

**PROPOSTA E AVALIAÇÃO DE UM ESTIMADOR
ROBUSTO EM DADOS BINOMIAIS INFLACIONADOS
DE ZERO**

AUGUSTO MACIEL DA SILVA

2009

AUGUSTO MACIEL DA SILVA

**PROPOSTA E AVALIAÇÃO DE UM ESTIMADOR ROBUSTO EM
DADOS BINOMIAIS INFLACIONADOS DE ZERO**

Dissertação apresentada à Universidade Federal de Lavras como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Orientador

Prof. Dr. Marcelo Ângelo Cirillo

LAVRAS
MINAS GERAIS-BRASIL

2009

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Silva, Augusto Maciel da.

Proposta e avaliação de um estimador robusto em dados
binomiais inflacionados de zero / Augusto Maciel da Silva. –
Lavras : UFLA, 2009.

47 p. : il.

Dissertação (Mestrado) – Universidade Federal de Lavras, 2009.
Orientador: Marcelo Ângelo Cirillo.
Bibliografia.

1. Distribuição binomial. 2. Estimação robusta. 3. Simulação
Monte Carlo. 4. Inferência estatística 5. Modelo binomial
inflacionado de zeros I. Universidade Federal de Lavras. II. Título.

CDD – 519.54

AUGUSTO MACIEL DA SILVA

**PROPOSTA E AVALIAÇÃO DE UM ESTIMADOR ROBUSTO EM
DADOS BINOMIAIS INFLACIONADOS DE ZERO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

APROVADA em 28 de fevereiro de 2009.

Prof. Dra. Carla Regina Guimarães Brighenti UFSJ

Prof. Dr. Mário Javier Ferrua Vivanco UFLA

Prof. Dr. Renato Ribeiro de Lima UFLA

Prof. Dr. Marcelo Ângelo Cirillo
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

À memória de:

Meus avós, José Augusto de Oliveira e

Conceição Maria de Oliveira.

José Maciel da Silva e

Dolores Maria Ribeiro.

Ofereço.

À minha esposa Liliam, pela paciência e auxílio.

Aos meus pais, Bebel e Nenzinho, pelo apoio, carinho e dedicação.

Ao meu irmão André, pelo exemplo de força e determinação.

Dedico.

AGRADECIMENTOS

A Deus, por iluminar o meu caminho e colocar pessoas tão especiais ao meu lado.

À minha esposa, Liliam, pela força durante toda a caminhada rumo à vitória.

Aos meus pais, Nenzinho e Bebel, que não mediram esforços para que o sonho se concretizasse.

Ao meu irmão André, pessoa mais forte que eu conheço, exemplo de determinação.

À minha tia, Maria Efigênia Guimarães, parte essencial desta conquista.

Aos meus primos, Luiz Flávio Guimarães e Valéria Cristina Guimarães, por todos os momentos felizes que passamos juntos.

Aos meus "priminhos", Matheus, Paulinha e Gilbran, por tornarem tão prazerosos os últimos anos.

Aos meus sogros, José Roberto e Marilena, pelo carinho e força.

À minha cunhada, Cristiane, pela força e companheirismo.

À Universidade Federal de Lavras (UFLA), pela oportunidade do mestrado.

Ao meu orientador, Marcelo Ângelo Cirillo, pelo conhecimento compartilhado, pela compreensão e sobretudo pela amizade.

Ao professor Paulo Dattori (USP), pela enorme contribuição e auxílio no desenvolvimento deste trabalho.

A todos os professores do DEX.

A todos funcionários do DEX, em especial, Josi, Maria, Selminha e Edila, pela atenção e carinho dispensados aos alunos.

À CAPES, pela bolsa de estudo concedida.

Aos meus amigos de longa data, Carlos Eduardo, Leandro Ferreira, Giselle

Borges, Mariana Martins, Danilo Germano, pelos prazerosos momentos que passamos juntos.

Ao Igor Chalfoun pela amizade e horas de lazer em frente ao Wii.

À minha amiga, Ana Patrícia, por todos os momentos felizes, pelo estudo, atenção e dedicação durante todos esse tempo.

À minha "irmãzinha" de mestrado, Tania, pelo carinho, auxílio e paciência.

Ao meu amigo Paulo César, pela alegria constante, auxílio nos estudos, piadas e por tornar a vida muito mais simples.

Ao meu amigo Edcarlos, pelos momentos de estudo e também de lazer, essenciais para que tudo fosse perfeito.

A todos meus companheiros de mestrado, Ana Paula, Ana Patrícia, Altemir, Edcarlos, Hiron, Richardson, Tânia, Sthefânia, Denise, Ricardo e Paulo.

Augusto Maciel da Silva

SUMÁRIO

LISTA DE TABELAS	i
LISTA DE FIGURAS	ii
RESUMO	iii
ABSTRACT	iv
1 INTRODUÇÃO	1
2 REFERENCIAL TEÓRICO	3
2.1 Estimaco do parmetro π de uma distribuico binomial	3
2.2 Estimador de mxima verossimilhanca para a binomial.	5
2.3 Modelo binomial com excesso de zeros.	6
2.4 Introduco  inferncia robusta.	8
2.5 Estimadores robustos para o modelo binomial.	9
2.5.1 Estimadores-M	9
2.5.2 Estimadores de mnima disparidade	10
2.5.3 Estimadores-E	11
3 METODOLOGIA	13
3.1 Gerao das amostras binomiais	13
4 RESULTADOS E DISCUSSO	15
4.1 Desenvolvimento do estimador	15
4.2 Avaliaco do estimador proposto <i>Pzib</i> por simulaco Monte Carlo	21
4.3 Exemplo de aplicaco do estimador <i>Pzib</i>	34
5 CONCLUSES	38
REFERNCIAS BIBLIOGRFICAS	39
ANEXOS	41

LISTA DE TABELAS

1	Valores de c_1 para o estimador $Pzib$ com abordagem $\rho_1(x)$ que minimizam a diferença em relação ao parâmetro π com proporção de zeros $\gamma=0,2$, $m=100$ e c_2 fixado em 1	23
2	Valores de c_1 para o estimador $Pzib$ com abordagem $\rho_1(x)$ que minimizam a diferença em relação ao parâmetro π com proporção de zeros $\gamma=0,5$, $m=100$ e c_2 fixado em 1	25
3	Valores de c_1 para o estimador $Pzib$ com abordagem $\rho_1(x)$ que minimizam a diferença em relação ao parâmetro π com proporção de zeros $\gamma=0,7$, $m=100$ e c_2 fixado em 1	27
4	Valores de α para o estimador $Pzib$ com abordagem $\rho_2(x)$ que minimizam a diferença em relação ao parâmetro de referência π com proporção de zeros $\gamma=0,2$, $m=100$ e c_2 fixado em 1	29
5	Valores de α para os estimador $Pzib$ com abordagem $\rho_2(x)$ que minimizam a diferença em relação ao parâmetro de referência π com proporção de zeros $\gamma=0,5$, $m=100$ e c_2 fixado em 1	31
6	Valores de α para os estimador $Pzib$ com abordagem $\rho_2(x)$ que minimizam a diferença em relação ao parâmetro de referência π com proporção de zeros $\gamma=0,7$, $m=100$ e c_2 fixado em 1	32
7	Valores assumidos para ilustrar a estimação de π utilizando o estimador $Pzib$	34
8	Tabela de cálculos para $Pzib$ sob $\rho_2(x)$	37

LISTA DE FIGURAS

1	Gráfico das estimativas para $\pi = 0,3, \gamma = 0,2$	42
2	Gráfico das estimativas para $\pi = 0,5, \gamma = 0,2$	42
3	Gráfico das estimativas para $\pi = 0,8, \gamma = 0,2$	42
4	Gráfico das estimativas para $\pi = 0,3, \gamma = 0,5$	42
5	Gráfico das estimativas para $\pi = 0,5, \gamma = 0,5$	43
6	Gráfico das estimativas para $\pi = 0,8, \gamma = 0,5$	43
7	Gráfico das estimativas para $\pi = 0,3, \gamma = 0,7$	43
8	Gráfico das estimativas para $\pi = 0,5, \gamma = 0,7$	43
9	Gráfico das estimativas para $\pi = 0,8, \gamma = 0,7$	44

RESUMO

SILVA, Augusto Maciel da. **Proposta e Avaliação de um Estimador Robusto em Dados Binomiais Inflacionados de Zero**. 2009. 47p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras. *

Os métodos usuais de inferência quando aplicados em populações binomiais com excesso de zeros ou observações de outras populações estão sujeitos a "erros grosseiros" de estimação. Observado esse problema, realiza-se uma estimação robusta no sentido de minimizar esses erros. O presente trabalho tem por objetivo propor um novo estimador robusto, no sentido de que o mesmo seja capaz de estimar uma proporção binomial em amostras inflacionadas de zeros. Com esse enfoque, duas abordagens matemáticas na construção desse estimador foram propostas e avaliadas por simulação Monte Carlo, na qual diferentes configurações descritas pela combinação de valores paramétricos representados pelos fatores taxa de valores nulos presentes na amostra (γ); probabilidades de sucesso (π) e diferentes tamanhos de amostra foram simulados. Concluiu-se que o estimador *Pzib* proposto neste trabalho, nas duas abordagens matemáticas caracterizadas pelas funções $\rho_1(x)$ e $\rho_2(x)$ apresentou-se robusto nas situações em que as amostras apresentaram baixa ($\gamma=0,20$) e média ($\gamma=0,50$) concentrações de zero.

* **Orientador:** Marcelo Ângelo Cirillo - UFLA

ABSTRACT

SILVA, Augusto Maciel da. **Proposal and Evaluation of a Robust Estimator of Zero Inflated Binomial Data.** 2009. 47p. Dissertation (Master in Statistics and Agricultural Experimentation) Federal University of Lavras, Lavras.*

The usual methods of inference when applied to binomial populations with excess of zeros or observations of other populations are subject to "gross errors" of estimation. Observed this problem, there will be an estimation robust in order to minimize these errors. This study aims to propose a new robust estimator in that it is capable of estimating a binomial proportion on zero inflated samples. With this purpose, two mathematical approaches in the construction of this estimator have been proposed and evaluated by Monte Carlo simulation, in which different configurations described by the combination of parametric values represented by the rate of null values present in the sample (γ), probabilities of success (π) and different sample sizes were simulated. It was concluded that the *Pzib* estimator proposed in this work, on the two mathematical approaches characterized by the functions $\rho_1(x)$ and $\rho_2(x)$ is presented robust in situations where the samples had low ($\gamma=0.20$) and mean ($\gamma=0.50$) concentrations of zero.

