

# DEMANDA POR SERVIÇOS DE SAÚDE: UMA ANÁLISE BASEADA EM DADOS CONTÁVEIS

Victor Rodrigues de Oliveira\*  
Flávio de Oliveira Gonçalves\*\*

## Resumo

O objetivo deste estudo é analisar quais os determinantes da demanda por cuidados médicos, diante das transformações dos sistemas de saúde, de forma a alocar os recursos de maneira mais eficiente e organizar a infraestrutura dos serviços. Para atender tal necessidade, a estratégia empírica consiste na estimação de modelos de dados contáveis. A partir dos microdados da Pesquisa Nacional por Amostra de Domicílios de 2008, a parametrização especificada pelo modelo *Finite Mixture* mostrou-se a mais adequada considerando-se o número médio previsto de consultas como critério de escolha. Os resultados indicaram duas subpopulações com comportamentos distintos: o grupo “*low-users*” e o grupo “*heavy-users*”. A renda contribui de forma positiva nos dois grupos, apesar de seu baixo valor em módulo. As mulheres, os negros e os indígenas são os maiores demandantes de cuidados médicos, assim como os indivíduos que possuem plano de saúde. A escolaridade apresentou sinal positivo em ambas as populações para a *dummy* que indica 15 anos ou mais de estudo. Verificou-se, também, a existência de desigualdades sociais no acesso aos serviços de saúde favorecendo os indivíduos de renda elevada. Neste sentido, o provimento gratuito de planos ou seguros de saúde e de um maior fornecimento de medicamentos para esses grupos se apresenta como uma opção de política pública.

**Palavras-chave:** demanda por saúde; modelos de dados contáveis; *Finite Mixture*.

**Classificação JEL:** C10, I19.

## Abstract

The objective of this study is to analyze the determinants of the demand for medical care in order to allocate resources more efficiently and organize the infrastructure of services. The empirical strategy consists in the estimation of count data models. From the microdata from the National Household Sample Survey to 2008, Finite Mixture model proved to be the most appropriate considering the average number of consultations provided as a criterion of choice. The results indicated two subpopulations with distinct behaviors: low-users and heavy-users. Income contributes positively in both groups, despite its low value in the module. Women, blacks and indigenous people are demanding more medical care, as well as individuals who have health insurance. Educational level showed positive signal in both populations for the dummy that indicates 15 or more years of study. There was also the existence of social inequalities in access to health services favoring high-income individuals. In this sense, the provision of free plans or health insurance and a greater supply of medicines for these groups are presented as a public policy option.

**Key words:** demand for health; count data models; Finite Mixture.

**JEL Classification:** C10, I19.

## Área 11 - Economia Social e Demografia Econômica

\* Aluno do Curso de Mestrado em Desenvolvimento Econômico, Universidade Federal do Paraná (UFPR), Bolsista CNPq, Brasil. E-mail: victor5491@gmail.com

\*\* Professor adjunto do Departamento de Economia da UFPR. E-mail: flaviogonsalves@hotmail.com

# DEMANDA POR SERVIÇOS DE SAÚDE: UMA ANÁLISE BASEADA EM DADOS CONTÁVEIS

## 1. Introdução

Um dos principais debates em economia da saúde, atualmente, diz respeito aos condicionantes da utilização dos serviços de saúde. Este processo é resultado da interação entre a necessidade do indivíduo e a sua decisão de procurar cuidados médicos. Os determinantes desta utilização estão associados às características dos indivíduos, do sistema de saúde e ao padrão de prática médica (RIBEIRO, 2005). Como apontado por Rodrigues et al. (2010, p. 6), diversos países têm passado por mudanças quanto à estrutura etária e ao padrão de mortalidade e morbidade, com consequências sobre custos e utilização de serviços médicos e ambulatoriais. Além disso, as transformações nos sistemas de saúde – reorganização dos modelos de atenção – e a incorporação de novas tecnologias, dentre outros fatores, têm alterado os padrões de utilização destes serviços. Desta forma, estas mudanças têm impacto sobre a alocação dos recursos e a organização da infraestrutura dos serviços.

A importância da provisão dos bens e serviços tem se tornado um tema central na agenda dos *policymakers*, o que pode ser observado na Tabela 1. Nota-se o aumento expressivo dos gastos totais *per capita* com saúde para todos os países selecionados, com exceção da Argentina e do pequeno aumento para o Chile. Concomitantemente, o gasto total com saúde como porcentagem do PIB e a expectativa de vida também aumentaram. Os Estados Unidos permanecem desde 2000 como o país que tem o maior gasto em saúde como porcentagem do PIB, uma mudança que representou um crescimento de 34% no período analisado. Além disso, verifica-se, também, que possuem o maior gasto total *per capita* com saúde nos dois anos analisados.

**Tabela 1 - Gasto total<sup>1</sup> com saúde como % do PIB, gasto total *per capita* com saúde (US\$ PPP) e expectativa de vida (em anos) em países selecionados de 2000 a 2010**

País	Gasto total com saúde como % do PIB		Gasto total <i>per capita</i> com saúde		Expectativa de vida (em anos) <sup>2</sup>	
	2000	2010	2000	2010	2000	2009
Estados Unidos	13,4	17,9	4.704	8.362	79	79
França	10,1	11,9	2.546	4.691	79	81
Alemanha	10,4	11,6	2.679	4.668	78	80
Portugal	9,3	11,0	1.654	2.366	77	79
Costa Rica	6,5	10,9	471	811	77	79
Reino Unido	7,0	9,6	1.828	3.503	78	80
Brasil	7,2	9,0	502	990	70	73
Finlândia	7,2	9,0	1.853	3.984	78	80
Argentina	9,2	8,1	839	742	75	75
Chile	8,3	8,0	768	947	77	79

Fonte: World Health Organization (2012)

Nota: <sup>1</sup> Soma do gasto despendido pelo setor público e privado

<sup>2</sup> Dados disponíveis somente para o ano de 2009

No Brasil, o gasto *per capita* com o setor de saúde cresceu a uma taxa de 7% ao ano no período de 2000 a 2010, indicando que o gasto absoluto nesse setor cresceu a taxas superiores que a taxa de crescimento populacional (1,17% ao ano). Além disso, a fração do PIB gasto em saúde aumentou de 7,2% para 9,0% neste período.

Assim, a importância do investimento em saúde pode ser vista em termos de bem estar social. Do ponto de vista individual o consumo de bens e serviços de saúde afeta diretamente o bem estar na medida em que o estado de saúde determina o grau de felicidade dos indivíduos. O provimento de serviços de saúde impacta o crescimento econômico uma vez que compõe o estoque de capital humano e determina a capacidade produtiva da economia. Além disso, é um setor intensivo em mão-de-obra para os

serviços de baixa e média complexidade e intensivo em tecnologia para os setores de alta complexidade (ANDRADE et al., 2011, p. 6).

A despeito da relevância do tema poucos estudos têm explorado os determinantes da demanda por saúde. A maioria dos estudos se preocupa em investigar os condicionantes dos gastos totais em saúde a partir de informações agregadas (ver GERDTHAM e JONSSON (1988), HITIRIS e POSNETT (1992)). Na literatura nacional os estudos que procuram investigar tais determinantes destacam a renda como um determinante importante dos gastos privados em saúde (ANDRADE et al., 2006).

O estudo de Kassouf (2005) utilizando os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), de 1998, analisou os determinantes da procura por atendimento à saúde devido à doença, nas áreas urbana e rural do Brasil. Através de um modelo de escolha binária, o fator mais importante na determinação da procura por atendimento à saúde foi o estado de saúde do indivíduo, tanto na área urbana quanto no meio rural. Na área urbana, o nível de escolaridade também teve grande impacto em reduzir a procura por serviços de saúde.

Brito (2005), por sua vez, estudou a demanda por serviços de saúde para a população de titulares e beneficiários da Fundação Sabesp de Seguridade (SABESP/REV) nos anos de 2001 a 2003. Este estudo investigou os determinantes dos gastos considerando uma população coberta por um plano de saúde. Os principais resultados encontrados indicaram grande importância do sexo, da idade e do estado de saúde passada na determinação dos gastos ambulatoriais. Para os gastos hospitalares apenas duas variáveis foram determinantes: a idade e os indivíduos que faleceram. Os resultados também indicaram que a elasticidade dos gastos aumenta com a idade e que os indivíduos que faleceram possuem elevada elasticidade.

