

## MODELOS INFLACIONADOS DE ZEROS: APLICAÇÕES NA ANÁLISE DE UM QUESTIONÁRIO DE FREQUÊNCIA ALIMENTAR

Giovana FUMES<sup>1</sup>  
José Eduardo CORRENTE<sup>2</sup>

- RESUMO: O objetivo do presente trabalho foi utilizar os modelos inflacionados de zero na análise de um questionário de frequência alimentar (QFA) de idosos de uma cidade de porte médio do interior do Estado de São Paulo, investigando os modelos que melhor se ajustam aos dados segundo a porcentagem de zeros encontrada e os fatores que influenciam tanto o consumo como o não consumo de alguns alimentos. A variável resposta utilizada foi a frequência mensal do consumo de alimentos provenientes do QFA e as covariáveis associadas eram vinculadas às características sociodemográficas, de prevenção e morbidades referidas. Foram utilizados os modelos de Poisson, Binomial Negativo, Poisson Inflacionado de Zeros e Binomial Negativo Inflacionado de Zeros. Para um conjunto de dados com até 10% de zeros o modelo usual Binomial Negativo se mostrou adequado. Devido à amplitude dos dados, os modelos inflacionados apresentaram-se adequados somente na porcentagem de 10% a 50% de zeros. Pode-se evidenciar ainda que, a probabilidade de consumo dos alimentos selecionados foi associada ao sexo, ao estado nutricional, ao uso de medicamentos e a prática de atividade física. A probabilidade de não consumo apareceu associada ao sexo, ao funcionamento do intestino e ao consumo diário de água.
- PALAVRAS-CHAVE: Distribuição binomial negativa; distribuição de Poisson; distribuição inflacionada de zero; questionário de frequência alimentar; superdispersão.

---

<sup>1</sup>Programa de Pós-graduação em Bioestatística, Instituto de Biociências, Universidade Estadual Paulista - UNESP, CEP: 18618-970, Botucatu, São Paulo, Brasil. E-mail: [gifumesbtu@hotmail.com](mailto:gifumesbtu@hotmail.com)

<sup>2</sup>Departamento de Bioestatística, Instituto de Biociências, Universidade Estadual Paulista - UNESP, CEP: 18618-970, Botucatu, São Paulo, Brasil. E-mail: [jecorren@ibb.unesp.br](mailto:jecorren@ibb.unesp.br)

## 1 Introdução

A proposta de um levantamento nutricional é avaliar o consumo dietético de um grupo ou população em relação a algum padrão, a respeito de adequação nutricional e prevenção de doenças crônicas (Dodd et al., 2006). Para tal finalidade vários instrumentos tem sido propostos, como o recordatório de 24 horas, o registro alimentar, o questionário de frequência alimentar, dentre outros.

O Questionário de Frequência Alimentar (QFA) contém basicamente duas ferramentas: uma lista de alimentos e a frequência habitual com a qual o indivíduo consome esses alimentos. A frequência de consumo é registrada em unidades de tempo: dias, semanas, meses ou anos. O formato é de perguntas simples com respostas fechadas, com opções de zero a dez vezes de consumo para cada alimento (Slater et al., 2003).

Os dados obtidos de um QFA são, dessa forma, contagens originárias da frequência com que cada alimento é consumido habitualmente pelo indivíduo. Uma estratégia de análise é supor que a distribuição desses dados siga uma distribuição de Poisson. Entretanto, dados de contagem podem apresentar uma variabilidade maior do que a esperada. Esse fenômeno se chama superdispersão e, nesses casos, opta-se em utilizar uma distribuição que melhor acomode essa extra variabilidade.

Outro fenômeno usualmente verificado na análise deste tipo de dados é o não consumo de determinados alimentos. Este não consumo pode ocorrer pelo fato de, no período de coleta, o indivíduo não ter consumido determinado alimento, ou porque tal alimento não faça parte da dieta habitual do indivíduo. Isso gera um excesso de zeros, podendo levar a uma superdispersão nos dados. Para contornar este problema, modelos inflacionados de zeros podem ser utilizados na análise desses dados (Lambert et al, 1992; Slymen et al., 2006).

O objetivo do presente trabalho foi utilizar os modelos inflacionados de zero na análise de um questionário de frequência alimentar de idosos de uma cidade de porte médio do interior do Estado de São Paulo, investigando os modelos que melhor se ajustaram aos dados segundo a porcentagem de zeros encontrada e os fatores que influenciaram tanto o consumo como o não consumo de alguns alimentos.

## 2 Metodologia

Os dados do Questionário de Frequência Alimentar (QFA) utilizados são provenientes de um estudo transversal para avaliar a qualidade e satisfação com a vida de idosos do município de Avaré, localizado no interior do Estado de São Paulo, Brasil.

Inicialmente foi elaborado um banco de dados dos idosos a partir do cadastro das Unidades Básicas de Saúde do município de Avaré, São Paulo. Como não se conhecia a prevalência de satisfação com a vida desses idosos no município, esta foi considerada como sendo de 50% com uma margem de erro amostral de 5% e confiabilidade de 95% bilateral. Com isto, o tamanho da amostral inicialmente calculado foi de 365 idosos. O sorteio dos participantes da amostra foi feito

estratificando por faixa etária de dez em dez anos. Para o presente estudo, foram sorteados 20% da amostra inicial, perfazendo um total de 73 indivíduos de ambos os sexos. O sorteio dos participantes foi novamente estratificado por faixa etária. Essa subamostra foi considerada para um estudo de validação sobre o Questionário de Frequência Alimentar (Projeto CNPq nº402533/2007-0), no qual se recomenda um grupo de 50 a 100 indivíduos (Slater et al., 2003).

