MODELO POISSON ZERO INFLACIONADO APLICADO AO NÚMERO DE DEFEITOS EM VEÍCULOS

Professora Deise Deolindo Silva
deisedeolindo@hotmail.com
Estatística

Resumo
Este trabalho teve como objetivo aplicar o modelo Poisson Zero Inflacionado - ZIP para o ajuste do número de defeitos em veículos. Como as empresas possuem linhas de produção com rígidas especificações de qualidade, os itens produzidos apresentam um número muito reduzido de imperfeições. Diante desse contexto, os conjuntos de dados apresentam grande quantidade de valores zero, o que dificulta a elaboração de uma análise precisa. Por esse motivo, estudou-se o modelo ZIP, pois ele considera uma distribuição degenerada no ponto zero e a de Poisson para os outros valores. Na interpretação dos resultados compararam-se as estimativas para os modelos de Poisson e ZIP e, como critério de seleção, utilizou-se o Teste de Significância Completamente Bayesiano – FBST, o qual comprovou que a distribuição ZIP é mais eficaz.

Palavras-chave: Inferência Bayesiana; Poisson Zero Inflacionado; Seleção de Modelos.

Introdução

O bom ajuste dos dados depende diretamente dos modelos probabilísticos atribuídos a eles. Nas aplicações envolvendo dados reais sobre contagens, geralmente atribuem-se distribuições discretas, que são amplamente desenvolvidas na literatura.

No entanto, é comum encontrar uma grande quantidade de zeros nos conjuntos de dados. Esse excesso dificulta a elaboração de uma análise estatística precisa para o problema, já que os modelos usuais desenvolvidos não ajustam bem tal situação.

Diante deste cenário é relevante pesquisar quais as origens desses zeros. Martin *et al*. (2005), ressaltaram que o valor zero pode acontecer de quatro maneiras diferentes: duas delas podem ser definidas como zeros *verdadeiros* e duas como *aleatórios ou falsos*.

Os zeros verdadeiros podem surgir da baixa frequência de ocorrência do evento. Por exemplo, se o interesse for estudar o número de faltas de funcionários em uma determinada empresa durante um período de tempo, neste caso, o excesso de zeros significa que os empregados faltam pouco ao serviço.

Outra situação considerada como zero verdadeiro é quando realmente no local não havia nenhum indivíduo presente. Podemos citar as aplicações que envolvem controle de qualidade que utilizam processo de fabricação moderno e, por isso, esperam que os itens não apresentem defeitos (estado perfeito).

Os zeros aleatórios ou falsos podem ser resultado de erros de amostragem ou um vício visual, ou seja, o indivíduo existe, ocupa o local, mas não estava presente durante a realização da pesquisa ou o elemento ocupa o local, está presente, mas o pesquisador não o encontra. Esse tipo de zero ocorre geralmente em estudos ecológicos, principalmente de vida selvagem ou aquática.

Então, os zeros ocorridos em um dos conjuntos de dados podem ter sido resultado de um zero verdadeiro, de um erro humano, ou ser um zero de amostragem. Infelizmente, a distinção desses tipos de zeros é, na maioria das situações, uma tarefa impossível de ser realizada.

A produção dos zeros excessivos nas amostras pode ser classificada de duas formas. Na primeira, o excesso é resultado da superdispersão, ou seja, a variância dos dados é maior que a assumida pelo modelo. Na segunda forma, o excesso é formado por subpopulações distintas e podem estar relacionadas a alguma intervenção natural ou truncamento nos dados. (PAULA, 2004).

Uma metodologia eficaz na modelagem de dados resultantes de contagens com zeros excessivos é a mistura de modelos, através das distribuições zero inflacionadas. Essas requerem a definição de uma distribuição discreta e outra degenerada no ponto zero. Como caso particular foi estudado o modelo Poisson Zero Inflacionado – ZIP.

Rodrigues (2003) apresentou a abordagem bayesiana para distribuições zero inflacionadas utilizando um procedimento baseado em dados ampliados. O objetivo, neste caso, foi tornar a *posteriori* conhecida, facilitando o tratamento computacional.

Como exemplificação desta metodologia foi objeto de estudo um conjunto de dados referentes ao número de defeitos em veículos. Para selecionar o modelo que melhor se ajusta aos dados foi utilizada a medida de evidência proposta por Pereira e Stern (1999), designada por *Full Bayesian Significance Test* (FBST).