Os estudos sobre demanda por saúde têm se concentrado em modelos de escolha binária. Pouca atenção tem sido dada para os modelos de dados contáveis, com destaque para Noronha e Andrade (2005).

Uribe (2008) a partir dos dados da PNAD 2003 estimou vários modelos de dados contáveis e determinou que a especificação mais adequada para estudar a demanda sobre saúde é o modelo *Hurdle* Binomial Negativo. Este modelo é estimado em duas etapas, em que na primeira determinam-se os condicionantes da procura médica e na segunda considera-se somente os usuários de serviços de saúde. Os resultados indicaram que existe grande desigualdade no acesso à saúde da população favorecendo os indivíduos que possuem renda elevada.

Desta forma, o objetivo deste estudo é investigar os condicionantes da demanda por serviços de saúde, compreendido como o número de consultas médicas realizadas no ano. Para atender tal necessidade diversos aspectos devem ser considerados, tais como idade, sexo, renda, escolaridade, hábitos de vida e qualidade de vida, uma vez que tem efeito sobre o perfil epidemiológico, as necessidades e as preferências pelos serviços de saúde.

Utilizando-se modelos de dados contáveis aplicados aos microdados da PNAD de 2008, estimou-se o número de consultas médicas realizadas por indivíduo ao ano. A decisão de ir ao médico está diretamente relacionada ao estado de saúde individual, o que justifica a adoção do indivíduo como unidade básica de estudo. A escolha das variáveis explicativas, isto é, a parametrização do modelo, assim como a interpretação dos resultados teve como base a teoria do capital humano aplicada à economia saúde (ver GROSSMAN, 1972).

O trabalho será dividido como segue. Após esta introdução, a próxima seção apresentará os modelos de dados contáveis enfatizando suas vantagens e desvantagens, além da base de dados utilizada. A seção 3 apresentará algumas estatísticas descritivas e o resultado dos modelos estimados, com destaque para a especificação escolhida. A última seção apresentará as considerações finais e os desdobramentos destes resultados.

## **2. Modelos Empíricos**

Esta seção tem por objetivo apresentar os métodos de dados contáveis que serão utilizados para estimar os determinantes da demanda por serviços médicos. Apresentar-se-á as distribuições de

probabilidade e as hipóteses utilizadas em cada modelo de forma a evidenciar as vantagens e desvantagens destes em conformidade com Cameron e Trivedi (2005).

## 2.1 Modelos de Dados Contáveis

Em muitas aplicações a variável dependente é um número inteiro não negativo ou contável. Diferentemente do modelo de regressão clássico, a variável dependente em um modelo de dados contáveis é discreta, com uma distribuição de probabilidade que assume somente valores inteiros não negativos, sendo um evento a realização desta variável estocástica. Assim, um evento pode ser interpretado como a ocorrência de um ponto dentro de um processo em que há uma taxa de ocorrência específica daquele evento. Desta forma, o modelo padrão de dados contáveis é aquele que utiliza a distribuição Poisson, também conhecida como distribuição dos eventos raros<sup>1</sup>.

O modelo básico de dados contáveis utiliza a distribuição Poisson. Esta distribuição pode ser inadequada quando há uma grande proporção de zeros coexistindo com um grande número de dados positivos, criando uma dificuldade de modelar uma mudança da variável dependente. Além disso, pode não se ajustar bem aos dados quando há sobredispersão, isto é, a razão entre a variância condicional e a média condicional é maior que um.

Seja  $Y$  uma variável discreta e  $\mu$  a taxa de ocorrência de um evento. Assim,  $Y$  segue uma distribuição Poisson com a seguinte função densidade de probabilidade (fdp):

$$Pr[Y = y] = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots, n \quad (1)$$

onde  $E[Y] = V[Y] = \mu$ , ou seja, a média é igual à variância. Esta propriedade é conhecida como equidispersão, porém é frequentemente violada. O modelo de regressão Poisson é obtido parametrizando-se a relação entre  $\mu$  e os regressores ( $x$ ) como segue:

$$\mu_i = \exp(x_i' \beta), \quad i = 1, \dots, n \quad (2)$$

onde  $\beta$  é um vetor de parâmetros a ser estimado. Além disso,  $E(y_i | x_i) = \exp(x_{1i} \beta_1) \exp(x_{2i} \beta_2) \dots \exp(x_{ki} \beta_k)$ .

Para verificar se há sobredispersão, especificou-se a seguinte forma funcional para a variância:  $V[y_i | x_i] = \mu_i + \alpha g(\mu_i)$  onde  $\alpha$  é um parâmetro desconhecido e  $g(\mu_i) = \mu_i^2$ . O teste estatístico para  $H_0: \alpha = 0$  versus  $H_1: \alpha > 0$  pode ser computado estimando a seguinte regressão por *Ordinary Least Squares* (OLS):  $[(y_i - \hat{\mu}_i)^2 - y_i] / \hat{\mu}_i = \alpha [g(\hat{\mu}_i) / \hat{\mu}_i] + u_i$ , onde  $u_i$  é um termo de erro. A estatística- $t$  é assintoticamente normal sobre a hipótese nula de que não há sobredispersão (Cameron e Trivedi, 2005, p. 670).

A regressão Poisson é muito restritiva. Em primeiro lugar, é frequente que os dados contáveis exibam sobredispersão como consequência da heterogeneidade não observada, dificultando, assim, a estimação da taxa de ocorrência do evento. Em segundo lugar, a distribuição Poisson assume que os eventos ocorrem de maneira independente ao longo de tempo, isto é, a probabilidade do paciente consultar o médico pela  $j$ -ésima vez é independente da  $(j + 1)$ -ésima e  $(j - 1)$ -ésima consulta. Por conseguinte, aplica-se de maneira muito limitada ao contexto econômico.

<sup>1</sup> A lei dos eventos raros estabelece que o número total de eventos ocorridos em um dado intervalo de tempo segue aproximadamente a distribuição Poisson se qualquer evento pode acontecer em um número de tentativas grande e com uma probabilidade de ocorrência pequena. Seja  $Y_{n,\pi}$  o número total de sucessos em um número  $n$  de tentativas de Bernoulli com probabilidade de ocorrência  $\pi$ . No limite quando  $n \rightarrow \infty, \pi \rightarrow 0$  e  $n\pi = \mu > 0$  tem-se:

$$Pr = [Y_{n,\pi} = k] = \left[ \binom{n}{k} \pi^k (1 - \pi)^{n-k} \right] = \frac{\mu^k e^{-\mu}}{k!}$$

Sabe-se que os modelos do tipo Binomial Negativo (BN) apoiam-se em hipóteses mais plausíveis, uma vez que a razão entre a variância e a média condicional é proporcional à média condicional.

Considere a distribuição Poisson novamente como  $f(y|\lambda) = \exp(-\lambda) \lambda^y / y!$ . Suponha que  $\lambda$ , a taxa de ocorrência do evento, é aleatória, em vez de uma função determinística do vetor de regressores  $x$ . Assuma que  $\lambda = \mu v$ , onde  $\mu$  é uma função determinística de  $x$ , por exemplo,  $\exp(x'\beta)$ , e  $v > 0$  é iid com fdp  $g(v|\alpha)$ . A distribuição Binomial Negativa é mais adequada quando há sobredispersão, adicionando  $\alpha$  como um parâmetro para capturar a heterogeneidade ( $\lambda$ ) não observada ( $v$ ). A densidade marginal de  $y$  não condicionada ao parâmetro aleatório  $v$ , mas condicionada aos parâmetros  $\mu$  e  $\alpha$ , é obtida integrando-se com relação à  $v$ . Assim

$$h[y|\mu, \alpha] = \int f(y|\mu, v)g(v|\alpha)dv \quad (3)$$

Se  $f(y|\lambda)$  é a densidade Poisson e  $g(v) = v^{\delta-1} e^{-v\delta} \delta^\delta / \Gamma(\delta)$ ,  $v, \delta > 0$ , é a função densidade de probabilidade Gama<sup>2</sup> com  $E[v] = 1$  e  $V[v] = 1/\delta$ , a distribuição BN pode ser obtida como a mistura entre a Poisson e a Gama como segue:

$$h[y_i|\mu_i, \delta] = \int_0^\infty \frac{e^{-\mu_i v} (\mu_i v)^{y_i}}{y_i!} \frac{v^{\delta-1} e^{-v\delta} \delta^\delta}{\Gamma(\delta)} dv, \quad y_i = 0, 1, 2, \dots, n$$

$$h[y_i|\mu_i, \delta] = \frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left( \frac{\mu_i}{\mu_i + \alpha^{-1}} \right)^{y_i}, \quad y = 0, 1, 2, \dots, n \quad (4)$$

onde  $\alpha = 1/\delta$ . Da mesma forma que para o modelo Poisson,  $E(y_i|x_i) = \exp(x_{1i}\beta_1) \exp(x_{2i}\beta_2) \dots \exp(x_{ki}\beta_k) \exp(\varepsilon_i)$ , com o acréscimo de  $\varepsilon_i$ , onde  $\exp(\varepsilon_i) = \delta$  representa um termo para capturar a heterogeneidade latente.

