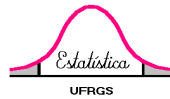




UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE MATEMÁTICA  
DEPARTAMENTO DE ESTATÍSTICA



# **Comparação de metodologias para estimar o coeficiente de correlação intraclasse no modelo log-binomial misto**

Autor: Mauricio Dalcin Zorzi

Orientador: Prof. Dra. Vanessa Bielefeldt Leotti Torman

Co-Orientadora: Prof. Dra. Suzi Alvez Camey

Porto Alegre, 25 de novembro de 2015.

Universidade Federal do Rio Grande do Sul  
Instituto de Matemática  
Departamento de Estatística

# Comparação de metodologias para estimar o coeficiente de correlação intraclasse no modelo log-binomial misto

Autor: Mauricio Dalcin Zorzi

Trabalho de Conclusão de Curso  
apresentado para obtenção  
do grau de Bacharel em Estatística.

Banca Examinadora:  
Professora Dra. Vanessa Bielefeldt Leotti Torman (orientadora)  
Professora Dra. Suzi Alvez Comey (co-orientadora)  
Professora Dra. Patrícia Klarmann Ziegelmann

Porto Alegre, 25 de novembro de 2015.



*Aos meus pais que me mostram diariamente o significado do amor incondicional.*

*“Franklin’s illness...gave him strength and courage he had not had before. He Had to think out the fundamentals of living and learn the greatest of all lessons – infinite patience and never ending persistence.”*

*Eleanor Roosevelt*

# Resumo

No contexto atual dos estudos epidemiológicos, os ensaios aleatorizados por *clusters* ganham popularidade devido a inviabilidade de alguns estudos de randomização individualizada. Apesar desta vantagem, estudos por *clusters* possuem menor poder do que os individualizados e necessitam de estimativas confiáveis para o coeficiente de correlação intraclasse (ICC), dado sua relação direta com o efeito do delineamento experimental que influi no tamanho da amostra dos estudos por *clusters*. Por isso, é importante se obter estimativas sólidas para este parâmetro.

Neste trabalho, realizamos uma comparação entre três metodologias de estimação para o ICC para o modelo log-binomial misto. Este modelo é adequado para estimar o risco relativo ou a razão de prevalência quando existem observações correlacionadas. As metodologias utilizadas neste trabalho foram propostas por Torman e Camey (1) e Yelland *et al* (2) e abordam maneiras distintas de se calcular o ICC. Torman e Camey (1) propõem um ICC utilizando a abordagem da variável latente baseada na lógica de Snijders e Bosker (3), enquanto que Yelland *et al* (2) propõem duas metodologias, uma baseada na aproximação de séries de Taylor e outra nas propriedades da distribuição log-normal.

O estudo foi conduzido através de simulação onde foram simulados bancos de dados com determinadas características controladas. Após a geração dos dados, estimou-se bayesianamente os coeficientes do modelo. Para o estimador pontual da moda, foram obtidos o vício, erro quadrático médio, percentual de observações que superestimaram ou subestimaram o ICC alvo em vinte por cento. Para os intervalos de credibilidade centrais, calculou-se a amplitude e o percentual de cobertura.

Observou-se performance diferenciada entre os estimadores propostos sendo que o estimador baseado na metodologia da série de Taylor obteve melhores resultados e o estimador da variável latente obteve as estimativas mais distantes do ICC alvo.

# Sumário

1.Introdução .....	9
2.Metodologia.....	11
2.1.1.Metodologia da Variável Latente .....	12
2.1.2.Metodologia das Séries de Taylor .....	12
2.1.3.Metodologia da Log-Normal.....	13
3.Simulação .....	14
4.Resultados.....	18
5.Discussão .....	22
Agradecimentos .....	24
Referências Bibliográficas.....	24
Anexos .....	26

Este artigo será submetido a “REVISTA *Clinical Trials*”



## 1.Introdução

No contexto dos estudos epidemiológicos, os ensaios aleatorizados por *clusters* ganham cada vez mais popularidade frente aos ensaios individualizados. Ensaios em *clusters* caracterizam-se pelo fato de a intervenção ser realizada no grupo e não somente no indivíduo e esta característica é uma vantagem deste delineamento, pois o torna muito mais praticável. Contudo, ao se realizar a intervenção em grupos, existe a possibilidade de que os indivíduos destes grupos apresentem uma correlação entre si e esta correlação diminui o poder destes estudos frente aos estudos individualizados.

A existência de correlação entre as unidades do mesmo grupo afeta a realização de um estudo principalmente no que se refere ao tamanho amostral, pois para compensar a correlação entre as unidades amostrais deve-se aumentar o tamanho da amostra para que o poder do estudo seja preservado. Consequentemente, é necessário que alguma medida quantifique a correlação dos indivíduos do mesmo *cluster* para que a amostra contenha o tamanho ideal, não punindo financeiramente o estudo, e nem diminuindo o poder do mesmo.

O coeficiente de correlação intraclasses (ICC) é o parâmetro que mede a correlação entre os indivíduos do mesmo *cluster*. O ICC é definido formalmente como a correlação de Pearson entre dois indivíduos do mesmo *cluster* ou o percentual de variância total atribuída à variância entre *clusters* (4). Devido ao aumento do uso dos ensaios por *clusters*, o ICC ganha um peso cada vez maior na comunidade científica. Citam-se Wu, Crespi e Wonk (4), Yelland *et al* (2), Torman e Camey (1), Pendergast *et al* (5), Ridou *et al* (6), Zou e Donner (7), Evans *et al* (8) e Turner *et al* (9) como trabalhos que avaliam ou propõem metodologias de estimação do ICC. Além do crescente número de publicações voltadas ao ICC, o CONSORT para ensaios clínicos em *clusters* – lista das melhores práticas a serem utilizadas nestes estudos - recomenda que o ICC seja avaliado para cada desfecho considerado, tornando-o uma medida fundamental nos ensaios por *cluster* (10).

Neste trabalho, estamos interessados, especificamente, no ICC voltado para o modelo log-binomial misto. Este modelo é adequado para a análise de ensaios clínicos randomizados por *cluster* e estudos multiníveis que são estudos onde se observa dependência entre as unidades amostrais.

O modelo log-binomial misto pertence à classe dos modelos lineares generalizados e supõe que o logaritmo da probabilidade de sucesso possui uma relação linear com as variáveis preditoras. Esta peculiaridade o torna um modelo ideal para a estimação do risco

relativo em delineamentos longitudinais ( ensaios clínicos, coorte) e da razão de prevalências no caso dos estudos transversais, pois basta exponencializar os coeficientes das covariáveis do modelo para obtermos as estimativas dos riscos relativos ou razão de prevalências associados a estas covariáveis. Destaca-se também que o modelo misto é um modelo adequado para os estudos em *clusters*, pois há a adição de um termo de erro gerado por um fator de dependência entre as unidades amostrais (1) .

Contudo, este modelo também apresenta pontos negativos, sendo que sua principal desvantagem é advinda dos problemas de convergência de seu algoritmo de estimação. Ressalta-se que Torman e Camey(1), através do uso de uma abordagem bayesiana, obtiveram sucesso na correção deste problema.

No que se refere ao ICC para o modelo log-binomial misto, este trabalho foca-se em três estimadores: um proposto por Torman e Camey – estimador da variável latente, e dois por Yelland *et al* – estimador das séries de Taylor e da log-normal. O interesse específico nestes três é dado pelo fato de que são os únicos estimadores que apresentam a capacidade de serem ajustados pelas covariáveis consideradas no modelo log-binomial misto. Esta característica é um ponto a seu favor e pode ser da maior utilidade. Por exemplo, em ensaios clínicos onde deseja-se retirar a influência do tratamento para medir a correlação entre as unidades dentro do mesmo cluster estes estimadores seriam os indicados. Contudo, se esta capacidade coloca estes três estimadores no mesmo grupo, esta é a única característica que os une. Torman e Camey (1) e Yelland *et al* (2) abordam o processo de obtenção dos estimadores de uma maneira diferenciada e obtêm estimadores distintos. Torman e Camey (1) utilizam a abordagem da variável latente enquanto que Yelland *et al* (2) apresenta duas metodologias, uma utilizando aproximações de séries de Taylor e a outra provinda das propriedades da distribuição log-normal.