* **Adviser:** Marcelo Ângelo Cirillo - UFLA.

1 INTRODUÇÃO

As inferências sobre o parâmetro da proporção de uma população binomial, em geral, são realizadas sobre a suposição de que as unidades amostrais sejam independentes e provenientes da mesma população. No entanto, ao tratar uma amostra contaminada, no sentido de que unidades amostrais são provenientes de outras populações, os estimadores usuais, como por exemplo, máxima verossimilhança e momentos, não podem ser utilizados, o que necessariamente motiva a pesquisa de outros métodos inferenciais, mais especificamente métodos robustos de estimação.

O estimador é dito robusto pela sua capacidade de considerar, em uma análise, a presença de desvios ou outliers em um determinado conjunto de dados, fornecendo assim uma estimativa condizente do parâmetro que se deseja estimar apesar da presença de dados discrepantes. Sendo assim, de grande utilidade, quando pressuposições exigidas por determinados modelos paramétricos não podem ser atendidas.

Estimadores robustos são estimadores que podem ser utilizados quando há a existência de desvios, superdispersão, outliers em um determinado conjunto de dados, fornecendo assim uma estimativa condizente do parâmetro que se deseja estimar apesar da presença de dados discrepantes.

Outra situação em que se faz necessária a realização de estimação robusta é quando a amostra apresenta um excesso de observações iguais a zero. Copas (1988) advertiu que, mesmo assumindo o modelo adequado, certo número de observações poderá vir a ser detectado como outlier, devido ao fato de que o valor da proporção " π " encontra-se próximo a 0 ou 1. O autor ressalta também que o método de estimação de máxima verossimilhança é sensível à presença de outliers

e sugere uma correção nas estimativas visto que diferentes modelos utilizados para resposta apresentam sensibilidade variada na detecção de outliers.

Segundo Ronchetti (2006), muitos procedimentos estatísticos são conhecidos pela sua não robustez, ou seja, resultados obtidos de dados que não seguem exatamente o modelo estatístico proposto não estariam representando bem a amostra ou a população dos quais vieram, sendo passível de utilização de métodos robustos para inferência.

São várias classes de estimadores robustos que podem ser utilizados para dados binomiais, por exemplo, os estimadores-M (Simpson, 1987), estimadores de mínima disparidade (Lindsay, 1994) e estimadores-E (Ruckstuhl & Welsh, 2001).

Os estimadores-E, como são citados por Ruckstuhl & Welsh (2001), foram construídos a partir de uma modificação na função de verossimilhança com a finalidade de reduzir o efeito das observações outliers na “cauda” da distribuição. Os autores afirmam que o desempenho do estimador-E depende de determinados valores assumidos pelas constantes de afinidade, referenciadas por c_1 e c_2 .

A motivação para realização deste trabalho se deu com o objetivo de propor e avaliar o comportamento de um estimador, denominado *Pzib*, utilizado para estimar a probabilidade de sucesso de uma população binomial inflacionada por zeros. As situações a que esses estimador será submetido serão feitas através de simulação Monte Carlo, e através das estimativas encontradas, será avaliada a sua robustez.

2 REFERENCIAL TEÓRICO

2.1 Estimação do parâmetro π de uma distribuição binomial

Uma distribuição binomial tem parâmetros (m, π) , em que m é o número de repetições de experimentos Bernoulli e π a probabilidade de sucessos. Assim, a distribuição binomial tem sua função de probabilidade dada por (Mood *et al.*, 1974):

$$f(Y = y) = \binom{m}{y} \pi^y (1 - \pi)^{m-y}, \quad y = 0, \dots, m \quad (2.1)$$

Algumas particularidades em relação ao modelo binomial merecem destaque, dentre elas, destaca-se que a distribuição binomial pertence à família exponencial, pois sua função de probabilidade pode ser descrita como:

$$f(y|\pi) = e^{c(\pi)T(y)+d(\pi)+S(y)}, \quad y \in A \quad (2.2)$$

em que c, d são funções reais de π ; T, S são funções reais de y e A não depende de π . Convém salientar que o fato da distribuição binomial poder ser representada na família exponencial, propicia estruturar uma grande variedade de modelos com a "mesma forma básica" e os mesmos procedimentos poderão ser usados para fazer inferências, pois todas as funções obtidas em 2.2 são matematicamente tratáveis. A título de ilustração, o modelo binomial (2.1) escrito na família exponencial é

dado por:

$$\begin{aligned} f(Y = y) &= \exp \left\{ \ln \binom{m}{y} + y \ln \pi + (m - y) \ln(1 - \pi) \right\} \\ &= \exp \left\{ \ln \left(\frac{\pi}{1 - \pi} \right)^y + m \ln(1 - \pi) + \ln \binom{m}{y} \right\} \end{aligned} \quad (2.3)$$

Contextualizando as propriedades de um estimador para o parâmetro π , Bolfarine & Sandoval, (2001) comentam que um bom estimador deve possuir as seguintes propriedades:

- não-viesado: ocorre quando a esperança é igual ao próprio valor do parâmetro que se deseja estimar;
- eficiência: ocorre quando a variância do estimador atinge o limite inferior da variância dos estimadores não-viesados. Vale salientar que os estimadores somente serão eficientes quando pertencerem à distribuições da família exponencial (Bolfarine & Sandoval, 2001). A eficiência de um estimador pode ser representada pelo quociente:

$$e(\hat{\pi}) = \frac{LI(\pi)}{Var[\hat{\pi}]} \quad (2.4)$$

em que LI é o limite inferior da variância dos estimadores não viciados de π . Salienta-se que quando $e(\hat{\pi}) = 1$, então $LI(\pi) = Var[\hat{\pi}]$.

- consistência: a idéia de consistência parte do pressuposto de que, à medida que o tamanho da amostra aumenta, os estimadores ficam tão próximos do

parâmetro que está sendo estimado. Seja Y_1, \dots, Y_n uma amostra aleatória da distribuição da variável aleatória Y , dependente do parâmetro π , então o estimador $\hat{\pi} = \hat{\pi}(Y_1, \dots, Y_n)$ é consistente para o parâmetro π , se,

$$\lim_{n \rightarrow \infty} P(|\hat{\pi} - \pi| > \varepsilon) = 0 \quad (2.5)$$

2.2 Estimador de máxima verossimilhança para a binomial.

Seja Y_1, \dots, Y_n variáveis aleatórias independentes e identicamente distribuídas de uma população binomial com função densidade de probabilidade dada por $f(y|\pi_1, \dots, \pi_n)$, a função de verossimilhança de acordo com Casella & Berger (1990) é definida por:

$$L(\pi|y) = L(\pi_1, \dots, \pi_n|y_1, \dots, y_n) = \prod_{i=1}^n f(y|\pi_1, \dots, \pi_n) \quad (2.6)$$

O estimador de máxima verossimilhança de π é o valor que maximiza a função dada em (2.6). Tomando a primeira derivada da função em relação ao parâmetro e igualando a zero, tem-se os pontos que são possíveis candidatos a máximo da função (2.6). Quando encontrada a expressão que a maximiza, encontra-se o estimador de máxima verossimilhança.

Por conveniência, tomando o logaritmo $L^*(\pi; y)$ a obtenção do estimador de máxima verossimilhança é matematicamente mais fácil de ser tratada, uma vez que o $\ln(L^*(\pi; y))$ é uma função crescente em y . Assim sendo, pode-se considerar $L^*(\pi; y)$ dado por

$$L^*(\pi; y) = \ln[L(\pi|y_1, \dots, y_n)] \quad (2.7)$$

Desse modo, para a distribuição binomial, tem-se:

$$\begin{aligned}
L^* &= \sum_{i=1}^n [y_i \ln(\pi) + (m - y_i) \ln(1 - \pi)] \\
L^* &= \ln(\pi) \sum_{i=1}^n y_i + \ln(1 - \pi) \sum_{i=1}^n (m - y_i) \\
\frac{dL^*}{d\pi} &= \left(\frac{1}{\pi}\right) \sum_{i=1}^n y_i - \left(\frac{1}{1 - \pi}\right) \sum_{i=1}^n (m - y_i) \\
\left(\sum_{i=1}^n y_i\right) (1 - \pi) &= \sum_{i=1}^n (m - y_i) \pi \\
\sum_{i=1}^n y_i - \pi \sum_{i=1}^n y_i &= nm\pi - \pi \sum_{i=1}^n y_i \\
\hat{\pi}_{emv} &= \frac{\sum_{i=1}^n y_i}{nm} = \frac{1}{m} \sum_{y=0}^m y f_n(y)
\end{aligned} \tag{2.8}$$

Obtem-se então o estimador de máxima verossimilhança para o parâmetro π da distribuição binomial, em que m representa o tamanho da amostra e $y = 0, \dots, m$. Ainda na expressão 2.8, pode se observar que $f_n(y)$ é obtido por:

$$f_n(y) = n^{-1} \sum_{i=1}^n I(Y_i = y), \quad y = 0, \dots, m \tag{2.9}$$

2.3 Modelo binomial com excesso de zeros.

Na prática, comumente é observada uma grande ocorrência de zeros em um conjunto de dados, sendo que nesses casos o número de zeros é maior que o predito, utilizando o modelo binomial básico. Esse fenômeno é chamado de “inflação de zeros”.

De acordo com Kemp & Kemp (1988), o estimador de máxima verossimilhança clássico de distribuições discretas tem muitas características desejáveis,

mas no caso da distribuição binomial, deve-se atentar para casos que não sigam exatamente essa distribuição, por exemplo, a ocorrência de superdispersão nos dados. Mesmo quando um método leva a estimadores explícitos para determinadas distribuições, estes podem não ser desejáveis. Estudos teóricos mostram que um método pode ser eficiente em determinada região do espaço paramétrico e ter comportamento contrário em outras regiões.

De acordo com Borgatto (2004), em caso de dados proporcionais, um modelo a ser utilizado é o modelo binomial inflacionado de zeros. Esse modelo então pode ser representado como uma distribuição de mistura de dois componentes, sendo que uma componente supõe que a ocorrência de zeros seja dada com probabilidade γ , enquanto que outra componente é caracterizada envolvendo uma distribuição binomial com probabilidade $(1-\gamma)$.