Para este modelo  $E[y|\mu, \alpha] = \mu$  e  $V[y|\mu, \alpha] = \mu(1 + \alpha\mu)$  e, portanto, a variância excede a média, uma vez que  $\alpha > 0$  e  $\mu > 0$ . Para verificar se há sobredispersão, recorre-se a um teste de Razão de Verossimilhança (LR) da hipótese  $H_0: \alpha = 0$  versus  $H_1: \alpha > 0$ , considerando-se a seguinte forma funcional para a variância:  $V[y_i|x_i] = \omega_i = \mu_i + \alpha\mu_i^p$ , onde  $p$  especifica o tipo da distribuição BN: se  $p = 1$  tem-se a distribuição Binomial Negativa do tipo 1 (BN1); se  $p = 2$  tem-se a distribuição Binomial Negativa do tipo 2 (BN2). Além disso, se  $y|\lambda$  segue uma distribuição Poisson e  $\lambda = \mu v$  com  $E[v] = 1$  o modelo sempre apresentará sobredispersão.

A utilização dos serviços de saúde ocorre em duas etapas: na primeira, o paciente inicialmente escolhe se vai ou não procurar tratamento, sendo esta uma escolha individual; posteriormente, se a escolha for positiva, o médico influenciaria a frequência das consultas.

Para atender tal necessidade, estima-se o modelo *Hurdle*. O modelo *Hurdle* ou *Two-part Model* relaxa o pressuposto de que os zeros e os demais dados positivos têm o mesmo processo gerador dos dados. Para tanto, os zeros são determinados pela densidade  $f_1(\cdot)$ , de tal forma que  $Pr[y_i = 0] = f_1(0)$ . Os valores positivos são determinados pela densidade truncada  $f_2(y_i | y_i > 0) = f_2(y_i) / (1 - f_2(0))$ , que é multiplicada por  $Pr[y_i > 0] = 1 - f_1(0)$  para garantir que a soma das probabilidades seja igual a 1. Portanto, a densidade pode representada por:

$$g(y_i) \sim \begin{cases} f_1(0) & \text{se } y_i = 0 \\ \frac{1 - f_1(0)}{1 - f_2(0)} f_2(y_i) & \text{se } y_i \geq 1 \end{cases} \quad (5)$$

<sup>2</sup> Note que isso é uma imposição ao modelo. Modelos que assumem a própria distribuição das variáveis seriam mais aceitáveis.

O primeiro estágio do modelo *Hurdle* é um modelo *Logit* para determinar a probabilidade de um indivíduo demandar consultas. O segundo estágio do modelo *Hurdle*, onde se estima o número esperado médio de consultas, pode ser aquele que considera a distribuição Poisson ou a BN truncada em zero<sup>3</sup>. Nesta etapa considera-se apenas a sub-amostra de indivíduos que utilizaram serviços de saúde. A especificação dos dois processos de decisão pode ser feita, por sua vez, com as mesmas variáveis explicativas, sendo interpretadas de maneiras diferentes, dependendo da etapa do processo.

O modelo *Hurdle* é construído a partir da especificação de duas funções de verossimilhança parametricamente independentes, cada uma relativa a uma etapa do processo de estimação. Desta forma, estimações simultâneas não gerariam ganhos de eficiência à estimação (PALERMO, PORTUGAL e SOUZA, 2005).

Para o modelo *Hurdle Poisson* pode-se parametrizar (5) como:

$$Pr[Y = y_i] = \begin{cases} \eta_i + (1 - \eta_i)e^{-\mu_{i1}} & \text{se } y_i = 0 \\ \frac{\{1 - [\eta_i + (1 - \eta_i)e^{-\mu_{i2}}]\} e^{-\mu_{i2}} \mu_{i2}^{y_i}}{(1 - e^{-\mu_{i2}}) y_i!} & \text{se } y_i > 0 \end{cases} \quad (6)$$

onde  $\eta_i$  representa a probabilidade de obter-se  $y_i = 0$ .

Considerando  $\mu_{i1} = \exp(x_i' \beta_1)$  e  $\mu_{i2} = \exp(x_i' \beta_2)$  como os parâmetros da distribuição para o caso de zeros e para um conjunto de números positivos  $J = \{1, 2, \dots\}$ , respectivamente, pode-se parametrizar (5) como uma distribuição BN:

$$y_i \sim \begin{cases} (1 + \alpha_1 \mu_{i1})^{-1/\alpha_1} & \text{se } y_i = 0 \\ \frac{\Gamma(\alpha_2^{-1} + y_i)}{\Gamma(\alpha_2^{-1}) \Gamma(y_i + 1)} \left( \frac{1}{(1 + \alpha_2 \mu_{i2})^{1/\alpha_2} - 1} \right)^{-1/\alpha_2} \left( \frac{\mu_{i2}}{\mu_{i2} + \alpha_2^{-1}} \right)^{y_i} & \text{se } y_i > 0 \end{cases} \quad (7)$$

definindo uma função indicadora  $I[y_i \in J] = 1$  se  $y_i \in J$  e  $I[y_i \in J] = 0$  se  $y_i = 0$ .

Os modelos *Finite Mixture*, entretanto, não necessariamente trabalham com a dicotomia usuários e não usuários do modelo *Hurdle*. Estes modelos se caracterizam por constituírem uma adição de densidades de  $M$  distintas populações para analisar a heterogeneidade não observada. Considere  $y$  como a combinação linear de  $M$  diferentes densidades, onde a  $j$ -ésima densidade é  $f_j(y_i | \theta_j)$ ,  $j = 1, 2, \dots, m$ . Portanto, um modelo *Finite Mixture* com  $m$  componentes é dado por:

$$f(y_i | \theta, \pi) = \sum_{j=1}^M \pi_j f_j(y_i | \theta_j), \quad 0 < \pi_j < 1, \quad \sum_{j=1}^M \pi_j = 1 \quad (8)$$

onde é possível parametrizar  $\pi_j$  como uma função logística do vetor de covariadas  $z_i$ ,  $\pi_j = \frac{\exp(z_i \phi_j')}{1 + \exp(z_i \phi_j')}$ .

A partir de (8), o objetivo é estimar o vetor de parâmetros  $\theta_j$ ,  $j = 1, 2, \dots, M$ . Se o modelo BN2 é utilizado, então  $f_j(y_i | \theta_j)$  é uma fdp do tipo BN2 (equação (3)) com parâmetros  $\mu_j = \exp(x' \beta_j)$  e  $\alpha_j$  e, portanto,  $\theta_j = (\beta_j, \alpha_j)$ . O número de densidades  $M$ , na prática, deve ser estimado, porém sejam  $d_i = (d_{i1}, d_{i2}, \dots, d_{iM})$  *dummies* tal que  $d_{ij} = 1$  e  $\sum_{j=1}^M d_{ij} = 1$ , indicando que  $y_i$  foi gerado pela  $j$ -ésima classe latente para  $i = 1, 2, \dots, n$  observações. Este modelo especifica que  $(y_i | d_i, \theta_j, \pi_j)$  é independentemente distribuído com densidade igual a:

<sup>3</sup> O modelo Binomial Negativo Truncado em Zero é utilizado para se distinguir situações que não têm zeros de situações que não podem ter zeros, como no caso a frequência ao médico ou o tempo de internação.

$$\sum_{j=1}^M d_{ij} f_j(y_i | \theta_j) = \prod_{j=1}^M f_j(y_i | \theta_j)^{d_{ij}} \quad (9)$$

onde  $d_{ij}$  são *iid* com distribuição multinomial.