Como não existe um QFA para idosos, foi utilizado um questionário validado para adultos, com uma lista composta de 67 alimentos ou preparações organizados nos grupos: sopas e massas, carnes e peixes, leite e derivados, verduras e legumes, molhos e temperos, frutas e bebidas (Fisberg et al., 2008).

Inicialmente, a variável resposta utilizada foi a frequência mensal do consumo desses 67 alimentos ou preparações para o ajuste dos modelos nulos e a comparação dos valores preditos pelos modelos com a proporção de zeros observada. Após escolhido o modelo mais próximo da proporção observada, foi feita a escolha das covariáveis e de alguns alimentos para estudo dos fatores que influenciam no consumo ou não dos itens do QFA. Profissionais da área nutricional foram consultados para selecionar esses alimentos ou preparações dentre os 67 contidos no QFA.

O ajuste dos modelos utilizados para verificar a associação entre as covariáveis de interesse e a frequência de consumo seguiu o enfoque dos modelos lineares generalizados, utilizando a variável resposta como contagens e os modelos Poisson, Binomial Negativo, Poisson Inflacionado de Zeros e Binomial Negativo Inflacionado de Zeros. As covariáveis escolhidas foram: sexo (masculino ou feminino), estado nutricional (desnutrido, eutrófico ou obeso), uso de medicamentos (anti hipertensivos ou outros), funcionamento do intestino (normal ou constipação e diarreia), prática de atividade física (sim ou não), idade (em anos) e consumo diário de água (em ml). Na sequência, um breve resumo dos modelos de regressão utilizados é apresentado.

**Modelo de Poisson:** Seja  $y_i$  o representante da contagem da  $i$ -ésima observação,  $i = 1, 2, \dots, n$ . Seja  $x_j$ ,  $j = 1, 2, \dots, p$ , um vetor de  $p$  covariáveis associadas e  $\beta$  um vetor de coeficientes da regressão a serem estimados. Uma variável resposta que segue uma distribuição de Poisson é dada por (Meyer, 1974):

$$P(y_i/x_j) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \quad \text{para } y_i = 0, 1, 2, \dots$$

em que  $\mu_i = \exp(x_j' \beta)$ . O valor esperado e a variância são dados respectivamente por:  $E(y_i/x_j) = Var(y_i/x_j) = \mu_i$ , com  $i = 1, 2, \dots, n$  e  $j = 1, 2, \dots, p$ .

**Modelo Binomial Negativo:** Um modelo de regressão no qual a variável resposta  $y_i$  apresenta uma distribuição binomial negativa (McCullagh e Nelder, 1989), com parâmetros,  $\mu_i$ ,  $i = 1, 2, \dots, n$  e  $k$ , segue a forma:

$$f(y_i/x_j) = \frac{\Gamma(y_i + k)}{y_i! \Gamma(k)} \left(1 + \frac{\mu_i}{k}\right)^{-k} \left(\frac{\mu_i}{k + \mu_i}\right)^{y_i} \quad \text{para } y_i = 0, 1, 2, \dots, n.$$

em que  $k$  é o parâmetro de dispersão. Esta distribuição tem valor esperado  $E(y_i/x_j) = \mu_i$  e a variância  $Var(y_i/x_j) = \mu_i \left(1 + \frac{\mu_i}{k}\right)$ , para  $i = 1, 2, \dots, n$  e  $j = 1, 2, \dots, p$ . Nota-se que a média é a mesma encontrada na distribuição de Poisson, porém, esta distribuição apresenta uma modificação na função de variância dependendo do parâmetro de dispersão, podendo acomodar, desse modo, uma extra-variância.

**Modelo de Poisson Inflacionado de Zeros (ZIP):** O modelo ZIP foi proposto por Lambert(1992), o qual apresenta um parâmetro de média  $\mu_i$ , com probabilidade  $p_i$  para os chamados zeros estruturais, que caracterizam, para o exemplo de aplicação, os alimentos que não fazem parte da dieta habitual dos indivíduos e  $(1 - p_i)$  para os zeros amostrais, que caracterizam os alimentos que não foram consumidos no período de estudo, com  $i = 1, 2, \dots, n$ . Esse modelo é dado por:

$$P(Y = y_i) = \begin{cases} p_i + (1 - p_i)e^{-\mu_i} & \text{para } y = 0 \\ \frac{(1 - p_i)e^{-\mu_i} \mu_i^{y_i}}{y_i!} & \text{para } y \geq 1. \end{cases}$$

O valor esperado e a variância de  $y_i$  no modelo de Poisson Inflacionado de Zeros são dados, respectivamente, por  $E(y_i/x_j) = (1 - p_i)\mu_i$  e  $Var(y_i/x_j) = \mu_i(1 - p_i)(1 + p_i\mu_i)$ , para  $i = 1, 2, \dots, n$  e  $j = 1, 2, \dots, p$ .

