frac{a_1}{a_1 + \beta_1} \\ S_X^2 &= \frac{a_1 \beta_1}{(a_1 + \beta_1 + 1)(a_1 + \beta_1)^2} \end{cases} \quad (5)$$

Da primeira equação do sistema de equações (5) tem-se que:

$$\begin{aligned} a_1 &= (a_1 + \beta_1)\bar{X} \\ \beta_1\bar{X} &= a_1 - a_1\bar{X} \\ \beta_1 &= \frac{a_1}{\bar{X}} - a_1 = a_1 \left( \frac{1 - \bar{X}}{\bar{X}} \right) \end{aligned} \quad (6)$$

Da segunda equação do sistema de equações (5), por sua vez, tem-se que:

$$a_1 \beta_1 = (a_1 + \beta_1)^2 (a_1 + \beta_1 + 1) S_X^2 \quad (7)$$

e substituindo-se (6) na equação (7), obtém-se:

$$\begin{aligned}
 a_1^2 \left( \frac{1 - \bar{X}}{\bar{X}} \right) &= \left( a_1 + \frac{a_1}{\bar{X}} - a_1 \right)^2 \left( a_1 + \frac{a_1}{\bar{X}} - a_1 + 1 \right) S_X^2 \\
 a_1^2 \left( \frac{1 - \bar{X}}{\bar{X}} \right) &= a_1^2 \left( \frac{1}{\bar{X}^2} \right) \left( \frac{a_1}{\bar{X}} + 1 \right) S_X^2 \\
 \left( \frac{1 - \bar{X}}{\bar{X}} \right) \left( \frac{\bar{X}}{S_X^2} \right) &= \frac{a_1}{\bar{X}} + 1 \\
 \frac{(1 - \bar{X})\bar{X}}{S_X^2} - 1 &= \frac{a_1}{\bar{X}} \\
 a_1 &= \bar{X} \left[ \frac{(1 - \bar{X})\bar{X}}{S_X^2} - 1 \right]. \tag{8}
 \end{aligned}$$

Por fim, substituindo-se (8) em (6), obtém-se:

$$\begin{aligned}
 \beta_1 &= a_1 \left[ \frac{(1 - \bar{X})\bar{X}}{\bar{X}} \right] \\
 &= \bar{X} \left[ \frac{(1 - \bar{X})\bar{X}}{S_X^2} - 1 \right] \left[ \frac{(1 - \bar{X})\bar{X}}{\bar{X}} \right] \\
 &= (1 - \bar{X}) \left[ \frac{(1 - \bar{X})\bar{X}}{S_X^2} - 1 \right].
 \end{aligned}$$

Logo, os estimadores de  $a_1$  e  $\beta_1$ , obtidos pelo método dos momentos são dados, respectivamente, por:

$$\hat{a}_1 = \bar{X} \left[ \frac{(1 - \bar{X})\bar{X}}{S_X^2} - 1 \right]$$

e

$$\hat{\beta}_1 = (1 - \bar{X}) \left[ \frac{(1 - \bar{X})\bar{X}}{S_X^2} - 1 \right] = \frac{1 - \bar{X}}{\bar{X}} \hat{a}_1$$

para  $S_X^2 < \bar{X}(1 - \bar{X})$ . Analogamente, se  $Y \sim Beta(a_2, \beta_2)$ , pode-se provar que:

$$\hat{a}_2 = \bar{Y} \left[ \frac{(1 - \bar{Y})\bar{Y}}{S_Y^2} - 1 \right]$$

e

$$\hat{\beta}_2 = (1 - \bar{Y}) \left[ \frac{(1 - \bar{Y})\bar{Y}}{S_Y^2} - 1 \right] = \frac{1 - \bar{Y}}{\bar{Y}} \hat{a}_2$$

para  $S_Y^2 < \bar{Y}(1 - \bar{Y})$ , em que  $Y_1, Y_2, \dots, Y_n$  é uma amostra aleatória de tamanho  $n$  extraída de uma população com distribuição beta com parâmetros  $a_2$  e  $\beta_2$ , ou seja,  $Y \sim Beta(a_2, \beta_2)$ ,  $\bar{Y} = \sum_{i=1}^n Y_i/n$  e  $S_Y^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2/(n - 1)$ .

Outro método comumente utilizado para a estimação de parâmetros de dis-

tribuições é o método da máxima verossimilhança, que utiliza os resultados obtidos pela amostra para determinar qual a distribuição, assumindo quaisquer possíveis valores de seus parâmetros, possui maior possibilidade de ter gerado tal amostra. Seja  $X$  uma variável aleatória com distribuição beta com parâmetros  $a_1$  e  $\beta_1$ , ou seja,  $X \sim Beta(a_1, \beta_1)$  e  $X_1, \dots, X_n$  uma amostra aleatória extraída dessa população,  $x_1, \dots, x_n$  uma amostra observada e  $\tilde{\mathbf{x}} = (x_1, \dots, x_n)^t$  um vetor de observações. Sabendo-se que a função densidade de probabilidades de  $X$  é dada pela equação (2), a função de verossimilhança e seu logaritmo são, respectivamente,

$$\begin{aligned} L(a_1, \beta_1; \tilde{\mathbf{x}}) &= \prod_{i=1}^n \frac{1}{B(a_1, \beta_1)} x_i^{a_1-1} (1-x_i)^{\beta_1-1} \\ &= \frac{1}{[B(a_1, \beta_1)]^n} \left[ \prod_{i=1}^n x_i \right]^{a_1-1} \left[ \prod_{i=1}^n (1-x_i) \right]^{\beta_1-1} \end{aligned}$$

e

$$\begin{aligned} l(a_1, \beta_1; \tilde{\mathbf{x}}) &= \ln L(a_1, \beta_1; \tilde{\mathbf{x}}) \\ &= (a_1 - 1) \sum_{i=1}^n \ln x_i + (\beta_1 - 1) \sum_{i=1}^n \ln(1 - x_i) - n \ln B(a_1, \beta_1). \end{aligned} \quad (9)$$

Derivando e igualando a zero o logaritmo da função de verossimilhança dada pela expressão (9), em relação a  $a_1$  e  $\beta_1$ , obtém-se o sistema de equações:

$$\begin{cases} \frac{\partial l(a_1, \beta_1; \tilde{\mathbf{x}})}{\partial a_1} = \sum_{i=1}^n \ln x_i - n \frac{\partial \ln B(a_1, \beta_1)}{\partial a_1} = 0 \\ \frac{\partial l(a_1, \beta_1; \tilde{\mathbf{x}})}{\partial \beta_1} = \sum_{i=1}^n \ln(1 - x_i) - n \frac{\partial \ln B(a_1, \beta_1)}{\partial \beta_1} = 0. \end{cases} \quad (10)$$

Entretanto,  $B(a_1, \beta_1) = \frac{\Gamma(a_1)\Gamma(\beta_1)}{\Gamma(a_1+\beta_1)}$ . Logo, tem-se que  $\ln B(a_1, \beta_1) = \ln \Gamma(a_1) + \ln \Gamma(\beta_1) - \ln \Gamma(a_1 + \beta_1)$  e, consequentemente,

$$\frac{\partial \ln B(a_1, \beta_1)}{\partial a_1} = \frac{\partial \ln \Gamma(a_1)}{\partial a_1} - \frac{\partial \ln \Gamma(a_1 + \beta_1)}{\partial a_1} = \psi(a_1) - \psi(a_1 + \beta_1) \quad (11)$$

e

$$\frac{\partial \ln B(a_1, \beta_1)}{\partial \beta_1} = \frac{\partial \ln \Gamma(\beta_1)}{\partial \beta_1} - \frac{\partial \ln \Gamma(a_1 + \beta_1)}{\partial \beta_1} = \psi(\beta_1) - \psi(a_1 + \beta_1), \quad (12)$$

uma vez que a função digama é definida por  $\psi(x) = \frac{d}{dx} \ln \Gamma(x)$ . Assim, substituindo as expressões (11) e (12) no sistema de equações (10), segue que:

$$\begin{cases} \sum_{i=1}^n \ln x_i - n [\psi(a_1) - \psi(a_1 + \beta_1)] = 0 \\ \sum_{i=1}^n \ln(1 - x_i) - n [\psi(\beta_1) - \psi(a_1 + \beta_1)] = 0 \end{cases}$$

ou

$$\begin{cases} \psi(a_1) - \psi(a_1 + \beta_1) - \frac{1}{n} \sum_{i=1}^n \ln x_i = 0 \\ \psi(\beta_1) - \psi(a_1 + \beta_1) - \frac{1}{n} \sum_{i=1}^n \ln(1 - x_i) = 0 \end{cases} \quad (13)$$

O sistema de equações (13), no entanto, não possui solução com forma analítica fechada. Assim, para encontrar  $\hat{a}_1$  e  $\hat{\beta}_1$  são utilizados métodos iterativos, como o método de Newton-Raphson:

$$\underset{\sim}{\boldsymbol{\theta}}^{(j+1)} = \underset{\sim}{\boldsymbol{\theta}}^{(j)} - \mathbf{G}^{-1(j)} \underset{\sim}{\mathbf{g}}^{(j)}$$

em que:

$$\begin{aligned} \underset{\sim}{\boldsymbol{\theta}}^{(j)} &= (a_1^{(j)} \quad \beta_1^{(j)})^t \\ \underset{\sim}{\mathbf{g}}^{(j)} &= \begin{pmatrix} g_1^{(j)} \\ g_2^{(j)} \end{pmatrix} = \begin{pmatrix} \psi(a_1^{(j)}) - \psi(a_1^{(j)} + \beta_1^{(j)}) - \frac{1}{n} \sum_{i=1}^n \ln(x_i) \\ \psi(\beta_1^{(j)}) - \psi(a_1^{(j)} + \beta_1^{(j)}) - \frac{1}{n} \sum_{i=1}^n \ln(1 - x_i) \end{pmatrix} \\ \mathbf{G}^{(j)} &= \begin{pmatrix} \frac{\partial g_1^{(j)}}{\partial a_1} & \frac{\partial g_1^{(j)}}{\partial \beta_1} \\ \frac{\partial g_2^{(j)}}{\partial a_1} & \frac{\partial g_2^{(j)}}{\partial \beta_1} \end{pmatrix} = \begin{pmatrix} \psi'(a_1^{(j)}) - \psi'(a_1^{(j)} + \beta_1^{(j)}) & -\psi'(a_1^{(j)} + \beta_1^{(j)}) \\ -\psi'(a_1^{(j)} + \beta_1^{(j)}) & \psi'(\beta_1^{(j)}) - \psi'(a_1^{(j)} + \beta_1^{(j)}) \end{pmatrix} \end{aligned}$$

Para iniciar o processo iterativo precisa-se dos valores iniciais  $a_1^{(0)}$  e  $\beta_1^{(0)}$  para  $a_1$  e  $\beta_1$ . Uma possibilidade é utilizar as estimativas encontradas pelo método dos momentos. Outra possibilidade é utilizar a aproximação  $\psi(\alpha) \approx \ln(\alpha - 1/2)$ . Consequentemente, utilizando

as propriedades de logaritmos tem-se que:

$$\frac{1}{n} \sum_{i=1}^n \ln(x_i) = \ln \left[ \prod_{i=1}^n (x_i) \right]^{\frac{1}{n}} = \ln M_X$$

e

$$\frac{1}{n} \sum_{i=1}^n \ln(1 - x_i) = \ln \left[ \prod_{i=1}^n (1 - x_i) \right]^{\frac{1}{n}} = \ln M_{1-X},$$

em que  $M_X$  é a média geométrica dos valores observados de  $X$  e  $M_{1-X}$  é a média geométrica de 1 menos os valores observados de  $X$ . Logo:

$$\ln \left( \frac{a_1 - 1/2}{a_1 + \beta_1 - 1/2} \right) \approx \ln M_X$$

e

$$\ln \left( \frac{\beta_1 - 1/2}{a_1 + \beta_1 - 1/2} \right) \approx \ln M_{1-X}.$$

Pode-se utilizar, então, como valores iniciais para  $a_1$  e  $\beta_1$  no processo iterativo:

$$a_1^{(0)} \approx \frac{1}{2} + \frac{M_X}{2(1 - M_X - M_{1-X})}$$

e

$$\beta_1^{(0)} \approx \frac{1}{2} + \frac{M_{1-X}}{2(1 - M_X - M_{1-X})},$$

respectivamente.

De forma análoga à exposta até aqui, podem-se estimar os parâmetros  $a_2$  e  $\beta_2$  da distribuição marginal de  $Y$ .

Até o momento, foram obtidos estimadores dos parâmetros  $a_1$ ,  $\beta_1$ ,  $a_2$  e  $\beta_2$ , ou seja, das distribuições marginais beta da distribuição beta bivariada de NADARAJAH e KOTZ (2005). Resta, portanto, estimar os demais parâmetros da mesma, ou seja,  $b_1$ ,  $b_2$ ,

$b$  e  $c$ , que obedecem ao seguinte sistema de equações:

$$\begin{cases} b_1 + b = \beta_1 \\ b_2 + b = \beta_2 \\ c - b_1 = a_1 \\ c - b_2 = a_2. \end{cases} \quad (14)$$

O sistema de equações (14) pode, entretanto, ser reescrito matricialmente da seguinte forma:

$$\underset{\sim}{\mathbf{A}} \underset{\sim}{\boldsymbol{\theta}_2} = \underset{\sim}{\boldsymbol{\theta}_1}, \quad (15)$$

em que:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 1 \end{pmatrix}, \quad \underset{\sim}{\boldsymbol{\theta}_2} = \begin{pmatrix} b_1 \\ b_2 \\ b \\ c \end{pmatrix}, \quad \underset{\sim}{\boldsymbol{\theta}_1} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ a_1 \\ a_2 \end{pmatrix}.$$

A matriz:

$$\underset{\sim}{(\mathbf{A} | \boldsymbol{\theta}_1)} = \left( \begin{array}{cccc|c} 1 & 0 & 1 & 0 & \beta_1 \\ 0 & 1 & 1 & 0 & \beta_2 \\ -1 & 0 & 0 & 1 & a_1 \\ 0 & -1 & 0 & 1 & a_2 \end{array} \right)$$

tem, por sua vez, sua forma escalonada dada por:

$$\underset{\sim}{(\mathbf{A} | \boldsymbol{\theta}_1)} = \left( \begin{array}{cccc|c} 1 & 0 & 1 & 0 & \beta_1 \\ 0 & 1 & 1 & 0 & \beta_2 \\ 0 & 0 & 1 & 1 & a_1 + \beta_1 \\ 0 & 0 & 0 & 0 & a_2 + \beta_2 - (a_1 + \beta_1) \end{array} \right).$$

Consequentemente, para que o sistema de equações (14) possua infinitas so-

luções é preciso que  $a_2 + \beta_2 - a_1 - \beta_1 = 0$ , ou seja:

$$a_1 + \beta_1 = a_2 + \beta_2.$$

Como, na prática, é quase impossível que  $\hat{a}_1 + \hat{\beta}_1 = \hat{a}_2 + \hat{\beta}_2$ , será abordada uma outra metodologia para a solução deste sistema.