A probabilidade de que  $y_i$  pertença a subpopulação  $m$  é dada por:

$$Pr[y_i \in \text{a subpopulação } m | x_i, y_i, \theta_j] = \frac{\pi_m f_m(y_i | x_i, \theta_m)}{\sum_{j=1}^M \pi_j f_j(y_i | x_i, \theta_j)} \quad (10)$$

A densidade dos  $M$  componentes do modelo *Finite Mixture* é especificada como:

$$f(y_i | \theta_1, \dots, \theta_M; \pi_1, \dots, \pi_M) = \sum_{j=1}^M \pi_j \frac{\Gamma(y_i + \psi_{j,i})}{\Gamma(\psi_{j,i}) \Gamma(y_i + 1)} \left( \frac{\psi_{j,i}}{\mu_{j,i} + \psi_{j,i}} \right)^{\psi_{j,i}} \left( \frac{\mu_{j,i}}{\mu_{j,i} + \psi_{j,i}} \right)^{y_i} \quad (11)$$

onde  $\mu = \exp(\mathbf{x}\beta)$  e  $\psi = (1/\alpha) \mu^{k^4}$ .

O modelo *Finite Mixture* é semiparamétrico uma vez que não é necessário fazer hipóteses sobre as distribuições de cada subpopulação. Assim, uma amostra aleatória desenhada a partir da população irá conter proporções  $\pi_1, \pi_2, \dots, \pi_j$  das  $j$  subpopulações, embora não se pode observar que caso pertence a qual subpopulação, isto é, as classes são latentes. Considerando-se algumas condições de regularidade é possível estimar  $(\pi_j, \theta_j), j = 1, 2, \dots, M$ .

Finalmente, os modelos *Zero-Inflated*, assim como o modelo *Hurdle*, também tratam da sobredisposição e do excesso de zeros, porém adotam outra estratégia ao invés da truncagem. Para cada observação existem dois processos possíveis de geração de dados, o resultado de um ensaio de Bernoulli determina qual o processo é usado. Para cada  $i$  o processo 1 é escolhido com probabilidade  $\rho_i$  e o processo 2 é escolhido com probabilidade  $1 - \rho_i$ . O primeiro processo gera somente zeros, enquanto o segundo processo,  $g(y_i | x_i)$ , gera os demais dados contáveis considerando uma distribuição Poisson ou BN, de tal forma que:

$$y_i \sim \begin{cases} 0 & \text{com probabilidade } \rho_i \\ g(y_i | x_i) & \text{com probabilidade } 1 - \rho_i \end{cases} \quad (12)$$

A probabilidade de  $\{Y = y_i | \mathbf{x}_i\}$  é

$$Pr(Y = y_i | \mathbf{x}_i, \mathbf{z}_i) = \begin{cases} \rho_i (\mathbf{z}_i \gamma') + \{1 - \rho_i (\mathbf{z}_i \gamma')\} g(0 | \mathbf{x}_i) & \text{se } y_i = 0 \\ \{1 - \rho_i (\mathbf{z}_i \gamma')\} g(y_i | \mathbf{x}_i) & \text{se } y_i > 0 \end{cases} \quad (13)$$

onde  $\mathbf{z}_i$  representa o vetor de regressores do primeiro estágio do modelo *Zero-Inflated*.

A média e a variância do modelo *Zero-Inflated Poisson (ZIP)* são:

$$\begin{cases} E(y_i | \mathbf{x}_i, \mathbf{z}_i) = \mu_i (1 - \rho_i) & (14.1) \\ V(y_i | \mathbf{x}_i, \mathbf{z}_i) = \mu_i (1 - \rho_i) (1 + \mu_i \rho_i) & (14.2) \end{cases}$$

A média e a variância do modelo *Zero-Inflated Negative Binomial (ZINB)* são:

$$\begin{cases} E(y_i | \mathbf{x}_i, \mathbf{z}_i) = \mu_i (1 - \rho_i) & (15.1) \\ V(y_i | \mathbf{x}_i, \mathbf{z}_i) = \mu_i (1 - \rho_i) (1 + \mu_i (\rho_i + \alpha)) & (15.2) \end{cases}$$

<sup>4</sup> Se  $k = 1$  tem-se a distribuição BN2; se  $k = 2$  tem-se a distribuição BN1; se  $k = 0$  tem-se o *fits best*.

Nota-se que o excesso de zeros implica sobredispersão em ambos os modelos.

Finalmente,  $\rho_i$  pode ser parametrizado por uma função logística do vetor de covariadas  $\mathbf{z}_i$ . Portanto:

$$\rho_i = \frac{\exp(\mathbf{z}_i\gamma')}{1 + \exp(\mathbf{z}_i\gamma')} \quad (16)$$

Esta distribuição é um caso específico de *Finite Mixture* com distribuição degenerada cuja massa está concentrada no zero. Ressalta-se que os modelos *Zero-Inflated* são variações das regressões Poisson e BN, onde os zeros provêm parcialmente destas distribuições (ZIP ou ZINB) com probabilidade  $(1 - \rho_i)$  e parcialmente da distribuição de não usuários  $\rho_i$ .

## 2.2 Base de Dados

A base de dados utilizada neste estudo são os microdados oriundos da Pesquisa Nacional por Amostra de Domicílios (PNAD) realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A PNAD é uma pesquisa por amostragem probabilística de domicílios, realizada em todo o território nacional. A pesquisa, realizada desde o final da década de 1960, conta com um questionário de caráter permanente, envolvendo perguntas relativas a características domiciliares e pessoais, tais como: tamanho da família, renda domiciliar, nível educacional dos moradores entre outras. Em alguns anos são investigadas características socioeconômicas e demográficas com caráter suplementar, tais como: migração, saúde, segurança alimentar e outros temas.

As informações foram obtidas para o ano de 2008 que apresenta os resultados do levantamento suplementar de saúde. A Pesquisa Suplementar de Saúde compreendeu entre os tópicos pesquisados os seguintes temas: morbidade, cobertura de plano de saúde, acesso a serviço de saúde, utilização a serviços de saúde, internação, atendimento de urgência no domicílio, violência, acidente de trânsito, sedentarismo entre outros.

A modelagem clássica está baseada no pressuposto de amostras independentes e identicamente distribuídas e amostragem aleatória simples. Estas hipóteses não são válidas para dados obtidos através de pesquisas amostrais como a PNAD, uma vez que as estimativas das variâncias (precisão dos estimadores) são influenciadas pela conglomeração, pela estratificação e pelos pesos de cada unidade amostrada. Assim, o estudo promoveu o ajustamento da amostra visando respeitar o plano amostral que lhe deu origem.

## 3. Determinantes da Demanda por Serviços de Saúde<sup>5</sup>

Na PNAD 2008 existe um conjunto de variáveis que são números inteiros não negativos, de tal forma que podem ser utilizadas como variável dependente em modelos de dados contáveis: o número de consultas realizadas no ano, o número de atendimentos realizados nas duas semanas anteriores à semana de referência, entre outras. Este estudo adotou o número de consultas realizadas no ano como variável dependente, uma vez que é mais abrangente em termos de horizonte temporal e de frequência. A Tabela 2 apresenta as variáveis utilizadas neste estudo e algumas de suas estatísticas descritivas.

**Tabela 2 – Descrição das variáveis utilizadas**

Variáveis	Descrição	Média	Desvio-padrão	Min	Max
	Variável dependente				
consultou	Número de consultas realizadas	2,804	4,761	0	98

(continua)

<sup>5</sup> Devido ao grande número de resultados obtidos estes foram omitidos e podem ser solicitados junto aos autores.