Destaca-se que Yelland *et al* (2) já realizou uma comparação por simulação entre os métodos da log-normal e da série de Taylor. Em seu trabalho, utiliza uma metodologia frequentista para a estimação dos parâmetros do modelo e analisa 36 cenários de simulação e conclui que o estimador do ICC a ser utilizado para o modelo log-binomial misto é o estimador da log normal. Contudo com a existência de mais um estimador que permite o uso de covariáveis, surge o espaço para a comparação entre os três.

Este trabalho se propõe a comparar as metodologias de estimação do ICC propostos por Torman e Camey (1) e Yelland *et al* (2). Além das medidas de qualidade de estimadores pontuais utilizadas por Yelland *et al* (2), foi possível calcular os intervalos de credibilidade de todas as metodologias, e medidas de desempenho associadas, como o percentual de cobertura e amplitude.

Inicialmente, explicita-se o suporte teórico e a metodologia de simulação. Após este item, serão apresentados os resultados obtidos seguidos de uma discussão sobre os mesmos. Um anexo será apresentado contendo os códigos utilizados para a obtenção dos resultados.

## 2. Metodologia

O modelo log-binomial misto é um modelo linear generalizado em que a função de ligação é o logaritmo e a variável independente é a probabilidade de sucesso do evento de interesse, sendo que este evento é de natureza binária. Diferencia-se o modelo misto do modelo tradicional pelo acréscimo de um termo de erro associado ao *cluster*. Neste modelo, estima-se a variância associada ao fator de dependência de uma maneira direta, sendo que esta variância encontra-se na mesma escala de medida da função de ligação.

Neste estudo, os dados serão simulados de acordo com o modelo (1) contendo duas covariáveis e um intercepto.

$$\log(\pi_{ij}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + a_i \quad (1)$$

O modelo da equação 1 é idêntico (modelo com intercepto e duas covariáveis) ao utilizado em Yelland *et al* (2) para ser possível comparar o estimador de Torman e Camey (1) com os resultados obtidos naquele estudo.

Na equação 1,  $\pi_{ij}$  representa a probabilidade de sucesso para a *j*-ésima observação do *i*-ésimo *cluster*;  $x_{1i}$  é uma variável indicadora da presença de tratamento, isto é, 1 se o *cluster* *i* é tratado e 0 caso contrário;  $x_{2ij}$  uma variável confundidora binária correlacionada, dependente do *cluster* e do indivíduo; e  $a_i \sim N(0, \sigma_a^2)$  o termo de erro associado ao *i*-ésimo *cluster*.

O ICC intraclasses é definido como a correlação de Pearson entre dois indivíduos do mesmo *cluster* ou a proporção da variância total atribuída à variância entre *clusters* (4). A expressão do ICC é dada através da equação 2:

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \quad (2)$$

onde,  $\sigma_b^2$  representa a variância entre os *clusters* e  $\sigma_w^2$  a variância dentro dos *clusters*.

Os três estimadores do ICC partem da equação 2, mas o que os distingue é a maneira de se definir  $\sigma_b^2$  e  $\sigma_w^2$ . Cada uma das metodologias utiliza um princípio para obter a variância dentro e entre *clusters*. Torman e Camey (1) partem do princípio da variável latente para obter  $\sigma_b^2$  e  $\sigma_w^2$ , enquanto que Yelland *et al* (2) utiliza as propriedades da distribuição binomial para  $\sigma_w^2$ , e para  $\sigma_b^2$  utiliza aproximações da série de Taylor e propriedades da distribuição log normal.

### 2.1.1. Metodologia da Variável Latente

A estimativa do ICC obtido através da metodologia da variável latente baseia-se no princípio de que os resultados binários observados estão relacionados à alguma variável contínua que não é diretamente observada. Snijders e Bosker (3) usaram o princípio da variável latente para encontrar o ICC para o modelo logístico misto, assumindo que o desfecho binário era proveniente da dicotomização de uma variável latente contínua.

Aplicando-se a metodologia da variável latente para o modelo log-binomial misto, Torman e Camey (1) encontram que esta variável latente poderia ser explicada através de um modelo com erros exponenciais de média igual a 1. Assim no ICC obtido pela variável latente, a variância dentro dos *clusters* é igual a 1 e a variância entre *clusters* é igual a variância do termo de erro associado ao *cluster*. Com isto, a definição do coeficiente para o modelo log-binomial misto é dada da seguinte maneira:

$$\rho_{vl} = \frac{\sigma_a^2}{\sigma_a^2 + 1} \quad (3)$$

Pelo fato de usarmos o termo  $\sigma_a^2$  na equação 3,  $\rho_{vl}$  será referente ao logaritmo da probabilidade de sucesso.

Nos estimadores da série de Taylor e da log-normal, que serão definidos nos itens 2.1.2 e 2.1.3 respectivamente, obtém-se a variância dentro dos *clusters* pela distribuição binomial fazendo com que esta seja estimada na escala de probabilidade. Contudo, a variância entre *clusters* é calculada na escala da função de ligação da probabilidade de sucesso e, com isso, faz-se necessário transformar a variância entre *clusters* também esteja na escala de probabilidade. Conseqüentemente, os estimadores da série de Taylor e da log-normal referem-se a probabilidade de sucesso.

### 2.1.2. Metodologia das Séries de Taylor

O estimador do ICC dado pela metodologia da aproximação da séries de Taylor, é obtido através de uma abordagem similar a Murray (11) que mostra como obter a variância entre *clusters* usando o erro associado ao *cluster* para o modelo logístico misto. Yelland et al (2) demonstra que para o modelo log-binomial, a variância entre *clusters* é aproximada por,

$$\sigma_b^2 \approx \sigma_a^2 * p^2 \quad (4)$$

onde  $p$  é a prevalência média dos *clusters* dada pela equação 5.

$$\bar{p} = \frac{1}{B} \sum_i \sum_j \frac{X_{ij}}{n_i} \quad (5)$$

onde  $B$  representa o número de *clusters*,  $Y_{ij}$  representa a variável binária indicando a ocorrência do desfecho do  $j$ -ésimo indivíduo do  $i$ -ésimo cluster e  $n_i$  é o tamanho do  $i$ -ésimo *clusters*.

A variância dentro dos *clusters*, é estimada por:

$$\hat{\sigma}_w^2 = [\bar{p}(1 - \bar{p})] \quad (6)$$

Substituindo (4) e (6) em (2) temos que o estimador do ICC da série de Taylor é pela equação 7:

$$\hat{\rho}_{st} = \frac{\hat{\sigma}_a^2 * \bar{p}^2}{\hat{\sigma}_a^2 * \bar{p}^2 + [\bar{p}(1 - \bar{p})]} \quad (7)$$

### 2.1.3. Metodologia da Log-Normal

Para o estimador do ICC define-se a variância entre os *clusters* utilizando uma propriedade da distribuição log-normal. A partir da equação (1), fazendo-se  $\mu = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij}$  pode-se obter que  $\log(\pi_{ij}) \sim N(\mu, \sigma_a^2)$ , e conseqüentemente  $\pi_{ij}$  tem distribuição log-normal com média  $\exp(\mu + (\sigma_a^2/2))$  e variância  $\exp(\sigma_a^2) * E[\pi_{ij}]^2$ . Usando  $\bar{p}$  para estimar  $E[\pi_{ij}]$ , isto implica que

$$\hat{\sigma}_b^2 = \exp(\hat{\sigma}_a^2 - 1) * \bar{p}^2 \quad (8)$$

Para a variância dentro dos *clusters* tem-se que esta é calculada como no estimador da aproximação da série de Taylor, i.e,  $\sigma_w^2 = [\bar{p}(1 - \bar{p})]$ . Imputando (8) e (6) em (2) temos que o estimador da log-normal é pela equação 9:

$$\hat{\rho}_m = \frac{\exp(\hat{\sigma}_a^2 - 1) * \bar{p}^2}{\exp(\hat{\sigma}_a^2 - 1) * \bar{p}^2 + [\bar{p}(1 - \bar{p})]} \quad (9)$$

onde  $\bar{p}$  é a prevalência média dos *clusters* dada na equação 5 e  $\sigma_a^2$  é a variância do termo de erro associado ao *cluster*.