Propõe-se, aqui, três métodos diferentes para encontrar os estimadores, que contemplam algumas das restrições ou equações apresentadas no sistema de equações (15). Após a apresentação dos métodos, serão feitas simulações para avaliar as vantagens e desvantagens de cada um.

Para obter os estimadores dos parâmetros  $b_1$ ,  $b_2$ ,  $b$  e  $c$  será utilizada a correlação amostral, entre as variáveis  $X$  e  $Y$ , dada por:

$$r_{XY} = \frac{1}{n-1} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S_X} \right) \left( \frac{Y_i - \bar{Y}}{S_Y} \right)$$

em que  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ,  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ ,  $S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  e  $S_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$ .

#### 4.1.1 Método 1

Neste primeiro método, manipula-se algebraicamente algumas das equações descritas no sistema de equações (15), com o intuito de reescrever as relações entre os parâmetros. Tem-se então que:

$$\beta_1 = b_1 + b \Leftrightarrow b_1 = \beta_1 - b,$$

$$c = a_1 + b_1 \Leftrightarrow b_1 = c - a_1$$

e portanto,

$$\beta_1 - b = c - a_1 \Leftrightarrow b + c = a_1 + \beta_1.$$

Seguindo o mesmo raciocínio para as demais equações,

$$\beta_2 = b_2 + b \Leftrightarrow b_2 = \beta_2 - b,$$

$$c = a_2 + b_2 \Leftrightarrow b_2 = c - a_2$$

e portanto,

$$\beta_2 - b = c - a_2 \Leftrightarrow b + c = a_2 + \beta_2.$$

Adicionalmente, uma vez que  $X$  e  $Y$  possuem, marginalmente, distribuições beta, tem-se que:

$$E(X)E(Y) = \frac{a_1}{a_1 + b_1 + b} \frac{a_2}{a_2 + b_2 + b} = \frac{a_1}{b+c} \frac{a_2}{b+c} = \frac{a_1 a_2}{(b+c)^2}, \quad (16)$$

e consequentemente:

$$Cor(X, Y) = \frac{b}{c} \sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}}. \quad (17)$$

Ao igualar o produto das médias populacionais (16) com o produto das amostrais (17) e a correlação populacional com a amostral obtém-se, respectivamente,

$$E(X)E(Y) = \bar{X}\bar{Y} \Leftrightarrow \frac{a_1 a_2}{(b+c)^2} = \bar{X}\bar{Y} \Leftrightarrow b+c = \sqrt{\frac{a_1 a_2}{\bar{X}\bar{Y}}}$$

e

$$Cor(X, Y) = r_{XY} \Leftrightarrow \frac{b}{c} \sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}} = r_{XY} \Leftrightarrow \frac{b}{c} = \frac{r_{XY}}{\sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}}}.$$

Logo, dados  $a_1$ ,  $a_2$ ,  $\beta_1$ ,  $\beta_2$  e  $r_{XY}$ , chega-se ao sistema de equações:

$$\begin{cases} b+c = \sqrt{\frac{a_1 a_2}{\bar{X}\bar{Y}}} \\ \frac{b}{c} = \frac{r_{XY}}{\sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}}} \end{cases} \quad (18)$$

cuja solução em termos de  $b$  e  $c$  é:

$$b = \frac{\sqrt{\frac{a_1 a_2}{\bar{X}\bar{Y}}} r_{XY}}{r_{XY} + \sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}}} = \frac{\frac{r_{XY}}{\sqrt{\bar{X}\bar{Y}}}}{\frac{r_{XY}}{\sqrt{a_1 a_2}} + \frac{1}{\sqrt{\beta_1 \beta_2}}} \quad (19)$$

e

$$c = \frac{\sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}} \sqrt{\frac{a_1 a_2}{\bar{X} \bar{Y}}}}{r_{XY} + \sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}}} = \sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}} \frac{b}{r_{XY}}. \quad (20)$$

Considerando as estimativas de  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$  obtidas pelos métodos descritos na seção 4.1 e substituindo-as nas equações (19) e (20), obtém-se os estimadores dos parâmetros  $b$  e  $c$  dados, respectivamente, por:

$$\hat{b} = \frac{\frac{r_{XY}}{\sqrt{\bar{X} \bar{Y}}}}{\frac{r_{XY}}{\sqrt{\hat{a}_1 \hat{a}_2}} + \frac{1}{\sqrt{\hat{\beta}_1 \hat{\beta}_2}}} \quad (21)$$

e

$$\hat{c} = \sqrt{\frac{\hat{a}_1 \hat{a}_2}{\hat{\beta}_1 \hat{\beta}_2}} \frac{\hat{b}}{r_{XY}}. \quad (22)$$

Adicionalmente, das igualdades  $b_1 = c - a_1$  e  $b_2 = c - a_2$ , chega-se a:

$$\hat{b}_1 = \hat{c} - \hat{a}_1 \quad (23)$$

e

$$\hat{b}_2 = \hat{c} - \hat{a}_2. \quad (24)$$

O problema deste método, entretanto, é que não são respeitadas as igualdades:

$$\hat{\beta}_1 = \hat{b}_1 + \hat{b}$$

e

$$\hat{\beta}_2 = \hat{b}_2 + \hat{b}.$$

#### 4.1.2 Método 2.1

Neste método, são utilizados três das quatro restrições, para que o sistema de equações (14) passe a ter solução única. Para isso iguala-se a equação da correlação populacional, dada por (17), à correlação amostral, ou seja,

$$\frac{b}{c} \sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}} = r_{XY}.$$

Consequentemente, chega-se ao sistema de equações:

$$\begin{cases} b_1 + b = \beta_1 \\ b_2 + b = \beta_2 \\ c - b_1 = a_1 \\ \frac{b}{c} = \frac{r_{XY}}{\sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}}}, \end{cases} \quad (25)$$

cuja solução em termos de  $b_1$ ,  $b_2$ ,  $b$  e  $c$  é:

$$b_1 = \frac{\beta_1 K - a_1 r_{XY}}{r_{XY} + K}, \quad (26)$$

$$b_2 = \frac{\beta_2(r_{XY} + K) - \beta_1 r_{XY} - a_1 r_{XY}}{r_{XY} + K}, \quad (27)$$

$$b = \frac{\beta_1 r_{XY} + a_1 r_{XY}}{r_{XY} + K}, \quad (28)$$

e

$$c = \frac{\beta_1 K + a_1 r_{XY}}{r_{XY} + K}, \quad (29)$$

em que

$$K = \sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}}. \quad (30)$$

Assim, substituindo-se as estimativas de  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$ , obtidas pelos métodos descritos na seção 4.1, nas equações (26) a (30), obtém-se os estimadores de  $b_1$ ,  $b_2$ ,  $b$  e  $c$  dados, respectivamente, por:

$$\begin{aligned}\hat{b}_1 &= \frac{\hat{\beta}_1 \hat{K} - \hat{a}_1 r_{XY}}{r_{XY} + \hat{K}}, \\ \hat{b}_2 &= \frac{\hat{\beta}_2(r_{XY} + \hat{K}) - \hat{\beta}_1 r_{XY} - \hat{a}_1 r_{XY}}{r_{XY} + \hat{K}}, \\ \hat{b} &= \frac{\hat{\beta}_1 r_{XY} + \hat{a}_1 r_{XY}}{r_{XY} + \hat{K}}, \\ \hat{c} &= \frac{\hat{\beta}_1 \hat{K} + \hat{a}_1 r_{XY}}{r_{XY} + \hat{K}},\end{aligned}$$

e

$$\hat{K} = \sqrt{\frac{\hat{a}_1 \hat{a}_2}{\hat{\beta}_1 \hat{\beta}_2}}.$$

em que  
Neste método, entretanto, não se consolidou uma restrição para resolver o sistema de equações (14), consequentemente, não é respeitada a igualdade:

$$\hat{c} = \hat{a}_2 + \hat{b}_2.$$

#### 4.1.3 Método 2.2

Assim como no método 2.1, utiliza-se três das quatro restrições, para que o sistema de equações (14) passe a ter solução única. No método 2.1 retirou-se a restrição  $c = a_2 + b_2$ , desta vez será retirada a equação  $c = a_1 + b_1$ . Para isso iguala-se a equação da correlação populacional, dada por (17), à correlação amostral, ou seja,

$$\frac{b}{c} \sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}} = r_{XY}.$$

Consequentemente, chega-se ao sistema de equações:

$$\begin{cases} b_1 + b = \beta_1 \\ b_2 + b = \beta_2 \\ c - b_2 = a_2 \\ \frac{b}{c} = \frac{r_{XY}}{\sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}}}, \end{cases} \quad (31)$$

cuja solução em termos de  $b_1$ ,  $b_2$ ,  $b$  e  $c$  é:

$$b_1 = \frac{\beta_1(r_{XY} + K) - \beta_2 r_{XY} - a_2 r_{XY}}{r_{XY} + K} \quad (32)$$

$$b_2 = \frac{K\beta_2 - a_2 r_{XY}}{r_{XY} + K}, \quad (33)$$

$$b = \frac{\beta_2 r_{XY} + a_2 r_{XY}}{r_{XY} + K}, \quad (34)$$

e

$$c = \frac{K\beta_2 + a_2 K}{r_{XY} + K}, \quad (35)$$

em que

$$K = \sqrt{\frac{a_1 a_2}{\beta_1 \beta_2}}. \quad (36)$$

Assim, substituindo-se as estimativas de  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$ , obtidas pelos métodos descritos na seção 4.1, nas equações (32) a (36), obtêm-se os estimadores de  $b_1$ ,  $b_2$ ,  $b$  e  $c$  dados, respectivamente, por:

$$\hat{b}_1 = \frac{\hat{\beta}_1(r_{XY} + \hat{K}) - \hat{\beta}_2 r_{XY} - \hat{a}_2 r_{XY}}{r_{XY} + \hat{K}},$$

$$\hat{b}_2 = \frac{\hat{K}\hat{\beta}_2 - \hat{a}_2 r_{XY}}{r_{XY} + \hat{K}},$$

$$\hat{b} = \frac{\hat{\beta}_2 r_{XY} + \hat{a}_2 r_{XY}}{r_{XY} + \hat{K}},$$

e

$$\hat{c} = \frac{\hat{K} \hat{\beta}_2 + \hat{a}_2 \hat{K}}{r_{XY} + \hat{K}},$$

em que

$$\hat{K} = \sqrt{\frac{\hat{a}_1 \hat{a}_2}{\hat{\beta}_1 \hat{\beta}_2}}.$$

Neste método, entretanto, não se consolidou uma restrição para resolver o sistema de equações (14), consequentemente, não é respeitada a igualdade:

$$\hat{c} = \hat{a}_1 + \hat{b}_1.$$

## 4.2 Avaliação dos estimadores

Para comparar os métodos descritos até aqui, serão utilizadas algumas medidas de precisão e viés dos estimadores. Serão apresentadas, ainda, simulações com a finalidade de encontrar o melhor método para diferentes situações. Seja  $\theta$  o parâmetro de interesse (pode ser  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ ,  $b$  ou  $c$ , por exemplo, sem perda de generalidades) e  $\hat{\theta}$ , seu estimador. Serão consideradas, assim, as seguintes medidas:

(i) Erro quadrático médio, dado por:

$$EQM(\hat{\theta}) = E \left[ (\hat{\theta} - \theta)^2 \right]. \quad (37)$$

(ii) Variância do estimador, dada por:

$$Var(\hat{\theta}) = E \left[ (\hat{\theta} - E(\hat{\theta}))^2 \right]. \quad (38)$$

(iii) Viés, dado por:

$$Viés(\hat{\theta}) = E(\hat{\theta}) - \theta. \quad (39)$$

(iv) Coeficiente de variação, dado por:

$$CV(\hat{\theta}) = \frac{\sqrt{Var(\hat{\theta})}}{\theta} 100\% \quad (40)$$

(v) Viés relativo: Assim como o coeficiente de variação, o viés relativo é uma padronização do viés do estimador, usado para os mesmos fins, mas podendo comparar estimadores com diferentes escalas de grandeza.

$$VR(\hat{\theta}) = \frac{Viés(\hat{\theta})}{\theta} 100\% \quad (41)$$

Note que há uma relação importante entre as medidas apresentadas nos itens

(i) a (v), dada por:

$$\begin{aligned} EQM(\hat{\theta}) &= E \left[ (\hat{\theta} - \theta)^2 \right] \\ &= E \left[ (\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2 \right] \\ &= E \left[ (\hat{\theta} - E(\hat{\theta}))^2 + 2(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) + (E(\hat{\theta}) - \theta)^2 \right] \\ &= E \left[ (\hat{\theta} - E(\hat{\theta}))^2 \right] + 2E \left[ (\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta) \right] + E \left[ (E(\hat{\theta}) - \theta)^2 \right] \\ &= E \left[ (\hat{\theta} - E(\hat{\theta}))^2 \right] + 2(E(\hat{\theta}) - \theta) E \left[ \hat{\theta} - E(\hat{\theta}) \right] + (E(\hat{\theta}) - \theta)^2 \\ &= E \left[ (\hat{\theta} - E(\hat{\theta}))^2 \right] + 2(E(\hat{\theta}) - \theta)(E(\hat{\theta}) - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)^2 \\ &= E \left[ (\hat{\theta} - E(\hat{\theta}))^2 \right] + (E(\hat{\theta}) - \theta)^2 \\ &= Var(\hat{\theta}) + (Viés(\hat{\theta}))^2 \end{aligned} \quad (42)$$

Assim, pode-se dizer que o erro quadrático médio é dado pela soma da variância com o viés ao quadrado, ou seja, pela soma de uma medida de imprecisão com o quadrado do viés.

Note, ainda, que se ambos os lados da equação (42) forem divididos por  $\theta^2$ , obtém-se uma medida de erro quadrático médio relativo, dado por:

$$\frac{EQM(\hat{\theta})}{\theta^2} = \frac{Var(\hat{\theta})}{\theta^2} + \left( \frac{Viés(\hat{\theta})}{\theta} \right)^2$$

ou, ainda,

$$\begin{aligned} \frac{EQM(\hat{\theta})}{\theta^2} (100\%)^2 &= \left( \frac{\sqrt{Var(\hat{\theta})}}{\theta} 100\% \right)^2 + \left( \frac{Viés(\hat{\theta})}{\theta} 100\% \right)^2 \\ &= [CV(\hat{\theta})]^2 + [VR(\hat{\theta})]^2, \end{aligned} \quad (43)$$

ou seja, o erro quadrático médio relativo é a soma do quadrado do coeficiente de variação com o quadrado do viés relativo.

Na prática, para avaliar a qualidade do estimador  $\hat{\theta}$  de um parâmetro  $\theta$  serão geradas  $B$  amostras de tamanho  $n$  da distribuição de interesse e obtidas as estimativas  $\hat{\theta}_b$ ,  $b = 1, \dots, B$  de cada uma das amostras. Após feito isso avalia-se a qualidade dos estimadores da seguinte forma:

$$\widehat{EQM}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b - \theta)^2,$$

$$\widehat{Var}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b - \bar{\theta})^2$$

e

$$\widehat{Viés}(\hat{\theta}) = \bar{\theta} - \theta$$

em que  $\bar{\theta} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b$ .