**Modelo Binomial Negativo Inflacionado de Zeros (ZINB):** Analogamente ao modelo ZIP (Cheung, 2002), com parâmetro de média  $\mu_i$ , probabilidade  $p_i$  para os zeros estruturais e  $1 - p_i$  para os zeros amostrais, com  $i = 1, 2, \dots, n$  e um adicional parâmetro de dispersão  $k$ , esse modelo é dado por:

$$P(Y = y_i) = \begin{cases} p_i + (1 - p_i) \left(\frac{k}{k + \mu_i}\right)^k & \text{para } y = 0 \\ (1 - p_i) \frac{\Gamma(y_i + k)}{\Gamma(k) y_i!} \left(\frac{k}{k + \mu_i}\right)^k \left(\frac{\mu_i}{k + \mu_i}\right)^{y_i} & \text{para } y \geq 1 \end{cases}$$

em que o seu valor esperado é definido por  $E(y_i/x_j) = (1 - p_i)\mu_i$  e sua variância por  $Var(y_i/x_j) = (1 - p_i)\mu_i \left(\mu_i p_i + 1 + \frac{\mu_i}{k}\right)$  para  $i = 1, 2, \dots, n$  e  $j = 1, 2, \dots, p$ . Novamente, nota-se que a média é a mesma encontrada na distribuição de Poisson Inflacionada de Zeros, porém, esta distribuição apresenta uma modificação na função de variância dependendo do parâmetro de dispersão, podendo ser uma alternativa para os casos com superdispersão.

Na associação com covariáveis, os modelos inflacionados apresentam funções de ligações distintas. Para a parte inflacionada de zeros, a ligação utilizada é a logística, dada por  $logit(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = g_j' \gamma$ , em que  $g_j$  é um vetor de covariáveis associadas e  $\gamma$  um vetor dos parâmetros desconhecidos,  $i = 1, 2, \dots, n$  e  $j = 1, 2, \dots, p$ . A razão de chances (*odds ratio* - OR) para as covariáveis escolhidas

foram calculadas. A razão de chances estabelece uma relação entre a razão de ocorrência de um dado evento e a razão de não ocorrência do mesmo evento (Vieira, 2004). Para parte não inflacionada de zeros, tem-se a ligação logarítmica, dada por  $\ln(\mu) = b'_j\beta$ , em que  $b_j$  é um vetor de covariáveis associadas e  $\beta$  um vetor de parâmetros desconhecidos, para  $j = 1, 2, \dots, p$ . No caso da ligação logarítmica, foi calculada também a razão de prevalência (RP) das covariáveis envolvidas no modelo. A razão de prevalência é uma medida que expressa a relação entre duas magnitudes de mesma natureza, em que o numerador corresponde a uma categoria que exclui o denominador, (Mérchan-Hamann et al., 2000) gerando um fator de comparação entre as categorias apresentadas.

Na associação com covariáveis, os modelos inflacionados apresentam funções de ligações distintas. Para a parte inflacionada de zeros, a ligação utilizada é a logística, dada por  $\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = g'_j\gamma$ , em que  $g_j$  é um vetor de covariáveis associadas e  $\gamma$  um vetor dos parâmetros desconhecidos,  $i = 1, 2, \dots, n$  e  $j = 1, 2, \dots, p$ . A razão de chances (*odds ratio* - OR) para as covariáveis escolhidas foram calculadas. A razão de chances estabelece uma relação entre a razão de ocorrência de um dado evento e a razão de não ocorrência do mesmo evento (Vieira, 2004). Para parte não inflacionada de zeros, tem-se a ligação logarítmica, dada por  $\ln(\mu) = b'_j\beta$ , em que  $b_j$  é um vetor de covariáveis associadas e  $\beta$  um vetor de parâmetros desconhecidos, para  $j = 1, 2, \dots, p$ . No caso da ligação logarítmica, foi calculada também a razão de prevalência (RP) das covariáveis envolvidas no modelo. A razão de prevalência é uma medida que expressa a relação entre duas magnitudes de mesma natureza, em que o numerador corresponde a uma categoria que exclui o denominador, (Mérchan-Hamann et al., 2000) gerando um fator de comparação entre as categorias apresentadas.

Para a comparação entre os modelos inflacionados, os critérios de *AIC - Akaike Information Criterion* e *BIC - Bayes Information Criterion* (Kuha, 2004) foram utilizados.

Uma medida também utilizada para verificar qual o modelo a ser utilizado foi o cálculo de uma taxa de extra-variação proposta por Bohning et al. (1997), que explicita o quanto da extra-variação presente nos dados pode ser atribuída ao excesso de zeros. Esta taxa baseia-se na variância do modelo de Poisson Inflacionado de Zeros (ZIP) e na diferença entre a média e a variância da variável em questão, e é dada por:

$$\frac{\widehat{E(Y)}(\widehat{\mu} - \widehat{E(Y)})}{s^2 - \bar{y}},$$

em que  $\widehat{E(Y)} = (1 - \widehat{p})\widehat{\mu}$  com  $\widehat{\mu} = \bar{y}$ , sendo  $\widehat{p}$  a estimativa do parâmetro  $p$  dada pela proporção de zeros observados na amostra,  $\bar{y}$  a estimativa da média e  $s^2$  a estimativa da variância de uma variável resposta.

A partir das contagens nulas, dos valores preditos e da taxa de extra-variação, os alimentos foram organizados em ordem crescente de não consumo, subdivididos

em três faixas: de 0% a 10%, de 10% a 50% e acima de 50%. Essas faixas foram fixadas de acordo com o observado na amostra.

O cálculo da taxa de extra-variação foi feito através do programa Microsoft *Excel 2007*. Os ajustes dos modelos de Poisson e Binomial Negativo usuais foram obtidos através do PROC GENMOD do SAS *for Windows*, versão 9.1.3. O programa STATA versão 9.0 foi utilizado para ajustar as distribuições de Poisson e Binomial Negativa inflacionadas de zeros, através das rotinas ZIP e ZINB, comparar os modelos usuais e inflacionados pelo teste de Vuong, comparar os modelos inflacionados através dos critérios de AIC e BIC e gerar os gráficos dos modelos ajustados.