1 Número Excessivo de Zeros em Contagens

Em muitas aplicações envolvendo dados reais sobre contagens são atribuídos os modelos discretos que são largamente desenvolvidos na literatura, podemos citar as distribuições de Poisson, binomial e binomial negativa.

Geralmente, alguns conjuntos de dados contêm um número excessivo de zeros que não são descritos pelo modelo assumido. Esses zeros podem ter origem em diferentes fontes, Martin *et al*. (2005) ressaltaram que o valor zero acontece de quatro modos, dois podem ser definidos como zeros verdadeiros e dois como falsos (aleatórios).

No primeiro caso, os zeros verdadeiros surgem de uma baixa frequência de ocorrência ou realmente o local não havia nenhum indivíduo presente. No segundo caso, o indivíduo existe, ocupa o local, mas não estava presente durante a pesquisa ou o elemento ocupa o local, está presente, mas o pesquisador não o encontra- como mencionado anteriormente e enfatizado neste momento para melhor fixação da informação.

Uma possível solução, para explicar zeros verdadeiros ou falsos, é utilizar as distribuições zero inflacionadas, que exige em sua estrutura conhecimentos sobre mistura de modelos. Neste caso, o modelo é obtido através da média ponderada de duas distribuições, uma degenerada no ponto zero e outra que se adequaria aos dados, caso não existisse zeros excessivos.

Quando os zeros inflacionados são resultados de excesso de zeros verdadeiros e falsos, não há nenhuma discussão formal na literatura de como modelar tais conjuntos de dados, justamente porque é difícil distinguir a origem de tais valores. Quando há a incerteza sobre a sua origem nas observações, um procedimento usual é utilizar distribuições truncadas.

Por exemplo, ao considerar uma linha de produção na qual se aplica controle de qualidade, a contagem de defeitos de um produto apresenta-se cada vez menor, ou seja, há um grande número de zeros, isto devido à modernização dos processos de fabricação. Neste caso, esses zeros correspondem a zeros determinísticos.

Conforme Martin *et al*. (2005), nas aplicações ecológicas os zeros podem ocorrer devido à espécie ser totalmente ausente na área amostrada ou quando a espécie está presente, mas não foi observada pelo pesquisador, neste caso, os zeros são aleatórios (falsos). Nessas aplicações o problema de zeros aleatórios ocorre frequentemente devido a erros humanos ou vícios no método de amostragem.

Sob a ótica estatística, a produção excessiva de zeros pode ser classificada de duas formas: *superdispersão ou subpopulações distintas*.

Conforme Hinde e Demétrio (1998) *apud* Paula (2004), superdispersão é um fenômeno comum que ocorre na modelagem de dados discretos e cuja ocorrência é caracterizada quando a variância observada excede aquela assumida pelo modelo.

Saito (2005) ressalta que os zeros podem ser produzidos por subpopulações distintas, ou seja, pode estar relacionada a alguma intervenção natural ou truncamento dos dados.

2 Distribuições Série de Potências Inflacionadas

O sucesso de uma modelagem depende, substancialmente, dos modelos probabilísticos adotados. Para dados discretos uma classe geral foi desenvolvida - *Distribuições Série de Potências* (PSD). Esta classe engloba tanto os modelos simples como os generalizados e pode ser considerada em diversas aplicações, obtendo bons resultados. (GUPTA *et al*. 1995).

Na análise de dados discretos existem frequentemente valores inflacionados, como, por exemplo, o ponto zero é observado com uma frequência significativamente maior que o admitido pelo modelo assumido; consequentemente, a classe de distribuições série de potências pode ser estendida para distribuições inflacionadas, e a denominamos de *Classe de Distribuições Série de Potências Inflacionadas* (IPSD).

Murat e Szynal (1998) relataram que ao modelar dados inflacionados é comum considerar mistura de modelos. Neste caso, propõe-se uma média ponderada de duas distribuições, ou seja, uma degenerada para o valor em excesso, enquanto os outros valores seguem um modelo conveniente.

Considere a seguinte aplicação sobre a produção de duas máquinas: a I produz itens perfeitos e a II produz defeitos de acordo com o modelo de Poisson. Ao observar a produção final não é possível identificar se o produto é oriundo da máquina I ou II, neste caso, o valor zero torna-se inflacionado, ou seja, é resultado de uma subpopulação que produz contagens zero. Uma modelagem adequada seria a distribuição de Poisson Zero Inflacionada.