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Simulações

**Cenário 1:** Neste primeiro caso considera-se que existe uma concordância entre os valores de severidade da doença estimados visualmente por um avaliador, e o valor real avaliado pelo computador, mas que, como espera-se, existe uma maior discrepância conforme a proporção da severidade se aproxima de 0,5. Utilizou-se os seguintes valores dos parâmetros  $a_1=2,0$ ,  $a_2=2,0$ ,  $b_1=0,5$ ,  $b_2=0,5$ ,  $b=2,0$  e  $c=2,5$  para gerar mil bancos de dados com estes parâmetros e, para cada banco gerado, calculou-se todas as estimativas utilizando os diferentes métodos propostos. Além disso, foi variado o tamanho da amostra em 10, 30, 100 e 300 pares de observações, com a finalidade de avaliar os estimadores quanto ao viés e precisão. A Tabela 6 apresenta medidas de precisão e viés dos estimadores, utilizando somente a primeira parte do processo estimativo, ou seja, por meio das distribuições marginais, das mil simulações feitas para amostras de tamanho 10, 30, 100 e 300.

Por meio desta, conforme aumenta-se o tamanho da amostra, o viés relativo diminui, ou seja, as estimativas se aproximam do verdadeiro valor dos parâmetros, indicando a convergência de ambos os métodos. Entretanto o viés relativo do estimador de máxima verossimilhança foi sempre maior, máximo ao dobro, do viés relativo do estimador por meio do método dos momentos. Além disso o estimador por meio do método dos momentos foi mais preciso (menor coeficiente de variação) do que o método da máxima verossimilhança. Portanto, o estimador via método dos momentos parece ser melhor. Sendo assim, partindo do valor destas estimativas, continua-se as simulações para os demais parâmetros. Os resultados da segunda parte das simulações podem ser vistos na Tabela 7. Através desta tabela compara-se os métodos descritos neste trabalho.

Para os valores atribuídos aos parâmetros nestas simulações, procurando reproduzir um caso que os pares de observações fossem concordantes com maior discordância na proporção 0,5, o método 1 teve um melhor desempenho, com estimativas mais próximas em média e com menor viés relativo.

Tabela 6 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$  no cenário 1, via método dos momentos e da máxima verossimilhança, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300

Método	Momentos				Verossimilhança			
Parâmetro	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.
<i>n=10</i>								
$a_1=2,0$	2,382	2,510	0,769	0,191	2,782	3,575	0,861	0,391
$a_2=2,0$	2,404	2,597	0,780	0,202	2,790	3,610	0,864	0,395
$\beta_1=2,5$	2,985	3,706	0,745	0,194	3,488	5,298	0,832	0,395
$\beta_2=2,5$	3,059	4,376	0,806	0,224	3,552	6,127	0,896	0,421
<i>n=30</i>								
$a_1=2,0$	2,099	0,344	0,289	0,050	2,204	0,380	0,291	0,102
$a_2=2,0$	2,091	0,368	0,300	0,046	2,204	0,412	0,304	0,102
$\beta_1=2,5$	2,638	0,554	0,292	0,055	2,773	0,623	0,296	0,109
$\beta_2=2,5$	2,623	0,579	0,300	0,049	2,769	0,655	0,305	0,108
<i>n=100</i>								
$a_1=2,0$	2,027	0,094	0,153	0,014	2,054	0,090	0,147	0,027
$a_2=2,0$	2,033	0,094	0,152	0,016	2,067	0,091	0,147	0,034
$\beta_1=2,5$	2,544	0,148	0,153	0,018	2,579	0,144	0,148	0,032
$\beta_2=2,5$	2,548	0,146	0,152	0,019	2,590	0,145	0,148	0,036
<i>n=300</i>								
$a_1=2,0$	2,003	0,029	0,085	0,002	2,011	0,027	0,082	0,006
$a_2=2,0$	2,009	0,027	0,082	0,004	2,016	0,025	0,079	0,008
$\beta_1=2,5$	2,503	0,046	0,086	0,001	2,513	0,043	0,083	0,005
$\beta_2=2,5$	2,512	0,042	0,082	0,005	2,521	0,039	0,079	0,008

Tabela 7 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros  $b_1$ ,  $b_2$ ,  $b$  e  $c$  no cenário 1, via método 1, método 2.1 e método 2.2, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300

Método	Método 1				Método 2.1				Método 2.2			
Parâmetro	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.
<i>n=10</i>												
$b_1=0,5$	0,883	3,218	3,505	0,766	0,872	2,512	3,081	0,744	1,185	5,445	4,461	1,370
$b_2=0,5$	0,882	2,831	3,277	0,764	1,226	4,695	4,083	1,452	0,942	4,740	4,264	0,884
$b=2,0$	2,144	3,134	0,882	0,072	2,193	3,251	0,896	0,097	2,250	4,605	1,066	0,125
$c=2,5$	3,190	6,809	1,007	0,276	3,287	8,279	1,107	0,315	3,369	11,141	1,289	0,348
<i>n=30</i>												
$b_1=0,5$	0,580	0,256	0,999	0,160	0,594	0,186	0,843	0,188	0,676	0,493	1,360	0,352
$b_2=0,5$	0,590	0,241	0,964	0,180	0,666	0,490	1,360	0,332	0,599	0,210	0,896	0,198
$b=2,0$	2,013	0,275	0,262	0,006	2,038	0,374	0,305	0,019	2,024	0,357	0,299	0,012
$c=2,5$	2,687	0,669	0,319	0,075	2,716	0,814	0,350	0,086	2,707	0,883	0,367	0,083
<i>n=100</i>												
$b_1=0,5$	0,504	0,049	0,443	0,008	0,514	0,032	0,354	0,028	0,533	0,124	0,702	0,066
$b_2=0,5$	0,513	0,049	0,442	0,026	0,520	0,115	0,676	0,040	0,515	0,029	0,338	0,030
$b=2,0$	2,014	0,067	0,129	0,007	2,023	0,091	0,150	0,012	2,015	0,087	0,147	0,008
$c=2,5$	2,527	0,138	0,148	0,011	2,538	0,173	0,166	0,015	2,529	0,171	0,165	0,012
<i>n=300</i>												
$b_1=0,5$	0,496	0,015	0,242	-0,008	0,499	0,009	0,192	-0,002	0,500	0,042	0,412	0,000
$b_2=0,5$	0,500	0,015	0,245	0,000	0,500	0,041	0,406	0,000	0,501	0,009	0,191	0,002
$b=2,0$	2,004	0,020	0,071	0,002	2,006	0,027	0,083	0,003	2,005	0,026	0,081	0,002
$c=2,5$	2,503	0,042	0,082	0,001	2,506	0,054	0,093	0,002	2,504	0,051	0,091	0,002

**Cenário 2:** Neste caso, assim como no anterior, considera-se que existe uma concordância entre os valores de severidade da doença estimados por um avaliador, e o valor real avaliado pelo computador, mas que, existe uma menor discrepância conforme a proporção da severidade se aproxima de 0,5. Utilizou-se os seguintes valores dos parâmetros  $a_1=2,0$ ,  $a_2=2,0$ ,  $b_1=0,2$ ,  $b_2=0,2$ ,  $b=2,0$  e  $c=2,2$  para gerar mil bancos de dados com estes parâmetros e, para cada banco gerado, calculou-se todas as estimativas utilizando os diferentes métodos propostos. Da mesma forma como no caso anterior, foi variado o tamanho da amostra em 10, 30, 100 e 300 pares de observações, com a finalidade de avaliar os estimadores quanto ao viés e precisão. A Tabela 8 apresenta medidas de precisão e viés

dos estimadores, utilizando somente a primeira parte do processo estimativo, ou seja, por meio das distribuições marginais, das mil simulações feitas para amostras de tamanho 10, 30, 100 e 300.

Tabela 8 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$  no cenário 2, via método dos momentos e da máxima verossimilhança, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300

Método	Momentos				Verossimilhança			
	Parâmetro	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.
<i>n=10</i>								
$a_1=2,0$	2,448	3,990	0,973	0,224	2,850	5,418	1,083	0,425
$a_2=2,0$	2,404	2,361	0,741	0,202	2,794	3,338	0,823	0,397
$\beta_1=2,2$	2,761	4,864	0,970	0,255	3,212	6,583	1,072	0,460
$\beta_2=2,2$	2,693	3,067	0,764	0,224	3,125	4,265	0,839	0,420
<i>n=30</i>								
$a_1=2,0$	2,100	0,361	0,296	0,050	2,219	0,411	0,301	0,109
$a_2=2,0$	2,094	0,367	0,299	0,047	2,212	0,405	0,300	0,106
$\beta_1=2,2$	2,310	0,445	0,299	0,050	2,440	0,502	0,303	0,109
$\beta_2=2,2$	2,308	0,480	0,311	0,049	2,437	0,531	0,313	0,108
<i>n=100</i>								
$a_1=2,0$	2,035	0,088	0,147	0,018	2,068	0,086	0,143	0,034
$a_2=2,0$	2,038	0,092	0,150	0,019	2,064	0,086	0,143	0,032
$\beta_1=2,2$	2,236	0,110	0,150	0,016	2,272	0,106	0,145	0,033
$\beta_2=2,2$	2,242	0,116	0,154	0,019	2,272	0,110	0,147	0,033
<i>n=300</i>								
$a_1=2,0$	2,016	0,028	0,083	0,008	2,029	0,027	0,081	0,014
$a_2=2,0$	2,018	0,029	0,084	0,009	2,029	0,027	0,080	0,014
$\beta_1=2,2$	2,210	0,033	0,082	0,005	2,225	0,031	0,079	0,011
$\beta_2=2,2$	2,214	0,033	0,082	0,006	2,227	0,031	0,079	0,012

Da mesma forma que o caso anterior, conforme aumenta-se o tamanho da

amostra, o viés relativo diminui, ou seja, as estimativas se aproximam do verdadeiro valor dos parâmetros, indicando a convergência de ambos os métodos. Entretanto o viés relativo do estimador de máxima verossimilhança foi maior, se comparado ao do estimador por meio do método dos momentos. Além disso o estimador por meio do método dos momentos foi mais preciso (menor coeficiente de variação) do que o método da máxima verossimilhança. Portanto, o estimador via método dos momentos parece ser melhor novamente. Sendo assim, partindo do valor destas estimativas, continua-se as simulações para os demais parâmetros. Os resultados da segunda parte das simulações podem ser vistos na Tabela 9, que compara os métodos descritos neste trabalho.

Tabela 9 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros  $b_1$ ,  $b_2$ ,  $b$  e  $c$  no cenário 2, via método 1, método 2.1 e método 2.2, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300

Método	Método 1				Método 2.1				Método 2.2			
Parâmetro	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.
<i>n=10</i>												
$b_1=0,2$	0,423	0,712	4,069	1,115	0,385	0,943	4,767	0,925	0,625	1,652	6,066	2,125
$b_2=0,2$	0,432	0,895	4,585	1,160	0,636	1,684	6,111	2,180	0,375	0,577	3,694	0,875
$b=2,0$	2,281	1,871	0,669	0,141	2,322	2,101	0,707	0,161	2,327	2,428	0,762	0,163
$c=2,2$	2,766	4,106	0,884	0,257	2,835	5,465	1,023	0,289	2,819	4,870	0,963	0,281
<i>n=30</i>												
$b_1=0,2$	0,248	0,065	1,255	0,240	0,233	0,036	0,940	0,165	0,319	0,169	1,969	0,595
$b_2=0,2$	0,245	0,066	1,261	0,225	0,323	0,164	1,933	0,615	0,232	0,038	0,956	0,160
$b=2,0$	2,079	0,288	0,265	0,040	2,088	0,345	0,291	0,044	2,090	0,336	0,286	0,045
$c=2,2$	2,319	0,421	0,290	0,054	2,328	0,482	0,312	0,058	2,333	0,506	0,318	0,060
<i>n=100</i>												
$b_1=0,2$	0,210	0,016	0,640	0,050	0,207	0,007	0,418	0,035	0,224	0,036	0,943	0,120
$b_2=0,2$	0,201	0,015	0,617	0,005	0,230	0,041	0,995	0,150	0,204	0,007	0,429	0,020
$b=2,0$	2,000	0,074	0,136	0,000	1,999	0,085	0,146	0,000	2,005	0,083	0,144	0,002
$c=2,2$	2,210	0,096	0,141	0,005	2,209	0,108	0,149	0,004	2,216	0,110	0,151	0,007
<i>n=300</i>												
$b_1=0,2$	0,200	0,006	0,380	0,000	0,201	0,002	0,240	0,005	0,205	0,016	0,625	0,025
$b_2=0,2$	0,201	0,006	0,371	0,005	0,205	0,016	0,625	0,025	0,201	0,002	0,238	0,005
$b=2,0$	2,005	0,024	0,078	0,002	2,006	0,028	0,084	0,003	2,006	0,028	0,083	0,003
$c=2,2$	2,202	0,033	0,082	0,001	2,203	0,037	0,087	0,001	2,202	0,037	0,087	0,001

Verifica-se que, para os valores atribuídos aos parâmetros nestas simulações, procurando reproduzir um caso que os pares de observações fossem concordantes mas, desta vez, com menor discordância na proporção 0,5, os três métodos foram similares, mas optou-se pelo método 1 por ser mais simples computacionalmente.

**Cenário 3:** Neste outro caso considera-se algo que acontece com frequência, a superestimação das proporções da severidade da doença pelos avaliadores, isto é, o valor real avaliado pelo computador é inferior ao valor estimado pelo avaliador mas, além disso, espera-se uma maior discrepância conforme a proporção da severidade se aproxima de 0,5. Utilizou-se os seguintes valores dos parâmetros  $a_1=1,5$ ,  $a_2=2,0$ ,  $b_1=0,7$ ,  $b_2=0,2$ ,  $b=2,0$  e  $c=2,2$  para gerar mil bancos de dados com estes parâmetros e, para cada banco gerado, calculou-se todas as estimativas utilizando os diferentes métodos propostos. Além disso, foi variado o tamanho da amostra em 10, 30, 100 e 300 pares de observações, com a finalidade de avaliar os estimadores quanto ao viés e precisão. A Tabela 10 apresenta medidas de precisão e viés dos estimadores, utilizando somente a primeira parte do processo estimativo, ou seja, por meio das distribuições marginais, das mil simulações feitas para amostras de tamanho 10, 30, 100 e 300.

Assim como nos cenários anteriores, conforme aumenta-se o tamanho da amostra, o viés relativo diminui, ou seja, as estimativas se aproximam do verdadeiro valor dos parâmetros, indicando a convergência de ambos os métodos. Entretanto o viés relativo do estimador de máxima verossimilhança foi sempre maior, máximo ao dobro, do viés relativo do estimador por meio do método dos momentos. Além disso o estimador por meio do método dos momentos foi mais preciso (menor coeficiente de variação) do que o método da máxima verossimilhança. Portanto, o estimador via método dos momentos foi melhor mais uma vez. Sendo assim, partindo do valor destas estimativas, continua-se as simulações para os demais parâmetros. Os resultados da segunda parte das simulações podem ser vistos na Tabela 11, a fim de comparar os métodos descritos.