Para cada ajuste dos modelos correspondentes foram obtidos:

- as estimativas dos parâmetros de cada covariável e seu respectivo erro padrão;
- a significância do teste (valor de p), baseado na estatística Wald para testar o efeito considerado (Kodde et al., 1986);
- a razão de prevalência (RP), provenientes nos modelos que apresentam ligação logarítmica e a razão de chances (*odds ratio* - OR), apresentadas nos modelos que possuem ligação *logit*, com seus respectivos intervalos de confiança a 95% (IC95%) (Fletcher e Fletcher, 2006);
- para o modelo binomial negativo inflacionado de zeros, a estimativa do parâmetro de dispersão  $\theta$  e seu respectivo intervalo de confiança, sendo dado por  $\theta = \frac{1}{k}$  (Long e Freese, 2001);
- para os modelos inflacionados de zeros, a significância do teste de Vuong (valor de p);
- e os critérios de AIC e BIC para os modelos descritos.

Gráficos explicitando a diferença entre as probabilidades preditas nos quatro modelos e as proporções observadas foram feitos para verificar o ajuste. (Long e Freese, 2001).

### 3 Resultados

Dos 73 idosos avaliados, 41 (56,16%) eram do sexo feminino, 36 (50,70%) eram eutróficos e 26 (36,62%) eram obesos, 34 (58,62%) referiram tomar medicamentos somente anti hipertensivos, 55 (75,34%) declararam apresentar funcionamento normal do intestino, 47 (64,38%) declararam não praticar atividade física. A idade média foi de 71,51 anos (DP=6,48 anos) e o consumo médio de água por dia foi de 1130,48 ml (DP=588,99 ml).

Para os alimentos com porcentagens de não consumo que não ultrapassaram 10%, o valor predito do modelo nulo de Poisson usual mostrou-se mais próximo dos valores observados, conforme apresentado na Tabela 1.

Tabela 1 - Porcentagem de zeros observados (0% a 10%) e valores preditos nos modelos usuais de Poisson e Binomial Negativo. Avaré, 2009

Alimentos	Observados	Preditos	
	n (%)	Poisson	BN
Arroz branco ou integral cozido com óleo e temperos	0 (0)	0,00	0,20
Carne de boi (bife, cozida, assada), miúdos, vísceras	1 (1)	0,00	0,27
Pão francês, pão de forma, integral, pão doce, torrada	2 (3)	0,00	0,32
Frango (cozido, frito, grelhado, assado)	2 (3)	0,00	0,26
Feijão (carioca, roxo, preto, verde)	4 (5)	0,00	0,53
Tomate	4 (6)	0,00	0,24
Alface	6 (8)	0,00	0,46

Embora o modelo de Poisson tenha se apresentado mais próximo do valor observado (Tabela 1), ao analisar a *deviance* foi encontrada uma superdispersão e, como alternativa, ajustou-se o modelo binomial negativo. Assim, para o consumo de frango, carne de boi, tomate e alface não foram encontradas covariáveis significativamente associadas ao consumo desses alimentos. Já para o arroz branco, feijão e pão francês, as associações significativas ocorreram com as covariáveis sexo e atividade física, conforme mostra a Tabela 2.

Tabela 2 - Ajuste de um modelo Binomial Negativo para o consumo de arroz branco, feijão e pão francês. Avaré, 2009

Alimentos	Variáveis	Estimativa	Erro Padrão	valor p	RP (IC95%)
Arroz	Masculino	0,341	0,120	0,006	1,406 (1,112-1,777)
	Feminino	-	-	-	1,000
Feijão	Masculino	0,581	0,224	0,011	1,788 (1,153-2,775)
	Feminino	-	-	-	1,000
Pão Francês	Atividade Física - não	-0,610	0,230	0,008	0,544 (0,347-0,853)
	Atividade Física - sim	-	-	-	1,000

De acordo com a Tabela 2, para os alimentos mais consumidos como arroz branco (ou integral cozido com óleo e temperos) e feijão (carioca, roxo, preto, verde), o modelo binomial negativo caracterizou uma associação significativa que apresentou os idosos do sexo masculino com uma probabilidade de consumo maior do que os idosos do sexo feminino (Arroz: RP=1,406 (IC95%=1,112-1,777); Feijão: RP=1,788 (IC95%=1,153-2,775)). Para o consumo de pão francês, pão de forma, integral, pão doce e torrada, foi observada associação com a realização de atividade física. Idosos que não praticavam este tipo de atividade mostraram um fator menor de consumo (RP=0,544 (IC95%=0,347-

0,853)) em relação aqueles que as praticavam.

Dos 67 ítems do questionário de frequência alimentar, 32 alimentos possuíam de 10% a 50% de ausência no consumo. Para esses casos, os valores observados foram próximos dos valores preditos no Modelo de Poisson Inflacionado de Zeros (ZIP) e, na maioria dos casos, se aproximaram também do Modelo Binomial Negativo Inflacionado de Zeros (ZINB). Observou-se também que, a taxa de extra-variação foi considerável. Desse modo, para essa porcentagem de respostas nulas os modelos inflacionados apresentaram-se os mais adequados para os dados. A Tabela 3 apresenta alguns itens de 10% a 50% de não consumo.