A abordagem bayesiana foi utilizada para obter as estimativas dos parâmetros envolvidos no modelo e está descrita a seguir.

2.1 Abordagem Bayesiana para a Distribuição Poisson Zero Inflacionada

Para melhorar modelagem se faz necessário formular o excesso de zeros apresentados pelos dados. Rodrigues (2003) menciona que existem muitas formulações, e sugeriu a seguinte,

$$\Pr[Y = y \mid \Theta = \theta] = \begin{cases} I_{\{0\}}(y), \theta = 0 \\ p(y \mid \theta), \theta > 0 \end{cases}$$

em que, $I_{\{0\}}(y)$ é uma distribuição que está degenerada por zeros e $p(y|\theta)$ é uma função de probabilidade que se ajusta aos dados. Muitos autores propuseram utilizar mistura de distribuições, pois incorporam o excesso de zeros apresentados pelos dados. Para isto, considere peso ω ao evento *0* e peso $(1-\omega)$ a θ com $0 \leq \omega \leq 1$.

Esta problemática pode ser representado da seguinte forma:

$$p(y|\theta, \omega) = \omega I_{\{0\}}(y) + (1-\omega)p(y|\theta), y = 0, 1, 2, \ldots \qquad (1)$$

na qual, $p(y|\theta)$ é uma distribuição de probabilidade discreta com vetor de parâmetros $\boldsymbol{\theta}$, que teoricamente se adequaria aos dados caso não houvesse a presença excessiva de dados, nesse caso a de Poisson. O parâmetro ω é a proporção de zeros que excede o que seria predito através de $p(y|\theta)$.

2.1.1 Função de Verossimilhança baseada nos Dados Aumentados

Supondo que $Y=(Y_1,...,Y_n)$ seja um vetor de n variáveis aleatórias com um modelo Poisson Zero Inflacionado (ZIP). Seja $A=\{y_i: y_i=0, i=1,...,n\}$ e $m=n(A)$, então a função de probabilidade é $L(\theta,\omega)=\left[\omega+(1-\omega)e^{-\lambda}\right]^m(1-\omega)^{n-m}\prod_{yi\notin A}\frac{e^{-\lambda}\lambda^{yi}}{y_i!}$

Os elementos do conjunto A vem de qualquer um de dois grupos diferentes, da distribuição degenerada de zero ou de $p(0)$. Como o modelo ZIP é uma mistura de duas distribuições, então a função de verossimilhança pode ser simplificada com a utilização de um procedimento baseado em dados ampliados com variáveis latentes. Conforme Rodrigues (2003), neste tipo de situação, é natural definir a seguinte variável.

$$I_i=\begin{cases}1, & p(\theta,\omega)\\ 0, & 1-p(\theta,\omega)\end{cases} \quad i=1,...,m \quad \text{e}$$

$$p(\theta,\omega)=\frac{\omega}{\omega+(1-\omega)p(0\mid\theta)}=\frac{\omega}{\omega+(1-\omega)e^{-\theta}}$$

Esta variável indica se o elemento da *i-ésima* posição de A é tirado do primeiro componente de (1) ou não. Assim a função de verossimilhança baseada nos dados aumentados $D=\{Y,I\}$, é:

$$L[\theta,\omega\mid D]=L[\theta,\omega]\prod_{i=1}^{m}p(\theta,\omega)^{Ii}(1-p(\theta,\omega))^{1-Ii}$$

$$=\left[\omega+(1-\omega)p(0\mid\theta)\right]^m(1-\omega)^{n-m}\prod_{i=1}^{m}p(y_i\mid\theta)\left[p(\theta,\omega)^{\Sigma Ii}(1-p(\theta,\omega))^{m-\Sigma Ii}\right]$$

$$=\frac{\omega^S+\left[(1-\omega)p(0\mid\theta)\right]^{m-S}}{\left[\omega+(1-\omega)p(0\mid\theta)\right]^{S+m-S}}\left[\omega+(1-\omega)p(0\mid\theta)\right]^m(1-\omega)^{n-m}\prod_{yi\notin A}p(y_i\mid\theta)$$