Tabela 10 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$  no cenário 3, via método dos momentos e da máxima verossimilhança, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300

Método	Momentos				Verossimilhança			
Parâmetro	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.
<i>n=10</i>								
$a_1=1,5$	1,808	1,452	0,777	0,205	2,070	1,953	0,851	0,380
$a_2=2,0$	2,417	3,847	0,958	0,208	2,810	4,976	1,039	0,405
$\beta_1=2,7$	3,323	5,280	0,819	0,231	3,806	6,931	0,885	0,410
$\beta_2=2,2$	2,631	3,648	0,846	0,196	3,061	4,927	0,930	0,391
<i>n=30</i>								
$a_1=1,5$	1,580	0,194	0,289	0,053	1,655	0,206	0,284	0,103
$a_2=2,0$	2,111	0,369	0,299	0,056	2,223	0,426	0,307	0,111
$\beta_1=2,7$	2,848	0,710	0,307	0,055	2,987	0,759	0,305	0,106
$\beta_2=2,2$	2,305	0,417	0,290	0,048	2,428	0,475	0,296	0,104
<i>n=100</i>								
$a_1=1,5$	1,517	0,045	0,141	0,011	1,537	0,040	0,131	0,025
$a_2=2,0$	2,035	0,081	0,141	0,018	2,064	0,078	0,136	0,032
$\beta_1=2,7$	2,736	0,150	0,143	0,013	2,774	0,143	0,137	0,027
$\beta_2=2,2$	2,240	0,102	0,144	0,018	2,272	0,102	0,141	0,033
<i>n=300</i>								
$a_1=1,5$	1,500	0,015	0,082	0,000	1,508	0,013	0,077	0,005
$a_2=2,0$	2,002	0,026	0,081	0,001	2,013	0,025	0,078	0,006
$\beta_1=2,7$	2,709	0,050	0,083	0,003	2,724	0,049	0,081	0,009
$\beta_2=2,2$	2,216	0,032	0,081	0,007	2,227	0,031	0,079	0,012

Tabela 11 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros  $b_1$ ,  $b_2$ ,  $b$  e  $c$  no cenário 3, via método 1, método 2.1 e método 2.2, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300

Método	Método 1				Método 2.1				Método 2.2			
Parâmetro	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.
<i>n=10</i>												
$b_1=0,7$	1,077	2,628	2,252	0,539	1,234	6,171	3,466	0,763	1,461	6,102	3,357	1,087
$b_2=0,2$	0,654	2,721	7,930	2,270	0,923	4,687	10,203	3,615	0,550	1,815	6,504	1,750
$b=2,0$	2,093	1,411	0,592	0,046	2,185	2,166	0,730	0,093	2,157	2,216	0,740	0,079
$c=2,2$	2,855	4,992	0,971	0,298	3,039	10,517	1,424	0,381	2,917	5,548	1,020	0,326
<i>n=30</i>												
$b_1=0,7$	0,758	0,220	0,666	0,083	0,773	0,197	0,625	0,104	0,817	0,565	1,060	0,167
$b_2=0,2$	0,284	0,128	1,738	0,420	0,402	0,302	2,556	1,010	0,260	0,100	1,551	0,300
$b=2,0$	2,047	0,266	0,257	0,024	2,066	0,364	0,300	0,033	2,065	0,348	0,293	0,032
$c=2,2$	2,326	0,527	0,325	0,057	2,340	0,626	0,354	0,064	2,354	0,701	0,374	0,070
<i>n=100</i>												
$b_1=0,7$	0,719	0,042	0,293	0,027	0,718	0,038	0,279	0,026	0,710	0,158	0,567	0,014
$b_2=0,2$	0,216	0,032	0,896	0,080	0,274	0,073	1,302	0,370	0,216	0,017	0,638	0,080
$b=2,0$	2,003	0,064	0,126	0,002	2,002	0,090	0,150	0,001	2,014	0,081	0,142	0,007
$c=2,2$	2,236	0,111	0,150	0,016	2,235	0,142	0,171	0,016	2,249	0,137	0,167	0,022
<i>n=300</i>												
$b_1=0,7$	0,706	0,015	0,173	0,009	0,708	0,012	0,157	0,011	0,710	0,050	0,319	0,014
$b_2=0,2$	0,205	0,011	0,532	0,025	0,218	0,027	0,817	0,090	0,205	0,005	0,358	0,025
$b=2,0$	2,001	0,021	0,072	0,000	2,004	0,030	0,087	0,002	2,002	0,025	0,079	0,001
$c=2,2$	2,212	0,035	0,085	0,005	2,214	0,043	0,094	0,006	2,214	0,044	0,095	0,006

Para os valores atribuídos aos parâmetros nestas simulações, procurando reproduzir um caso que o avaliador superestima as proporções de severidade da doença, com maior discordância na proporção 0,5, o método 2.1 teve o pior desempenho entre os três, basta verificar os valores do coeficiente de variação e viés relativo. Os métodos 1 e 2.2 não diferem muito entre si mas, uma vez que a dispersão dos dados possui uma assimetria, o método 2.2 é mais adequado por considerar a restrição  $c = a_2 + b_2$ .

**Cenário 4:** Neste último caso considera-se, de forma semelhante ao anterior, que existe uma tendência mas, neste caso, considera-se a subestimação das proporções de severidade da doença pelos avaliadores, isto é, o valor real avaliado pelo computador é

superior ao valor estimado pelo avaliador mas, da mesma forma, espera-se uma maior discrepância conforme a proporção da severidade se aproxima de 0,5. Utilizou-se os seguintes valores dos parâmetros  $a_1=2,0$ ,  $a_2=1,5$ ,  $b_1=0,2$ ,  $b_2=0,7$ ,  $b=2,0$  e  $c=2,2$  para gerar mil bancos de dados com estes parâmetros e, para cada banco gerado, calculou-se todas as estimativas utilizando os diferentes métodos propostos. Além disso, foi variado o tamanho da amostra em 10, 30, 100 e 300 pares de observações, com a finalidade de avaliar os estimadores quanto ao viés e precisão. A Tabela 12 apresenta medidas de precisão e viés dos estimadores, utilizando somente a primeira parte do processo estimativo, ou seja, por meio das distribuições marginais, das mil simulações feitas para amostras de tamanho 10, 30, 100 e 300.

Conforme todos os outros casos, ao aumentar o tamanho da amostra, o viés relativo diminui, ou seja, as estimativas se aproximam do verdadeiro valor dos parâmetros, indicando a convergência de ambos os métodos. Entretanto o viés relativo do estimador de máxima verossimilhança foi sempre maior, máximo ao dobro, do viés relativo do estimador por meio do método dos momentos. Além disso o estimador por meio do método dos momentos foi mais preciso (menor coeficiente de variação) do que o método da máxima verossimilhança. Portanto, o estimador via método dos momentos foi melhor novamente. Sendo assim, partindo do valor destas estimativas, continua-se as simulações para os demais parâmetros. Os resultados da segunda parte das simulações podem ser vistos na Tabela 13 para, mais uma vez comparar os métodos.

Para os valores atribuídos aos parâmetros nestas simulações, procurando reproduzir um caso que o avaliador subestima as proporções de severidade da doença, com maior discordância na proporção 0,5, os três métodos apresentam um desempenho parecido, no entanto, o método 2.2 é pior que os demais, basta verificar os valores do coeficiente de variação e viés relativo. Dentre os métodos 1 e 2.1, uma vez que os dados possuem assimetria, optou-se pelo método 2.1 por considerar a restrição  $c = a_1 + b_1$ .

Tabela 12 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$  no cenário 4, via método dos momentos e da máxima verossimilhança, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300

Método	Momentos				Verossimilhança			
Parâmetro	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.
<i>n=10</i>								
$a_1=2,0$	2,345	2,283	0,735	0,173	2,743	3,231	0,818	0,371
$a_2=1,5$	1,791	1,580	0,815	0,194	2,070	2,148	0,900	0,380
$\beta_1=2,2$	2,667	3,605	0,836	0,212	3,119	5,042	0,931	0,418
$\beta_2=2,7$	3,312	4,644	0,765	0,227	3,831	6,234	0,824	0,419
<i>n=30</i>								
$a_1=2,0$	2,087	0,350	0,293	0,044	2,200	0,393	0,297	0,100
$a_2=1,5$	1,571	0,221	0,310	0,047	1,648	0,226	0,301	0,099
$\beta_1=2,2$	2,284	0,416	0,291	0,038	2,408	0,469	0,297	0,095
$\beta_2=2,7$	2,858	0,839	0,334	0,059	3,005	0,899	0,332	0,113
<i>n=100</i>								
$a_1=2,0$	2,029	0,088	0,148	0,014	2,061	0,086	0,143	0,030
$a_2=1,5$	1,516	0,049	0,147	0,011	1,537	0,046	0,140	0,025
$\beta_1=2,2$	2,226	0,110	0,150	0,012	2,263	0,110	0,148	0,029
$\beta_2=2,7$	2,726	0,162	0,149	0,010	2,765	0,160	0,146	0,024
<i>n=300</i>								
$a_1=2,0$	2,006	0,026	0,081	0,003	2,017	0,025	0,078	0,008
$a_2=1,5$	1,508	0,014	0,080	0,005	1,514	0,012	0,073	0,009
$\beta_1=2,2$	2,212	0,034	0,083	0,005	2,224	0,032	0,080	0,011
$\beta_2=2,7$	2,716	0,052	0,084	0,006	2,727	0,046	0,079	0,010

Tabela 13 - Médias, erros quadráticos médios (E.Q.M.), coeficientes de variação (C.V.) e viéses relativos (V.R.) das estimativas dos parâmetros  $b_1$ ,  $b_2$ ,  $b$  e  $c$  no cenário 4, via método 1, método 2.1 e método 2.2, em simulações de 1000 amostras de tamanhos 10, 30, 100 e 300

Método	Método 1				Método 2.1				Método 2.2				
	Parâmetro	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.	Média	E.Q.M.	C.V.	V.R.
<i>n=10</i>													
$b_1=0,2$	0,690	5,251	11,192	2,450		0,626	5,083	11,070	2,130	0,984	7,576	13,193	3,920
$b_2=0,7$	1,148	6,109	3,472	0,640		1,535	8,199	3,913	1,193	1,245	7,404	3,808	0,779
$b=2,0$	2,079	1,084	0,519	0,040		2,141	1,908	0,687	0,070	2,184	1,772	0,659	0,092
$c=2,2$	2,877	7,976	1,246	0,308		2,991	10,144	1,402	0,360	3,006	10,301	1,412	0,366
<i>n=30</i>													
$b_1=0,2$	0,310	0,154	1,880	0,550		0,269	0,099	1,534	0,345	0,375	0,259	2,389	0,875
$b_2=0,7$	0,777	0,222	0,664	0,110		0,881	0,653	1,125	0,259	0,814	0,240	0,680	0,163
$b=2,0$	2,038	0,269	0,259	0,019		2,039	0,315	0,280	0,020	2,071	0,393	0,311	0,036
$c=2,2$	2,371	0,570	0,334	0,078		2,374	0,662	0,361	0,079	2,408	0,751	0,382	0,095
<i>n=100</i>													
$b_1=0,2$	0,228	0,031	0,864	0,140		0,213	0,014	0,596	0,065	0,254	0,067	1,270	0,270
$b_2=0,7$	0,708	0,038	0,277	0,011		0,739	0,156	0,561	0,056	0,721	0,034	0,263	0,030
$b=2,0$	1,996	0,061	0,124	-0,002		1,994	0,079	0,140	-0,003	2,009	0,088	0,148	0,004
$c=2,2$	2,222	0,096	0,141	0,010		2,220	0,122	0,158	0,009	2,236	0,127	0,161	0,016
<i>n=300</i>													
$b_1=0,2$	0,197	0,011	0,521	-0,015		0,201	0,005	0,343	0,005	0,218	0,026	0,802	0,090
$b_2=0,7$	0,705	0,014	0,167	0,007		0,699	0,048	0,312	-0,001	0,703	0,012	0,155	0,004
$b=2,0$	2,002	0,021	0,073	0,001		2,005	0,026	0,081	0,002	2,001	0,028	0,084	0,000
$c=2,2$	2,207	0,034	0,084	0,003		2,211	0,043	0,094	0,005	2,205	0,042	0,093	0,002

Em resumo, uma vez que não foi possível satisfazer todas as funções de restrições que envolvem os parâmetros a serem estimados, foram construídos diferentes métodos, em que, cada um deles, contempla um conjunto destas funções. Foi visto que cada um dos métodos foi mais apropriado para diferentes formas da distribuição. De forma geral, quando a maior parte dos pontos se encontra acima da reta  $y = x$ , ou seja, quando o avaliador superestima a severidade da doença, o melhor é o método 2.2. Quando, por outro lado, ele subestima a severidade, é indicado o método 2.1. Finalmente, quando ele não está viesado, ou seja, os valores estimados visualmente se aproximam em média dos observados, sugere-se a utilização do método 1.

A seguir, serão utilizados os bancos de dados apresentados nas Tabelas 1, 2 e

3 para ilustrar a metodologia apresentada.

## 5.2 Análise dos dados

Como pode-se ver, no gráfico de dispersão dos dados observados (ver Figura 7), houve uma tendência a superestimar a severidade da ferrugem do amendoim, se comparados aos valores reais apresentados pelo computador. Portanto, concluiu-se nas simulações utilizou-se o método dos momentos para estimar os parâmetros  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$  das distribuições marginais de  $X$  e  $Y$ , respectivamente. Com base nas estimativas obtidas encontram-se, por sua vez, as estimativas dos parâmetros restantes  $b_1$ ,  $b_2$ ,  $b$  e  $c$  utilizando o método 2.2. Os resultados podem ser vistos na Tabela 14.

Tabela 14 - Estimativas dos parâmetros, via método dos momentos para  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$ , e via método 2.2 para  $b_1$ ,  $b_2$ ,  $b$  e  $c$ , dos dados de severidade da ferrugem no amendoim

Parâmetro	Estimativa	Parâmetro	Estimativa
$a_1$	3,845	$b_1$	6,132
$a_2$	2,031	$b_2$	0,500
$\beta_1$	9,774	$b$	3,641
$\beta_2$	3,559	$c$	1,949

Com base nas estimativas dos parâmetros  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ ,  $b$  e  $c$ , foi gerado um banco de dados com 200 observações da distribuição BBNK cujo gráfico de dispersão é apresentado também na Figura 7, porém com pontos em vermelho.

Como se pode observar, aparentemente os dados tem distribuição BBNK, uma vez que a nuvem de pontos gerada é muito próxima da observada.

No gráfico de dispersão dos dados observados (ver Figura 8), diferente do anterior, o avaliador conseguiu estimativas bem próximas da severidade da mancha castanha, se comparados aos valores avaliados pelo computador. Portanto, conforme vimos nas simulações iremos utilizar o método dos momentos para estimar os parâmetros  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$ . Após feito isso, com base no resultado dessas estimativas, encontraremos as estimativas dos parâmetros restantes  $b_1$ ,  $b_2$ ,  $b$  e  $c$  utilizando o método 2.2. O resultado pode ser visto na Tabela 15.