(continuação)

Variáveis	Descrição	Média	Desvio-padrão	Min	Max
<b>Variáveis independentes</b>					
<b>Auto avaliação do estado de saúde</b>					
saude2	Se considera a saúde boa = 1	0,538	0,499	0	1
saude3	Se considera a saúde regular = 1	0,209	0,407	0	1
saude4	Se considera a saúde ruim = 1	0,035	0,184	0	1
saude5	Se considera a saúde muito ruim = 1	0,009	0,094	0	1
<b>Indicadores de qualidade de saúde</b>					
banheiro	Tem banheiro no domicílio = 1	0,957	0,203	0	1
esgoto	Se escoadouro do banheiro é esgoto = 1	0,521	0,500	0	1
lixo	Se lixo domiciliar é coletado diariamente = 1	0,801	0,399	0	1
luz	Se iluminação do domicílio é energia = 1	0,985	0,123	0	1
filtro	Se tiver filtro d'água = 1	0,513	0,500	0	1
agua	Se tiver água canalizada em pelo menos um cômodo do domicílio = 1	0,927	0,260	0	1
abaste	Se forma de abastecimento de água é rede geral com canalização interna = 1	0,833	0,373	0	1
<b>Atributos adquiridos</b>					
renda	Renda mensal domiciliar <i>per capita</i>	672,95	1067,08	0	150000
plano de saúde	Se tem direito a plano de saúde = 1	0,265	0,441	0	1
ler	Se sabe ler e escrever = 1	0,090	0,286	0	1
estudo1	Se 1 a 4 anos de estudo = 1	0,243	0,429	0	1
estudo2	Se 5 a 8 anos de estudo = 1	0,266	0,442	0	1
estudo3	Se 9 a 11 anos de estudo = 1	0,281	0,449	0	1
estudo4	Se 12 a 14 anos de estudo = 1	0,045	0,208	0	1
estudo5	Se 15 anos ou mais anos de estudo = 1	0,065	0,247	0	1
<b>Atributos natos</b>					
idade2	Se idade entre 14 e 19 anos = 1	0,128	0,334	0	1
idade3	Se idade entre 20 e 39 anos = 1	0,382	0,486	0	1
idade4	Se idade entre 40 e 49 anos = 1	0,159	0,366	0	1
idade5	Se idade entre 50 e 64 anos = 1	0,155	0,362	0	1
idade6	Se idade maior que 65 anos = 1	0,088	0,284	0	1
mulher	Se mulher = 1	0,518	0,500	0	1
preto	Se preto = 1	0,073	0,261	0	1
amarelo	Se amarelo = 1	0,006	0,076	0	1
pardo	Se pardo = 1	0,431	0,495	0	1
indigena	Se indígena = 1	0,003	0,055	0	1
<b>Morbidade</b>					
cancer	Se possui câncer = 1	0,993	0,081	0	1
diabetes	Se possui diabetes = 1	0,958	0,202	0	1
bronquite	Se possui bronquite = 1	0,657	0,098	0	1
hiperte	Se possui hipertensão = 1	0,165	0,371	0	1
coracao	Se possui problemas de coração = 1	0,954	0,210	0	1

(continua)

(continuação)

Variáveis	Descrição	Média	Desvio-padrão	Min	Max
Variáveis independentes					
Morbidade					
rim	Se possui doença renal = 1	0,015	0,120	0	1
depress	Se possui depressão = 1	0,950	0,217	0	1
tuber	Se possui tuberculose = 1	0,998	0,039	0	1
tend	Se possui tendinite = 1	0,968	0,175	0	1
cirrose	Se possui cirrose hepática = 1	0,998	0,040	0	1
acamado	Se esteve acamado recentemente = 1	0,045	0,207	0	1
Variáveis relacionadas à moradia					
morador	Total de moradores	3,368	1,667	1	21
regiao2	Se mora na Região Nordeste = 1	0,276	0,447	0	1
regiao3	Se mora na Região Sudeste = 1	0,427	0,495	0	1
regiao4	Se mora na Região Sul = 1	0,154	0,361	0	1
regiao5	Se mora na Região Centro-Oeste = 1	0,072	0,259	0	1
urbano	Se reside em área urbana = 1	0,862	0,345	0	1

Fonte: Elaborado pelos autores com base na PNAD/2008

### 3.1 Resultados do modelo Poisson e Binomial Negativo

Os modelos Poisson (PRM) e Binomial Negativo (NBRM) apresentaram resultados semelhantes em termos dos coeficientes estimados. Em geral, houve poucas diferenças com relação aos sinais dos coeficientes, referente às variáveis esgoto, idade2 e morador, que apresentaram sinal positivo para a especificação PRM e negativo para a NBRM, respectivamente. A variável morador apresentou baixo valor absoluto para a especificação NBRM. Todos os coeficientes foram significantes a 1%.

Em ambos os modelos as variáveis relacionadas ao estado de saúde, que refletem a avaliação do indivíduo quanto as suas condições de saúde, apresentaram o sinal esperado – indivíduos que se consideram menos saudáveis vão mais ao médico. Os efeitos destas variáveis são positivos e aumentam à medida que o indivíduo se auto avalia com um estado de saúde pior.

As *dummies* indicadoras da presença de banheiro no domicílio, de coleta de lixo diária, de iluminação através de energia elétrica e de água canalizada em pelo menos um cômodo do domicílio, relacionadas à qualidade da saúde do indivíduo, apresentaram sinal positivo em ambas as especificações. Este resultado é decorrente do baixo nível de pessoas que procuram atendimento médico por prevenção, uma vez que apenas 21,68% das pessoas que possuem banheiro no domicílio, coleta de lixo diária, iluminação através de energia elétrica e água canalizada o fizeram por este motivo.

A variável esgoto, por sua vez, apresentou sinal positivo para a especificação PRM e negativo para a NBRM, respectivamente. Este resultado indica que a presença de rede coletora de esgoto reduz a demanda dos indivíduos por serviços médicos considerando a especificação NBRM, apesar de apenas 52% dos domicílios terem rede coletora de esgoto como forma de escoamento do banheiro.

As variáveis relacionadas aos atributos adquiridos apresentaram os sinais esperados com exceção da variável estudo1, correspondente a 1 a 4 anos de estudo, que apresentou sinal negativo. As demais variáveis de escolaridade, tendo como base de comparação os indivíduos sem instrução e com menos de um ano de estudo, indicaram que quanto maior esta, maior será a demanda dos indivíduos por cuidados médicos. Entretanto, esta variável não exerce tanta influência na utilização dos serviços médicos, uma vez que apenas 20,97% dos usuários o fizeram por motivos preventivos vis-à-vis 49,23% que procuraram algum serviço de saúde por motivo de doença. A variável renda e plano de saúde apresentaram sinal positivo, apesar da baixa magnitude da primeira. Dos usuários de plano de saúde 26% destes procuraram

o médico por motivos preventivos. Assim, ter plano de saúde ou estimular o maior número de exames com fins de checagem do estado de saúde e/ou as pessoas são mais cuidadosas (URIBE, 2008).

No que tange aos atributos natos, em geral todos os coeficientes apresentaram o sinal esperado. As variáveis de idade, com relação ao grupo de indivíduos de 0 a 13 anos, demonstraram que as pessoas de 20 a 39 anos de idade e os idosos, respectivamente, tem maior necessidade de frequentar médicos. Da mesma forma, as mulheres são as maiores demandantes de serviços médicos. No que diz respeito à raça, os negros são os que demandam menos serviços de saúde e os indígenas são os que mais demandam com relação aos brancos.

Os coeficientes de morbidade apresentaram sinais e magnitudes diversas, com destaque para a *dummy* de hipertensão. O coeficiente que indica maior demanda por saúde se um indivíduo esteve acamado recentemente, isto é, a pessoa que, por motivo de saúde, ficou pelo menos meio dia de cama ou internada em hospital, mesmo que não tenha permanecido no leito, apresentou um coeficiente positivo e com alta magnitude. Finalmente, os indivíduos que residem na Região Sudeste e na área urbana são os maiores usuários de serviços de saúde.

O efeito marginal das variáveis relacionadas à auto avaliação do estado de saúde, à morbidade, à educação e à idade apresentaram um valor significativo considerando ambas as especificações. Considerando uma mudança discreta no número esperado da variável dependente, uma variação do valor mínimo para o valor máximo da variável independente, por exemplo, o resultado é semelhante ao do efeito marginal.

Conforme Long e Freese (2001), a análise das probabilidades estimadas e o teste de sobredispersão constituem informações úteis para a escolha do melhor modelo, uma vez que há grande proporção de zeros e alta variância no número de consultas.

A Tabela 3 apresenta as probabilidades estimadas das especificações PRM e NBRM. Como pode ser visto, as probabilidades estimadas são favoráveis a especificação NBRM, de forma especial para o evento zero consultas. Este resultado decorre da grande proporção de zeros.