### 3.Simulação

Para este trabalho, destaca-se que o código de geração dos bancos de dados foram gentilmente concedidos pela Professora Lisa N. Yelland, onde estavam inclusos todos os pormenores da geração dos dados, incluindo sementes aleatórias.

No que se refere especificamente ao processo da geração dos dados, foram considerados doze cenários. Cada cenário é o resultado da combinação entre os valores distintos assumidos para os riscos relativos do tratamento, do confundidor e do parâmetro  $\sigma_a^2$ , sendo que todos os cenários estão listados na tabela 1.

**Tabela 1 – Cenários de Simulação**

Cenário	Risco Relativo do Tratamento	Risco Relativo do Confundidor	$\sigma_a^2$
1	1,25	1,25	0,01
2	1,25	1,25	0,10
3	1,25	1,25	0,50
4	1,25	2,00	0,01
5	1,25	2,00	0,10
6	1,25	2,00	0,50
7	2,00	1,25	0,01
8	2,00	1,25	0,10
9	2,00	1,25	0,50
10	2,00	2,00	0,01
11	2,00	2,00	0,10
12	2,00	2,00	0,50

Para todos doze cenários analisados foram utilizadas as seguintes características: o valor de 0,1 para o risco basal esperado; 20 *clusters* sendo que cada um com 20 indivíduos totalizando um tamanho amostral de 400 indivíduos; prevalência 50% de *clusters* tratados

(variável  $x_{1i}$ ); geração da variável confundidora através do modelo beta binomial com 0,5 de prevalência média por *cluster* e 0,05 de ICC (variável  $x_{2ij}$ ). Por fim, para cada um dos possíveis cenários descritos acima, serão simulados 1000 conjuntos de dados.

Em seu trabalho, Yelland *et al* (2) propõe mais combinações entre número de *clusters* e de indivíduos por *cluster*. Porém, o caso que considera 20 *cluster* com 20 indivíduos apresentou os piores resultados (2). Com isso, decidiu-se realizar a simulação somente para este caso.

Para cada banco de dados simulado foi ajustado um modelo log-binomial misto utilizando uma abordagem bayesiana através de MCMC (*Markov chain Monte Carlo*) para se obter as estimativas dos parâmetros. Utilizou-se uma abordagem bayesiana devido ao fato de que o algoritmo de estimação tradicional do modelo log-binomial misto pode apresentar problemas de convergência. Este problema acontece quando as estimativas dos parâmetros do modelo possuem valores maiores que a unidade e menores que zero. O uso de uma abordagem bayesiana proposto por Torman e Camey (1), solucionou este problema de convergência fazendo com que essa abordagem se tornasse uma alternativa para a estimação do modelo log-binomial misto.

A estatística bayesiana combina o conhecimento prévio do parâmetro, dado pela priori, com a função de verossimilhança da variável, para a obtenção da distribuição a posteriori de onde são realizadas as inferências sobre o parâmetro. Contudo, para muitos modelos estatísticos, a distribuição a posteriori não pode ser obtida analiticamente. Para estes casos, a distribuição a posteriori pode ser aproximada numericamente através de MCMC (*Markov Chain Monte Carlo*). A metodologia MCMC constitui-se em uma família de algoritmos utilizada para gerar dados com uma distribuição de interesse, que para o caso da inferência bayesiana é a distribuição a posteriori. Contudo, para que os dados simulados possam ser utilizados, estes devem possuir determinadas propriedades sendo que uma destas é a ausência de autocorrelação nas simulações (12).

Para se obter melhores resultados referentes à autocorrelação das cadeias, foi realizada uma reparametrização do modelo baseada no modelo de Poisson usando equações de estimação generalizadas (13). Essa reparametrização se baseia na substituição dos parâmetros originais do modelo por novos parâmetros cuja matriz de variâncias e covariâncias será aproximadamente igual à matriz identidade. Contudo, no modelo log-binomial a matriz de covariância dos parâmetros pode não ser facilmente obtida, devido a problemas de convergência do algoritmo. Para este problema de convergência, Salmeron, Cano e Chirlaque (14) propuseram uma reparametrização baseada na decomposição de Cholesky da matriz de variâncias e covariâncias da regressão de Poisson. Esse método

pode melhorar a eficiência do MCMC fazendo com que sejam necessárias menos iterações.

Com isso o modelo da equação (1) reparametrizado é dado pela equação (10):

$$\log(\pi_{ij}) = \theta_0 Z_{0i} + \theta_1 Z_{1i} + \theta_2 Z_{2ij} + a_i \quad (10)$$

onde  $\theta_0 Z_{0i}$ ,  $\theta_1 Z_{1i}$  e  $\theta_2 Z_{2ij}$  são a reparametrização para  $\beta_0$ ,  $\beta_1 x_{1i}$  e  $\beta_2 x_{2ij}$  respectivamente. Ressalta-se que, mesmo com a reparametrização do modelo, os parâmetros de interesse deste estudo são apenas os ICC definidos através das três metodologias, que não mudam com este procedimento, já que esta não modifica o termo de erro do modelo. Destaca-se que a interpretabilidade dos riscos relativos para os parâmetros não é perdida com a reparametrização, pois, após a geração das observações, é possível converter os coeficientes reparametrizados ( $\theta_0, \theta_1$  e  $\theta_2$ ) para os originais ( $\beta_0, \beta_1$  e  $\beta_2$ ). Para maiores detalhes ver Salmeron (14) e o anexo contendo os códigos.

Devido a abordagem bayesiana de estimação, deve-se atribuir distribuições a priori para os parâmetros do modelo. Supôs-se que os parâmetros  $\theta$  possuem distribuição normal com média zero e variância 10.000 e que  $\sigma_a$  possui distribuição uniforme com parâmetros 0,01 e 100 o que caracteriza prioris não informativas para todos os parâmetros. Foram utilizadas estas distribuições devido ao desconhecimento inicial do comportamento dos parâmetros. Além disto, como Yelland *et al* (2) utiliza uma metodologia frequentista de estimação dos coeficientes, a incorporação de um conhecimento prévio sobre os parâmetros faria com que os resultados deste trabalho não fossem diretamente comparáveis aos de Yelland *et al* (2).

Para cada parâmetro de interesse, geraram-se 2 cadeias de 30.000 simulações, utilizando um *burn-in* de 2.000, com *thin* igual a 5. Todos estes critérios foram definidos através da análise visual das cadeias e com suporte na estatística de Gelman e Rubin. Foram guardadas as cadeias simuladas e calculou-se para cada amostra, a média, a moda – dada pelo máximo da densidade estimada de forma não paramétrica, a mediana e o tamanho efetivo da cadeia relativo a cada parâmetro.