Assim o modelo ZIB (Zero Inflated Binomial) é representado por

$$f(Y = y) = \begin{cases} \gamma + (1 - \gamma)(1 - \pi)^m, & y = 0 \\ (1 - \gamma) \binom{m}{y} \pi^y (1 - \pi)^{m-y}, & y = 1, 2, \dots, m \end{cases} \quad (2.10)$$

com $E(Y) = (1 - \gamma)m_i\pi$ e $\text{Var}(Y) = \{[(1 - \gamma)m_i\pi][(1 - \pi)(1 - \gamma m)]\}$. O parâmetro γ , conforme dito anteriormente, representa a probabilidade de ocorrência de zeros, existindo a restrição $0 \leq \gamma < 1$. No caso de $\gamma > 0$, observa-se uma inflação na $\text{Var}(Y)$, ocorrendo assim superdispersão, devido ao excesso de zeros. Por outro lado, se $\gamma = 0$, $\text{Var}(Y)$ será exatamente a mesma de um modelo binomial padrão.

2.4 Introdução à inferência robusta.

Os principais métodos de inferência baseiam-se em modelos paramétricos. As inferências baseadas nesses modelos são condicionadas a determinadas pressuposições, as quais, quando violadas, comprometem a confiabilidade dos resultados. Nesse caso, métodos robustos são uma alternativa viável para solucionar esses problemas.

Segundo Hampel (1971), a inferência robusta trabalha justamente com essa questão, ou seja, preocupa-se com a construção de procedimentos que forneçam resultados confiáveis, em situações nas quais um modelo não esteja em conformidade com os dados que apresentam algum tipo de desvio, como por exemplo:

- arredondamento de observações;
- ocorrência de erros grosseiros;
- o modelo por si só poderá ser somente uma aproximação sobre o mecanismo de inferência. Ou seja, existem dados que não podem ser descritos pelo modelo paramétrico adotado para inferência.

Nesse mesmo contexto, o autor ressalta que modelos são apenas uma aproximação da realidade e quando não forem totalmente apropriados para descrever os dados, requerem mudanças na distribuição do estimador a ser utilizado. Isso será possível propondo certa condição de robustez ao modelo. Para tal, é preciso especificar os tipos de desvios que podem estar incorrendo.

Os chamados “erros grosseiros” podem ser ocasionados pela presença de outliers. De acordo com Bustos (1986), os métodos utilizados em análise que levam em conta a presença de outliers, por exemplo, as regras de rejeição, pertencem à classe denominada Estatística Robusta num sentido amplo. Assim, são apresentados alguns dos objetivos da estatística robusta:

- descrever uma estrutura que melhor se ajusta a uma massa de dados;
- identificar pontos que se desviam da maioria, que tenham grande influência sobre o restante do conjunto e propor métodos para análises.

Somente com o advento dos computadores, as técnicas robustas puderam ser estudadas mais profundamente, com ferramentas adequadas para descrever matematicamente esses problemas (Staudte & Sheather, 1990).

A teoria da robustez, aliada a métodos Monte-Carlo e métodos numéricos, auxilia no entendimento de problemas de natureza estatística e rapidez em sua solução (Bustos, 1986).

2.5 Estimadores robustos para o modelo binomial.

A literatura em estimação robusta para dados discretos apresenta inúmeras soluções para o modelo binomial. Ruckstuhl & Welsh (2001) discutiram o desempenho dos Estimadores-M e Estimadores de Mínima Disparidade, que podem ser aplicados ao modelo binomial. Estas classes de estimadores serão descritas resumidamente a seguir. Um outra classe, chamada Estimadores-E, também discutida pelos autores, será apresentada com mais detalhes, por ser fundamental para o entendimento do trabalho.

2.5.1 Estimadores-M

Os Estimadores-M, originalmente propostos por Huber (1964), para estimar parâmetros robustamente, têm sido aplicados com sucesso a uma grande variedade de problemas de estimação em que a estabilidade dos estimadores é fonte de constante preocupação. Um dos questionamentos que levou à criação dos Estimadores-M foi justamente o problema de se inferir sobre parâmetros de uma distribuição, assumindo normalidade, quando na verdade essa normalidade não existisse. Isso

resultaria em médias com baixa performance que poderiam causar variâncias errôneas.

Conforme exposto por Huber (1972), encontravam-se muitos problemas para estender esse método para modelar ou truncar determinados modelos de parâmetros. Com isso, precisou-se de estudos que criassem novos procedimentos que atendessem a situações de maior dificuldade.

Uma teoria de otimização dos Estimadores-M foi desenvolvida por Hampel (1971), para o caso uniparamétrico. Esse estudo facilitou a construção de estimadores robustos que apresentavam acurácia para determinados modelos, incluindo o binomial. O autor apontou ainda que eram necessários melhores estimadores para o caso de amostras pequenas e grandes.

2.5.2 Estimadores de mínima disparidade

Esta classe de estimadores, apresentada por Lindsay (1994), surgiu de um questionamento sobre robustez e eficiência. A questão era que alguns métodos poderiam ser robustos, mas não eficientes, por exemplo, o estimador de mínima distância de Hellinger (Simpson, 1987), que permitia pesos menores às observações mais discrepantes no modelo, tornando-o equivalente ao estimador de máxima verossimilhança em modelos conhecidos.

Em contrapartida, Lindsay (1994) mostrou que esse estimador, baseado em suas propriedades, possui uma curva de influência que não contempla algumas situações de robustez, afirmando que essa curva muitas vezes leva a situações de robustez enganosas.

Com isso, o autor propôs estudar outros métodos de distâncias mínimas para investigar esse fenômeno, utilizando a chamada “função de ajuste residual” cuja configuração determina completamente a robustez e eficiência de determinado es-

timador. A partir daí foi feito um estudo e chegou-se então ao Estimador de Mínima Disparidade, que é uma outra classe de estimadores que representa bem um modelo binomial sob condições adversas.

2.5.3 Estimadores-E

A idéia central desses estimadores é desenvolver uma metodologia para manipular uma possível contaminação grosseira, e consiste em modificar o log verossimilhança para reduzir o efeito das observações na cauda das distribuições (Ruckstuhl & Welsh, 2001). Entretanto, há também uma preocupação com os efeitos de uma contaminação distribuída que não é restrita somente na cauda e tende a deflacionar ou inflacionar várias classes de modos arbitrários. Isto sugere que é mais plausível trabalhar na frequência de escala do que na escala de observações individuais. O modo para decidir que classes controlar é relacionar as frequências relativas para a função de probabilidade do modelo binomial, um conhecimento que nos leva em direção da estimação de mínima distância na escala de frequência relativa (Ruckstuhl & Welsh, 2001).

A escolha da divergência ou disparidade é arbitrária e será confirmada pelas propriedades e desempenho do estimador. Faz sentido trabalhar com a disparidade que produza o estimador de máxima verossimilhança quando sob influência do modelo binomial.

É trabalhado a disparidade de verossimilhança $H(\pi, f_n)$:

$$H(\pi, f_n) = \sum_{y=0}^m \rho(x) p_{\pi}(y), \text{ em que } x = \frac{f_n(y)}{p_{\pi}(y)} \quad (2.11)$$

$f_n(y) = n^{-1} \sum_{i=1}^n I(Y_i = y)$, $y = 0, \dots, m$, é computada como a proporção das observações iguais a y em uma amostra de tamanho n , $p_{\pi}(y)$ a probabilidade con-

siderando uma proporção π e $\rho(x) = x \ln(x)$. A disparidade de verossimilhança é minimizada pelo estimador de máxima verossimilhança ($\hat{\pi}_{emv}$) para o modelo binomial e assume a mesma relação do log-verossimilhança como os mínimos quadrados para a máxima verossimilhança em modelos Gaussianos. Ruckstuhl & Welsh (2001) substitui $x \ln(x)$ por uma função linear que tem o efeito de reduzir a taxa em que $\rho(x)$ tende ao infinito e reduzir o efeito de classes para as quais $\frac{f_n(y)}{p_\pi(y)}$ é grande. O mesmo deve ser feito nas classes para as quais $\frac{f_n(y)}{p_\pi(y)}$ é pequeno para prevenir que $x \ln(x)$ tenda a zero quando x decresça para zero. Isto é, os autores substituíram $x \ln(x)$ por uma função linear quando x fosse menor que 1. Impondo a exigência que o ρ modificado tenha uma derivação contínua, levará à seguinte função:

$$\rho(x) = \begin{cases} (\ln(c_1) + 1)x - c_1, & \text{se } x < c_1 \\ x \ln(x), & \text{se } c_1 < x < c_2 \\ (\ln(c_2) + 1)x - c_2, & \text{se } x > c_2 \end{cases} \quad (2.12)$$

É importante salientar que a função $\rho(x)$ é garantida sobre determinados valores assumidos para constantes c_1 e c_2 e por recomendação de Ruckstuhl & Welsh (2001), utiliza-se $c_1 < c_2 = 1$. O Estimador $\hat{\pi}$ que minimiza H é dado por:

$$\hat{\pi} = \hat{\pi}(f) = \underset{\pi}{\operatorname{argmin}} H(\pi, f) \quad (2.13)$$

3 METODOLOGIA

3.1 Geração das amostras binomiais

Para a condução deste trabalho, as amostras inflacionadas de valores nulos foram geradas seguindo o modelo binomial inflacionado de zeros, conforme descrito na seção 2.3 (expressão 2.10).

Seguindo o modelo mencionado, os valores paramétricos assumidos no processo de simulação foram resultantes da combinação dos seguintes fatores: probabilidade de sucesso (π), fixada nos valores 0,3, 0,5 e 0,8, taxa de valores nulos (γ), representando a porcentagem de zeros, fixada em 0,2, 0,5 e 0,7 e por fim diferentes tamanhos amostrais (n), determinados pelos valores de 20, 30, 50, 70 e 90.