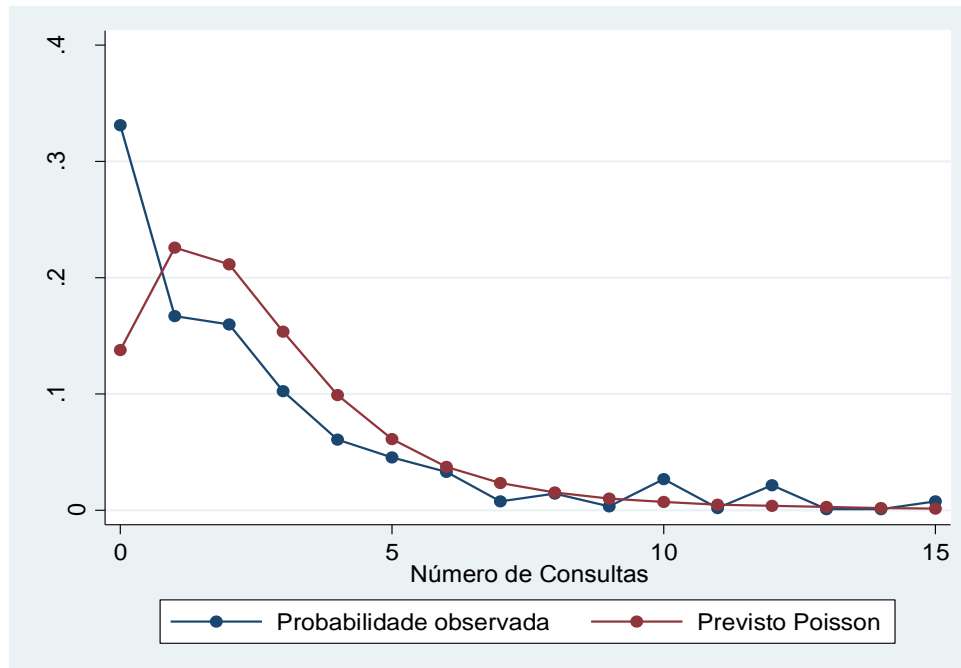
**Tabela 3 – Probabilidades estimadas pelo modelo Poisson e Binomial Negativo**

Número de consultas	% que se consultou	Observado	Previsto Poisson	Previsto Binomial Negativo
0	32,93	0,3311	0,1379	0,3295
1	49,57	0,1669	0,2257	0,2032
2	65,57	0,1598	0,2113	0,1315
3	75,85	0,1023	0,1536	0,0880
4	82,00	0,0605	0,0990	0,0605
5	86,53	0,0454	0,0609	0,0427
6	89,90	0,0331	0,0374	0,0308
7	90,63	0,0074	0,0234	0,0227
8	92,07	0,0143	0,0151	0,0171
9	92,37	0,0031	0,0101	0,0130
10	95,05	0,0265	0,0069	0,0101
11	95,24	0,0019	0,0048	0,0080
12	97,35	0,0212	0,0034	0,0064
13	97,44	0,0009	0,0025	0,0052
14	97,53	0,0009	0,0018	0,0042
15	98,27	0,0075	0,0014	0,0035
Média		0,0614	0,0621	0,0610

Fonte: Elaborado pelos autores com base na PNAD/2008

A Figura 1 apresenta o ajuste gráfico das probabilidades previstas pelo modelo Poisson. O ponto crítico deste modelo é a previsão do evento zero consultas. Nota-se que este modelo subestima o valor observado.

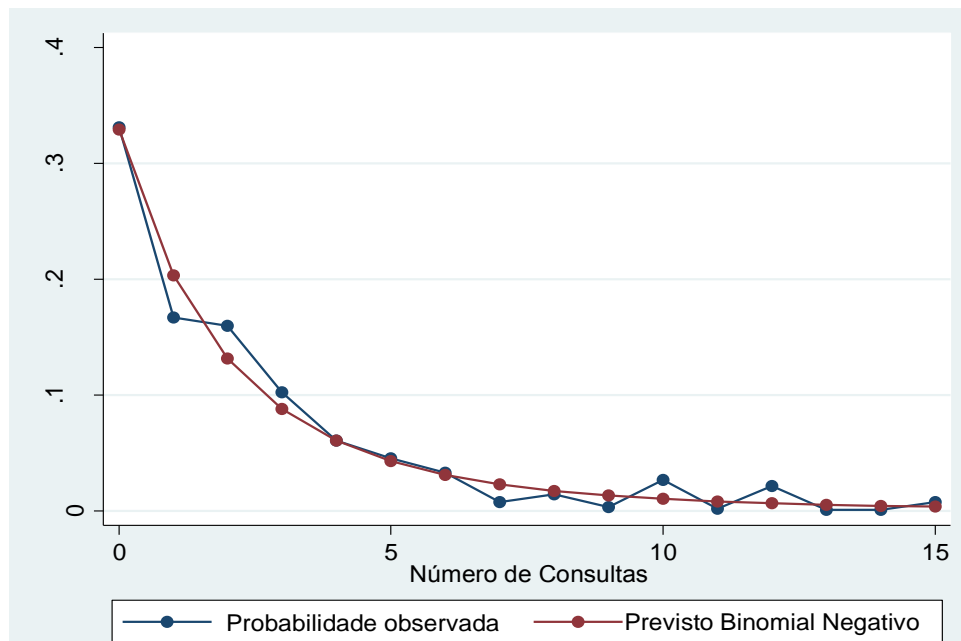
**Figura 1 – Ajuste gráfico das probabilidades: Modelo Poisson**



Fonte: Elaborado pelos autores com base na PNAD/2008

A Figura 2, por sua vez, apresenta o ajuste gráfico das probabilidades previstas pelo modelo Binomial Negativo. De forma semelhante ao modelo anterior, o ponto crítico é a previsão do evento zero consultas. Diferentemente do modelo anterior este apresenta um ajuste melhor aos dados sobre consultas médicas.

**Figura 2 – Ajuste gráfico das probabilidades: Modelo Binomial Negativo**



Fonte: Elaborado pelos autores com base na PNAD/2008

A Tabela 4 apresenta o resultado do teste de sobredispersão.

**Tabela 4 - Testes de sobredispersão**

$\ln(\alpha)$	0,04255
$\alpha$	1,04347
teste LR $\alpha = 0$	1,8e+08
prob > qui-quadrado (01)	0,0000
log pseudo-verossimilhança Poisson	-701.966,11
log pseudo-verossimilhança Binomial Negativo	-511.337,92
Observações	121.927.361

Fonte: Elaborado pelos autores com base na PNAD/2008

O teste LR indicou a presença de sobredispersão, rejeitando a hipótese nula de que  $\alpha$  é igual a zero, favorecendo a especificação NBRM. Portanto, esta especificação, segundo os critérios apresentados, é a que melhor se adequa aos dados vis-à-vis a especificação PRM.

### 3.2 Resultados dos modelos *Hurdle*

Como exposto anteriormente a utilização dos serviços de saúde ocorre em duas etapas. O paciente inicialmente escolhe se vai ou não procurar tratamento, sendo esta uma escolha individual; posteriormente, se a escolha for positiva, o médico influenciaria a frequência das consultas. Assim, o modelo *Hurdle* seria mais adequado.

As estimações dos modelos *Hurdle* Poisson (HP) e *Hurdle* Binomial Negativo (HBN) geraram coeficientes semelhantes. Os coeficientes do modelo *logit*, primeiro estágio do modelo *Hurdle*, apresentaram em sua maioria o sinal esperado. As variáveis relacionadas ao estado de saúde do indivíduo indicaram que à medida que o mesmo considera sua saúde pior maior será a probabilidade de recorrer aos serviços de saúde.

Observa-se, também, que a demanda por cuidados médicos é uma função crescente do nível de escolaridade e da idade dos indivíduos. Da mesma forma, os demais coeficientes apresentaram padrão semelhante ao exposto anteriormente – PRM e NBRM. Todos os coeficientes foram significativos a 1%, exceto a *dummy* urbano que não é significativa a 10%. Observando-se a *odds ratio* nota-se que a demanda por consultas médicas quando o indivíduo considera a sua saúde muito ruim é aproximadamente 3,4 vezes maior em comparação com o caso padrão – saúde muito boa. A posse de plano de saúde aumenta a demanda por serviços médicos em 2,15 vezes em comparação aos indivíduos que não possuem plano.