Normalmente em estudos realizados através de simulação, sabemos de antemão o parâmetro de interesse e, com isso, podemos obter conclusões mais assertivas sobre os resultados obtidos. Contudo, o ICC é um parâmetro que não pode ser definido previamente, graças ao fato de que não temos uma expressão fechada para a variância dentro dos *clusters* para o modelo log-binomial na escala das probabilidades. Yelland *et al*



(2) propõem a utilização do ICC alvo, uma medida calculada através da combinação de um processo de simulação e a definição do efeito do delineamento experimental.

Para obter-se o ICC alvo, em cada replicação da simulação foram simulados dois bancos de dados: um considerando a dependência das unidades amostrais ( caracterizado pela correlação na variável confundidora e pelo termo de erro associado ao cluster inserido no modelo) e outro sem correlação na variável confundidora e sem o termo de erro associado ao cluster. Com estes dois bancos de dados simulados podemos aproximar o efeito do delineamento experimental ( DEFF ) pela razão da média das variâncias dos dois bancos simulados, i.e,  $DEFF \approx \sigma_{dep}^2 / \sigma_{ind}^2$  onde  $\sigma_{ind}^2$  é a média das variâncias do estimador do coeficiente da variável indicadora do tratamento nos bancos supondo independência e  $\sigma_{dep}^2$  é a média das variâncias do estimador do coeficiente da variável indicadora do tratamento nos bancos supondo dependência. Como  $DEFF = 1 + (\bar{m} - 1)\rho$ , onde  $\bar{m}$  é o tamanho médio do cluster e  $\rho$  o ICC, basta isolar o  $\rho$  para se obter a estimativa do ICC alvo. Sabendo que este não é o verdadeiro valor do coeficiente, estipula-se que o mesmo é um valor razoável para o parâmetro verdadeiro.

Ressaltamos que o ICC alvo é um método interessante para se obter uma ideia do verdadeiro valor do ICC, pois mede diretamente o impacto da dependência nos dados na estimação do coeficiente do tratamento podendo-se medir a inflação da variância. Contudo, não é uma medida direta de correlação, além de ser uma aproximação dada pelo processo de simulação o que a torna um valor apenas aproximado. Também ressalta-se que o ICC alvo encontra-se em uma escala diferente do estimador de Torman e Camey (1).

Os resultados pontuais obtidos para cada uma das metodologias de cálculo do ICC foram comparadas utilizando-se o erro quadrático médio, vício e vício mediano. Além disto, apurou-se o percentual de observações que ficaram acima e abaixo de vinte por cento do ICC alvo. Para os intervalos de credibilidade calculou-se a amplitude média e percentual de cobertura (percentual de vezes que o ICC alvo está incluído entre os de credibilidade e o intervalo limite superior do intervalo de credibilidade). Os valores dos ICC's alvos que serão utilizados para as medidas comparativas foram obtidos em Yelland *et al* (2).

Destaca-se que o software utilizado para a análise das simulações foi o R. 3.0.2, especificamente os pacotes *sandwich* (17) para se obter as estimativas da regressão de Poisson que serão utilizadas para a reparametrização do modelo, e para a geração das cadeias através de MCMC utilizou-se o JAGS 3.4.0 em conjunto com o *rjags*. Para a geração dos bancos de dados foi usado o SAS 9.0.

## 4.Resultados

Através da simulação realizada, observou-se diferença na performance dos estimadores dos três estimadores. Optou-se por apenas apresentar os resultados referentes ao estimador pontual da moda, pois foi o estimador que apresentou menores valores de erro quadrático médio (EQM). Na Tabela 2 é apresentado o vício médio dos estimadores para cada cenário. Observa-se que o estimador da variável latente sempre superestima o ICC alvo, enquanto que os estimadores log-normal e série de Taylor sempre o subestimam. Ressalta-se que comparativamente, o estimador  $\rho_{ln}$  sempre apresentou melhores resultados para o vício médio. Observa-se a tendência de piora de resultados para os estimadores  $\rho_{vl}$  e  $\rho_{st}$  a medida que a variância entre cluster aumenta sendo que este padrão não é observado para o estimador  $\rho_{ln}$ .

**Tabela 2 – Vício médio da moda dos estimadores para os cenários considerados**

Cenário	Risco Relativo do Tratamento	Risco Relativo do Confundidor	Variância entre Clusters (Escala Logarítmica)	ICC Alvo	Vício Médio da	Vício Médio	Vício Médio da
					Moda	da Moda	Moda
					$\rho_{vl}$	$\rho_{st}$	$\rho_{ln}$
1	1,25	1,25	0,01	0,0081	0,0156	-0,0049	-0,0048
2	1,25	1,25	0,10	0,0187	0,0486	-0,0094	-0,0093
3	1,25	1,25	0,50	0,0746	0,2398	-0,0177	-0,0047
4	1,25	2,00	0,01	0,0082	0,0091	-0,0049	-0,0049
5	1,25	2,00	0,10	0,0240	0,0404	-0,0111	-0,0110
6	1,25	2,00	0,50	0,1016	0,2111	-0,0252	-0,0116
7	2,00	1,25	0,01	0,0074	0,0108	-0,0040	-0,0040
8	2,00	1,25	0,10	0,0211	0,0434	-0,0077	-0,0077
9	2,00	1,25	0,50	0,0879	0,2235	-0,0134	-0,0007
10	2,00	2,00	0,01	0,0072	0,0054	-0,0037	-0,0036
11	2,00	2,00	0,10	0,0274	0,0358	-0,0089	-0,0086
12	2,00	2,00	0,50	0,1131	0,1695	-0,0278	-0,0156

As tabelas 3 e 4 apresentam o percentual de vezes em que os estimadores subestimaram ou superestimaram o ICC alvo em 20%, respectivamente. Na tabela 3, observa-se que o estimador da variável latente apresenta melhores resultados no quesito subestimação em mais de 20% do ICC alvo comparados aos estimadores obtidos pela log-normal e pela série de Taylor. Ressalta-se que nos estimadores log-normal e séries de Taylor, no mínimo, 46,90% das amostras apresentam subestimação em mais de 20% do ICC alvo para os cenários analisados. A medida que a variância entre *clusters* aumenta, os estimadores  $\rho_{st}$  e  $\rho_{ln}$  mostraram uma diminuição no percentual de casos que subestimaram o ICC alvo em mais de 20%. Para o estimador da variável latente, esse padrão não foi observado, e destaca-se que os melhores resultados foram para a maior variância.

Para o percentual de estimativas que superaram o ICC alvo em mais de 20%, nota-se que o estimador variável latente sempre apresenta um percentual maior de superestimação comparativamente aos outros dois, combinando com o comportamento observado do vício. Notou-se que o percentual de observações que superaram o *alvo* em 20% para os estimador  $\rho_{ln}$  sempre aumenta à medida que a variância entre *clusters* aumenta. Para  $\rho_{st}$  este comportamento não foi observado. Nesse quesito, observa-se o mesmo comportamento para o  $\rho_{vl}$ , exceto para o caso onde os riscos relativos do tratamento e do confundidor são ambos iguais a 1,25.