$$=\omega^S(1-\omega)^{n-S}\prod_{yi\notin A}p(y_i\mid\theta)$$

Assim, $S=\sum_{i=1}^{m}I_i\sim Bin\left[m,p(\theta,\omega)\right]$. A *priori* conjunta é dada por $\pi(\theta,\omega)$ e a *posteriori* conjunta (θ,ω), dado D é

$$\pi(\theta,\omega|D)=L(\theta,\omega|D)\pi(\theta,\omega) \qquad (2)$$

Considerando o modelo de Poisson, temos a seguinte função de verossimilhança baseado em dados aumentados D é

$$L(\theta,\omega)\ \alpha\ \omega^S(1-\omega)^{n-S}\theta^Z e^{-(n-S)\theta},\ Z=\sum_{y\notin A}y_i \qquad (3)$$

A função de verossimilhança sugere *prioris* independentes da seguinte forma:

$$\theta\sim Gama(a,b)\ e\ \omega\sim Beta(c,d)$$

dessa forma, a distribuição a *posteriori* conjunta para (θ,ω), dado D, é

$$\pi(\theta,\omega|D)=\ \omega^{S+c+1}(1-\omega)^{n-S+d+1}\theta^{Z+a+1}e^{-(n-S+b)\theta} \qquad (4)$$

Para a obtenção das distribuições a *posteriori* dos parâmetros envolvidos no modelo, faz-se necessário utilizar métodos computacionais para aproximar a distribuição a *posteriori*.

3 Seleção de Modelos

O processo para a seleção de modelos é reconhecidamente essencial em inferência, pois o ajuste de um modelo a um conjunto de dados envolve a discriminação entre diferentes modelos competitivos. A seleção bayesiana de modelos, em geral, é baseada no cálculo de probabilidades a *posteriori* para os modelos em questão e não apresenta dificuldades na comparação entre modelos com estruturas diferentes.

Pereira e Stern (2008) propuseram a medida de evidência bayesiana denominada Teste de Significância Completamente Bayesiano - *Full Bayesian Significanc Test* (FBST)- aplicada para hipóteses precisas.

3.1 Teste de Significância Completamente Bayesiano

Pereira e Stern (2008) desenvolveram o FBST para testar a significância de uma hipótese precisa. Este teste mostra a qualidade da teoria de decisão bayesiana. Um dos benefícios em utilizá-la é por não apresentar problemas quando são atribuídas distribuições a *priori* impróprias.

Para o cálculo da medida de evidência *Ev* é necessário somente conhecer a distribuição a *posteriori*, que não apresenta complicações quando as dimensionalidades do parâmetro e do espaço amostral são grandes. Computacionalmente *Ev* utiliza somente a otimização e a integração numérica, não se baseando em resultados assintóticos.

Para determinar *Ev* é necessário considerar uma hipótese precisa H_0: $\theta \in \Theta_0$ então:

$$g^* = \sup_{H0} g_x(\theta) \text{ e } T = \{ \theta \in \Theta_0 : g_x(\theta) > g^* \},$$

pois, $g_x(\theta) = g(\theta|x) \propto L_x(\theta)g(\theta)$.

O valor da média de evidência bayesiana contra H_0 é definido como a probabilidade a *posteriori* do conjunto tangencial, isto é, $Ev^C = \Pr(\theta \in T \mid x) = \int_T g_x(\theta) d\theta$.

O valor da medida de evidência que apóia H_0, $Ev = 1 - Ev^C$, não é uma evidência contra a hipótese alternativa. Equivalentemente, *Ev* não é evidência a favor da alternativa, embora esteja contra H_0. Ou seja, o Teste de Significância Completamente Bayesiano é o procedimento que rejeita H_0 sempre que *Ev* é pequeno, ou similarmente, não rejeita H_0 quando *Ev* for grande (PEREIRA E STERN, 2008).

3.1.1 Teste de Significância Completamente Bayesiano para a Distribuição de Poisson Zero Inflacionada

Rodrigues (2006) apresentou que o Teste de Significância Completamente Bayesiano pode ser aplicado para situações onde os dados são ampliados. Esta medida de evidência é encontrada em dois passos. O primeiro é obtido através da otimização e o outro através da integração numérica da distribuição a *posteriori* do parâmetro de locação θ.

Para isto considere, H_0: $\omega = 0$, ou seja, o modelo M_0 é adequado e a hipótese H_1: $\omega > 0$, o modelo M_1 é adequado. Se a fatoração da função de verossimilhança for obtida, o teste de significância completamente bayesiano é baseado na verossimilhança marginal de θ.