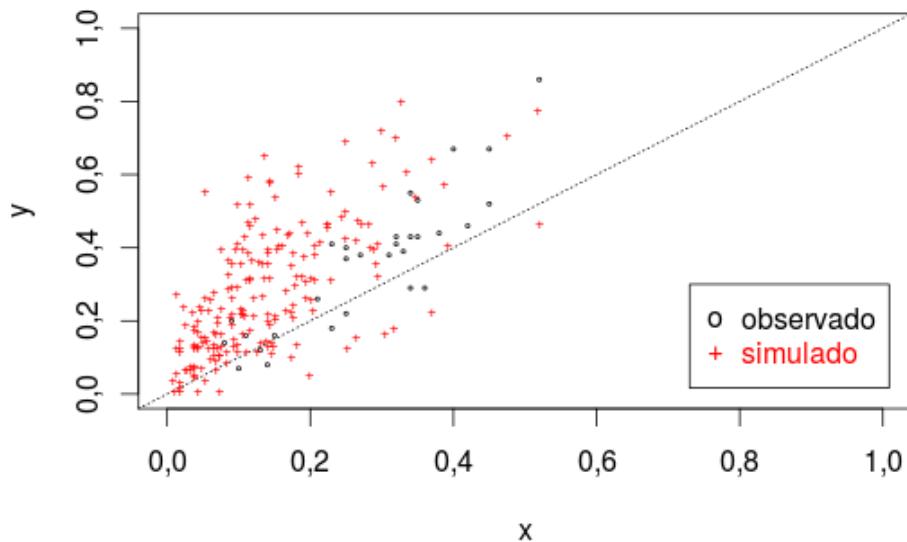


Figura 7 - Gráficos de dispersão: (i) dos pares de observações  $(x_i, y_i)$ , ( $i = 1, \dots, 30$ ) em que  $X$  é a severidade real de ferrugem e  $Y$  é a severidade estimada visualmente por um avaliador; (ii) de uma amostra de 200 observações de  $(X, Y)$  com distribuição BBNK com parâmetros iguais às estimativas obtidas pelo método 2.2

Tabela 15 - Estimativas dos parâmetros, via método dos momentos para  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$ , e via método 1 para  $b_1$ ,  $b_2$ ,  $b$  e  $c$ , dos dados de severidade da mancha castanha no amendoim

Valor	Estimativa	Valor	Estimativa
$a_1$	1,524	$b_1$	0,500
$a_2$	1,805	$b_2$	0,500
$\beta_1$	3,335	$b$	3,684
$\beta_2$	4,238	$c$	1,734

Baseando-se nas estimativas dos parâmetros  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ ,  $b$  e  $c$ , foi gerado um banco de dados com 200 observações da distribuição BBNK cujo gráfico de dispersão é apresentado também na Figura 8, porém com pontos em vermelho.

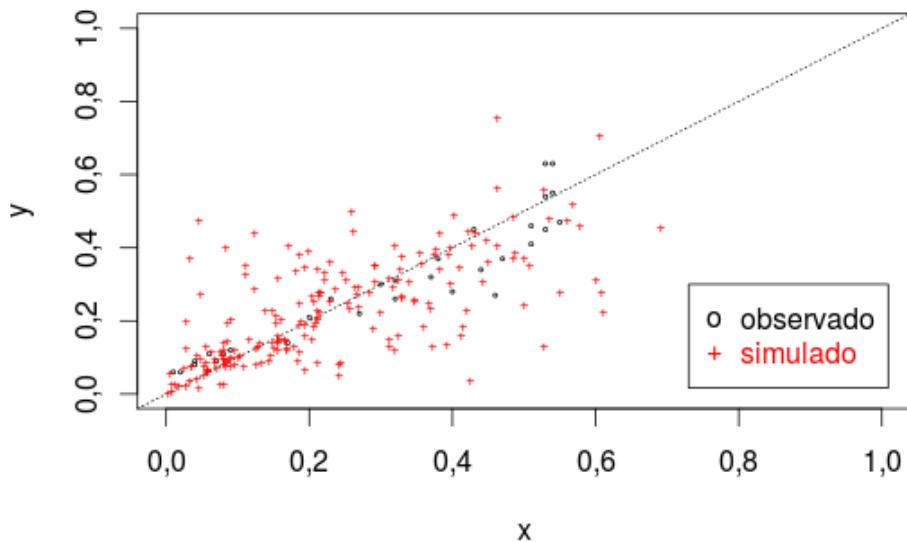


Figura 8 - Gráficos de dispersão: (i) dos pares de observações  $(x_i, y_i)$ , ( $i = 1, \dots, 30$ ) em que  $X$  é a severidade real de mancha castanha e  $Y$  é a severidade estimada visualmente por um avaliador; (ii) de uma amostra de 200 observações de  $(X, Y)$  com distribuição BBNK com parâmetros iguais às estimativas obtidas pelo método 1

Observa-se que, aparentemente, os dados tem distribuição BBNK, uma vez que a nuvem de pontos simulados é próxima da observada.

No último banco de dados, pode-se ver no gráfico de dispersão dos dados observados (ver Figura 9) que houve uma tendência a subestimar a severidade da mancha preta, se comparados aos valores avaliados pelo computador. Portanto, concluiu-se nas simulações utilizou-se o método dos momentos para estimar os parâmetros  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$  das distribuições marginais de  $X$  e  $Y$ , respectivamente. Com base nas estimativas obtidas encontram-se, por sua vez, as estimativas dos parâmetros restantes  $b_1$ ,  $b_2$ ,  $b$  e  $c$  utilizando o método 2.1. Os resultados podem ser vistos na Tabela 16.

Com base nas estimativas dos parâmetros  $a_1$ ,  $a_2$ ,  $b_1$ ,  $b_2$ ,  $b$  e  $c$ , foi gerado um banco de dados com 200 observações da distribuição BBNK cujo gráfico de dispersão é apresentado também na Figura 9, porém com pontos em vermelho.

Tabela 16 - Estimativas dos parâmetros, via método dos momentos para  $a_1$ ,  $a_2$ ,  $\beta_1$  e  $\beta_2$ , e via método 2.1 para  $b_1$ ,  $b_2$ ,  $b$  e  $c$ , dos dados de severidade da mancha castanha no amendoim

Valor	Estimativa	Valor	Estimativa
$a_1$	2.344	$b_1$	0,500
$a_2$	1.532	$b_2$	0,500
$\beta_1$	5.881	$b$	5.974
$\beta_2$	5.119	$c$	2.250

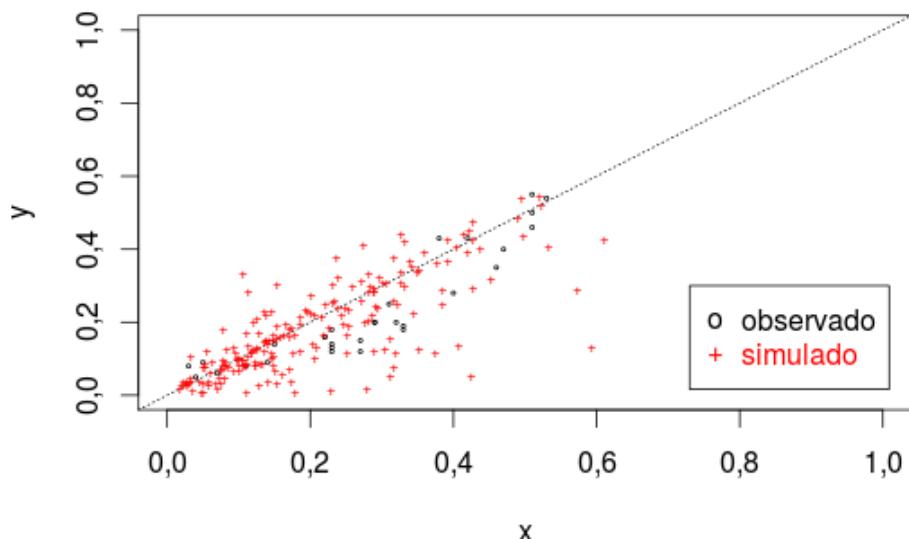


Figura 9 - Gráficos de dispersão: (i) dos pares de observações  $(x_i, y_i)$ , ( $i = 1, \dots, 30$ ) em que  $X$  é a severidade real de mancha preta e  $Y$  é a severidade estimada visualmente por um avaliador; (ii) de uma amostra de 200 observações de  $(X, Y)$  com distribuição BBNK com parâmetros iguais às estimativas obtidas pelo método 2.1

Nota-se que, aparentemente, os dados tem distribuição BBNK, uma vez que a nuvem de pontos simulados é próxima da observada.

Como trabalhos futuros, pretende-se aperfeiçoar o método para incluir todas as restrições paramétricas da distribuição. Além disso, pretende-se propor um teste para

verificar a adequação da distribuição ajustada aos dados. Parece viável também, reparametrizar a distribuição BBNK e verificar suas propriedades de estimação. É preciso estudar a utilização de métodos bayesianos, como o ABC, para a estimação dos parâmetros e obtenção dos respectivos intervalos de credibilidade. Por fim, parece apropriado aprofundar o conhecimento em distribuições beta bivariadas alternativas para a modelagem dos dados de severidade de doenças do amendoinzeiro.

## 6 CONCLUSÃO

Neste trabalho foram analisados três bancos de dados, que tratam da severidade de doenças foliares no amendoim, com o objetivo de estimar os parâmetros da distribuição a qual acreditamos modelar bem esses dados, a beta bivariada. Segundo algumas simulações feitas, chegamos a alguns critérios para encontrar as estimativas destes parâmetros. Por fim, esses critérios pareceram ser bastante satisfatórios uma vez que conseguimos estimativas aparentemente bem consistentes, levando em consideração as limitações que o método proposto apresenta. A metodologia aqui apresentada pode ser utilizada também em outros casos, não somente para analisar severidade em plantas.

Esse estudo mostra que é importante a continuidade da construção de novas distribuições, em particular as beta bivariadas. Além disso, novos métodos de estimação também são importantes para tentar criar estimadores cada vez mais precisos de modo a encontrar aquele que melhor responda aos objetivos propostos. Um tema bastante interessante para pesquisas futuras seria encontrar um estimador através da função de densidade conjunta dessa distribuição, mas o desafio é que provavelmente será necessária a utilização de métodos iterativos, uma vez que ela não possui uma forma fechada.

Por fim, este trabalho proporcionou uma oportunidade singular de trabalhar com inferência estatística, em especial com estimação de parâmetros em distribuições bivariadas, além de poder me aprofundar no estudo das distribuições beta.



## REFERÊNCIAS

- ARNOLD, B.C.; NG, H.K.T. Flexible bivariate beta distributions. **Journal of Multivariate Analysis**, Amsterdam, v. 102, n. 8, p. 1194–1202, 2011.
- BALAKRISHNAN, N.; LAI, C. **Continuous bivariate distributions**. Springer Science & Business Media, New York, 2009. 683 p.
- BERGAMIN FILHO, A.; KIMATI, H.; AMORIM, L. **Manual de fitopatologia**. Agronômica Ceres, Piracicaba, v. 1, 1995. 919 p.
- CAMOCHENA, R.C.; SANTOS, I. dos; MAZARO, S.M. Escala diagramática para avaliação da severidade da Mancha Ocular em milho causada por *Kabatiella zeae*. **Ciência Rural**, Santa Maria, v. 38, n. 8, p. 2124–2131, 2008.
- BROOKS COUNTY Peanut Disease Control. 2014. Acesso em: 16/abr/2015. Disponível em:  
[<http://blog.extension.uga.edu/brooksag/2014/07/peanut-disease-control>](http://blog.extension.uga.edu/brooksag/2014/07/peanut-disease-control)
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA Avaliação para resistência a Mancha Preta e Ferrugem em híbridos entre o amendoim e espécies silvestres de *Arachis*. **Boletim de Pesquisa e Desenvolvimento**, v. 96, 2005. Acesso em: 12/mar/2015. Disponível em:  
[<http://ainfo.cnptia.embrapa.br/digital/bitstream/CENARGEN/26835/1/bp096.pdf>](http://ainfo.cnptia.embrapa.br/digital/bitstream/CENARGEN/26835/1/bp096.pdf)
- FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, Abingdon, v. 31, n. 7, p. 799–815, 2004.
- GUPTA, A.K.; NADARAJAH, S. **Handbook of beta distribution and its applications**. CRC Press, New York, 2004. 579 p.
- GUPTA, A.K.; WONG, C.F. On three and five parameter bivariate beta distributions. **Metrika**, Berlin, v. 32, n. 1, p. 85–91, 1985.
- INSTITUTO AGRONÔMICO. **Monitoramento das doenças foliares do amendoim e avisos climáticos para indicar as pulverizações com fungicida**. 1999. Acesso em: 21/mai/2015. Disponível em:  
[<http://www.iac.sp.gov.br/publicacoes/agronomico/monitoramento\\_doenças\\_foliares.php>](http://www.iac.sp.gov.br/publicacoes/agronomico/monitoramento_doenças_foliares.php)
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Produção Agrícola Municipal**. 2013. Acesso em: 23/jun/2015. Disponível em:  
[<http://www.sidra.ibge.gov.br>](http://www.sidra.ibge.gov.br)
- JOHNSON, N.L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous univariate distributions**. Wiley & Sons, Michigan, v. 2, 1995. 752 p.
- JOHNSON, N.L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous Multivariate Distributions, Models and Applications**. John Wiley & Sons, New York, v. 1, 2002. 619 p.
- JONES, M.C. Multivariate t and beta distributions associated with the multivariate f distribution. **Metrika**, Berlin, v. 54, n. 3, p. 215–231, 2002.

- KIMATI, H. et al. **Manual de fitopatologia: doenças das plantas cultivadas.** Agronômica Ceres, São Paulo, v. 2, 1997. 663 p.
- MORAES, S.A. de **Quantificação de doenças de plantas.** 2007. Acesso em: 20/jul/2015. Disponível em: <[http://www.infobibos.com/Artigos/2007\\_1/doencas](http://www.infobibos.com/Artigos/2007_1/doencas)>
- MORGENSTERN, D. Einfache beispiele zweidimensionaler verteilungen. **Mitteilungsblatt für Mathematische Statistik**, Heidelberg, v. 8, n. 1, p. 234–235, 1956.
- NADARAJAH, S. The bivariate F 3-beta distribution. **Communications of the Korean Mathematical Society**, Seoul, v. 21, n. 2, p. 363, 2006.
- NADARAJAH, S. The bivariate Gauss hypergeometric beta distribution. **Integral Transforms and Special Functions**, Abingdon, v. 19, n. 12, p. 859–868, 2008.
- NADARAJAH, S.; KOTZ, S. Some bivariate beta distributions. **Statistics**, Abingdon, v. 39, n. 5, p. 457–466, 2005.
- NELSEN, R.B. **An introduction to copulas.** Springer Science & Business Media, New York, 1999. 218 p.
- OAKES, D.; RITZ, J. Regression in a bivariate copula model. **Biometrika**, Oxford, v. 87, n. 2, p. 345–352, 2000.
- OLKIN, I.; LIU, R. A bivariate beta distribution. **Statistics & Probability Letters**, Amsterdam, v. 62, n. 4, p. 407–412, 2003.
- OLKIN, I.; TRIKALINOS, T.A. Constructions for a bivariate beta distribution. **Statistics & Probability Letters**, Amsterdam, v. 96, p. 54–60, 2015.
- TOMERLIN, J.R.; HOWELL T.A. Distrain: A computer program for training people to estimate disease severity on cereal leaves. **Plant Disease**, Beltsville, v. 72, n. 5, p. 455–459, 1988.
- TRIVEDI, P.K.; ZIMMER, D.M. **Copula modeling: an introduction for practitioners.** Now Publishers, 2007. 115 p.
- VALE, F.X.R.; FERNANDES FILHO, E.I.; LIBERATO, J.R. Quant: A software for plant disease severity assessment. **8th International Congress of Plant Pathology**, Christchurch, v. 8, p. 105, 2003.