Considerando-se o segundo estágio do modelo *Hurdle*, somente as pessoas que utilizaram os serviços de saúde, os coeficientes do modelo HBN são maiores em módulo do que os do modelo HP. Os resultados obtidos indicaram que a presença de rede coletora de esgoto no domicílio, o abastecimento de água através de rede geral com canalização interna e a posse de filtro d'água reduzem a demanda por consultas médicas. Da mesma forma, o efeito da educação e da idade sobre a demanda é maior do que o previsto pelos modelos PRM e NBRM. Com relação às variáveis de morbidade merece destaque a *dummy* indicadora de doença renal.

O efeito marginal e a mudança discreta no número esperado da variável dependente permaneceram semelhantes aos modelos anteriores, porém com menor magnitude.

Assim como o modelo PRM e NBRM, o teste de sobredispersão foi favorável ao modelo HBN, conforme a tabela a seguir.

**Tabela 5 - Testes de sobredispersão e previsão dos modelos**

$\ln(\alpha)$	0,17707
$\alpha$	1,19372
teste LR $\alpha = 0$	2,55e+05
prob > qui-quadrado (01)	0,0000
log pseudo-verossimilhança <i>Logit</i>	-138.239,34

(continua)

(continuação)

log pseudo-verossimilhança <i>Hurdle</i> Poisson	-501.317,94
log pseudo-verossimilhança <i>Hurdle</i> Binomial Negativo	-367.746,18
Observações	121.927.361

Fonte: Elaborado pelos autores com base na PNAD/2008

### 3.3 Resultados dos modelos *Zero-Inflated*

A estimação dos modelos *Zero-Inflated* ou *Always-Zero group*, que tem por objetivo estimar a probabilidade de obterem-se observações iguais à zero, gerou conclusões semelhantes aos dos modelos anteriores. Todos os coeficientes foram significativos a 1%, tanto para o modelo *Zero-Inflated* Poisson (ZIP) quanto para o modelo *Zero-Inflated* Binomial Negativo (ZINB). Os coeficientes associadas às variáveis de auto avaliação do estado de saúde apresentaram sinal negativo, indicando que à medida que o indivíduo torna-se mais pessimista menor é a probabilidade dele pertencer ao grupo que nunca realiza consultas médicas. A proporção de indivíduos que considera sua saúde muito boa e não consultou médico é de 27,31% e reduz-se para 0,25% quando seu estado de saúde torna-se muito ruim.

As estimativas de morbidade, em sua maioria, apresentaram o sinal esperado para estes modelos – coeficiente negativo. Este resultado é consistente, uma vez que indivíduos portadores de certas doenças – câncer, diabetes, bronquite, hipertensão, doença renal, tuberculose, tendinite e cirrose – são grupos em que o número de consultas é alto, sendo baixa a probabilidade de estarem no grupo que nunca realiza consultas médicas. Apenas 8,48% dos indivíduos portadores destas doenças não recorreram a atendimento médico. Dos pacientes que consultaram pelo menos 1 vez o médico 74,93% o procuraram por motivo de doença. Os coeficientes associados às demais variáveis apresentaram o sinal esperado. Foram computados os efeitos marginais e a mudança discreta no número esperado da variável dependente que permaneceram semelhantes aos modelos anteriores.

Como realizado para as especificações PRM e NBRM, foram comparadas as probabilidades previstas pelos modelos ZIP e ZINB e foi computado o teste de sobredispersão. Conforme as Tabela 6 e 7 o modelo *Zero-Inflated* Binomial Negativo foi escolhido.

**Tabela 6 – Probabilidades estimadas pelos modelos *Zero-Inflated* Poisson e *Zero-Inflated* Binomial Negativo**

Número de consultas	Observado	% que se consultou	Previsto <i>Zero-Inflated</i> Poisson	Previsto <i>Zero-Inflated</i> Binomial Negativo
0	0,3311	32,93	0,3218	0,3346
1	0,1669	49,57	0,0984	0,2005
2	0,1598	65,57	0,1334	0,1294
3	0,1023	75,85	0,1301	0,0867
4	0,0605	82,00	0,1039	0,0599
5	0,0454	86,53	0,0734	0,0425
6	0,0331	89,90	0,0483	0,0309
7	0,0074	90,63	0,0309	0,0229
8	0,0143	92,07	0,0197	0,0173
9	0,0031	92,37	0,0127	0,0133
10	0,0265	95,05	0,0083	0,0104
11	0,0019	95,24	0,0056	0,0082
12	0,0212	97,35	0,0038	0,0066
13	0,0009	97,44	0,0026	0,0053
14	0,0009	97,53	0,0019	0,0043
15	0,0075	98,27	0,0013	0,0036
Média	0,0614		0,0623	0,0610

Fonte: Elaborado pelos autores com base na PNAD/2008

Para os modelos Poisson e Binomial Negativa o fator crítico era a previsão dos zeros. Entretanto, para os modelos *Always-Zero group*, que tratam o evento zero consulta adequadamente, o problema é a previsão do evento uma consulta médica realizada no ano.

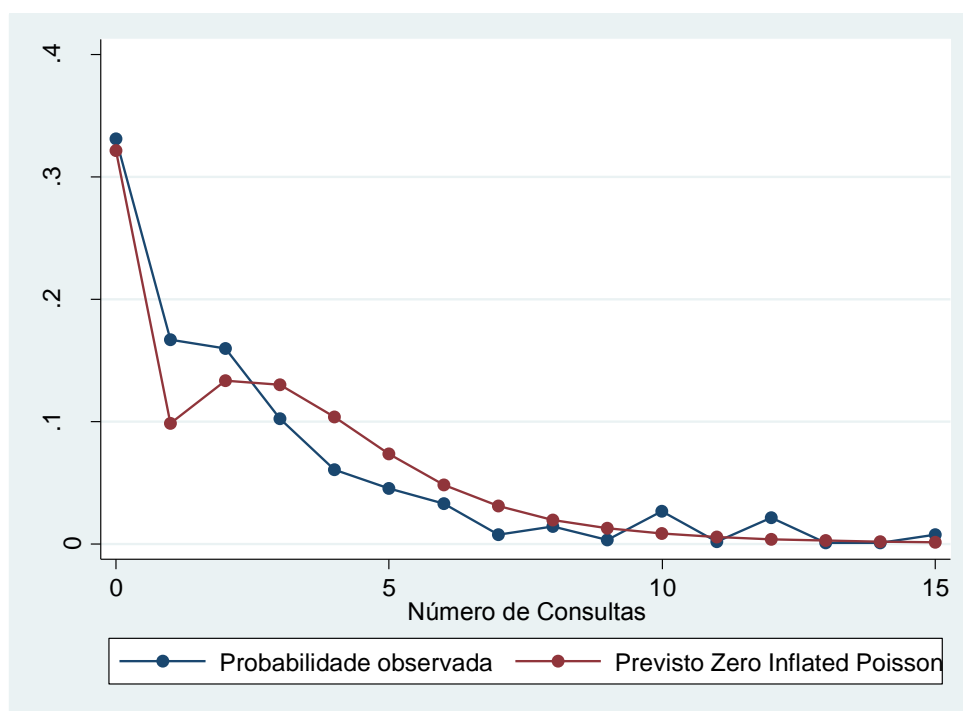
**Tabela 7 - Testes de sobredispersão**

$\ln(\alpha)$	0,04255
$\alpha$	1,04347
teste LR $\alpha = 0$	3,7e+06
prob > qui-quadrado (01)	0,0000
log pseudo-verossimilhança <i>Zero-Inflated</i> Poisson	-648.736,9
log pseudo-verossimilhança <i>Zero-Inflated</i> Binomial Negativo	-511.337,9
Observações	121.927.361