Passo de Otimização – Encontre a moda θ_0 da distribuição a *posteriori* $\pi_0(\theta|Y)$ sob H_0: $\theta \in \Theta_0$, onde $\Theta_0 = \{\theta: \omega = 0\}$, em que a densidade a *posteriori* $\pi_0(\theta|Y)$ é dada por

$$\pi_0(\theta \mid Y) = \pi_0(\theta, \omega = 0) \prod_{i=1}^{n} p(y \mid \theta)$$

Equivale a encontrar a moda θ_0 de $\pi_0(\theta|Y)$ sob H_0 e pode ser calculada através de

$$\theta_0 = \frac{Z + a - 1}{n + b}$$

em que $\pi_0(\theta|Y)$ é a distribuição $Gama(Z + a, n + b)$.

Passo de Integração – a medida de evidência é obtida através de

$$Ev(H_0 \mid D) = 1 - \Pr[\theta \in Z^*(D) \mid D] = 1 - \int_0^{\infty} \int_{Z^*(D)}^{\infty} \pi(\theta, \omega \mid D) d\theta d\omega$$

com, $[Z^*(D)] = \{\theta: \pi(\theta|D) \geq \pi_0(\theta|D)\}$.

A densidade marginal $\pi(\theta|D)$ corresponde à distribuição Gama com parâmetros $\left(\sum_{yi\notin A} y_i + a, n - S + b\right)$ com máximo em $\frac{Z + a - 1}{n - S + b}$.

Na situação de dados ampliados Rodrigues (2006) disse que é equivalente testar

$$H_0 : \theta = \frac{Z + a - 1}{n + b}.$$

A medida de evidência é $Ev(H_0|D)$ é igual a 1 se, e somente se, $\omega = 0$, ou seja, não se rejeita a hipótese nula de que os dados se ajustam melhor ao modelo de Poisson. Se $Ev(H_0|D)$ for pequena, tem-se que $\omega > 0$, logo rejeita-se H_0 em favor de H_1 e o modelo ZIP representa melhor os dados. (Rodrigues (2006) implementou o *Full Bayesian Significance Test* – FBST no *Software Winbugs* e este foi utilizado neste contexto).

4 Número de Defeitos em Veículos

Silva (2009), apresentou um conjunto de dados sobre o número de defeitos em veículos. Em determinado período foram verificados 54 carros e classificados quanto ao número de não conformidades. Ajustou-se os modelos de Poisson e ZIP com o objetivo de verificar quais deles ajustam melhor dados com zeros em excesso. As informações estão dispostas a seguir.

Valores	**0**	**1**	**2**	**3**	**Total**
Frequência	42	8	2	2	**54**

Tabela 1 – Número de Defeitos em Veículos

Verifica-se que, aproximadamente, 78% dos carros são considerados conformes. Esses dados poderiam ser ajustados pelo modelo de Poisson. No entanto, essa distribuição não conseguiria ajustar essa quantidade excessiva de zeros e, como alternativa, foi proposto verificar a adequabilidade do modelo ZIP.

Para este trabalho, a sugestão foi encontrar as estimativas das *posterioris* através do *Software* R e obter a probabilidade de cobertura para os parâmetros.

É necessário, no entanto, especificar as *prioris* para os respectivos parâmetros. Para λ do modelo de Poisson foi atribuído uma *priori* Gama difusa com hiperparâmetros $a = b = 0{,}0001$. A *posteriori* é, também, uma distribuição Gama com parâmetros atualizados. Foram geradas via R duas cadeias com *50000* amostras, utilizando um período de aquecimento de *5000* observações e saltos de iterações entre as amostras. A convergência do procedimento Monte Carlo via Cadeia de Markov (MCMC) deste exemplo foram monitorados pelo diagnóstico de Gelman-Rubin, que consiste basicamente em uma análise de variância intra e entre as cadeias geradas.

Os resumos a *posteriori* para λ estão na Tabela 2.

Parâmetro	**Média**	**Desvio Padrão**	**2.5%**	**97.5%**
θ	0,3332	0,0791	0,1930	0,5056

Tabela 2 – Resumo Estatístico para o Parâmetro Theta do Modelo Poisson

A figura 1 apresenta as distribuições estimadas e o comportamento das cadeias ao longo das interações para os parâmetros do modelo Poisson.