Tabela 3 - Porcentagem de zeros observados (10% a 50%), valores preditos nos modelos inflacionados de zeros de Poisson (ZIP) e binomial negativo (ZINB) e cálculo da taxa de extra-variação. Avaré, 2009

Alimentos	Observados n (%)	Preditos		
		ZIP	ZINB	Taxa(%)
Ovo (cozido, frito)	7 (10)	0,10	0,11	0,15
Banana	8 (11)	0,12	0,20	0,13
Laranja	12 (17)	0,17	0,20	0,12
Carne de porco (lombo, bisteca)	13 (20)	0,20	0,21	0,24
Café ou chá com açúcar	17 (23)	0,23	0,23	0,30
Leite integral	19 (26)	0,27	0,80	0,19
Embutidos (presunto, mortadela, salsicha)	19 (26)	0,27	0,31	0,19
Peixe (cozido, frito) e frutos do mar	16 (27)	0,27	0,27	0,10
Bolo (simples, recheado)	21 (32)	0,34	0,73	0,21
Batata ou mandioca frita	26 (37)	0,37	0,53	0,18
Refrigerante comum	29 (40)	0,40	0,43	0,08
Maçã, pêra	33 (46)	0,46	0,79	0,13

Para exemplificar os ajustes dos modelos inflacionados, as Tabelas 4 e 5 apresentam, respectivamente, as covariáveis significativas associadas ao consumo ou não de leite integral e embutidos (presunto, mortadela, salsicha).

Tabela 4 - Ajuste de um modelo Binomial Negativo Inflacionado de Zeros para o consumo de leite integral. Avaré, 2009

Parte	Variáveis	Estimativa	Erro Padrão	valor p	OR (IC95%)
Inflacionada	Água ml	0,002	0,001	0,043	1,002 (1,001-1,003)
$\theta$		0,585	0,144		0,585 (0,361-0,948)
Teste Vuong	ZINB vs BN AIC=447,989	z=3,09 BIC=486,807	P(z)=0,001		

De acordo com a Tabela 4, o modelo ZINB mostrou uma única associação significativa, dada pela probabilidade de não consumo de leite integral (parte

inflacionada) associada ao consumo de água em ml por dia. Idosos que consumiram mais água diariamente apresentaram uma probabilidade maior de não consumo de leite integral (OR=1,002 (IC95%=1,001-1,003)).

A Figura 1 mostra os quatro modelos ajustados para frequência do consumo de leite integral associado à covariável significativa. Destacaram-se os modelos inflacionados de Poisson (ZIP) e binomial negativo (ZINB) com probabilidades médias próximas das proporções observadas (a diferença entre as probabilidades médias dos modelos e as proporções observadas apresentou-se em torno de zero), em particular quanto se tratou da frequência do não consumo de leite integral. Os modelos de Poisson e Binomial Negativo apresentaram probabilidades médias mais distantes das proporções observadas.

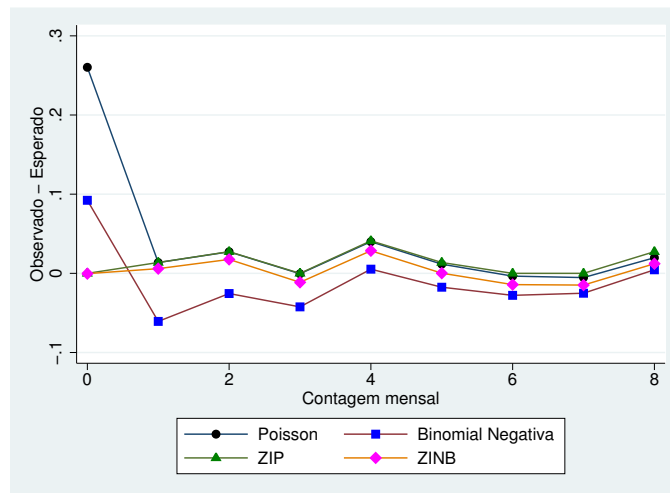


Figura 1 - Diferença entre a proporção observada e as probabilidades médias provenientes dos quatro modelos ajustados para o consumo de leite integral. Avaré, 2009.

A Tabela 5 refere-se ao ajuste do modelo ZIP na associação das covariáveis significativas para o consumo de embutidos (presunto, mortadela e salsicha). Para o não consumo (parte inflacionada) tem-se que idosos que referiram funcionamento normal do intestino apresentaram uma probabilidade menor de não consumo (OR=0,242 (IC95%=0,062-0,945)) em relação aqueles que referiram problemas de constipação e diarreia. Para o consumo (parte não inflacionada), destacaram-se idosos desnutridos os quais apresentaram maior probabilidade (RP=2,045 (IC95%=1,372-3,048)) em relação aos idosos obesos. O uso de medicamentos anti hipertensivos em relação as demais categorias também foi um fator significativo (RP=0,638 (IC95%=0,502-0,811)).