A principal motivação para a construção do estimador proposto neste trabalho para estimar a proporção de sucessos em uma amostra, proveniente de uma população binomial inflacionada de zeros se deve ao fato de que o estimador de máxima verossimilhança quando aplicado nesta situação apresenta "erros grosseiros". Em contrapartida, os estimadores robustos pertencente a classe dos estimadores-E apresentam uma certa sensibilidade em relação às constantes de afinidades assumidas. Além do mais, no que tange à sua praticidade de aplicação, em muitas situações torna-se inviável, uma vez que o argumento desse estimador que minimiza a função de disparidade requer que todo o campo paramétrico ($0 < \pi < 1$) seja "percorrido". O estimador de máxima verossimilhança é dado por:

$$\hat{\pi}_{emv} = 1/m \sum_{y=0}^m y f_n(y) \quad (3.1)$$

em que m representa o número de ensaios fixado em $m=100$, $f_n(y)$ a proporção de observações iguais a y em uma amostra de tamanho n , conforme a função

indicadora em (3.2).

$$f_n(y) = n^{-1} \sum_{i=1}^n I(Y = y_i), \quad y = 0, \dots, m \quad (3.2)$$

Assim sendo, a proposta de um novo estimador, cujo desenvolvimento encontra-se descrito na seção 4.1, em síntese envolve a combinação do estimador de máxima verossimilhança (3.1) com a função de disparidade (3.3), a qual caracteriza a classe dos estimadores-E, sendo essa definida por

$$H(\pi, f_n) = \sum_{y=0}^m \rho(x) p_\pi(y), \text{ em que } x = \frac{f_n(y)}{p_\pi(y)} \quad (3.3)$$

em que

$$\rho(x) = \begin{cases} (\ln(c_1) + 1)x - c_1, & \text{se } x < c_1 \\ x \ln(x), & \text{se } c_1 < x < c_2 \\ (\ln(c_2) + 1)x - c_2, & \text{se } x > c_2 \end{cases} \quad (3.4)$$

Em uma segunda abordagem a função ρ será modificada de tal forma, que a constante de afinidade c_1 , seja fixada em 0,1, dado $c_2=1$.

Como poderá ser observado na seção seguinte, o estimador, apresentado em duas abordagens, necessita do conhecimento de constantes que o torne robusto. Assim foram construídas tabelas e gráficos que representam a escolha correta das constantes para um número de ensaios Bernoulli fixado em $m=100$.

Convém salientar que para outros valores de m , para a obtenção das constantes que propiciam a robustez no estimador, encontra-se no Anexo B o programa construído pelo software R (R Development Core Team, 2008), com a utilização do pacote VGAM, necessário para geração das amostras com excesso de zero.

4 RESULTADOS E DISCUSSÃO

De acordo com a estrutura, esta seção é subdividida nos tópicos: 4.1 - Desenvolvimento do estimador; 4.2 - Avaliação do estimador proposto por simulação Monte Carlo.

4.1 Desenvolvimento do estimador

A motivação para o desenvolvimento do estimador robusto, proposto neste trabalho para proporções binomiais inflacionadas de zero, é dada na argumentação de que, na presença de excesso de zeros, o estimador de máxima verossimilhança para proporções binomiais (4.1) apresenta erros grosseiros (Ruckstuhl & Welsh, 2001).

$$\hat{\pi}_{emv} = \frac{1}{m} \sum_{y=0}^m y f_n(y) \text{ em que, } f_n(y) = \frac{1}{n} \sum_{i=1}^n I(Y = y_i) \quad (4.1)$$

e Y corresponde a cada unidade amostral, proveniente de um população binomial (m, π) inflacionadas de zeros. Para reduzir esse erro, uma alternativa já conhecida na literatura (Simpson, 1987, Lindsay, 1994) é dada pelo uso de um estimador pertencente à classe dos estimadores-E, o qual, em síntese, consiste em minimizar a função H (4.2) definida por:

$$H(\pi, f_n) = \sum_{y=0}^m \rho(x) p_{\pi(y)} \text{ em que, } x = \frac{f_n(y)}{p_{\pi(y)}} \text{ e } \rho(x) = x \ln(x) \quad (4.2)$$

Modificando a função $\rho(x)$ (4.2) impondo funções lineares nas extremidades,

definidas pelas constantes c_1 e c_2 , tem-se

$$\rho_1(x) = \begin{cases} (\ln(c_1) + 1)x - c_1 & \text{se } x < c_1 \\ x \ln(x) & \text{se } c_1 \leq x \leq c_2 \\ (\ln(c_2) + 1)x - c_2 & \text{se } x > c_2 \end{cases} \quad (4.3)$$

Em (4.3), deduz-se que a modificação é justificada pela redução na rapidez com que a função $\rho(x) = x \ln(x)$ tende a infinito quando x tende a infinito. De forma comparativa, pode-se observar, ao considerar $\rho(x) = ax + b$ com $a \neq 0$, podendo verificar através do limite (4.4)

$$\lim_{x \rightarrow \infty} \frac{x \ln x}{ax + b} \stackrel{L'H}{=} \lim_{x \rightarrow \infty} \frac{\ln x + 1}{a} = \infty \quad (4.4)$$

que de fato $x \ln x$ tende ao infinito mais rapidamente do que uma função linear.

O estimador que minimiza $H(\pi, f_n)$, substituindo $\rho(x) = x \ln(x)$ por $\rho_1(x)$, segundo Ruckstuhl & Welsh (2001) é dado por:

$$\hat{\pi} = \arg \min H(\pi) \text{ para } 0 \leq \pi \leq 1 \quad (4.5)$$

Em (4.5) nota-se que a obtenção de $\hat{\pi}$, em termos práticos e computacionais, depara-se com um problema no mecanismo gerador de dados, em virtude de que, a função $H(\pi, f_n)$, descrita em (4.2), deverá ser "bombardeada" com infinitos valores pertencentes ao intervalo $[0,1]$. Assim, entendemos que um problema de natureza contínua está sendo tratado de forma discretizada no processo de simulação, o que possivelmente, poderá acarretar em estimativas que minimizam a função $H(\pi, f_n)$ localmente, favorecendo a não obtenção de um mínimo global. Em consequência desse problema, nossa proposta consistiu em obter um estimador que seja robusto a amostras inflacionadas por zero e que contorne o problema mencio-

nado, propiciando uma fácil implementação computacional.

Uma particularidade, inerente ao estimador-E, é ressaltada por Ruckstuhl & Welsh (2001)) em relação aos valores das constantes c_1 e c_2 assumidos. Nesse contexto, os autores mencionam que, assumindo $c_1 = 0$ e $c_2 \rightarrow \infty$, $\hat{\pi}$ (4.5), resulta no mínimo relativo do estimador entropia, idêntico ao estimador de máxima verossimilhança do modelo binomial. Os autores recomendam que com $c_1 < c_2 = 1$ a robustez do estimador é melhorada. Entretanto, afirmam que um resultado inesperado é que sob algumas condições, o Estimador-E pode ser viesado. Para contornar esse problema, propõe uma função para substituir c_2 , que segundo os autores será discutida em trabalhos futuros.

Com essa motivação propõe-se o estimador *Pzib* (4.6) em uma primeira abordagem, considerando $\rho_1(x)$ (4.3)

$$Pzib = \sum_{y=0}^m \rho_1(x) p_{\hat{\pi}_{emv}}(Y = y) \quad (4.6)$$

Uma outra modificação na função $\rho(x) = x \ln(x)$, também foi proposta, de tal forma que o estimador *Pzib* foi construído em uma outra abordagem. Para um melhor entendimento, enunciam-se as seguintes definições:

Definição 1: Dizemos que uma função $f : X \subset \mathfrak{R} \rightarrow \mathfrak{R}$ é contínua em $x_0 \in X$ se dado $\xi > 0$, existir um $\delta > 0$ tal que $x \in X$, $|x - x_0| < \delta \Rightarrow |f(x) - f(x_0)| < \xi$. Se f é contínua para qualquer $x_0 \in X$, então dizemos que simplesmente f é contínua.

Exemplo 1: Considerando a função linear descrita em (4.7)

$$\begin{aligned} f : \mathfrak{R} &\rightarrow \mathfrak{R} \\ x &\rightarrow f(x) = ax + b \end{aligned} \quad (4.7)$$

de fato, pois se fixarmos $x_0 \in \mathfrak{R}$ e dado um $\xi > 0$ tem-se que $\delta = \frac{\xi}{|a|}$. Assim sendo,

resulta que

$$\begin{aligned} |x - x_0| < \delta &\Rightarrow |f(x) - f(x_0)| = \\ |ax + b - ax_0 - b| &= \\ |a(x - x_0)| = |a| \times |x - x_0| < |a| \times \frac{\xi}{|a|} = \xi \quad p/a \neq 0. \end{aligned}$$

Definição 2: dizemos que uma função $f : X \subset \mathfrak{R} \rightarrow \mathfrak{R}$ é uniformemente contínua em X se dado $\xi > 0$, existir um $\delta > 0$ tal que

$$x_0, x_1 \in X, |x_0 - x_1| < \delta \Rightarrow |f(x_0) - f(x_1)| < \xi$$

Procedem as seguintes observações:

- (1) $f(x)$ é uniformemente contínua em X, então $f(x)$ é contínua em X.
- (2) Estendendo a demonstração do exemplo 1, pode-se demonstrar que a função é uniformemente contínua.

Definição 3: dizemos que uma função $f : X \subset \mathfrak{R} \rightarrow \mathfrak{R}$ é Lspchitz contínua em X se existe um $L > 0$ tal que $|f(x_0) - f(x_1)| < L|x_0 - x_1|$, $\forall x_0, x_1 \in X$.

Procedem as seguintes observações:

- (1) A função do exemplo (1) é Lspchitz contínua com $L = |a|$
- (2) Um caso particular da função Lspchitz forma uma importante classe de funções uniformemente contínuas definida como a classe das funções Hölder contínuas (Definição 4).

Definição 4: dizemos que uma função $f : X \subset \mathfrak{R} \rightarrow \mathfrak{R}$ é Hölder contínua com expoente $0 < \alpha < 1$ se existir uma função $H > 0$ tal que

$$|f(x_0) - f(x_1)| < H |x_0 - x_1|^\alpha, \quad \forall x_0, x_1 \in X.$$

Exemplo 2: Considere a função (4.8)

$$\begin{aligned} f &: (0, +\infty) \subset \mathfrak{R} \rightarrow \mathfrak{R} \\ x &\rightarrow f(x) = \sqrt{x} \end{aligned} \tag{4.8}$$

Provaremos que a função descrita em (4.8) é Hölder contínua com expoente $\alpha = \frac{1}{2}$. Para isso, assumiremos que $x_0, x_1 \in (0, +\infty)$, logo podemos obter os seguintes resultados.