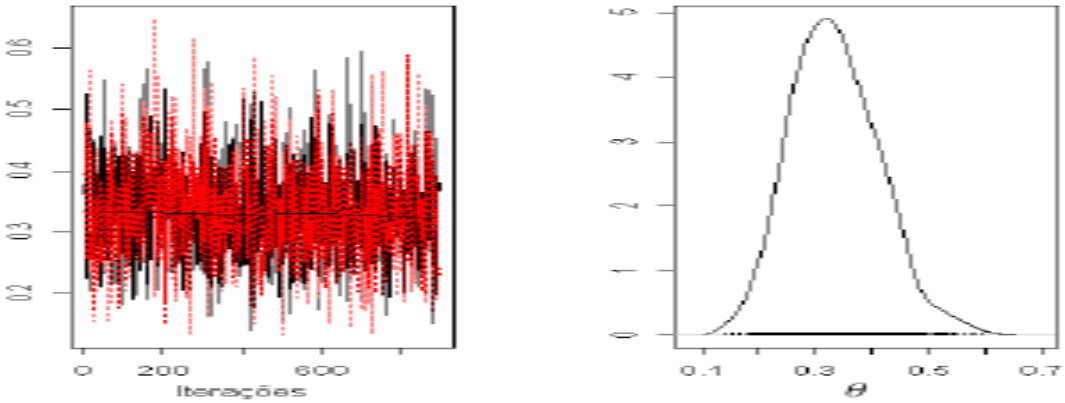

Figura 1 - Comportamento das Cadeias ao longo das Iterações e Distribuições Estimadas via Amostrador de *Gibbs* e o para o Parâmetro λ do Modelo Poisson.

Para a obtenção das *posterioris* do Modelo ZIP também foi considerada para λ uma *priori* Gama difusa com os mesmos hiperparâmetros atribuídos ao modelo Poisson, e para ω foi atribuída uma *priori* Beta (de *Jeffreys*). Considerou-se o mesmo período de aquecimento e tamanho de amostra. Além disso, através do critério de Gelman-Rubin, pode-se dizer que houve convergência para os parâmetros do modelo de Poisson e ZIP.

Os resumos a *posteriori* para θ e ω estão na Tabela 2.

Parâmetro	Média	Desvio Padrão	2.5%	97.5%
θ	0.5176	0.0202	0.4791	0.5577
ω	0.2242	0.0272	0.1711	0.2768

Tabela 2 – Resumo Estatístico para o Parâmetro Theta e Peso do Modelo ZIP

A figura 2 apresenta as distribuições estimadas e o comportamento das cadeias ao longo das interações para os parâmetros θ e ω do modelo ZIP.

Considerando os gráficos apresentados para as amostras geradas, observa-se que existe um indício de convergência que comprovou-se através do critério de *Gelman-Rubin*.

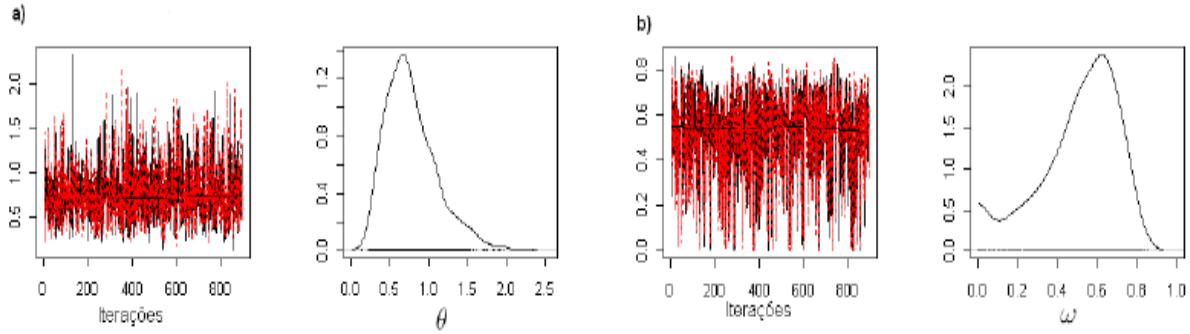

Figura 2 – **a)** Comportamento das Cadeias ao longo das Iterações e Distribuições Estimadas via Amostrador de *Gibbs* e o para o Parâmetro θ do Modelo ZIP. **b)** Comportamento das Cadeias ao longo das Iterações e Distribuições Estimadas via Amostrador de *Gibbs* e o para o Parâmetro ω do Modelo ZIP.