**Tabela 3 – Percentual de observações que subestimaram o ICC Alvo em mais de vinte por cento para os cenários considerados .**

Cenário	Risco Relativo do Tratamento	Risco Relativo do Confundidor	Variância entre <i>Clusters</i> (Escala Logarítmica)	ICC Alvo	% $\rho_{vl}$	% $\rho_{st}$	% $\rho_{ln}$
1	1,25	1,25	0,01	0,0081	25,1	93,4	91,0
2	1,25	1,25	0,10	0,0187	46,8	79,4	81,7
3	1,25	1,25	0,50	0,0746	17,5	60,0	56,1
4	1,25	2,00	0,01	0,0082	44,2	93,4	92,4
5	1,25	2,00	0,10	0,0240	60,8	72,4	74,6
6	1,25	2,00	0,50	0,1016	8,3	60,2	54,7
7	2,00	1,25	0,01	0,0074	40,6	91,5	89,9
8	2,00	1,25	0,10	0,0211	54,3	69,8	71,9
9	2,00	1,25	0,50	0,0879	9,5	51,2	46,9
10	2,00	2,00	0,01	0,0072	57,8	91,6	91,3
11	2,00	2,00	0,10	0,0274	50,7	62,9	64,1
12	2,00	2,00	0,50	0,1131	7,2	60,1	53,8

**Tabela 4 – Percentual de observações que superestimaram o ICC Alvo em mais de vinte por cento para os cenários considerados .**

Cenário	Risco Relativo do Tratamento	Risco Relativo do Confundidor	Variância entre <i>Clusters</i> (Escala Logarítmica)	ICC Alvo	% $\rho_{vl}$	% $\rho_{st}$	% $\rho_{ln}$
1	1,25	1,25	0,01	0,0081	49,8	4,7	5,3
2	1,25	1,25	0,10	0,0187	30,8	16,5	14,1
3	1,25	1,25	0,50	0,0746	82,4	20,0	26,6
4	1,25	2,00	0,01	0,0082	31,6	5,4	5,2
5	1,25	2,00	0,10	0,024	33,7	18,6	18
6	1,25	2,00	0,50	0,1016	89,1	16,9	24,7
7	2,00	1,25	0,01	0,0074	36,7	6,8	6,9
8	2,00	1,25	0,10	0,0211	37,5	22,3	20,8
9	2,00	1,25	0,50	0,0879	88,7	24,8	32,2
10	2,00	2,00	0,01	0,0072	20,9	7,5	7,5
11	2,00	2,00	0,10	0,0274	48,1	21,6	21,4
12	2,00	2,00	0,50	0,1131	87,2	17,0	24,3

Para o erro quadrático médio (tabela 5), destaca-se que em nenhum dos casos o estimador da variável latente superou os outros dois estimadores. Na comparação entre  $\rho_{st}$  e  $\rho_{ln}$ , somente nos cenários 1,4,7 e 10 o EQM do  $\rho_{ln}$  foi menor que o EQM do  $\rho_{st}$ . Com o aumento da variância entre *clusters*, observa-se o aumento do EQM para todos os três estimadores.

**Tabela 5 – Erro Quadrático Médio para os cenários considerados**

Cenário	Risco Relativo do Tratamento	Risco Relativo do Confundidor	Variância entre Clusters (Escala Logarítmica)	ICC Alvo	EQM $\rho_{vt}$	EQM $\rho_{st}$	EQM $\rho_{ln}$
1	1,25	1,25	0,01	0,0081	0,0031760	0,0000661	0,0000610
2	1,25	1,25	0,10	0,0187	0,0132916	0,0003053	0,0003101
3	1,25	1,25	0,50	0,0746	0,0910592	0,0025032	0,0078482
4	1,25	2,00	0,01	0,0082	0,0014411	0,0000663	0,0000637
5	1,25	2,00	0,10	0,0240	0,0091595	0,0004498	0,0004736
6	1,25	2,00	0,50	0,1016	0,0670533	0,0032209	0,0057454
7	2,00	1,25	0,01	0,0074	0,0016158	0,0000583	0,0000545
8	2,00	1,25	0,10	0,0211	0,0096727	0,0003921	0,0004138
9	2,00	1,25	0,50	0,0879	0,0748289	0,0026427	0,0050914
10	2,00	2,00	0,01	0,0072	0,0007988	0,0000660	0,0000657
11	2,00	2,00	0,10	0,0274	0,0060415	0,0005311	0,0005720
12	2,00	2,00	0,50	0,1131	0,0464904	0,0033904	0,0047472

Em relação a amplitude média (tabela 6) e cobertura dos intervalos de credibilidade (tabela 7) mantém-se a tendência de que o estimador da variável latente apresenta resultados inferiores aos outros dois, sendo que o  $\rho_{st}$  apresenta maior percentual de cobertura comparado aos outros estimadores com exceção do cenário 12. Para a amplitude, o estimador  $\rho_{st}$  obtêm os menores valores em todos os cenários. A medida que a variância entre *clusters* aumenta, aumenta a amplitude e diminui a cobertura para todos os estimadores.

**Tabela 6 – Amplitude média dos intervalos de credibilidade com 95% de probabilidade para os cenários considerados**

Cenário	Risco Relativo do Tratamento	Risco Relativo do Confundidor	Variância entre <i>Clusters</i> (Escala Logarítmica)	ICC Alvo	Amplitude $\rho_{vt}$	Amplitude $\rho_{st}$	Amplitude $\rho_{ln}$
1	1,25	1,25	0,01	0,0081	0,3426	0,0736	0,1048
2	1,25	1,25	0,10	0,0187	0,4174	0,1024	0,1673
3	1,25	1,25	0,50	0,0746	0,5448	0,2108	0,4752
4	1,25	2,00	0,01	0,0082	0,2622	0,0686	0,0862
5	1,25	2,00	0,10	0,0240	0,3512	0,1049	0,1497
6	1,25	2,00	0,50	0,1016	0,4954	0,2254	0,4552
7	2,00	1,25	0,01	0,0074	0,2682	0,0707	0,0899
8	2,00	1,25	0,10	0,0211	0,3608	0,1092	0,1586
9	2,00	1,25	0,50	0,0879	0,5001	0,2289	0,4648
10	2,00	2,00	0,01	0,0072	0,1944	0,0653	0,0760
11	2,00	2,00	0,10	0,0274	0,2984	0,1128	0,1477
12	2,00	2,00	0,50	0,1131	0,4586	0,2319	0,4227

**Tabela 7 – Percentual de cobertura dos Intervalos de Credibilidade com 95% de probabilidade para os cenários considerados**

Cenário	Risco Relativo do Tratamento	Risco Relativo do Confundidor	Variância entre <i>Clusters</i> (Escala Logarítmica)	ICC Alvo	Cobertura $\rho_{vt}$	Cobertura $\rho_{st}$	Cobertura $\rho_{ln}$
1	1,25	1,25	0,01	0,0081	97,5	99,5	99,4
2	1,25	1,25	0,10	0,0187	90,8	99,1	98,8
3	1,25	1,25	0,50	0,0746	48,4	93,0	92,1
4	1,25	2,00	0,01	0,0082	97,4	99,8	99,8
5	1,25	2,00	0,10	0,0240	88,0	98,3	98,1
6	1,25	2,00	0,50	0,1016	44,5	93,6	93,1
7	2,00	1,25	0,01	0,0074	97,6	99,6	99,6
8	2,00	1,25	0,10	0,0211	87,0	98,0	97,7
9	2,00	1,25	0,50	0,0879	41,0	93,7	91,7
10	2,00	2,00	0,01	0,0072	97,6	99,1	99,1
11	2,00	2,00	0,10	0,0274	85,1	96,7	96,5
12	2,00	2,00	0,50	0,1131	48,2	94,2	94,7

Na Figura 1, mostram-se os *boxplots* das estimativas do ICC obtidos para cada um dos estimadores para cada cenário analisado. Fica evidente a tendência de superestimação do ICC alvo pelo estimador da variável latente além da maior dispersão do estimador comparado ao estimador das séries de Taylor e da log-normal. Destaca-se a menor variabilidade do estimador da série de Taylor frente aos demais nos cenários 3,6,9 e 12 onde a variância entre *clusters* é maior.