Para $x_0 > x_1$ tem-se:

$$\begin{aligned} \sqrt{x_0} &= \sqrt{x_0 - x_1 + x_1} \leq \sqrt{x_0 - x_1} + \sqrt{x_1} = \sqrt{|x_0 - x_1|} + \sqrt{x_1} \\ \rightarrow \sqrt{x_0} - \sqrt{x_1} &\leq \sqrt{|x_0 - x_1|} \end{aligned} \tag{*}$$

Para $x_1 > x_0$ tem-se:

$$\begin{aligned} \sqrt{x_1} &= \sqrt{x_1 - x_0 + x_0} \leq \sqrt{x_1 - x_0} + \sqrt{x_0} = \sqrt{|x_1 - x_0|} + \sqrt{x_0} \\ \rightarrow \sqrt{x_1} - \sqrt{x_0} &\leq \sqrt{|x_1 - x_0|} \end{aligned} \tag{**}$$

Em (*) e (**) conclui-se que

$$|\sqrt{x_0} - \sqrt{x_1}| \leq \sqrt{|x_0 - x_1|}, \text{ isto é } |f(x_0) - f(x_1)| \leq |x_0 - x_1|^{\frac{1}{2}}$$

De um modo geral, mostra-se que dado $0 < \alpha < 1$ e uma função definida em (4.9)

$$\begin{aligned} f &: (0, +\infty) \subset \mathfrak{R} \rightarrow \mathfrak{R} \\ x &\rightarrow f(x) = x^\alpha \end{aligned} \tag{4.9}$$

é do tipo Hölder contínua com expoente α . De acordo com esse resultado, pode-se enunciar uma importante propriedade.

Propriedade 1 - Dados $0 < \alpha < \beta \leq 1$ tem-se que x^α tende a infinito quando x tende a infinito "menos rapidamente" do que x^β . Nota-se que $\beta - \alpha > 0$, assim, escreve-se

$$\lim_{x \rightarrow \infty} \frac{x^\beta}{x^\alpha} = \lim_{x \rightarrow \infty} x^{\beta - \alpha} = \infty \quad (4.10)$$

Com a finalidade de diminuir a rapidez com que a função $\rho(x) = x \ln(x)$ tende a infinito quando x tende a infinito, a proposta do estimador *Pzib*, nessa segunda abordagem, implica em modificar a função $\rho(x)$, de modo a diminuir seu crescimento quando x tende a infinito. Assim, generalizando a função $\rho_1(x)$ (4.3) para uma função que satisfaça a propriedade 1, utilizou-se a função Hölder contínua definida em (4.11). Note que para $\alpha = 1$, $\rho_2(x)$ é equivalente a $\rho_1(x)$

$$\rho_2(x) = \begin{cases} \begin{cases} \left\{ c_1^{1-\alpha} \ln(c_1) + [(1-\alpha) \ln(c_1) + 1] \frac{c_1^{1-\alpha}}{\alpha} \right\} x^\alpha & \text{se } x < c_1 \\ -[(1-\alpha) \ln(c_1) + 1] \frac{c_1}{\alpha} & \end{cases} \\ \begin{cases} x \ln(x) & \text{se } c_1 \leq x \leq c_2 \\ \left\{ c_2^{1-\alpha} \ln(c_2) + [(1-\alpha) \ln(c_2) + 1] \frac{c_2^{1-\alpha}}{\alpha} \right\} x^\alpha & \text{se } x > c_2 \\ -[(1-\alpha) \ln(c_2) + 1] \frac{c_2}{\alpha} & \end{cases} \end{cases} \quad (4.11)$$

Chamamos a atenção que a função $\rho_2(x)$ tem a mesma classe de diferenciabilidade que a função $\rho_1(x)$, a saber, $\rho_1(x)$ e $\rho_2(x) \in C^1(\mathfrak{R}^+)$, isto é, são funções contínuas e a primeira derivada também é contínua. Logo, o estimador *Pzib* na segunda abordagem resultante é dado por

$$Pzib = \sum_{y=0}^m \rho_2(x) p_{\hat{\pi}_{emv}}(Y = y) \quad (4.12)$$

4.2 Avaliação do estimador proposto *Pzib* por simulação Monte Carlo

O estimador *Pzib* proposto nas duas abordagens matemáticas $\rho_1(x)$ (4.6) e $\rho_2(x)$ (4.11) difere apenas na rapidez de convergência quando $x \rightarrow \infty$. Neste contexto, suspeita-se que o conhecimento a priori do escalar α ($0 < \alpha < 1$) inerente a classe de funções Hölder (4.11), em função da escolha apropriada para os valores das constantes de afinidade c_1 e c_2 , poderá agregar resultados importantes em relação a questão relativa da "velocidade" de convergência. Dessa forma, por meio de um estudo empírico, mantendo-se as configurações representadas pelas combinações dos valores paramétricos mencionadas na Seção 3, foram obtidos os resultados através de Simulação Monte Carlo, na qual 5000 realizações experimentais do modelo binomial inflacionado (Seção 2.3), foram feitas com o intuito de gerar amostras e posteriormente aplicar o estimador *Pzib* para cada uma nas duas abordagens.

Isso exposto, considerando o estimador *Pzib* caracterizado pela função $\rho_1(x)$, para a situação em que a amostra representou uma taxa de zeros em média de 20%, 50% e 70% ($\gamma=0,20$, $\gamma=0,50$ e $\gamma = 0,70$), os resultados descritos nas Tabelas 1-3 foram obtidos de tal forma que o valor da constante c_1 pesquisado propiciasse um menor viés em relação ao parâmetro π inferido. Dado que a intenção é justamente reproduzir uma tabela de valores para α , quando *Pzib* é dado por $\rho_2(x)$ e outra tabela de valores de c_1 quando *Pzib* é dado por $\rho_1(x)$, nas quais o pesquisador possa utilizar o estimador *Pzib* em uma inferência estatística, tendo por base, a estimativa de máxima verossimilhança de π em uma amostra com excesso de zeros, julgou-se apropriado computar os desvios entre a estimativa π_{emv} e a estimativa de *Pzib* dado por $|Pzib - \pi_{emv}| < k$, em que k indica um valor tolerável para esta diferença.

Evidentemente, em situações reais esse desvio não é possível de ser calculado,

pois o valor paramétrico π não é conhecido. Porém, pelo fato de que o estimador de máxima verossimilhança usual submetido a amostras com excesso de zeros apresenta erros grosseiros (Ruckstuhl & Welsh, 2001), o pesquisador pode utilizá-lo como referência de consulta às tabelas, apresentadas no decorrer da discussão deste trabalho onde os desvios são apresentados.

Com esse enfoque, procede-se a discussão dos resultados, avaliados separadamente para cada taxa de excesso de zeros avaliada, isto é, $\gamma=0,20$, $\gamma=0,50$ e $\gamma = 0,70$, respectivamente descritos nas Tabelas 1, 2 e 3.

TABELA 1: Valores de c_1 para o estimador $Pzib$ com abordagem $\rho_1(x)$ que minimizam a diferença em relação ao parâmetro π com proporção de zeros $\gamma=0,2$, $m=100$ e c_2 fixado em 1

n		$\pi=0,3$	$\pi=0,5$	$\pi=0,8$
20	c_1	0,4700	0,3000	0,1100
	π_{emv}	0,2354	0,4013	0,6353
	$Pzib$	0,3334	0,5088	0,7975
	$ Pzib - \pi_{emv} $	0,0980	0,1075	0,1622
	Viés	0,1115	0,0177	-0,0031
30	c_1	0,4400	0,3000	0,1100
	π_{emv}	0,2429	0,4032	0,6372
	$Pzib$	0,2997	0,4842	0,7992
	$ Pzib - \pi_{emv} $	0,0568	0,0810	0,1620
	Viés	-0,0012	-0,0317	-0,0010
50	c_1	0,4400	0,2700	0,1100
	π_{emv}	0,2397	0,4025	0,6404
	$Pzib$	0,2927	0,4866	0,7933
	$ Pzib - \pi_{emv} $	0,0530	0,0841	0,1529
	Viés	-0,0242	-0,0269	-0,0084
70	c_1	0,4400	0,2600	0,1100
	π_{emv}	0,2384	0,3981	0,6393
	$Pzib$	0,2889	0,4921	0,8008
	$ Pzib - \pi_{emv} $	0,0505	0,0940	0,1615
	Viés	-0,0369	-0,0157	0,0010
90	c_1	0,4000	0,2500	0,1100
	π_{emv}	0,2409	0,4008	0,6371
	$Pzib$	0,2858	0,4819	0,8119
	$ Pzib - \pi_{emv} $	0,0449	0,0811	0,1748
	Viés	-0,0472	-0,0361	0,0148

Viés calculado em relação ao estimador $Pzib$ com precisão até a quarta casa decimal.

Com base nos resultados encontrados na Tabela 1, pode-se verificar que de um modo geral, os valores determinados na constante c_1 , mantendo fixo $c_2=1$ são recomendáveis para utilização no estimador $Pzib$, considerando $\rho_1(x)$. Tal recomendação é justificada pelos baixos valores do viés obtido em cada configuração paramétrica. Contudo, no que tange a praticidade em utilizar esta Tabela, pode-se observar por meio dos resultados específicos ao cálculo do desvio $|Pzib - \pi_{emv}| < k$, no qual considera-se que $k=0,15$ é um valor adequado para esse desvio. Convém salientar que este valor de 0,15 foi definido arbitrariamente, o que sugere que o pesquisador poderá procurar por melhores valores de c_1 utilizando o programa de simulação (Anexo B) de forma a propiciar estimativas mais refinadas.

TABELA 2: Valores de c_1 para o estimador $Pzib$ com abordagem $\rho_1(x)$ que minimizam a diferença em relação ao parâmetro π com proporção de zeros $\gamma=0,5$, $m=100$ e c_2 fixado em 1

n		$\pi=0,3$	$\pi=0,5$	$\pi=0,8$
20	c_1	0,6900	0,5000	0,2000
	π_{emv}	0,1496	0,2637	0,3810
	$Pzib$	0,2971	0,4923	0,8000
	$ Pzib - \pi_{emv} $	0,1475	0,2286	0,4190
	Viés	-0,0097	-0,0154	0,0000
30	c_1	0,6900	0,5000	0,2000
	π_{emv}	0,1513	0,2497	0,4018
	$Pzib$	0,2944	0,4974	0,8000
	$ Pzib - \pi_{emv} $	0,1431	0,2477	0,3982
	Viés	-0,0186	-0,0051	0,0000
50	c_1	0,6900	0,5000	0,2000
	π_{emv}	0,1543	0,2533	0,3870
	$Pzib$	0,2928	0,4970	0,8000
	$ Pzib - \pi_{emv} $	0,1385	0,2437	0,4130
	Viés	-0,0241	-0,0060	0,0000
70	c_1	0,6900	0,5000	0,2000
	π_{emv}	0,1494	0,2478	0,4027
	$Pzib$	0,2950	0,4977	0,8000
	$ Pzib - \pi_{emv} $	0,1456	0,2499	0,3973
	Viés	-0,0168	-0,0047	0,0000
90	c_1	0,6900	0,5000	0,2000
	π_{emv}	0,1514	0,2498	0,4017
	$Pzib$	0,2951	0,4975	0,8000
	$ Pzib - \pi_{emv} $	0,1437	0,2477	0,3983
	Viés	-0,0163	-0,0050	0,0000

Viés calculado em relação ao estimador $Pzib$ com precisão até a quarta casa decimal.