A Figura 2 mostra os quatro modelos ajustados para frequência do consumo de embutidos (presunto, mortadela e salsicha) associado às covariáveis

Tabela 5 - Ajuste de um modelo de Poisson Inflacionado de Zeros para o consumo de embutidos (presunto, mortadela e salsicha). Avaré, 2009

Parte	Variáveis	Estimativa	Erro Padrão	valor p	RP (IC95%)
Não Inflacionada	Desnutrido	0,715	0,204	0,000	2,045 (1,372-3,048)
	Eutrófico	0,240	0,132	0,068	1,271 (0,982-1,645)
	Obeso	-	-	-	1,000
	Uso anti hipertensivo	-0,450	0,122	0,000	0,638 (0,502-0,811)
	Outros medicamentos	-	-	-	1,000
Parte	Variáveis	Estimativa	Erro Padrão	valor p	OR (IC95%)
Inflacionada	Masculino	-1,271	0,772	0,100	0,280 (0,006-1,270)
	Feminino	-	-	-	1,000
	Intestino - normal	-1,420	0,696	0,041	0,242 (0,062-0,945)
	Intestino - C/D	-	-	-	1,000
Teste Vuong	ZIP vs Poisson	z=3,49	P(z)=0,000		
	AIC=331,795	BIC=345,972			

significativas. Destacaram-se os modelos inflacionados de Poisson (ZIP) e binomial negativo(ZINB) com probabilidades médias próximas das proporções observadas (novamente a diferença apresentou-se em torno de zero), em particular quanto se tratou da frequência do não consumo de embutidos. Novamente, os modelos de Poisson e Binomial Negativo apresentaram probabilidades médias mais distantes das proporções observadas.

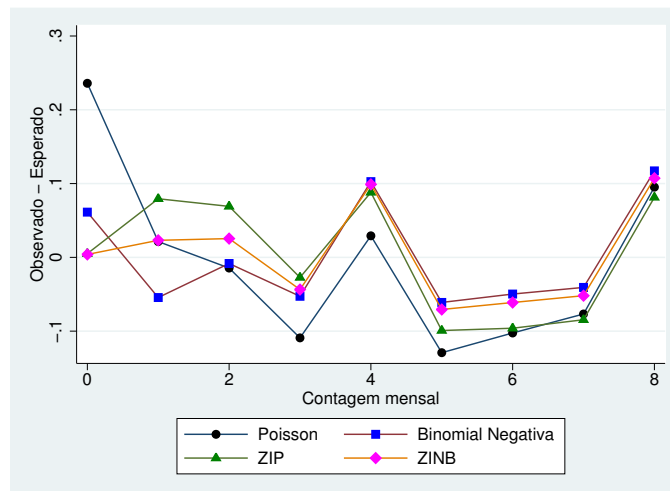


Figura 2 - Diferença entre a proporção observada e as probabilidades médias provenientes dos quatro modelos ajustados para o consumo de embutidos (presunto, mortadela e salsicha). Avaré, 2009.

Os alimentos com mais de 50% de não consumo representam 41,79% do total de itens desse questionário de frequência alimentar. Os valores preditos dos

modelos inflacionados de zeros apresentaram um ajuste novamente mais próximo dos observados, porém, em muitos casos na associação com as covariáveis os processos iterativos para estimação dos parâmetros não convergiram. Além disso, a taxa de extra-variação diminuiu na medida em que se aumentou a porcentagem de respostas nulas. Desse modo, para porcentagem acima de 50% de observações nulas, os modelos inflacionados não foram ajustados. Alguns alimentos dessa faixa estão apresentados na Tabela 6.

Tabela 6 - Porcentagem de zeros observados (50% a 100%), valores preditos nos modelos inflacionados de zeros de Poisson (ZIP) e binomial negativo (ZINB) e cálculo da taxa de extra-variação. Avaré, 2009

Alimentos	Observados n (%)	Preditos		
		ZIP	ZINB	Taxa(%)
Queijos diversos	40 (56)	0,56	0,57	0,07
Melão, melância	39 (57)	0,58	0,88	0,07
Sobremesas, doces, tortas, pudins	41 (60)	0,60	0,64	0,06
Manteiga ou margarina comum	45 (62)	0,65	0,84	0,07
Chocolate, bombom, brigadeiro	44 (63)	0,63	0,78	0,04
Iogurte com frutas	52 (73)	0,75	0,96	0,03
Macarrão com molho sem carne	54 (75)	0,75	0,82	0,08
Leite desnatado	58 (81)	0,81	0,86	0,03
Biscoito recheado, amanteigado	61 (85)	0,85	0,88	0,02
Abacate	59 (87)	0,87	0,89	0,01
Manteiga ou margarina light	66 (90)	0,90	0,90	0,01
Maionese, molho, patê, chantilly	72 (99)	0,99	0,99	0,00

## 4 Discussão

Os dados provenientes de um QFA, em geral, apresentam um número excessivo de zeros, devido ao não consumo de alguns alimentos.

Para os alimentos com baixa porcentagem de não consumo (de 0% a 10%) o modelo mais próximo do observado foi o modelo de Poisson, porém, devido a uma superdispersão presente nos dados, o melhor ajuste foi dado pelo modelo binomial negativo. McCullagh e Nelder(1989) propõe a distribuição binomial negativa como uma alternativa para modelos de contagem nos casos de alta variabilidade. Navarro et al. (2001), estudando os fatores associados ao número de internações recorrentes de pacientes idosos afirma que na comparação entre os resíduos de *deviance* dos modelos de Poisson e binomial negativo, 67,9% das observações mal ajustadas pelo modelo de Poisson são ajustadas pelo modelo binomial negativo. Desse modo, no estudo presente esta foi a alternativa utilizada para contornar a alta variabilidade quando os dados não apresentaram uma porcentagem considerável de zeros.