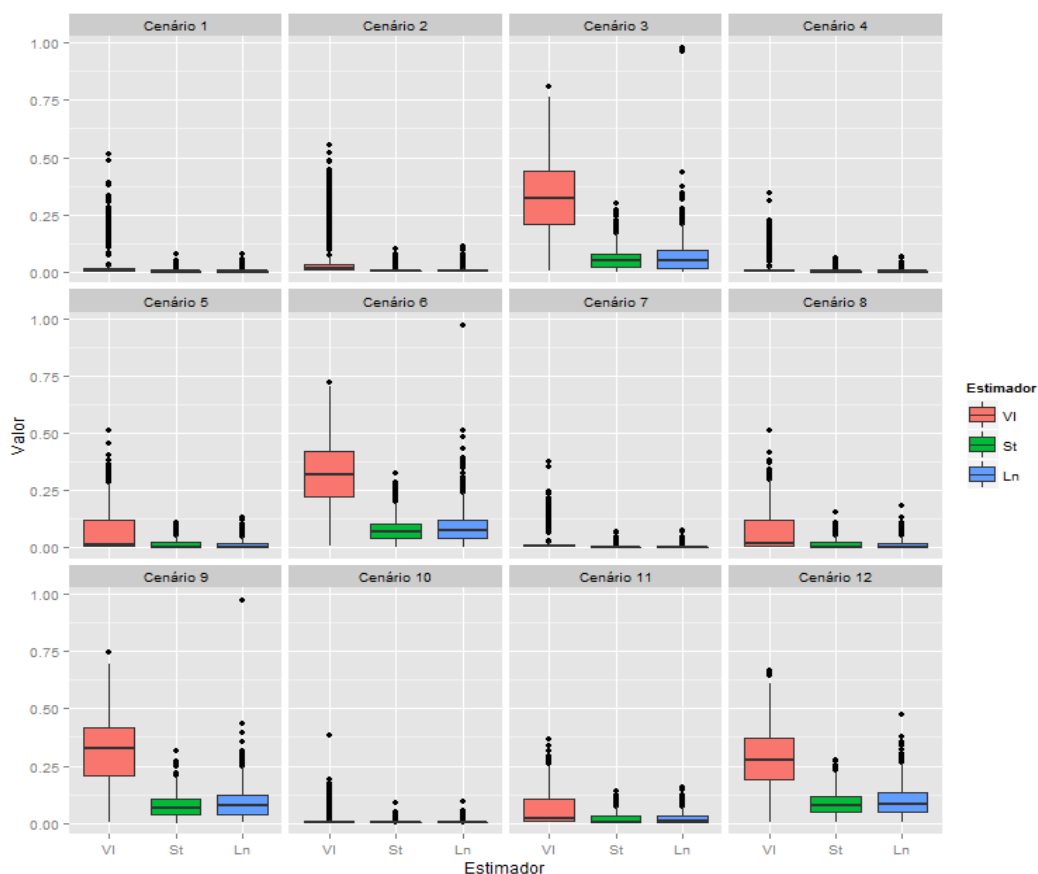


Figura 1 – Boxplots dos estimadores para cada cenário

## 5. Discussão

Neste trabalho, foi realizada uma comparação entre três estimadores do ICC do modelo log-binomial misto utilizando uma abordagem bayesiana para o ajuste do modelo. Destaca-se que os estimadores obtiveram performances distintas para os critérios considerados.

O estimador da séries de Taylor obteve os melhores resultados em relação aos demais. Foi o que apresentou maior precisão entre os três (menor erro quadrático médio), além de obter maior percentual de cobertura, menor amplitude dos intervalos de credibilidade. Destaca-se que para o percentual de observações que superestimaram o ICC alvo, este estimador teve performance equivalente ao estimador da log-normal, ficando acima do estimador da variável latente.

O estimador da log-normal obteve performance inferior ao da séries de Taylor. Mesmo com essa performance inferior, o estimador apresentou maior acurácia entre os três estimadores comparados além de ter empatado com o  $\rho_{st}$  no percentual de estimativas que superestimaram o ICC alvo em mais de 20%.

Por fim, o estimador da variável latente apresentou performance inferior aos demais. Obteve os piores resultados nos quesitos vício, erro quadrático médio, amplitude dos intervalos de credibilidade, percentual de cobertura e percentual de estimativas que superestimaram o ICC alvo em 20%. Contudo, foi o melhor estimador no quesito percentual de estimativas que subestimaram o ICC alvo em mais de 20%.

Foram observados alguns padrões para os resultados obtidos nos quesitos analisados. De maneira geral, sempre que se observou um aumento na variância entre *clusters*, a acurácia dos estimadores nos critérios analisados foi menor, exceto para o percentual de vezes em que os estimadores subestimam em mais de 20% do ICC alvo. Isto pode indicar que, nos três estimadores analisados, pode estar faltando a contabilização de algum fator.

Para futuros estudos que estejam interessados em estimar o ICC, recomenda-se que a metodologia de estimação a ser utilizada seja a metodologia da série de Taylor devido ao seu menor EQM em oito dos doze cenários comparados. Além do menor EQM, o percentual de cobertura do ICC alvo e a menor amplitude do estimador, reforçam sua escolha como estimador adequado para o ICC do modelo log-binomial misto.

Em seu trabalho, Yelland *et al* (2) conclui a comparação dos seus estimadores de maneira distinta, recomendando  $\rho_{ln}$ . Destaca-se que esta diferença de resultados pode ser devida ao menor número de cenários analisados, neste trabalho foram comparados doze cenários enquanto que em Yelland *et al* (2) comparam-se trinta e seis; a metodologia de estimação bayesiana utilizada para estimação dos parâmetros do modelo, pois as estimativas obtidas foram diferentes daquelas apresentadas em Yelland *et al* (2); e do fato de que Yelland *et al* (2) não utiliza o EQM como critério de comparação em seu trabalho.

Em relação a performance muito inferior do estimador da variável latente, levanta-se a hipótese de que se está comparando medidas de escalas distintas, pois no modelo log-binomial estima-se a variância associada ao fator de dependência de uma maneira direta, sendo que esta pertence a mesma escala de medida da função de ligação. Nota-se que ambos os estimadores da log-normal e da série de Taylor encontram-se na escala de probabilidades enquanto que o estimador da variável latente está na escala do logaritmo da probabilidade. Destaca-se que Wu *et al* (4) observou o mesmo comportamento para o estimador do ICC para o modelo logístico. Acredita-se que, com maneiras mais eficazes de se apurar o verdadeiro valor do ICC, essa discrepância de performance do estimador da variável latente seja sanada. Salienta-se que, mesmo com performance inferior para os casos analisados,  $\rho_{vl}$  apresenta embasamento teórico e por isto, não pode ser descartado.

Este trabalho apresenta pontos interessantes a serem considerados. Inicialmente, pela performance dos estimadores analisados, destaca-se que foi possível realizar uma recomendação clara de qual estimador utilizar, sendo que esta recomendação diverge com o que foi apresentado em Yelland *et al* (2). Além desta contribuição, o uso de uma abordagem bayesiana para a obtenção dos estimadores apresentou vantagens em pontos onde a abordagem frequentista falha, como no problema de convergência do algoritmo de estimação do modelo log-binomial misto. Também pelo uso de bayesiana, foram obtidas outras informações relevantes para as estatísticas calculadas. Cita-se os intervalos de credibilidade como o caso mais claro, informação que não foi obtida em Yelland *et al* (2). Como ponto final, ressalta-se que estudos futuros devem levar em consideração as metodologias aqui apresentadas e outras formas para a estimação do ICC.

## Agradecimentos

Faz-se um agradecimento à Dr. Lisa N. Yelland pela concessão dos códigos de geração das amostras.