Seguindo a mesma discussão na pesquisa dos valores a serem assumidos para c_1 , porém diferenciando na proporção de valores zeros iguais a 50 %, contidos na amostra, os resultados encontrados na Tabela 2 evidenciaram que para todos os tamanhos amostrais avaliados, dentro da diferença tolerável ($k=0,15$), verificou-se que para esta concentração de zero, as estimativas de π_{emv} são condizentes com o valor de $\pi=0,30$ e portanto os valores apropriados de c_1 são confiáveis e plausíveis de serem utilizados com um indicativo para consulta.

Para as demais situações, $\pi=0,5$ e $\pi=0,8$, embora os valores do viés sejam relativamente baixos, o uso da estimativa π_{emv} como indicativo para seleção destas constantes, poderá acarretar dúvida em virtude dos altos valores dos desvios $|Pzib - \pi_{emv}|$ serem superiores a 0,15. Portanto, neste sentido recomenda-se que o pesquisador tenha uma certa cautela para amostras que apresentam taxa de excesso de zeros próxima a 50% ($\gamma=0,50$).

Aumentando o valor da taxa de zeros para $\gamma=0,70$ conforme é observado nos resultados encontrados na Tabela 3, estende-se a mesma discussão e a mesma cautela incluindo os resultados observados para $\pi=0,3$.

TABELA 3: Valores de c_1 para o estimador $Pzib$ com abordagem $\rho_1(x)$ que minimizam a diferença em relação ao parâmetro π com proporção de zeros $\gamma=0,7$, $m=100$ e c_2 fixado em 1

n		$\pi=0,3$	$\pi=0,5$	$\pi=0,8$
20	c_1	0,7000	0,5000	0,2000
	π_{emv}	0,0929	0,1622	0,2428
	$Pzib$	0,2987	0,4999	0,7995
	$ Pzib - \pi_{emv} $	0,2058	0,3377	0,5567
	Viés	-0,0045	-0,0002	-0,0007
30	c_1	0,7000	0,5000	0,2000
	π_{emv}	0,0898	0,1522	0,2441
	$Pzib$	0,2989	0,5000	0,8000
	$ Pzib - \pi_{emv} $	0,2091	0,3478	0,5559
	Viés	-0,0038	-0,0001	0,000
50	c_1	0,7000	0,5000	0,2000
	π_{emv}	0,0857	0,1571	0,2392
	$Pzib$	0,2994	0,5000	0,8000
	$ Pzib - \pi_{emv} $	0,2137	0,3429	0,5608
	Viés	-0,0019	-0,0001	0,000
70	c_1	0,7000	0,5000	0,2000
	π_{emv}	0,0915	0,1475	0,2448
	$Pzib$	0,2992	0,5000	0,8000
	$ Pzib - \pi_{emv} $	0,2077	0,3525	0,5552
	Viés	-0,0025	0,000	0,000
90	c_1	0,7000	0,5000	0,2000
	π_{emv}	0,0905	0,1460	0,2390
	$Pzib$	0,2995	0,5000	0,8000
	$ Pzib - \pi_{emv} $	0,2090	0,3540	0,5610
	Viés	-0,0017	0,000	0,000

Viés calculado em relação ao estimador $Pzib$ com precisão até a quarta casa decimal.

Chama-se a atenção que o uso do estimador *Pzib* incorporando a função $\rho_1(x)$ em determinadas situações apresentou estimativas coerentes com o valor paramétrico (π). Contudo, vale ressaltar, que em cada situação o valor da constante c_1 foi diferenciado.

Partindo de uma situação na qual, a robustez do estimador *Pzib*, seja verificada para apenas um único valor de c_1 , mantendo $c_2=1$, como nos casos anteriores, bem como, o estudo da função $\rho_2(x)$ em relação a velocidade de convergência quando $x \rightarrow \infty$ os resultados encontrados nas Tabelas 4-6 corresponderam às estimativas de máxima verossimilhança, *Pzib* e desvios.

Em virtude do que foi mencionado, preliminarmente, este estudo foi realizado fixando $c_1=0,1$. Entretanto, chama-se a atenção que as estimativas robustas para amostras inflacionadas de zeros, com o uso do recurso computacional do programa disponível no Anexo B poderão ser obtidas para qualquer valor assumido em c_1 obedecendo $c_1 < c_2 = 1$ e a restrição $0 < \alpha < 1$. Nota-se que $\alpha=1$ resulta na função $\rho_1(x)$, cujas propriedades em relação a velocidade de convergência já foram discutidas no desenvolvimento. Para quaisquer valores de α , dado a restrição de α , $\rho_2(x)$ caracteriza-se como uma função pertencente à classe de funções Hölder.

Isso exposto, procedeu-se com a discussão dos seguintes resultados.

TABELA 4: Valores de α para o estimador $Pzib$ com abordagem $\rho_2(x)$ que minimizam a diferença em relação ao parâmetro de referência π com proporção de zeros $\gamma=0,2$, $m=100$ e c_2 fixado em 1

n		$\pi=0,3$	$\pi=0,5$	$\pi=0,8$
20	α	0,2600	0,1700	0,1200
	π_{emv}	0,2374	0,3958	0,6515
	$Pzib$	0,2897	0,5268	0,8106
	$ Pzib - \pi_{emv} $	0,0523	0,1310	0,1591
30	α	0,1900	0,1600	0,1200
	π_{emv}	0,2418	0,3986	0,6479
	$Pzib$	0,2827	0,4955	0,8118
	$ Pzib - \pi_{emv} $	0,0409	0,0968	0,1639
50	α	0,1300	0,1300	0,1200
	π_{emv}	0,2382	0,3990	0,6336
	$Pzib$	0,3063	0,5124	0,8156
	$ Pzib - \pi_{emv} $	0,0681	0,1135	0,1819
70	α	0,1000	0,1200	0,1200
	π_{emv}	0,2411	0,4027	0,6419
	$Pzib$	0,2957	0,4621	0,7958
	$ Pzib - \pi_{emv} $	0,0546	0,0594	0,1540
90	α	0,0900	0,1100	0,1200
	π_{emv}	0,2387	0,3993	0,6387
	$Pzib$	0,2985	0,4895	0,7803
	$ Pzib - \pi_{emv} $	0,0598	0,0902	0,1417

Esse estudo foi realizado com o intuito de verificar se com o conhecimento de π_{emv} de cada situação, a escolha de α poderia ser realizada corretamente. Esse resultado foi comprovado para o caso representado na Tabela 4, visto que as diferenças não apresentaram valores altos, considerados de um modo geral maiores que 0,15. Como o estimador de máxima verossimilhança pode ser obtido na amostra, uma vez de posse de sua estimativa, pode-se concluir sobre o valor correto do parâmetro de referência π utilizado e conseqüentemente utilizar a constante α adequada que resultaria em estimativas robustas, como por exemplo nas situações onde a estimativa seja próxima de 0,30. Esse resultado, garante que para baixa concentração de zeros ($\gamma=0,20$), a escolha de α para o cálculo da estimativa, poderia ser realizada observando o valor correto na tabela e aplicado no estimador *Pzib*, garantindo assim, estimativas robustas e acuradas.

Em se tratando de médias concentrações de zero (Tabela 5), sobre o critério de que $|Pzib - \pi_{emv}| < 0,15$, recomenda-se o uso do *Pzib* com a função $\rho_2(x)$ apenas para baixas proporções, isto é, $\pi=0,3$.

TABELA 5: Valores de α para os estimador $Pzib$ com abordagem $\rho_2(x)$ que minimizam a diferença em relação ao parâmetro de referência π com proporção de zeros $\gamma=0,5$, $m=100$ e c_2 fixado em 1

n		$\pi=0,3$	$\pi=0,5$	$\pi=0,8$
20	α	0,2700	0,1800	0,1300
	π_{emv}	0,1521	0,2537	0,3895
	$Pzib$	0,2800	0,5012	0,7720
	$ Pzib - \pi_{emv} $	0,1279	0,2475	0,3825
30	α	0,2600	0,1800	0,1300
	π_{emv}	0,1471	0,2442	0,4008
	$Pzib$	0,2925	0,5001	0,7720
	$ Pzib - \pi_{emv} $	0,1454	0,2558	0,3711
50	α	0,2500	0,1800	0,1300
	π_{emv}	0,1492	0,2527	0,3922
	$Pzib$	0,3053	0,4979	0,7717
	$ Pzib - \pi_{emv} $	0,1562	0,2452	0,3795
70	α	0,2500	0,1800	0,1300
	π_{emv}	0,1494	0,2537	0,4084
	$Pzib$	0,2922	0,4986	0,7718
	$ Pzib - \pi_{emv} $	0,1428	0,2449	0,3633
90	α	0,2400	0,1800	0,1300
	π_{emv}	0,1478	0,2452	0,3981
	$Pzib$	0,3068	0,4969	0,7717
	$ Pzib - \pi_{emv} $	0,1590	0,2517	0,3736

Em relação a altas concentrações de zero ($\gamma=0,70$) os resultados encontrados na Tabela 6 evidenciaram que o estimador $Pzib$, mesmo incorporando a função $\rho_2(x)$, não apresentaram resultados satisfatórios pelo critério estabelecido neste trabalho $|Pzib-\pi_{emv}|<0,15$.