Horton et al.(2007) apresentam a limitação da distribuição de Poisson no ajuste dos casos com superdispersão quando esta é ocasionada pela presença dos zeros. No caso do estudo presente, os modelos inflacionados de zeros ajustaram bem as variáveis que possuíam de 10% a 50% de zeros. Na literatura, diversos autores

fazem uso também desses modelos como alternativa para dados de contagem que apresentam uma quantidade considerável de zeros, como Slymen et al. (2006), no qual comparam cinco modelos: Poisson, Poisson superdisperso, binomial negativo, ZIP e ZINB para dados de atividade física praticada por mulheres latinas, relatando que o modelo ZIP apresenta o melhor ajuste para avaliar fatores associados à prática ou não de atividade física; Poston e Mckibben (2003) trabalham com modelagem de número de crianças nascidas numa população de mulheres, onde há uma predominância de mulheres sem filhos, explicitando através dos modelos inflacionados os fatores associados a esse baixo nascimento de crianças; Cheung (2002) também aplica os modelos inflacionados para verificar fatores que influenciam no desenvolvimento motor desde a gestação. Assim, os modelos inflacionados têm sido amplamente aplicados para estudo de fatores que influenciam ou não em determinada característica sobre uma população.

Para os alimentos com mais de 50% de zeros, os modelos inflacionados não foram ajustados, apresentando em muitos casos problemas no processo de convergência. Kipnis et al. (2009) propõem outros modelos mais adequados para caracterização e consumo envolvendo o Questionário de Frequência Alimentar. Os autores mostram o uso de modelos com erro de medida, nos quais a informação obtida pelo QFA é utilizada como uma variável instrumental para a correção do consumo alimentar feito através de outro instrumento utilizado na área nutricional, como por exemplo, o recordatório 24 horas.

Um fator que pode ser considerado como explicativo para o não ajuste dos modelos para altas porcentagens de não consumo é a amplitude da frequência dos dados gerados por esse Questionário de Frequência Alimentar, a qual variou da ausência para uma contagem de até cento e vinte vezes de consumo do alimento por mês, dificultando o ajuste dos modelos propostos para esse conjunto de dados. Sheu et al. (2004) ajustam o modelo ZINB para um estudo sobre os fatores associados aos fumantes, no qual a variável resposta é o número de cigarros fumados por dia com variação de consumo diário de zero a 20 cigarros, numa amostra com uma porcentagem de não fumantes de 82,1%. Em Chin e Quddus (2003), o modelo ZIP é utilizado para ajustar dados sobre acidentes envolvendo pedestres, no qual a porcentagem de não acidentes é superior a 80%, com uma variação de zero a 6 acidentes. Bohning et al. (1997) expõem em seu trabalho dados modelados pelo modelo ZIP, no qual o percentual de não acidentes de carro por motorista é de 82,98%, com uma amplitude de zero a 3 acidentes por motorista. Bohning et al. (1999) apresentam um trabalho para comparação de tratamentos distintos para saúde bucal de crianças em seis escolas distintas, nos quais comparou-se o número de problemas odontológicos, registrando uma amplitude de nenhuma até 8 ocorrências, com uma porcentagem considerável de não ocorrência. Desse modo, faz-se necessário estudos mais amplos para se inferir sobre a adequação de tais modelos e as porcentagens de zeros utilizados, não sendo possível estabelecer um padrão para o uso desses modelos, porém, comparando com a literatura observa-se que a amplitude da variável resposta é um fator relevante.

## Conclusões

Para os dados do Questionário de Frequência Alimentar, os modelos inflacionados de zeros se mostraram adequados para o ajuste na porcentagem de 10% a 50% de não consumo. O modelo usual Binomial Negativo apresentou-se coerente para uma porcentagem com até 10% de zeros. Acima de 50% de zeros, outros modelos podem ser testados, como modelos com erros de medidas e variáveis instrumentais. Todavia, outros estudos devem ser feitos para inferir sobre os modelos inflacionados de zeros e as faixas de zeros estabelecidas, através de simulação de dados, levando-se em conta a variação na porcentagem de zeros e a amplitude dos dados.

Este trabalho forneceu uma caracterização integrada do consumo e do não consumo de alimentos nesta população de idosos através do uso de modelos inflacionados de zero. Com isto, pode-se evidenciar que a probabilidade de consumo de arroz, feijão, pão, leite integral e embutidos está associada ao sexo, ao estado nutricional, ao uso de medicamentos e a prática de atividade física. A probabilidade de não consumo aparece associada ao sexo, ao funcionamento do intestino e ao consumo diário de água.

FUMES G.; CORRENTE J. E. Zero-Inflated Models: applications in a Food Frequency Questionnaire. *Rev. Bras. Biom.*, São Paulo, v.28, n.1, p.24-38, 2010.

- **ABSTRACT:** *This study aimed at using zero-inflated models to analyze a food frequency questionnaire (FFQ) applied to elderly individuals from a medium-sized city in inner São Paulo state, investigating the models that can be best fitted to the data according to the percent of zeros found and the factors influencing both intake and non-intake of certain food types. The response variable used was the monthly intake frequency of foods from the FFQ, and the associated co-variables were related to socio-demographic characteristics concerning prevention and reported morbidities. The Poisson, Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial models were used. For a data set with up to 10% of zeros, the usual Negative Binomial model showed to be adequate. Due to data amplitude, the inflated models showed to be adequate only at the percentage of 10% to 50% of zeros. It was also shown that the probability of consumption of the selected food types was associated with gender, nutritional status, medication use and physical activity practice. The probability of non-consumption was observed in association with gender, bowel functioning and daily water consumption.*
- **KEYWORDS:** *Negative binomial distribution; Poisson distribution; zero-inflated distribution; food frequency questionnaire; overdispersion.*

## Agradecimentos

Agradecemos ao CNPq (Processo no. 402533/2007-0) pelo auxílio recebido para o desenvolvimento desta pesquisa.