## APÊNDICE



## Linhas de comandos no R

```

#-----#
#          GRÁFICOS
#-----#



# Figura 4a - simetricos a=b a>=1 b>=1
a = 1
b = 1
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),ylab="f(x)",ylim=c(0,3))
a = 2
b = 2
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=2)
a = 4
b = 4
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=3)
a = 6
b = 6
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=4)
legend(0,3,legend=c("a = b = 1","a = b = 2","a = b = 4","a = b = 6"),
lty=c(1,2,3,4),lwd=1:4,y.intersp=0.5)

# Figura 4b - simetricos U a=b 0<a,b<1
a = 0.8
b = 0.8
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),ylab="f(x)",ylim=c(0,3))
a = 0.5
b = 0.5
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=2)
a = 0.3
b = 0.3
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=3)
a = 0.1
b = 0.1
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=4)
legend(0.3,3,legend=c("a = b = 0,8","a = b = 0,5","a = b = 0,3","a = b = 0,1"),
lty=c(1,2,3,4),lwd=1:4,y.intersp=0.5)

# Figura 5a - assimetricos a<b
a = 2
b = 5
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),ylab="f(x)",ylim=c(0,3))
a = 3
b = 4
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=2)
a = 1.5
b = 2
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=3)
a = 0.2
b = 0.8
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=4)
legend(0.6,3,legend=c("a = 2, b = 5","a = 3, b = 4","a = 1,5, b = 2","a = 0,2, b = 0,8"),
lty=c(1,2,3,4),lwd=1:4,y.intersp=0.5)

# Figura 5b - assimetricos a>b
a = 5
b = 3

```

```

curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),ylab="f(x)",ylim=c(0,3))
a = 2
b = 1
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=2)
a = 2
b = 0.5
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=3)
a = 0.6
b = 0.3
curve(x^(a-1)*(1-x)^(b-1)/beta(a,b),add=TRUE,lty=4)
legend(0,3,legend=c("a = 5, b = 3","a = 2, b = 1","a = 2, b = 0,5","a = 0,6, b = 0,3"),
lty=c(1,2,3,4),lwd=1:4,y.intersp=0.5)

#-----
#                         ESTIMAÇÃO
#-----

# GERANDO A BETA BIVARIADA
a1 = 2
a2 = 2
b = 2
c = 2.5
b1 = c - a1
b2 = c - a2
beta1 = b1 + b
beta2 = b2 + b
n = 100
U <- rbeta(n, a1, b1)
V <- rbeta(n, a2, b2)
W <- rbeta(n, c, b)
X = U*W
Y = V*W
# Grafico distribuicao
options(OutDec=",")
plot(X, Y, xlim=c(0, 1), ylim=c(0, 1), xlab="x", ylab="y", cex=0.2)
abline(a=0, b=1, lty=3)
plot(density(X),xlim=c(0,1))
curve(x^(a1-1)*(1-x)^(b1+b-1)/beta(a1,b1+b),add=TRUE,col="blue")
plot(density(Y),xlim=c(0,1))
curve(x^(a2-1)*(1-x)^(b2+b-1)/beta(a2,b2+b),add=TRUE,col="blue")

# Falsos 0s e 1s
arred = 6
arred_X <- round(X,arred)
arred_Y <- round(Y,arred)
zero_X <- (1:n)[arred_X==0]
zero_Y <- (1:n)[arred_Y==0]
um_X <- (1:n)[arred_X==1]
um_Y <- (1:n)[arred_Y==1]
if(length(zero_X)!=0 & length(zero_Y)!=0 & length(um_X)!=0 & length(um_Y)!=0){
sem_X=arred_X[-c(zero_X,um_X)]
sem_Y=arred_Y[-c(zero_Y,um_Y)]
menor_X=min(sem_X)
menor_Y=min(sem_Y)
maior_X=max(sem_X)
maior_Y=max(sem_Y)
subzero_X=(0+menor_X)/2
subzero_Y=(0+menor_Y)/2
}

```

```

subum_X=(maior_X+1)/2
subum_Y=(maior_Y+1)/2
X[zero_X]=subzero_X
X[um_X]=subum_X
Y[zero_Y]=subzero_Y
Y[um_Y]=subum_Y
}
if(sum(log(1-X))==-Inf){
stop("diminuir arred")
}

# PARTE 1: ESTIMANDO PARAMETROS DAS MARGINAIS

# Metodo dos Momentos
Xbarra = sum(X)/length(X)
Ybarra = sum(Y)/length(Y)
S2X = sum((X-Xbarra)^2)/(length(X)-1)
S2Y = sum((Y-Ybarra)^2)/(length(Y)-1)
rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(Y)-1)*sqrt(S2X)*sqrt(S2Y))
a1_mm = max(0.01,Xbarra*((1-Xbarra)*Xbarra/S2X)-1))
beta1_mm = max(0.01,(1-Xbarra)*((1-Xbarra)*Xbarra/S2X)-1))
a2_mm = max(0.01,Ybarra*((1-Ybarra)*Ybarra/S2Y)-1))
beta2_mm = max(0.01,(1-Ybarra)*((1-Ybarra)*Ybarra/S2Y)-1))

# Maxima Verossimilhanca: Newton-Raphson
n_it = 5
theta_X <- matrix(rep(NA, 2*(n_it+1)), ncol=2)
theta_X[1,1] = a1_mm
theta_X[1,2] = beta1_mm
for(j in 1:n_it){
g1 <- digamma(theta_X[j,1]) - digamma(theta_X[j,1]+theta_X[j,2]) - sum(log(X))/n
g2 <- digamma(theta_X[j,2]) - digamma(theta_X[j,1]+theta_X[j,2]) - sum(log(1-X))/n
dg1_a <- trigamma(theta_X[j,1]) - trigamma(theta_X[j,1]+theta_X[j,2])
dg1_b <- -trigamma(theta_X[j,1]+theta_X[j,2])
dg2_a <- -trigamma(theta_X[j,1]+theta_X[j,2])
dg2_b <- trigamma(theta_X[j,2]) - trigamma(theta_X[j,1]+theta_X[j,2])
g_X <- c(g1,g2)
G_X <- matrix(c(dg1_a,dg1_b,dg2_a,dg2_b), ncol=2, byrow=T)
theta_X[j+1,] <- c(theta_X[j,1],theta_X[j,2]) - solve(G_X) %*% g_X
}
theta_Y <- matrix(rep(NA, 2*(n_it+1)), ncol=2)
theta_Y[1,1] = a2_mm
theta_Y[1,2] = beta2_mm
for(k in 1:n_it){
g1 <- digamma(theta_Y[k,1]) - digamma(theta_Y[k,1]+theta_Y[k,2]) - sum(log(Y))/n
g2 <- digamma(theta_Y[k,2]) - digamma(theta_Y[k,1]+theta_Y[k,2]) - sum(log(1-Y))/n
dg1_a <- trigamma(theta_Y[k,1]) - trigamma(theta_Y[k,1]+theta_Y[k,2])
dg1_b <- -trigamma(theta_Y[k,1]+theta_Y[k,2])
dg2_a <- -trigamma(theta_Y[k,1]+theta_Y[k,2])
dg2_b <- trigamma(theta_Y[k,2]) - trigamma(theta_Y[k,1]+theta_Y[k,2])
g_Y <- c(g1,g2)
G_Y <- matrix(c(dg1_a,dg1_b,dg2_a,dg2_b), ncol=2, byrow=T)
theta_Y[k+1,] <- c(theta_Y[k,1],theta_Y[k,2]) - solve(G_Y) %*% g_Y
}
a1_mv = max(0.01,theta_X[n_it,1])
beta1_mv = max(0.01,theta_X[n_it,2])
a2_mv = max(0.01,theta_Y[n_it,1])
beta2_mv = max(0.01,theta_Y[n_it,2])

```

```

# Verossimilhanca
M = 50
alpha_X = seq(a1/9,3*a1,length=M)
alpha_Y = seq(a2/9,3*a2,length=M)
beta_X = seq(beta1/9,3*beta1,length=M)
beta_Y = seq(beta2/9,3*beta2,length=M)
soma1_X = sum(log(X))
soma1_Y = sum(log(Y))
soma2_X = sum(log(1-X))
soma2_Y = sum(log(1-Y))
La_X <- (alpha_X-1)*soma1_X+(beta1-1)*soma2_X-n*log(beta(alpha_X,beta1))
La_Y <- (alpha_Y-1)*soma1_Y+(beta2-1)*soma2_Y-n*log(beta(alpha_Y,beta2))
Lb_X <- (a1-1)*soma1_X+(beta_X-1)*soma2_X-n*log(beta(a1,beta_X))
Lb_Y <- (a2-1)*soma1_Y+(beta_Y-1)*soma2_Y-n*log(beta(a2,beta_Y))
beta.lik.x <- function(X,alpha,beta) {
soma1 = sum(log(X))
soma2 = sum(log(1-X))
(alpha-1)*soma1+(beta-1)*soma2-n*log(beta(alpha,beta))
}
beta.lik.y <- function(Y,alpha,beta) {
soma1 = sum(log(Y))
soma2 = sum(log(1-Y))
(alpha-1)*soma1+(beta-1)*soma2-n*log(beta(alpha,beta))
}
contour.beta.lik.x <- function(X, M=50) {
alpha=seq(a1/9,3*a1,length=M)
beta =seq(beta1/9,3*beta1,length=M)
lik= matrix(0,M,M)
for (i in 1:M){
for (j in 1:M){
lik[i,j] = beta.lik.x(X,alpha[i],beta[j])
}
}
contour(alpha,beta,lik,xlab=expression(alpha),ylab=expression(beta),nlevels=20)
}
contour.beta.lik.y <- function(Y, M=50) {
alpha=seq(a2/9,3*a2,length=M)
beta =seq(beta2/9,3*beta2,length=M)
lik= matrix(0,M,M)
for (i in 1:M){
for (j in 1:M){
lik[i,j] = beta.lik.y(Y,alpha[i],beta[j])
}
}
contour(alpha,beta,lik,xlab=expression(alpha),ylab=expression(beta),nlevels=20)
}
# Grafico verossimilhanca de X
contour.beta.lik.x(X)
plot(alpha_X, La_X, type='l')
plot(beta_X, Lb_X, type='l')
# Grafico verossimilhanca de Y
contour.beta.lik.y(Y)
plot(alpha_Y, La_Y, type='l')
plot(beta_Y, Lb_Y, type='l')

# Tabela Comparativa
data.frame("Valor"=c(a1,a2,beta1,beta2),
"MM"=c(a1_mm,a2_mm,beta1_mm,beta2_mm),

```

```

"MV"=c(a1_mv,a2_mv,beta1_mv,beta2_mv),
row.names=c("a1","a2","beta1","beta2"))

# PARTE 2: ESTIMANDO PARAMETROS RESTANTES

a1_est = a1_mv # <- escolher um metodo (a1_mm ou a1_mv)
a2_est = a2_mv # <- escolher um metodo (a2_mm ou a2_mv)
beta1_est = beta1_mv # <- escolher um metodo (beta1_mm ou beta1_mv)
beta2_est = beta2_mv # <- escolher um metodo (beta2_mm ou beta2_mv)

# Metodo 1
rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(X)-1)*sqrt(S2X)*sqrt(S2Y))
k1 = sqrt((a1_est*a2_est)/(Xbarra*Ybarra))
k2 = sqrt((a1_est*a2_est)/(beta1_est*beta2_est))
b_m1 = max(0.01,(k1*rXY)/(rXY+k2))
c_m1 = max(0.01,(k1*k2)/(rXY+k2))
b1_m1 = max(0.01,c_m1-a1_est)
b2_m1 = max(0.01,c_m1-a2_est)
# Testando Restricoes
data.frame("Estimativa"=c(c_m1,a1_est+b1_m1,a2_est+b2_m1),
row.names=c("c","a1+b1","a2+b2"))
data.frame("Estimativa"=c(beta1_est,b1_m1+b_m1,beta2_est,b2_m1+b_m1),
row.names=c("beta1","b1+b","beta2","b2+b"))

# Metodo 2.1
b1_m2.1 = max(0.01,(beta1_est*k2-a1_est*rXY)/(rXY+k2))
b2_m2.1 = max(0.01,(beta2_est*(rXY+k2)-beta1_est*rXY-a1_est*rXY)/(rXY+k2))
b_m2.1 = max(0.01,(beta1_est*rXY+a1_est*rXY)/(rXY+k2))
c_m2.1 = max(0.01,(beta1_est*k2+a1_est*k2)/(rXY+k2))
# Testando Restricoes
data.frame("Estimativa"=c(c_m2.1,a1_est+b1_m2.1,a2_est+b2_m2.1),
row.names=c("c","a1+b1","a2+b2"))
data.frame("Estimativa"=c(beta1_est,b1_m2.1+b_m2.1,beta2_est,b2_m2.1+b_m2.1),
row.names=c("beta1","b1+b","beta2","b2+b"))

# Metodo 2.2
b1_m2.2 = max(0.01,(beta1_est*(rXY+k2)-beta2_est*rXY-a2_est*rXY)/(rXY+k2))
b2_m2.2 = max(0.01,(beta2_est*k2-a2_est*rXY)/(rXY+k2))
b_m2.2 = max(0.01,(beta2_est*rXY+a2_est*rXY)/(rXY+k2))
c_m2.2 = max(0.01,(beta2_est*k2+a2_est*k2)/(rXY+k2))
# Testando Restricoes
data.frame("Estimativa"=c(c_m2.2,a1_est+b1_m2.2,a2_est+b2_m2.2),
row.names=c("c","a1+b1","a2+b2"))
data.frame("Estimativa"=c(beta1_est,b1_m2.2+b_m2.2,beta2_est,b2_m2.2+b_m2.2),
row.names=c("beta1","b1+b","beta2","b2+b"))

# Tabela Comparativa
data.frame("Valor"=c(b1,b2,b,c),
"Metodo1"=c(b1_m1,b2_m1,b_m1,c_m1),
"Metodo2.1"=c(b1_m2.1,b2_m2.1,b_m2.1,c_m2.1),
"Metodo2.2"=c(b1_m2.2,b2_m2.2,b_m2.2,c_m2.2),
row.names=c("b1","b2","b","c"))

#-----
#           SIMULAÇÃO
#-----

# Definindo Parametros