TABELA 6: Valores de α para os estimador $Pzib$ com abordagem $\rho_2(x)$ que minimizam a diferença em relação ao parâmetro de referência π com proporção de zeros $\gamma=0,7$, $m=100$ e c_2 fixado em 1

n		$\pi=0,3$	$\pi=0,5$	$\pi=0,8$
20	α	0,2600	0,1800	0,1300
	π_{emv}	0,0949	0,1484	0,2485
	$Pzib$	0,2935	0,4995	0,7725
	$ Pzib - \pi_{emv} $	0,1987	0,3511	0,5240
30	α	0,2600	0,1800	0,1300
	π_{emv}	0,0900	0,1510	0,2343
	$Pzib$	0,2843	0,4945	0,7721
	$ Pzib - \pi_{emv} $	0,1943	0,3435	0,5378
50	α	0,2500	0,1800	0,1300
	π_{emv}	0,0882	0,1452	0,2367
	$Pzib$	0,3021	0,4936	0,7717
	$ Pzib - \pi_{emv} $	0,2139	0,3484	0,5350
70	α	0,2500	0,1800	0,1300
	π_{emv}	0,0890	0,1500	0,2468
	$Pzib$	0,2991	0,4936	0,7717
	$ Pzib - \pi_{emv} $	0,2101	0,3436	0,5250
90	α	0,2500	0,1800	0,1300
	π_{emv}	0,0885	0,1506	0,2349
	$Pzib$	0,2981	0,4936	0,7717
	$ Pzib - \pi_{emv} $	0,2096	0,3430	0,5369

Para um melhor estudo do desempenho deste estimador considerando $\rho_2(x)$ em relação a outros tamanhos amostrais e viabilidade de utilizar π_{emv} como um indicativo para pesquisa do valor de α que garanta a robustez poderão ser observados nas Figuras 1-9, nas quais o viés poderá ser observado, dispostas no Anexo A.

4.3 Exemplo de aplicação do estimador *Pzib*

Para fins didáticos, ilustra-se a aplicação do estimador *Pzib* considerando a função $\rho_2(x)$ em uma amostra de tamanho n , na qual cada unidade amostral foi considerada independente e identicamente distribuída por uma binomial (m, π) . Dada esta especificação e para efeito de comparação simulou-se uma amostra, proveniente de um modelo binomial inflacionado com $m=100$ e $\pi=0,3$.

TABELA 7: Valores assumidos para ilustrar a estimação de π utilizando o estimador *Pzib*

Parâmetros	m	n	γ	Unidades Amostrais
$\pi = 0,3$	100	20	0,20	25, 30, 31, 25, 24, 27, 0, 26, 31, 32, 26, 28, 21, 27, 29, 30, 0, 0, 28, 32

Importante ressaltar que pelo fato de utilizar a função $\rho_2(x)$, fixou-se $c_1=0,1$ e $c_2=1$. Entretanto o valor de α adequado foi consultado na Tabela 3 que foi gerada através da rotina do Anexo B. Com essa informação, escolheu-se $\alpha=0,26$. A seguir são apresentados os passos para obtenção do estimador *Pzib*.

Primeiro passo: obtenção da estimativa de máxima verossimilhança:

$$\hat{\pi}_{emv} = 1/m \sum_{y=0}^m y f_n(y)$$

em que

$$f_n(y) = n^{-1} \sum_{i=1}^n I(Y = y_i), y = 0, \dots, m$$

$$f_n(0) = \frac{I(Y = 0) + I(Y = 0) + \cdots + I(Y = 0)}{20} = \frac{3}{20} = 0,15$$

$$\vdots$$

$$f_n(0) = \frac{I(Y = 0) + I(Y = 0) + \cdots + I(Y = 0)}{20} = \frac{2}{20} = 0,10$$

Lembrando que o mesmo procedimento deve ser feito para cada y , observando que para o caso em questão, só teremos valores não-nulos quando y for igual às unidades amostrais. Com isso pode-se calcular o estimador de máxima verossimilhança

$$\hat{\pi}_{emv} = \frac{0f_n(0) + 1f_n(1) + \cdots + 100f_n(100)}{100} = 0,236$$

Com base nessa estimativa, pode-se utilizar a tabela, procurando por um valor de π_{emv} próximo ao encontrado. Pelas unidades amostrais geradas, tem-se por base $\gamma=0,20$, o que corresponde a Tabela 4. Obedecendo a regra de que os desvios devem ser menor que 0,15, tem-se uma idéia da estimativa do parâmetro, que poderia ser 0,30.

Segundo Passo: Cálculo das probabilidades considerando o EMV

$$\begin{aligned}
 p_{\pi_{emv}}(0) &= \binom{100}{0} 0,236^0 (1 - 0,236)^{100-0} = 2,04 \times 10^{-12} \\
 p_{\pi_{emv}}(1) &= \binom{100}{1} 0,236^1 (1 - 0,236)^{100-1} = 6,30 \times 10^{-11} \\
 p_{\pi_{emv}}(2) &= \binom{100}{2} 0,236^2 (1 - 0,236)^{100-2} = 9,63 \times 10^{-10} \\
 &\vdots \\
 p_{\pi_{emv}}(100) &= \binom{100}{100} 0,236^{100} (1 - 0,236)^{100-100} = 1,96 \times 10^{-63}
 \end{aligned}$$

Terceiro Passo: Cálculo do Estimador *Pzib* utilizando $\rho_2(x)$

$$Pzib = \sum_{y=0}^m \rho^*(x) p_{\hat{\pi}_{MLE}}(Y = y) \mathbf{e} x = \frac{f_n(y)}{p_{\hat{\pi}_{emv}}(y)}$$

em que:

$$\rho_2(x) = \begin{cases} \left\{ \begin{array}{l} \left\{ c_1^{1-\alpha} \ln(c_1) + [(1-\alpha) \ln(c_1) + 1] \frac{c_1^{1-\alpha}}{\alpha} \right\} x^\alpha \\ - [(1-\alpha) \ln(c_1) + 1] \frac{c_1}{\alpha} \end{array} \right. & \text{se } x < c_1 \\ \left\{ \begin{array}{l} x \ln(x) \end{array} \right. & \text{se } c_1 \leq x \leq c_2 \\ \left\{ \begin{array}{l} \left\{ c_2^{1-\alpha} \ln(c_2) + [(1-\alpha) \ln(c_2) + 1] \frac{c_2^{1-\alpha}}{\alpha} \right\} x^\alpha \\ - [(1-\alpha) \ln(c_2) + 1] \frac{c_2}{\alpha} \end{array} \right. & \text{se } x > c_2 \end{cases}$$

TABELA 8: Tabela de cálculos para $Pzib$ sob $\rho_2(x)$

m	$p_{\hat{\pi}_{emv}}$	x	$\rho_2(x)$	$\rho_2(x)p_{\hat{\pi}_{emv}}$
0	$2,0386 \times 10^{-12}$	$7,3579 \times 10^{10}$	2568,8177	$5,2368 \times 10^{09}$
1	$6,2970 \times 10^{-11}$	0	0,2707	$1,7049 \times 10^{11}$
2	$9,6280 \times 10^{-10}$	0	0,2707	$2,6069 \times 10^{10}$
3	$9,7163 \times 10^{-09}$	0	0,2707	$2,6305 \times 10^{09}$
4	$7,2783 \times 10^{-08}$	0	0,2707	$1,9705 \times 10^{08}$
5	$4,3167 \times 10^{-07}$	0	0,2707	$1,1686 \times 10^{07}$
⋮	⋮	⋮	⋮	⋮
21	0,0804	0,6214	-0,2956	-0,0237
⋮	⋮	⋮	⋮	⋮
100	$1,9552 \times 10^{-63}$	0	0,2707	$5,2935 \times 10^{-64}$
$Pzib = \sum_{y=0}^m \rho_2(x)p_{\hat{\pi}_{MLE}}$				0,2900

Em se tratando do $Pzib$ considerando $\rho_1(x)$ o procedimento é análogo, assumindo $\alpha=1$. Entretanto, o valor de c_1 deverá ser escolhido de forma apropriada.

Comparando a estimativa de $Pzib$, com o valor paramétrico ($\pi = 0,30$), nota-se que o valor de α utilizado resultou em uma estimativa acurada, pois utilizando o critério estabelecido por $|Pzib - \pi_{emv}| < 0,15$ pode-se verificar que o valor α é adequado para a realização desta inferência.

5 CONCLUSÕES

O estimador *Pzib* proposto neste trabalho, nas duas abordagens matemáticas caracterizadas pelas funções $\rho_1(x)$ e $\rho_2(x)$ apresentou-se robusto nas situações em que as amostras apresentaram baixa ($\gamma=0,20$) e média ($\gamma=0,50$) concentrações de zero.

O programa (Anexo B) é de fácil implementação e pode ser utilizado para geração de tabelas com a finalidade de selecionar valores de c_1 e α apropriados para outras configurações paramétricas (m , n e π).

A acurácia e precisão das estimativas *Pzib* são flexíveis de serem melhoradas computacionalmente, adotando o critério $|Pzib - \pi_{emv}| < k$, em que k corresponde a um valor tolerável, especificado subjetivamente pelo pesquisador.

REFERÊNCIAS BIBLIOGRÁFICAS

BOLFARINE, H.; SANDOVAL, M.C. **Introdução à inferência estatística**. Rio de Janeiro: Sociedade Brasileira de Matemática, 2001. 123p.

BORGATTO, A.F. **Modelos para proporções com superdispersão e excesso de zeros**: um procedimento Bayesiano. 2004. 90p. Tese (Doutorado em Estatística e Experimentação Agronômica)–Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP.

BUSTOS, O. Algumas idéias de robustez aplicadas a estimação paramétrica em series temporais. In: SIMPOSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 7., 1986, Campinas, SP. **Anais...** Campinas: Unicamp, 1986. p.301-305.

CASELLA, G.; BERGER, R.L. **Statistical inference**. Belmont: Duxburg, 1990. 650p.

COPAS, J.B. Binary Regression Models for Contaminated Data. **Journal of the Royal Statistical Society**: Series B, Methodological, London, v.50, n.2, p.225-265, Apr. 1988.

HAMPEL, F.R. A general qualitative definition of robustness. **Annals of Mathematical Statistics**, Ann Arbor, v.42, n.6, p.1887-1896, Nov. 1971.

HUBER, P.J. Robust estimation of a Location Parameter. **Annals of Mathematical Statistics**, Ann Arbor, v.35, n.1, p.73-101, Jan. 1964.