## Referências Bibliográficas

1. Torman VBL, Camey SA. Bayesian models as a unified approach to estimate relative risk (or prevalence ratio) in binary and polytomous outcomes. *Emerg Themes Epidemiol* [Internet]. dezembro de 2015 [citado 17 de novembro de 2015];12(1). Recuperado de: <http://www.ete-online.com/content/12/1/8>
2. Yelland LN, Salter AB, Ryan P, Laurence CO. Adjusted intraclass correlation coefficients for binary data: methods and estimates from a cluster-randomized trial in primary care. *Clin Trials*. 1 de fevereiro de 2011;8(1):48–58.
3. Snijders TAB, Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. 2. ed. Los Angeles, Calif.: Sage; 2012. 354 p.
4. Wu S, Crespi CM, Wong WK. Comparison of methods for estimating the intraclass correlation coefficient for binary responses in cancer prevention cluster randomized trials. *Contemp Clin Trials*. setembro de 2012;33(5):869–80.
5. Pendergast JF, Gange SJ, Newton MA, Lindstrom MJ, Palta M, Fisher MR. A Survey of Methods for Analyzing Clustered Binary Response Data. *Int Stat Rev Rev Int Stat*. abril de 1996;64(1):89.
6. Ridout MS, Demétrio CG, Firth D. Estimating intraclass correlation for binary data. *Biometrics*. março de 1999;55(1):137–48.
7. Zou G, Donner A. Confidence interval estimation of the intraclass correlation coefficient for binary outcome data. *Biometrics*. setembro de 2004;60(3):807–11.
8. Evans BA, Feng Z, Peterson AV. A comparison of generalized linear mixed model procedures with estimating equations for variance and covariance parameter estimation in longitudinal studies and group randomized trials. *Stat Med*. 30 de novembro de 2001;20(22):3353–73.



9. Turner RM, Omar RZ, Thompson SG. Bayesian methods of analysis for cluster randomized trials with binary outcome data. *Stat Med.* 15 de fevereiro de 2001;20(3):453–72.
10. Campbell MK, Piaggio G, Elbourne DR, Altman DG, for the CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. *BMJ.* 4 de setembro de 2012;345(sep04 1):e5661–e5661.
11. Murray DM. Design and analysis of group-randomized trials. New York: Oxford University Press; 1998. 467 p.
12. Gilks WR, Richardson S, Spiegelhalter DJ, organizadores. Markov chain Monte Carlo in practice. Boca Raton, Fla: Chapman & Hall; 1998. 486 p.
13. Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *Am J Epidemiol.* 1 de abril de 2004;159(7):702–6.
14. Salmerón D, Cano JA, Chirlaque MD. Reducing Monte Carlo error in the Bayesian estimation of risk ratios using log-binomial regression models. *Stat Med.* 30 de agosto de 2015;34(19):2755–67.

# Anexos

## 1.Códigos

Geração das Cadeias

### Código do modelo reparametrizado

```
writeLines(
  "model{
    for (i in 1:N) {
      log(p[i]) <- theta0 * Z0[i] + theta1 * Z1[i] + theta2 * Z2[i] + u[cluster[i]]
      y[i] ~ dbern(p[i])
    }

    for(j in 1:nc1){
      u[j] ~ dnorm(0,tau)
    }

    theta0 ~ dnorm(0, 0.000001)
    theta1 ~ dnorm(0, 0.000001)
    theta2 ~ dnorm(0, 0.000001)

    sigma ~ dunif(0.01, 100)
    tau <- 1/(sigma*sigma)
    s2 <- sigma*sigma

    pb<-sum(p[])/N

    ICC <- s2/(s2+1) # ICC Variável Latente

    sigma2w<-pb*(1-pb)
    sigma2b1<-s2*pb*pb
    sigma2b2<-(exp(s2)-1)*pb*pb

    ICCY1<- sigma2b1/(sigma2b1+sigma2w)
    ICCY2<- sigma2b2/(sigma2b2+sigma2w)

    # Implementando as restrições para cada observação
    for (i in 1:N){
      ones[i] ~ dbern(C1[i])
      C1[i] <- step(1-p[i])
    }

  }", "model_yelland_reparam.txt")

# Função para analisar amostras através do JAGS
analise.jags = function(dados,n,nthin,niter,nburn,ncad,parameters){

  # Carregando pacotes necessários
  library(rjags)
  library(geepack)

  numero=dados[1,1]

  poigee <- geeglm(formula = outcome_clus ~ as.factor(treatment) + as.factor(bincov_clus),
    data = dados,
    family = poisson(link = "log"),
    id = cluster_id,
    corstr = "independence")

  S3 = poigee$geese$vbeta
```

```

X = model.matrix(poigee)
n = nrow(X)

Sigma <- S3
L <- chol(Sigma)
Z <- X%*%t(L)

dados.jags = list(N = n, ncl = 20, Z0 = Z[,1], Z1 = Z[,2], Z2 = Z[,3],
  y = dados$outcome_clus,
  cluster = dados$cluster_id)

th.ini1 = matrix(c(log(0.5),0,0),ncol=3)%*%solve(L)
th.ini2 = matrix(c(log(0.1),0,0),ncol=3)%*%solve(L)

inits.jags = list(
  list(theta0 = th.ini1[1], theta1 = th.ini1[2], theta2 = th.ini1[3], sigma=1, u=rep(-3,20)),
  list(theta0 = th.ini2[1], theta1 = th.ini2[2], theta2 = th.ini2[3], sigma=1, u=rep(-3,20))
)

m <- jags.model("model_yelland_reparam.txt", dados.jags, inits.jags, n.chains=ncad, n.adapt=(nburn/2), quiet=TRUE)
update(m, (nburn/2), progress.bar="none")
mcmc <- coda.samples(m, parameters, n.iter=niter,thin=nthin, progress.bar="none")

## Registrando estimativas de média, mediana e ic
## med = estimador media bayesiano, q50 = estimador mediana bayesiano, mod = estimador moda   ##bayesiano, bay
= ic bayesiano

sigma.med = summary(mcmc)$statistics['sigma','Mean']
sigma.q50 = summary(mcmc)$quantiles['sigma','50%']
ic1.sigma.bay = summary(mcmc)$quantiles['sigma','2.5%']
ic2.sigma.bay = summary(mcmc)$quantiles['sigma','97.5%']

ICC.med = summary(mcmc)$statistics['ICC','Mean']
ICC.q50 = summary(mcmc)$quantiles['ICC','50%']
ic1.ICC.bay = summary(mcmc)$quantiles['ICC','2.5%']
ic2.ICC.bay = summary(mcmc)$quantiles['ICC','97.5%']

ICCY1.med = summary(mcmc)$statistics['ICCY1','Mean']
ICCY1.q50 = summary(mcmc)$quantiles['ICCY1','50%']
ic1.ICCY1.bay = summary(mcmc)$quantiles['ICCY1','2.5%']
ic2.ICCY1.bay = summary(mcmc)$quantiles['ICCY1','97.5%']

ICCY2.med = summary(mcmc)$statistics['ICCY2','Mean']
ICCY2.q50 = summary(mcmc)$quantiles['ICCY2','50%']
ic1.ICCY2.bay = summary(mcmc)$quantiles['ICCY2','2.5%']
ic2.ICCY2.bay = summary(mcmc)$quantiles['ICCY2','97.5%']

## Registrando cadeias simuladas, para cálculo da moda
theta0sim = as.matrix(mcmc)[,'theta0']
theta1sim = as.matrix(mcmc)[,'theta1']
theta2sim = as.matrix(mcmc)[,'theta2']
sigmasim = as.matrix(mcmc)[,'sigma']
ICCsims = as.matrix(mcmc)[,'ICC']
ICCY1sim = as.matrix(mcmc)[,'ICCY1']
ICCY2sim = as.matrix(mcmc)[,'ICCY2']

tabelamcmc = data.frame(theta0sim, theta1sim, theta2sim, sigmasim, ICCsim, ICY1sim, ICY2sim)

eval(parse(text=paste('write.csv2(tabelamcmc, "mcmc_jags_dados_',numero, '.csv"), sep=")'))

## calculando estimativa da moda
kernel4 = density(sigmasim, bw="nrd0", n = 512, from=min(sigmasim), to=max(sigmasim))
posmo4 = which.max(kernel4$y)