```

```

a1 = 2
a2 = 1.5
b = 2
c = 2.2
b1 = c - a1
b2 = c - a2
beta1 = b1 + b
beta2 = b2 + b
n = 300

# Simulacao
max = 1000
theta_mm <- matrix(rep(NA, 4*max), ncol=4)
theta_mv <- matrix(rep(NA, 4*max), ncol=4)
theta_m1 <- matrix(rep(NA, 4*max), ncol=4)
theta_m2.1 <- matrix(rep(NA, 4*max), ncol=4)
theta_m2.2 <- matrix(rep(NA, 4*max), ncol=4)
for(i in 1:max){
  U <- rbeta(n, a1, b1)
  V <- rbeta(n, a2, b2)
  W <- rbeta(n, c, b)
  X = U*W
  Y = V*W
  # Falsos 0s e 1s
  arred = 6
  arred_X <- round(X,arred)
  arred_Y <- round(Y,arred)
  zero_X <- (1:n)[arred_X==0]
  zero_Y <- (1:n)[arred_Y==0]
  um_X <- (1:n)[arred_X==1]
  um_Y <- (1:n)[arred_Y==1]
  if(length(zero_X)!=0 & length(zero_Y)!=0 & length(um_X)!=0 & length(um_Y)!=0){
    sem_X=arred_X[-c(zero_X,um_X)]
    sem_Y=arred_Y[-c(zero_Y,um_Y)]
    menor_X=min(sem_X)
    menor_Y=min(sem_Y)
    maior_X=max(sem_X)
    maior_Y=max(sem_Y)
    subzero_X=(0+menor_X)/2
    subzero_Y=(0+menor_Y)/2
    subum_X=(maior_X+1)/2
    subum_Y=(maior_Y+1)/2
    X[zero_X]=subzero_X
    X[um_X]=subum_X
    Y[zero_Y]=subzero_Y
    Y[um_Y]=subum_Y
  }
  if(sum(log(1-X))==-Inf){
    stop("diminuir arred")
  }
  # Metodo dos Momentos
  Xbarra = sum(X)/length(X)
  Ybarra = sum(Y)/length(Y)
  S2X = sum((X-Xbarra)^2)/(length(X)-1)
  S2Y = sum((Y-Ybarra)^2)/(length(Y)-1)
  rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(Y)-1)*sqrt(S2X)*sqrt(S2Y))
  a1_mm = max(0.01,Xbarra*((1-Xbarra)*Xbarra/S2X)-1))
  beta1_mm = max(0.01,(1-Xbarra)*(((1-Xbarra)*Xbarra/S2X)-1))
}

```

```

a2_mm = max(0.01,Ybarra*((1-Ybarra)*Ybarra/S2Y)-1))
beta2_mm = max(0.01,(1-Ybarra)*((1-Ybarra)*Ybarra/S2Y)-1))
theta_mm[i,1] <- a1_mm
theta_mm[i,2] <- beta1_mm
theta_mm[i,3] <- a2_mm
theta_mm[i,4] <- beta2_mm
# Maxima Verossimilhanca: Newton-Raphson
n_it = 5
theta_X <- matrix(rep(NA, 2*(n_it+1)), ncol=2)
theta_X[1,1] = a1_mm
theta_X[1,2] = beta1_mm
for(j in 1:n_it){
g1 <- digamma(theta_X[j,1]) - digamma(theta_X[j,1]+theta_X[j,2]) - sum(log(X))/n
g2 <- digamma(theta_X[j,2]) - digamma(theta_X[j,1]+theta_X[j,2]) - sum(log(1-X))/n
dg1_a <- trigamma(theta_X[j,1]) - trigamma(theta_X[j,1]+theta_X[j,2])
dg1_b <- -trigamma(theta_X[j,1]+theta_X[j,2])
dg2_a <- -trigamma(theta_X[j,1]+theta_X[j,2])
dg2_b <- trigamma(theta_X[j,2]) - trigamma(theta_X[j,1]+theta_X[j,2])
g_X <- c(g1,g2)
G_X <- matrix(c(dg1_a,dg1_b,dg2_a,dg2_b), ncol=2, byrow=T)
theta_X[j+1,] <- c(theta_X[j,1],theta_X[j,2]) - solve(G_X) %*% g_X
}
theta_Y <- matrix(rep(NA, 2*(n_it+1)), ncol=2)
theta_Y[1,1] = a2_mm
theta_Y[1,2] = beta2_mm
for(k in 1:n_it){
g1 <- digamma(theta_Y[k,1]) - digamma(theta_Y[k,1]+theta_Y[k,2]) - sum(log(Y))/n
g2 <- digamma(theta_Y[k,2]) - digamma(theta_Y[k,1]+theta_Y[k,2]) - sum(log(1-Y))/n
dg1_a <- trigamma(theta_Y[k,1]) - trigamma(theta_Y[k,1]+theta_Y[k,2])
dg1_b <- -trigamma(theta_Y[k,1]+theta_Y[k,2])
dg2_a <- -trigamma(theta_Y[k,1]+theta_Y[k,2])
dg2_b <- trigamma(theta_Y[k,2]) - trigamma(theta_Y[k,1]+theta_Y[k,2])
g_Y <- c(g1,g2)
G_Y <- matrix(c(dg1_a,dg1_b,dg2_a,dg2_b), ncol=2, byrow=T)
theta_Y[k+1,] <- c(theta_Y[k,1],theta_Y[k,2]) - solve(G_Y) %*% g_Y
}
a1_mv = max(0.01,theta_X[n_it,1])
beta1_mv = max(0.01,theta_X[n_it,2])
a2_mv = max(0.01,theta_Y[n_it,1])
beta2_mv = max(0.01,theta_Y[n_it,2])
theta_mv[i,1] <- a1_mv
theta_mv[i,2] <- beta1_mv
theta_mv[i,3] <- a2_mv
theta_mv[i,4] <- beta2_mv
a1_est = a1_mm # <- escolher um metodo (a1_mm ou a1_mv)
a2_est = a2_mm # <- escolher um metodo (a2_mm ou a2_mv)
beta1_est = beta1_mm # <- escolher um metodo (beta1_mm ou beta1_mv)
beta2_est = beta2_mm # <- escolher um metodo (beta2_mm ou beta2_mv)
# Metodo 1
rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(X)-1)*sqrt(S2X)*sqrt(S2Y))
k1 = sqrt((a1_est*a2_est)/(Xbarra*Ybarra))
k2 = sqrt((a1_est*a2_est)/(beta1_est*beta2_est))
b_m1 = max(0.01,(k1*rXY)/(rXY+k2))
c_m1 = max(0.01,(k1*k2)/(rXY+k2))
b1_m1 = max(0.01,c_m1-a1_est)
b2_m1 = max(0.01,c_m1-a2_est)
theta_m1[i,1] <- b1_m1
theta_m1[i,2] <- b2_m1

```

```

theta_m1[i,3] <- b_m1
theta_m1[i,4] <- c_m1
# Metodo 2.1
b1_m2.1 = max(0.01,(beta1_est*k2-a1_est*rXY)/(rXY+k2))
b2_m2.1 = max(0.01,(beta2_est*(rXY+k2)-beta1_est*rXY-a1_est*rXY)/(rXY+k2))
b_m2.1 = max(0.01,(beta1_est*rXY+a1_est*rXY)/(rXY+k2))
c_m2.1 = max(0.01,(beta1_est*k2+a1_est*k2)/(rXY+k2))
theta_m2.1[i,1] <- b1_m2.1
theta_m2.1[i,2] <- b2_m2.1
theta_m2.1[i,3] <- b_m2.1
theta_m2.1[i,4] <- c_m2.1
# Metodo 2.2
b1_m2.2 = max(0.01,(beta1_est*(rXY+k2)-beta2_est*rXY-a2_est*rXY)/(rXY+k2))
b2_m2.2 = max(0.01,(beta2_est*k2-a2_est*rXY)/(rXY+k2))
b_m2.2 = max(0.01,(beta2_est*rXY+a2_est*rXY)/(rXY+k2))
c_m2.2 = max(0.01,(beta2_est*k2+a2_est*k2)/(rXY+k2))
theta_m2.2[i,1] <- b1_m2.2
theta_m2.2[i,2] <- b2_m2.2
theta_m2.2[i,3] <- b_m2.2
theta_m2.2[i,4] <- c_m2.2
}
warnings()

# Analise das Estimativas: Parte 1
dec = 3
# Media
media_a1_mm = round(sum(theta_mm[,1])/max,dec)
media_beta1_mm = round(sum(theta_mm[,2])/max,dec)
media_a2_mm = round(sum(theta_mm[,3])/max,dec)
media_beta2_mm = round(sum(theta_mm[,4])/max,dec)
media_a1_mv = round(sum(theta_mv[,1])/max,dec)
media_beta1_mv = round(sum(theta_mv[,2])/max,dec)
media_a2_mv = round(sum(theta_mv[,3])/max,dec)
media_beta2_mv = round(sum(theta_mv[,4])/max,dec)
# Erro Quadratico Medio
EQM_a1_mm = round(sum((theta_mm[,1]-a1)^2)/max,dec)
EQM_beta1_mm = round(sum((theta_mm[,2]-beta1)^2)/max,dec)
EQM_a2_mm = round(sum((theta_mm[,3]-a2)^2)/max,dec)
EQM_beta2_mm = round(sum((theta_mm[,4]-beta2)^2)/max,dec)
EQM_a1_mv = round(sum((theta_mv[,1]-a1)^2)/max,dec)
EQM_beta1_mv = round(sum((theta_mv[,2]-beta1)^2)/max,dec)
EQM_a2_mv = round(sum((theta_mv[,3]-a2)^2)/max,dec)
EQM_beta2_mv = round(sum((theta_mv[,4]-beta2)^2)/max,dec)
# Coeficiente de Variacao
CV_a1_mm = round(sqrt(sum((theta_mm[,1]-media_a1_mm)^2)/max)/a1,dec)
CV_beta1_mm = round(sqrt(sum((theta_mm[,2]-media_beta1_mm)^2)/max)/beta1,dec)
CV_a2_mm = round(sqrt(sum((theta_mm[,3]-media_a2_mm)^2)/max)/a2,dec)
CV_beta2_mm = round(sqrt(sum((theta_mm[,4]-media_beta2_mm)^2)/max)/beta2,dec)
CV_a1_mv = round(sqrt(sum((theta_mv[,1]-media_a1_mv)^2)/max)/a1,dec)
CV_beta1_mv = round(sqrt(sum((theta_mv[,2]-media_beta1_mv)^2)/max)/beta1,dec)
CV_a2_mv = round(sqrt(sum((theta_mv[,3]-media_a2_mv)^2)/max)/a2,dec)
CV_beta2_mv = round(sqrt(sum((theta_mv[,4]-media_beta2_mv)^2)/max)/beta2,dec)
# Vies Relativo
VR_a1_mm = round((media_a1_mm-a1)/a1,dec)
VR_beta1_mm = round((media_beta1_mm-beta1)/beta1,dec)
VR_a2_mm = round((media_a2_mm-a2)/a2,dec)
VR_beta2_mm = round((media_beta2_mm-beta2)/beta2,dec)
VR_a1_mv = round((media_a1_mv-a1)/a1,dec)

```

```

VR_beta1_mv = round((media_beta1_mv-beta1)/beta1,dec)
VR_a2_mv = round((media_a2_mv-a2)/a2,dec)
VR_beta2_mv = round((media_beta2_mv-beta2)/beta2,dec)
# Tabela
data.frame("Valor"=c(a1,a2,beta1,beta2),
"Media MM"=c(media_a1_mm,media_a2_mm,media_beta1_mm,media_beta2_mm),
"EQM MM"=c(EQM_a1_mm,EQM_a2_mm,EQM_beta1_mm,EQM_beta2_mm),
"CV MM"=c(CV_a1_mm,CV_a2_mm,CV_beta1_mm,CV_beta2_mm),
"VR MM"=c(VR_a1_mm,VR_a2_mm,VR_beta1_mm,VR_beta2_mm),
"Media MV"=c(media_a1_mv,media_a2_mv,media_beta1_mv,media_beta2_mv),
"EQM MV"=c(EQM_a1_mv,EQM_a2_mv,EQM_beta1_mv,EQM_beta2_mv),
"CV MV"=c(CV_a1_mv,CV_a2_mv,CV_beta1_mv,CV_beta2_mv),
"VR MV"=c(VR_a1_mv,VR_a2_mv,VR_beta1_mv,VR_beta2_mv),
"n"=c(n,"","",""),
"Rep"=c(max,"","",""),
row.names=c("a1","a2","beta1","beta2"))

# Analise das Estimativas: Parte 2
dec = 3
# Media
media_b1_m1 = round(sum(theta_m1[,1])/max,dec)
media_b2_m1 = round(sum(theta_m1[,2])/max,dec)
media_b_m1 = round(sum(theta_m1[,3])/max,dec)
media_c_m1 = round(sum(theta_m1[,4])/max,dec)
media_b1_m2.1 = round(sum(theta_m2.1[,1])/max,dec)
media_b2_m2.1 = round(sum(theta_m2.1[,2])/max,dec)
media_b_m2.1 = round(sum(theta_m2.1[,3])/max,dec)
media_c_m2.1 = round(sum(theta_m2.1[,4])/max,dec)
media_b1_m2.2 = round(sum(theta_m2.2[,1])/max,dec)
media_b2_m2.2 = round(sum(theta_m2.2[,2])/max,dec)
media_b_m2.2 = round(sum(theta_m2.2[,3])/max,dec)
media_c_m2.2 = round(sum(theta_m2.2[,4])/max,dec)
# Erro Quadratico Medio
EQM_b1_m1 = round(sum((theta_m1[,1]-b1)^2)/max,dec)
EQM_b2_m1 = round(sum((theta_m1[,2]-b2)^2)/max,dec)
EQM_b_m1 = round(sum((theta_m1[,3]-b)^2)/max,dec)
EQM_c_m1 = round(sum((theta_m1[,4]-c)^2)/max,dec)
EQM_b1_m2.1 = round(sum((theta_m2.1[,1]-b1)^2)/max,dec)
EQM_b2_m2.1 = round(sum((theta_m2.1[,2]-b2)^2)/max,dec)
EQM_b_m2.1 = round(sum((theta_m2.1[,3]-b)^2)/max,dec)
EQM_c_m2.1 = round(sum((theta_m2.1[,4]-c)^2)/max,dec)
EQM_b1_m2.2 = round(sum((theta_m2.2[,1]-b1)^2)/max,dec)
EQM_b2_m2.2 = round(sum((theta_m2.2[,2]-b2)^2)/max,dec)
EQM_b_m2.2 = round(sum((theta_m2.2[,3]-b)^2)/max,dec)
EQM_c_m2.2 = round(sum((theta_m2.2[,4]-c)^2)/max,dec)
# Coeficiente de Variacao
CV_b1_m1 = round(sqrt(sum((theta_m1[,1]-media_b1_m1)^2)/max)/b1,dec)
CV_b2_m1 = round(sqrt(sum((theta_m1[,2]-media_b2_m1)^2)/max)/b2,dec)
CV_b_m1 = round(sqrt(sum((theta_m1[,3]-media_b_m1)^2)/max)/b,dec)
CV_c_m1 = round(sqrt(sum((theta_m1[,4]-media_c_m1)^2)/max)/c,dec)
CV_b1_m2.1 = round(sqrt(sum((theta_m2.1[,1]-media_b1_m2.1)^2)/max)/b1,dec)
CV_b2_m2.1 = round(sqrt(sum((theta_m2.1[,2]-media_b2_m2.1)^2)/max)/b2,dec)
CV_b_m2.1 = round(sqrt(sum((theta_m2.1[,3]-media_b_m2.1)^2)/max)/b,dec)
CV_c_m2.1 = round(sqrt(sum((theta_m2.1[,4]-media_c_m2.1)^2)/max)/c,dec)
CV_b1_m2.2 = round(sqrt(sum((theta_m2.2[,1]-media_b1_m2.2)^2)/max)/b1,dec)
CV_b2_m2.2 = round(sqrt(sum((theta_m2.2[,2]-media_b2_m2.2)^2)/max)/b2,dec)
CV_b_m2.2 = round(sqrt(sum((theta_m2.2[,3]-media_b_m2.2)^2)/max)/b,dec)
CV_c_m2.2 = round(sqrt(sum((theta_m2.2[,4]-media_c_m2.2)^2)/max)/c,dec)