HUBER, P.J. Robust statistics: a review. **Annals of Mathematical Statistics**, Ann Arbor, v.43, n.4, p.1041-1067, Aug. 1972

KEMP, C.D.; KEMP, A.W. Rapid estimation for discrete distributions, **The Statistician**, London, v.37, n.2, p.243-255, June 1988.

LINDSAY, B.G.; Efficiency versus robustness: the case for minimum hellinger distance and related methods. **The Annals of Statistics**, Hayward, v.22, n.4, p.1081-1114, Sep. 1994

MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. **Introduction to the theory of statistics**. 3.ed. New York: J. Wiley, 1974. 564p.

R DEVELOPMENT CORE TEAM. **R**: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2008.

RONCHETTI, E. The historical development of robust statistics. In: INTERNATIONAL CONFERENCE OF TEACHING STATISTICAL, 7., 2006, Salvador. **Proceeding...**Salvador: ICOTS, 2006. 1 CD-Rom.

RUCKSTUHL, A.F.; WELSH A.H. Robust Fitting of the Binomial Model. **The Annals of Statistics**, Hayward, v.29, n.4, p.1117-1136, Nov. 2001.

SIMPSON, D.G. Minimum Hellinger Distance Estimation for the Analysis of Count Data. **Journal of the American Statistical Association**, New York, v.82, n.399, p.802- 807, Sep. 1987

STAUDTE, ROBERT G.; SHEATHER, SIMON J. **Robust estimation and testing**. New York: John Wiley, 1990. 351p.

ANEXOS

	Páginas
ANEXO A: Gráficos.....	42
ANEXO B: Rotina	45

ANEXO A - Gráficos

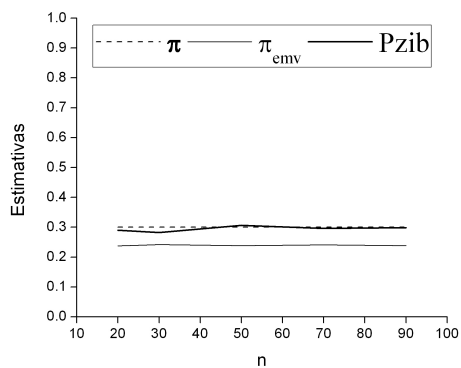


FIGURA 1: Gráfico das estimativas para $\pi = 0,3, \gamma = 0,2$

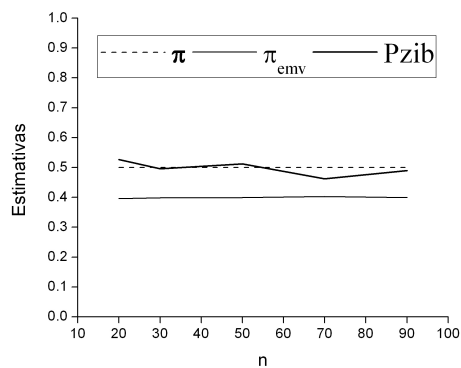


FIGURA 2: Gráfico das estimativas para $\pi = 0,5, \gamma = 0,2$

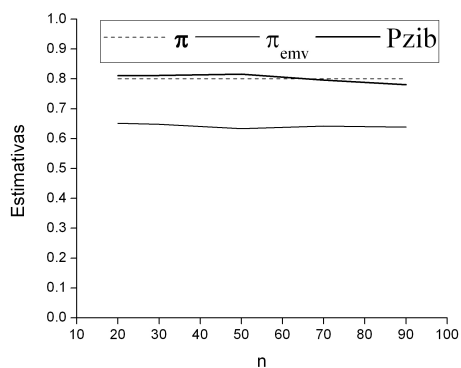


FIGURA 3: Gráfico das estimativas para $\pi = 0,8, \gamma = 0,2$

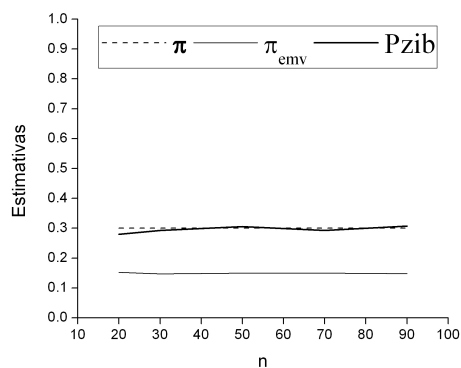


FIGURA 4: Gráfico das estimativas para $\pi = 0,3, \gamma = 0,5$

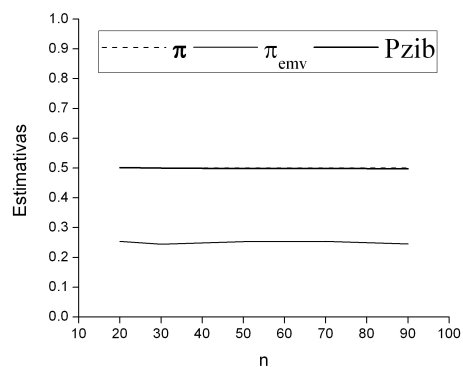


FIGURA 5: Gráfico das estimativas para $\pi = 0,5, \gamma = 0,5$

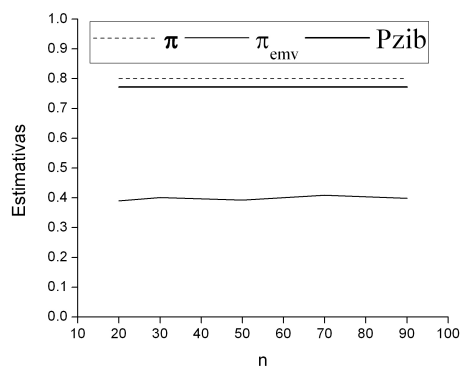


FIGURA 6: Gráfico das estimativas para $\pi = 0,8, \gamma = 0,5$

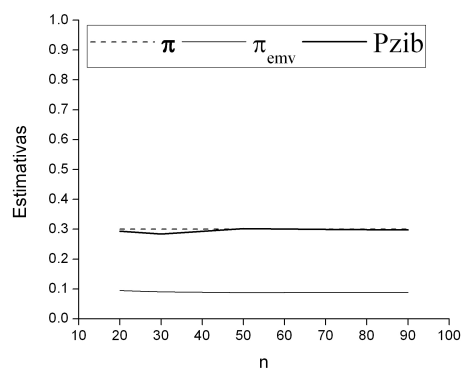


FIGURA 7: Gráfico das estimativas para $\pi = 0,3, \gamma = 0,7$

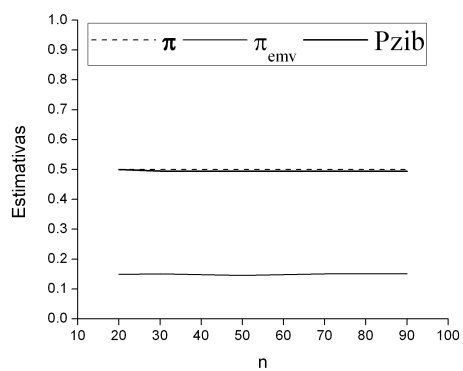


FIGURA 8: Gráfico das estimativas para $\pi = 0,5, \gamma = 0,7$

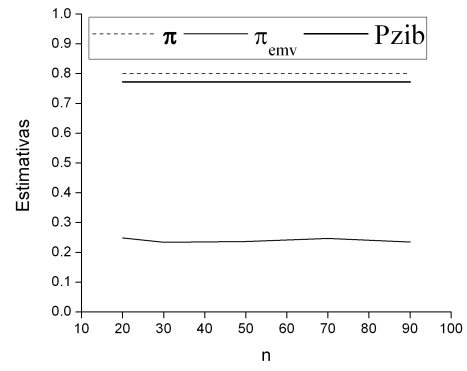


FIGURA 9: Gráfico das estimativas para $\pi = 0,8$, $\gamma = 0,7$

ANEXO B - Rotina

- Estimador PZIB sob $\rho_1(x)$ e $\rho_2(x)$

```
library(VGAM) ##pacote requerido##

### parametros de simulação #####
m=100
n=30
pr=0.5
gama=0.50
u=0.38 #utilizado somente em  $\rho_2(x)$ 
      #para utilização em  $\rho_1(x)$  basta##
      ##substituir por u=1###
nsim=5000
# ##### constantes de afinidade ##### #

c1=0.5### $c_1$  pode ser alterada aqui##
c2=1
# ##### DEFINIÇÃO DAS FUNÇÕES ##### #
vcont=c(rep(0,m+1))
y=seq(0,m)
rho=c(rep(0,m+1))
estimadores=matrix(0,nsim,3,dimnames=list
(c(rep("Simulação",nsim)),
c("j", "EMV", "Pzib")))

# ##### Calculo do fn(y) ##### #

fny=function(m,n,dados,vet)

{
  for (a in 1:(m))
  {

    prop=0
    aux=vet[a]

    for (b in 1:n)
```

```

{
  if (aux==dados[b]) prop=prop+1
}
vcont[a]=(prop)/n

}

return(vcont)
}

estimaPzib=function(x,p,c1,c2,alfa)

{
  estPzib=0

  for (b in 1:length(x))

  {

if (x[b]>=c1 && x[b]<=c2) rho[b]=x[b]*log(x[b])
if (x[b]<c1) rho[b]=((c1^(1-u))*log(c1)
+((1-u)*log(c1)+1)
*(c1^(1-u)/u))*x[b]^u
-(((1-u)*log(c1)+1)*c1/u)
if (x[b]>c2) rho[b]=((c2^(1-u))*log(c2)
+((1-u)*log(c2)+1)
*(c2^(1-u)/u))*x[b]^u
-(((1-u)*log(c2)+1)*c2/u)

auxPzib=rho[b]*p[b]

estPzib=auxPzib + estPzib

}

return (estPzib)

```

```

}

for (j in 1:nsim)
{
  amos1=rzibinom(n, m, pr, phi = gama)
  soma=sum(amos1)

  while (soma==0)
  {
    amos1=rzibinom(n, m, pr, phi = gama)
  }

  resfny=fny(m,n,amos1,y)

  estmv=sum(y*resfny)/m ##### Estima fn(y) #####
  prob=dbinom(y,m,estmv)
  x=(resfny/prob)
  Pzib=estimaPzib(x,prob,c1,c2,u)

  estimadores[j,1]=j
  estimadores[j,2]=estmv
  estimadores[j,3]=Pzib
}

m1= mean(estimadores[,2])
m2= mean(estimadores[,3])

vies=(m2-pr)/pr

```