```

```

sigma.mod = kernel4$x[posmo4]

kernel5 = density(ICCsim, bw="nrd0", n = 512, from=min(ICCsim), to=max(ICCsim))
posmo5 = which.max(kernel5$y)
ICC.mod = kernel5$x[posmo5]

kernel6 = density(ICC1sim, bw="nrd0", n = 512, from=min(ICC1sim), to=max(ICC1sim))
posmo6 = which.max(kernel6$y)
ICC1.mod = kernel6$x[posmo6]

kernel7 = density(ICC2sim, bw="nrd0", n = 512, from=min(ICC2sim), to=max(ICC2sim))
posmo7 = which.max(kernel7$y)
ICC2.mod = kernel7$x[posmo7]

# Estatística R
rhat = gelman.diag(mcmc)

# Tamanhos de amostra efetivos
effsize = effectiveSize(mcmc)

# Transformando os Thetas em Betas para obter os RR
theta.sim=cbind(tabelamcmc[,1], tabelamcmc[,2], tabelamcmc[,3])

beta.sim = theta.sim%%L
RR.sim = exp(beta.sim)
beta.sim<-as.data.frame(beta.sim)

is.atomic(beta.sim)
is.atomic(summary(beta.sim[,1]))

beta.asmcmc = as.mcmc.list(lapply(mcmc, function(x) as.mcmc(x[, -c(1:4)]%%L)))
effsizeBeta<-effectiveSize(beta.asmcmc)

beta0.med = mean(beta.sim[,1])
beta0.q50 = quantile(beta.sim[,1],0.5)
ic1.beta0.bay = quantile(beta.sim[,1],0.0275)
ic2.beta0.bay = quantile(beta.sim[,1],0.975)

beta1.med = mean(beta.sim[,2])
beta1.q50 = quantile(beta.sim[,2],0.5)
ic1.beta1.bay = quantile(beta.sim[,2],0.0275)
ic2.beta1.bay = quantile(beta.sim[,2],0.975)

beta2.med = mean(beta.sim[,3])
beta2.q50 = quantile(beta.sim[,3],0.5)
ic1.beta2.bay = quantile(beta.sim[,3],0.0275)
ic2.beta2.bay = quantile(beta.sim[,3],0.975)

RR1.med = mean(RR.sim[,1])
RR1.q50 = quantile(RR.sim[,1],0.5)
ic1.RR1.bay = quantile(RR.sim[,1],0.0275)
ic2.RR1.bay = quantile(RR.sim[,1],0.975)

RR2.med = mean(RR.sim[,2])
RR2.q50 = quantile(RR.sim[,2],0.5)
ic1.RR2.bay = quantile(RR.sim[,2],0.0275)
ic2.RR2.bay = quantile(RR.sim[,2],0.975)

RR3.med = mean(RR.sim[,3])
RR3.q50 = quantile(RR.sim[,3],0.5)
ic1.RR3.bay = quantile(RR.sim[,3],0.0275)
ic2.RR3.bay = quantile(RR.sim[,3],0.975)

```

```

## calculando estimativa da moda betas e RR's
kernel4 = density(RR.sim[,1], bw="nrd0", n = 400, from=min(RR.sim[,1]), to=max(RR.sim[,1]))
posmo4 = which.max(kernel4$y)
RR1.mod = kernel4$x[posmo4]

kernel5 = density(RR.sim[,2], bw="nrd0", n = 400, from=min(RR.sim[,2]), to=max(RR.sim[,2]))
posmo5 = which.max(kernel5$y)
RR2.mod = kernel5$x[posmo5]

kernel6 = density(RR.sim[,3], bw="nrd0", n = 400, from=min(RR.sim[,3]), to=max(RR.sim[,2]))
posmo6 = which.max(kernel6$y)
RR3.mod = kernel6$x[posmo6]

kernel7 = density(beta.sim[,1], bw="nrd0", n = 400, from=min(beta.sim[,1]), to=max(beta.sim[,1]))
posmo7 = which.max(kernel7$y)
beta0.mod = kernel7$x[posmo7]

kernel8 = density(beta.sim[,2], bw="nrd0", n = 400, from=min(beta.sim[,2]), to=max(beta.sim[,2]))
posmo8 = which.max(kernel8$y)
beta1.mod = kernel8$x[posmo8]

kernel9 = density(beta.sim[,3], bw="nrd0", n = 400, from=min(beta.sim[,3]), to=max(beta.sim[,3]))
posmo9 = which.max(kernel9$y)
beta2.mod = kernel9$x[posmo9]

# Saída da função
saida = list(sigma.med, sigma.q50, sigma.mod, ic1.sigma.bay, ic2.sigma.bay, #sigmasim,
            beta0.med, beta0.q50, beta0.mod, ic1.beta0.bay, ic2.beta0.bay, #beta0sim
            beta1.med, beta1.q50, beta1.mod, ic1.beta1.bay, ic2.beta1.bay, #beta1sim
            beta2.med, beta2.q50, beta2.mod, ic1.beta2.bay, ic2.beta2.bay, #beta2sim
            RR1.med, RR1.q50, RR1.mod, ic1.RR1.bay, ic2.RR1.bay, #RR1sim
            RR2.med, RR2.q50, RR2.mod, ic1.RR2.bay, ic2.RR2.bay, #RR2sim
            RR3.med, RR3.q50, RR3.mod, ic1.RR3.bay, ic2.RR3.bay, #RR3sim
            ICC.med, ICC.q50, ICC.mod, ic1.ICC.bay, ic2.ICC.bay, #ICCsims,
            ICCY1.med, ICCY1.q50, ICCY1.mod, ic1.ICCY1.bay, ic2.ICCY1.bay, # ICCY1sim, p1sim
            ICCY2.med, ICCY2.q50, ICCY2.mod, ic1.ICCY2.bay, ic2.ICCY2.bay, rhat, effsize, effsizeBeta) #ICCY2

names(saida) = c('sigma.med', 'sigma.q50', 'sigma.mod', 'ic1.sigma.bay', 'ic2.sigma.bay', '#rr2sim',
                'beta0.med', 'beta0.q50', 'beta0.mod', 'ic1.beta0.bay', 'ic2.beta0.bay', '#beta0sim',
                'beta1.med', 'beta1.q50', 'beta1.mod', 'ic1.beta1.bay', 'ic2.beta1.bay', '#beta1sim',
                'beta2.med', 'beta2.q50', 'beta2.mod', 'ic1.beta2.bay', 'ic2.beta2.bay', '#beta2sim',
                'RR1.med', 'RR1.q50', 'RR1.mod', 'ic1.RR1.bay', 'ic2.RR1.bay', '#RR1sim',
                'RR2.med', 'RR2.q50', 'RR2.mod', 'ic1.RR2.bay', 'ic2.RR2.bay', '#RR2sim',
                'RR3.med', 'RR3.q50', 'RR3.mod', 'ic1.RR3.bay', 'ic2.RR3.bay', '#RR3sim',
                'ICC.med', 'ICC.q50', 'ICC.mod', 'ic1.ICC.bay', 'ic2.ICC.bay', '#ICCsims',
                'ICCY1.med', 'ICCY1.q50', 'ICCY1.mod', 'ic1.ICCY1.bay', 'ic2.ICCY1.bay', # 'ICCY1sim', 'ICCY2sim',
                'ICCY2.med', 'ICCY2.q50', 'ICCY2.mod', 'ic1.ICCY2.bay', 'ic2.ICCY2.bay', 'rhat', 'effsize', 'effsizeBeta')

class(saida) = 'saidasim'
return(saida)
}

```