Através da Tabela 3 pode-se notar que há evidências que o modelo ZIP ajusta melhor os dados. Nela são apresentados os valores esperados segundo cada modelo e o respectivo valor do FBST.

Valores	$Freq_{Obs}$	$Esp_{Poisson}$	Esp_{zip}
0	42	38.6978	39.7264
1	8	12.8941	9.4521
2	2	2.1482	3.6608

3	2	0.2386	0.9452

Tabela 3 - Valores Observados e Esperados segundo os Modelos ZIP e Poisson

Para comprovar é necessário verificar, através da medida de evidência *Full Bayesian Significance Test* – FBST, qual modelo melhor se ajusta aos dados.

Para tal aplicação o valor obtido foi *Ev = 0.04*. Ou seja, o modelo ZIP ajusta melhor os dados sobre o número de defeitos em veículos. Isto era esperado, pois em empresas que utilizam controle de qualidade, o número de defeitos tende a ser o menor possível.

5 Considerações Finais

Em situações práticas é comum utilizar os modelos discretos para ajustar dados de contagem. Existe, porém, uma complicação na análise estatística para este tipo de dados, que é a presença excessiva de zeros. Uma solução para o problema é considerar na modelagem a Classe de Distribuições Série de Potências Inflacionadas.

Como caso particular estudou-se a distribuição Poisson Zero Inflacionada no contexto bayesiano e sua aplicabilidade em situações práticas. Esta metodologia foi aplicada a um conjunto de dados referentes ao número de defeitos em veículos.

Considerou-se como critério seleção de modelos o *Full Bayesian Significance Test* (FBST), proposto por Pereira e Stern (1999) *apud* Rodrigues (2006). Esta medida de evidência é de fácil interpretação, além de ser eficaz quando o tamanho amostral é grande, o que não acontece com outros critérios de seleção.

Através desta medida de evidência comprovou-se que a distribuição ZIP conseguiu modelar melhor os dados referentes ao número de defeitos em veículos. Ou seja, o ZIP é uma alternativa eficaz para o modelo de Poisson quando existem zeros excessivos.

Referências

DATTA, G. S., BAYARRY, S. e BERGER, J. **Model Selection for Count Data: ZIP It?** Apresentação no 6^0 Workshop de Inferência Bayesiana Objetiva. University of Rome: La Sapienza, 2007.

GUPTA, P. L., GUPTA, R. C. e TRIPATHI, R. C. **Inflated Modified Power Series Distributions with Applications**. Communications in Statistics - Theory and Methods, v. 24, p. 2355-2374, 1995.

MARTIN, T. G., Wintle, B. A., RHODES, J. R., KUHNERT, P. M., FIELD, S. A., LOW-CHOY, S. J., TYRE, A. J. e POSSINGHAM, H. P. **Zero Tolerance Ecology: Improving Ecological Inference by Modelling the Source of Zero Observations**. Ecology Letters, p. 12351246, 2005.

MURAT, M. e SZYNAL, D. **Non-Zero Inflated Modified Power Series Distributions, Communications in Statistics** - Theory and Methods, v. 27, p. 3047-3064, 1998.

PAULA, G. A. **Modelos de Regressão com Apoio Computacional**. Instituto de Matemática e Estatística da Universidade de São Paulo.São Paulo, 2004.

RODRIGUES, J. **Bayesian Analysis of Zero-Inflated Distributions**. Communications in Statistics - Vol 32, n 2, p. 281-289, 2003.

RODRIGUES, J. **Full Bayesian Significance Test for Zero-Inflated Distributions**. Communications in Statistics - Vol 35, p. 1-9, 2006.

SAITO M. Y. **Inferência Bayesiana para Dados Discretos com Excesso de Zero e Uns**. Dissertação (Mestrado), Programa Pós Graduação em Estatística - Departamento de Estatística, Universidade Federal de São Carlos. São Carlos, 2005.

SILVA, D. D. **Classe de Distribuições Série de Potências Inflacionadas com Aplicações**. Dissertação (Mestrado), Programa de Pós Graduação em Estatística - Departamento de Estatística, Universidade Federal de São Carlos. São Carlos, 2009.