## Referências

- BOHNING, D.; DIETZ, E.; SCHLATTMANN, P. Zero-Inflated count models and their applications in public health and social science. In: ROST, J.; LANGEHEINE, R.(Ed.). *Applications of latent trait and latent class models in the social sciences*. Munster: Waxmann, 1997. p.333-344.
- BOHNING, D.; DIETZ, E.; SCHLATTMANN, P.; MENDONCA, L.; KIRCHNER, U. The Zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *J. R. Stat. Soc. Series A; Stat. Soc.*, London, v.162, n.2, p.195-209, 1999.
- CHEUNG, Y. B. Zero-inflated models for regression analysis of count data: a study of growth and development. *Stat. Med.*, Hoboken, v.21, n.10, p.1461-1469, 2002.
- CHIN, H. C.; QUDDUS, M. A. Modeling count data with excess zeros: an empirical application to traffic accidents. *Sociol. Methods Res.*, Thousand Oaks-CA, v. 32, n.1, p. 90-116, 2003.
- DODD, K. W.; GUENTHER, P. M.; FREEDMAN, L. S.; SUBAR, A. F.; KIPNIS, V.; MIDTHUNE, D.; TOOZE J. A.; KREBS-SMITH, S. M. Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *J. Am. Diet. Assoc.*, Philadelphia, v.106, p.1640-1650, 2006.
- FLETCHER, R. H.; FLETCHER, S. W. Frequência. In: DUNCAN, M.S.(Ed.). *Epidemiologia clínica: elementos essenciais*. Porto Alegre: Artmed, 2006. p.82-97.
- FISBERG, R. M; COLUCCI, A. C. A.; MORIMOTO, J. M.; MARCHIONI, D. M. L. Questionário de frequência Alimentar para adultos com base em estudo populacional. *Rev. Saúde Pública*, São Paulo, v.3, n.43, p.550-554, 2008.
- HORTON, N. J.; KIM, E.; SAITZ R. A cautionary note regarding count models of alcohol consumption in randomized controlled trials. *BMC Med. Res. Methodol.*, London, v.7, n.9, p.1-9, 2007.
- KIPNIS, V.; MIDTHUNE, D.; BUCKMAN, D. W.; DODD, K. W.; GUENTHER, P. M.; KREBS-SMITH, S. M.; SUBAR, A. F.; TOOZE, J. A.; CARROL, R. J.; FREEDMAN, L. S. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics*, Maryland-USA, p.1-8, 2009. DOI:10.1111/J.1541-0420.2009.01223-X.
- KODDE, D. A.; PALM, F. C. Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*, Chicago, v.54, n.5, p.1243-1248, 1986.
- KUHA, J. AIC e BIC: comparisions of assumptions and performance. *Sociol. Methods Res.*, Thousand Oaks-CA, v.33, n.2, p.188-229, 2004.
- LAMBERT, D. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, Washington, v.34, n.1, p.1-14,1992.

- MCCULLAGH, P.; NELDER, J. A. *Generalized linear models*. Londres: Chapman e Hall, 1989. 511p.
- MÉRCHAN-HAMANN, E; TAUIL, P. L.; COSTA, M. P. Terminologia das medidas e indicadores em epidemiologia: subsídios para uma possível padronização da nomenclatura. *Inf. Epidemiol. SUS*, Brasília, v.9, n.4, p.273-284, 2000.
- MEYER, P. L. *Probabilidade: aplicações à estatística*. Rio de Janeiro: Livros Técnicos e Científicos, 1974, 426p.
- LONG, J. S.; FREESE, J. Predicted probabilities for count models. *Stata J.*, Texas-USA, v.1, n.1, p.51-57, 2001.
- NAVARRO, A.; UTZET, F.; CAMINAL, J.; MARTIN, M. La distribución binomial negativa frente a la de Poisson en el análisis de fenómenos recurrentes. *Gac. Sanit.*, Barcelona, v.15, n.5, p.447-452, 2001.
- POSTON, D.L.; MCKIBBEN, S.L. Using zero-inflated count regression models to estimate the fertility of U.S. women. *J. Mod. Appl. Stat. Methods*, Michigan, v.2, n.2, p.371-379, 2003.
- SHEU, M.; HU, T.; KEELER, T. E.; ONG, M.; SUNG, H. Y. The effect of a major cigarette price change on smoking behavior in California: a zero-inflated negative binomial model. *Health Econ.*, Australian, v.8, n.13, p.781-791, 2004.
- SLATER, B.; MARCHIONI, S. T.; FISBERG, R. M. Validação de questionários de frequência alimentar - QFA - considerações metodológicas. *Rev. Bras. Epidemiol.*, São Paulo, v.6, p.200-208, 2003.
- SLYMEN, D. J.; AYALA, G. A.; ARREDONDO, E. M.; ELDER, J. P. A demonstration of modeling count data with an application to physical activity. *Epidemiol. Persp. Innov.*, United Kingdom, v.3, n.3, p.1-9, 2006.
- VIERA, S. *Bioestatística - tópicos avançados*. Rio de Janeiro: Elsevier, 2004. 216p.
- VUONG, Q. H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, Chicago, v.57, n.2, p.307-333, 1989.

Recebido em 30.09.2009.

Aprovado após revisão em 07.02.2010.