```

```

# Vies Relativo
VR_b1_m1 = round((media_b1_m1-b1)/b1,dec)
VR_b2_m1 = round((media_b2_m1-b2)/b2,dec)
VR_b_m1 = round((media_b_m1-b)/b,dec)
VR_c_m1 = round((media_c_m1-c)/c,dec)
VR_b1_m2.1 = round((media_b1_m2.1-b1)/b1,dec)
VR_b2_m2.1 = round((media_b2_m2.1-b2)/b2,dec)
VR_b_m2.1 = round((media_b_m2.1-b)/b,dec)
VR_c_m2.1 = round((media_c_m2.1-c)/c,dec)
VR_b1_m2.2 = round((media_b1_m2.2-b1)/b1,dec)
VR_b2_m2.2 = round((media_b2_m2.2-b2)/b2,dec)
VR_b_m2.2 = round((media_b_m2.2-b)/b,dec)
VR_c_m2.2 = round((media_c_m2.2-c)/c,dec)
# Tabelas
data.frame("Valor"=c(b1,b2,b,c),
"Media M1"=c(media_b1_m1,media_b2_m1,media_b_m1,media_c_m1),
"EQM M1"=c(EQM_b1_m1,EQM_b2_m1,EQM_b_m1,EQM_c_m1),
"CV M1"=c(CV_b1_m1,CV_b2_m1,CV_b_m1,CV_c_m1),
"VR M1"=c(VR_b1_m1,VR_b2_m1,VR_b_m1,VR_c_m1),
"n"=c(n,"","","")),
"Rep"=c(max,"","","")),
row.names=c("b1","b2","b","c"))
data.frame("Valor"=c(b1,b2,b,c),
"Media M2.1"=c(media_b1_m2.1,media_b2_m2.1,media_b_m2.1,media_c_m2.1),
"EQM M2.1"=c(EQM_b1_m2.1,EQM_b2_m2.1,EQM_b_m2.1,EQM_c_m2.1),
"CV M2.1"=c(CV_b1_m2.1,CV_b2_m2.1,CV_b_m2.1,CV_c_m2.1),
"VR M2.1"=c(VR_b1_m2.1,VR_b2_m2.1,VR_b_m2.1,VR_c_m2.1),
"Media M2.2"=c(media_b1_m2.2,media_b2_m2.2,media_b_m2.2,media_c_m2.2),
"EQM M2.2"=c(EQM_b1_m2.2,EQM_b2_m2.2,EQM_b_m2.2,EQM_c_m2.2),
"CV M2.2"=c(CV_b1_m2.2,CV_b2_m2.2,CV_b_m2.2,CV_c_m2.2),
"VR M2.2"=c(VR_b1_m2.2,VR_b2_m2.2,VR_b_m2.2,VR_c_m2.2),
"n"=c(n,"","","")),
"Rep"=c(max,"","","")),
row.names=c("b1","b2","b","c"))

#-----
#          APLICAÇÃO
#-----


# Peanut Rust (X: Real, Y: Estimado)
X = c(.31,.42,.14,.45,.36,.23,.38,.45,.10,.34,.15,.35,.32,.25,.52,.35,.33,.08,.34,.27,.21,.13,.11,.34,.3
Y = c(.38,.46,.08,.67,.29,.18,.44,.52,.07,.29,.16,.53,.41,.22,.86,.43,.39,.14,.43,.38,.26,.12,.16,.55,.4
plot(X, Y, xlim=c(0, 1), ylim=c(0, 1), cex=0.1)
abline(a=0, b=1, col="Red", lty=2)
# Metodo dos Momentos
n = length(X)
epsilon = 0.5
Xbarra = sum(X)/length(X)
Ybarra = sum(Y)/length(Y)
S2X = sum((X-Xbarra)^2)/(length(X)-1)
S2Y = sum((Y-Ybarra)^2)/(length(Y)-1)
rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(Y)-1)*sqrt(S2X)*sqrt(S2Y))
a1_mm = max(epsilon,Xbarra*((1-Xbarra)*Xbarra/S2X)-1)
beta1_mm = max(epsilon,(1-Xbarra)*(((1-Xbarra)*Xbarra/S2X)-1))
a2_mm = max(epsilon,Ybarra*((1-Ybarra)*Ybarra/S2Y)-1)
beta2_mm = max(epsilon,(1-Ybarra)*(((1-Ybarra)*Ybarra/S2Y)-1))
a1_est = a1_mm
a2_est = a2_mm

```

```

beta1_est = beta1_mm
beta2_est = beta2_mm
# Metodo 2.2
rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(X)-1)*sqrt(S2X)*sqrt(S2Y))
k1 = sqrt((a1_est*a2_est)/(Xbarra*Ybarra))
k2 = sqrt((a1_est*a2_est)/(beta1_est*beta2_est))
b1_m2.2 = max(epsilon,(beta1_est*(rXY+k2)-beta2_est*rXY-a2_est*rXY)/(rXY+k2))
b2_m2.2 = max(epsilon,(beta2_est*k2-a2_est*rXY)/(rXY+k2))
b_m2.2 = max(epsilon,(beta2_est*rXY+a2_est*rXY)/(rXY+k2))
c_m2.2 = max(epsilon,(beta2_est*k2+a2_est*k2)/(rXY+k2))
# Testando Restricoes
data.frame("Estimativa"=c(c_m2.2,a1_est+b1_m2.2,a2_est+b2_m2.2),
row.names=c("c","a1+b1","a2+b2"))
data.frame("Estimativa"=c(beta1_est,b1_m2.2+b_m2.2,beta2_est,b2_m2.2+b_m2.2),
row.names=c("beta1","b1+b","beta2","b2+b"))
# Tabela
data.frame("MM"=c(a1_mm,a2_mm,beta1_mm,beta2_mm),
row.names=c("a1","a2","beta1","beta2"))
data.frame("Metodo2.2"=c(b1_m2.2,b2_m2.2,b_m2.2,c_m2.2),
row.names=c("b1","b2","b","c"))
# Gerando a Beta Bivariada
a1 = a1_est
a2 = a2_est
b = b_m2.2
c = c_m2.2
b1 = b1_m2.2
b2 = b2_m2.2
ng = 200
U <- rbeta(ng, a1, b1)
V <- rbeta(ng, a2, b2)
W <- rbeta(ng, c, b)
Xg = U*W
Yg = V*W
# Graficos
options(OutDec=",")
plot(X, Y, xlim=c(0, 1), ylim=c(0, 1), xlab="x", ylab="y", cex=0.3)
abline(a=0, b=1, lty=3)
points(Xg, Yg, pch="+", col="red", cex=0.6)
legend(0.73,0.3,legend=c("observado","simulado"),pch=c("o","+"),col=c("black","red"),
text.col=c("black","red"))

#-----
# Early Leaf Spot (X: Real, Y: Estimado)
X = c(.43,.54,.51,.04,.17,.44,.54,.01,.08,.53,.07,.47,.20,.55,.37,.51,.27,.04,.32,.32,.53,.09,.46,.
Y = c(.45,.63,.46,.08,.14,.34,.55,.06,.11,.63,.09,.37,.21,.47,.32,.41,.22,.09,.26,.31,.54,.12,.27,.
plot(X, Y, xlim=c(0, 1), ylim=c(0, 1), cex=0.1)
abline(a=0, b=1, col="Red", lty=2)
# Metodo dos Momentos
n = length(X)
epsilon = 0.5
Xbarra = sum(X)/length(X)
Ybarra = sum(Y)/length(Y)
S2X = sum((X-Xbarra)^2)/(length(X)-1)
S2Y = sum((Y-Ybarra)^2)/(length(Y)-1)
rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(Y)-1)*sqrt(S2X)*sqrt(S2Y))
a1_mm = max(epsilon,Xbarra*((1-Xbarra)*Xbarra/S2X)-1)
beta1_mm = max(epsilon,(1-Xbarra)*(((1-Xbarra)*Xbarra/S2X)-1))

```

```

a2_mm = max(epsilon,Ybarra*((1-Ybarra)*Ybarra/S2Y)-1))
beta2_mm = max(epsilon,(1-Ybarra)*(((1-Ybarra)*Ybarra/S2Y)-1))
a1_est = a1_mm # <- escolher um metodo (a1_mm ou a1_mv)
a2_est = a2_mm # <- escolher um metodo (a2_mm ou a2_mv)
beta1_est = beta1_mm # <- escolher um metodo (beta1_mm ou beta1_mv)
beta2_est = beta2_mm # <- escolher um metodo (beta2_mm ou beta2_mv)
# Metodo 1
rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(X)-1)*sqrt(S2X)*sqrt(S2Y))
k1 = sqrt((a1_est*a2_est)/(Xbarra*Ybarra))
k2 = sqrt((a1_est*a2_est)/(beta1_est*beta2_est))
b_m1 = max(epsilon,(k1*rXY)/(rXY+k2))
c_m1 = max(epsilon,(k1*k2)/(rXY+k2))
b1_m1 = max(epsilon,c_m1-a1_est)
b2_m1 = max(epsilon,c_m1-a2_est)
# Testando Restricoes
data.frame("Estimativa"=c(c_m1,a1_est+b1_m1,a2_est+b2_m1),
row.names=c("c","a1+b1","a2+b2"))
data.frame("Estimativa"=c(beta1_est,b1_m1+b_m1,beta2_est,b2_m1+b_m1),
row.names=c("beta1","b1+b","beta2","b2+b"))
# Tabela Comparativa
data.frame("MM"=c(a1_mm,a2_mm,beta1_mm,beta2_mm),
row.names=c("a1","a2","beta1","beta2"))
data.frame("Metodo1"=c(b1_m1,b2_m1,b_m1,c_m1),
row.names=c("b1","b2","b","c"))
# Gerando a Beta Bivariada
a1 = a1_est
a2 = a2_est
b = b_m1
c = c_m1
b1 = b1_m1
b2 = b2_m1
ng = 200
U <- rbeta(ng, a1, b1)
V <- rbeta(ng, a2, b2)
W <- rbeta(ng, c, b)
Xg = U*W
Yg = V*W
# Graficos
options(OutDec=",")
plot(X, Y, xlim=c(0, 1), ylim=c(0, 1), xlab="x", ylab="y", cex=0.3)
abline(a=0, b=1, lty=3)
points(Xg, Yg, pch="+", col="red", cex=0.6)
legend(0.73,0.3,legend=c("observado","simulado"),pch=c("o","+"),col=c("black","red"),
text.col=c("black","red"))

#-----
# Late Leaf Spot (X: Real, Y: Estimado) - Silvio
X = c(.04,.51,.47,.07,.14,.23,.51,.42,.32,.33,.33,.27,.15,.29,.51,.29,.05,.27,.53,.38,.03,.46,.31,.23,.21,.19,.17,.15,.13,.11,.09,.07,.05,.03,.01)
Y = c(.05,.55,.40,.06,.09,.12,.50,.43,.20,.19,.18,.12,.14,.20,.46,.20,.09,.15,.54,.43,.08,.35,.25,.13,.11,.09,.07,.05,.03,.01)
plot(X, Y, xlim=c(0, 1), ylim=c(0, 1), cex=0.1)
abline(a=0, b=1, col="Red", lty=2)
# Metodo dos Momentos
n = length(X)
epsilon = 0.5
Xbarra = sum(X)/length(X)
Ybarra = sum(Y)/length(Y)
S2X = sum((X-Xbarra)^2)/(length(X)-1)

```

```

S2Y = sum((Y-Ybarra)^2)/(length(Y)-1)
rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(Y)-1)*sqrt(S2X)*sqrt(S2Y))
a1_mm = max(epsilon,Xbarra*((1-Xbarra)*Xbarra/S2X)-1))
beta1_mm = max(epsilon,(1-Xbarra)*((1-Xbarra)*Xbarra/S2X)-1))
a2_mm = max(epsilon,Ybarra*((1-Ybarra)*Ybarra/S2Y)-1))
beta2_mm = max(epsilon,(1-Ybarra)*((1-Ybarra)*Ybarra/S2Y)-1))
a1_est = a1_mm # <- escolher um metodo (a1_mm ou a1_mv)
a2_est = a2_mm # <- escolher um metodo (a2_mm ou a2_mv)
beta1_est = beta1_mm # <- escolher um metodo (beta1_mm ou beta1_mv)
beta2_est = beta2_mm # <- escolher um metodo (beta2_mm ou beta2_mv)
# Metodo 2.1
rXY = sum((X-Xbarra)*(Y-Ybarra))/((length(X)-1)*sqrt(S2X)*sqrt(S2Y))
k1 = sqrt((a1_est*a2_est)/(Xbarra*Ybarra))
k2 = sqrt((a1_est*a2_est)/(beta1_est*beta2_est))
b1_m2.1 = max(epsilon,(beta1_est*k2-a1_est*rXY)/(rXY+k2))
b2_m2.1 = max(epsilon,(beta2_est*(rXY+k2)-beta1_est*rXY-a1_est*rXY)/(rXY+k2))
b_m2.1 = max(epsilon,(beta1_est*rXY+a1_est*rXY)/(rXY+k2))
c_m2.1 = max(epsilon,(beta1_est*k2+a1_est*k2)/(rXY+k2))
# Testando Restricoes
data.frame("Estimativa"=c(c_m2.1,a1_est+b1_m2.1,a2_est+b2_m2.1),
row.names=c("c","a1+b1","a2+b2"))
data.frame("Estimativa"=c(beta1_est,b1_m2.1+b_m2.1,beta2_est,b2_m2.1+b_m2.1),
row.names=c("beta1","b1+b","beta2","b2+b"))
# Tabela
data.frame("MM"=c(a1_mm,a2_mm,beta1_mm,beta2_mm),
row.names=c("a1","a2","beta1","beta2"))
data.frame("Metodo2.1"=c(b1_m2.1,b2_m2.1,b_m2.1,c_m2.1),
row.names=c("b1","b2","b","c"))
# Gerando a Beta Bivariada
a1 = a1_est
a2 = a2_est
b = b_m2.1
c = c_m2.1
b1 = b1_m2.1
b2 = b2_m2.1
ng = 200
U <- rbeta(ng, a1, b1)
V <- rbeta(ng, a2, b2)
W <- rbeta(ng, c, b)
Xg = U*W
Yg = V*W
# Graficos
options(OutDec=",")
plot(X, Y, xlim=c(0, 1), ylim=c(0, 1), xlab="x", ylab="y", cex=0.3)
abline(a=0, b=1, lty=3)
points(Xg, Yg, pch="+", col="red", cex=0.6)
legend(0.73,0.3,legend=c("observado","simulado"),pch=c("o","+"),col=c("black","red"),
text.col=c("black","red"))

```