Fonte: Elaborado pelos autores com base na PNAD/2008

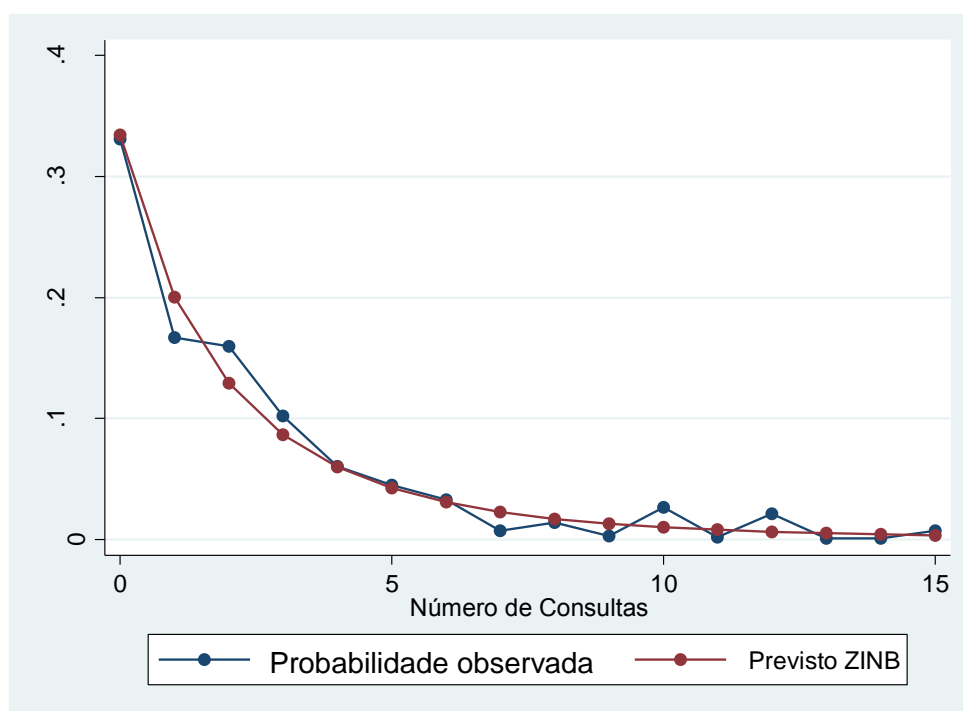
As Figuras 3 e 4 apresentam os resultados gráficos dos modelos *Always-Zero group*. Novamente, o modelo especificado pela distribuição BN foi o mais adequado.

**Figura 3 – Ajuste gráfico das probabilidades: Modelo *Zero-Inflated* Poisson**



Fonte: Elaborado pelos autores com base na PNAD/2008

**Figura 4 – Ajuste gráfico das probabilidades: Modelo *Zero-Inflated Binomial Negativo***



Fonte: Elaborado pelos autores com base na PNAD/2008

#### 3.4 Resultados do modelo *Finite Mixture*

Finalmente, a estimação do modelo *Finite Mixture* (FMM) com dois componentes<sup>6</sup> utilizando uma mistura da distribuição Binomial Negativa do tipo 2 gerou resultados diversos para cada uma das subpopulações. A primeira subpopulação, com  $\pi_1 = 0,7724$ , obteve um valor previsto médio de 1,97 consultas ao ano, enquanto a segunda, com  $\pi_2 = 0,2276$ , obteve um valor previsto médio de 5,67 consultas ao ano. Assim, a primeira subpopulação pode ser definida como sendo referente ao grupo das pessoas saudáveis ou “*low-users*” e a segunda como o das pessoas mais suscetíveis a adoecerem ou “*heavy-users*” (URIBE, 2008).

Como apontado por Cameron e Trivedi (2005), os modelos *Finite Mixture* apresentam duas vantagens em comparação aos demais modelos apresentados acima. Em primeiro lugar, a escolha do número de componentes permite determinar o número de tipos de subpopulação. Em segundo lugar, a escolha da distribuição das subpopulações permite um melhor tratamento da heterogeneidade não observada. Estes modelos são mais adequados para os estudos sobre demanda por consultas médicas, uma vez que a dicotomia entre “usuários” e “não usuários” não é necessariamente válida.

Os resultados da primeira subpopulação demonstraram que a piora do estado de saúde, o nível educacional, a idade, a posse de plano de saúde e a *dummy* acamado têm efeito positivo sobre a demanda por saúde. O efeito marginal destas variáveis também é significativo. A variável relacionada ao total de moradores e as *dummies* que indicam se o indivíduo sabe ler e escrever e se reside na área urbana têm efeito negativo sobre a demanda por serviços médicos. O efeito marginal destas variáveis em módulo é baixo.

Para a segunda subpopulação as *dummies* que indicam se o indivíduo sabe ler e escrever, se tem água canalizada em pelo menos um cômodo do domicílio e se tem de 12 a 14 anos de estudo não se mostraram significantes. As *dummies* que procuram capturar o efeito da qualidade da saúde do indivíduo sobre a demanda por serviços médicos têm menor impacto sobre o componente 1 do que sobre o componente 2. A *dummy* hipertensão, por sua vez, apresentou um sinal positivo e significativo.

<sup>6</sup> A adoção de dois componentes segue Uribe (2008). A utilização de três componentes inviabilizou a convergência do modelo.



As mulheres que pertencem a este grupo demandam menos serviços de saúde vis-à-vis as mulheres da subpopulação 1. Os amarelos em comparação aos brancos apresentaram um coeficiente negativo indicando sua menor procura por cuidados médicos.

Foram computados os efeitos marginais e observaram-se dois comportamentos antagônicos: o primeiro componente apresentou resultados semelhantes aos modelos anteriores; o segundo, por sua vez, apresentou resultados superiores a todos os modelos.

A Tabela 8 apresenta o resultado do teste de sobredispersão. Nota-se que o modelo *Finite Mixture* é o modelo que apresentou a menor sobredispersão em comparação com os demais modelos estimados.

**Tabela 8 – Resumo do modelo *Finite Mixture***

	Componente 1	Componente 2
$\ln(\alpha)$	-0,62949	-0,06141
$\alpha$	0,532860	0,940435
Proporção do componente ( $\pi$ )	0,772380	0,227619
Observações	121.927.361	121.927.361

Fonte: Elaborado pelos autores com base na PNAD/2008

Os testes de sobredispersão para todos os modelos estimados mostraram que aqueles parametrizados pela distribuição Binomial Negativa são superiores aos que utilizaram a distribuição Poisson, uma vez que há grande proporção de zeros e alta variância.

### 3.5 Comparações dos resultados obtidos

Observando a Tabela 9 o valor médio previsto pelo modelo *Finite Mixture* é o que mais se aproxima do valor observado, inferindo-se que este é o que melhor se ajusta aos dados sobre consultas médicas.

**Tabela 9 – Valores previstos para o número de consultas ao ano por modelo**

Modelo	Média	Desvio-padrão	Mínimo	Máximo
PRM	2,749	2,131	0,577	62,210
NBRM	2,849	2,875	0,533	154,806
HP	3,782	1,784	1,574	45,865
HBN	3,811	2,148	1,808	104,679
ZIP	2,662	1,786	0,740	44,966
ZINB	2,837	2,648	0,285	104,357
FMM-NB2	2,817	2,690	0,549	132,578
FMM-NB2 (componente 1)	1,976	2,097	0,248	72,095
FMM-NB2 (componente 2)	5,671	4,912	1,469	337,814
Observado	2,804	4,761	0	98

Fonte: Elaborado pelos autores com base na PNAD/2008

De acordo com o modelo escolhido o brasileiro realiza, em média, 2,82 consultas ao ano. Para o grupo *low-users* contribuíram de forma positiva para este resultado os seguintes condicionantes: avaliação pessimista do estado de saúde, direito a plano de saúde, ter banheiro em casa, ter acesso à coleta de lixo, ter acesso à água canalizada, ter esgoto, ter acesso à iluminação na forma de energia elétrica, renda domiciliar *per capita*, ser mulher (76% dos pacientes que consultaram o médico são mulheres), ser negro (50,32% das consultas foram realizadas por negros), ser indígena (talvez em função da oferta de médicos e/ou Programa Saúde da Família em área indígenas), ter mais de 20 anos de idade, anos de estudo, ter hipertensão, ter doença renal, acamado, e residir em qualquer região, com destaque para a Região Sudeste.

Para o grupo *heavy-users* contribuíram de forma positiva para este resultado as mesmas variáveis que contribuíram para o grupo *low-users*, com exceção de esgoto, ser pardo, anos de estudo. Além disso, tem impacto positivo residir em área urbana, o número de moradores e ter mais de 14 anos de idade.

Uma vez determinado o modelo mais adequado, restringiu-se a amostra para os indivíduos que não têm plano de saúde e que, por hipótese, utilizam o sistema público de saúde, e mantendo-se o mesmo modelo e a mesma especificação, observou-se que a renda passa a ser um determinante negativo para as duas subpopulações. A escolaridade tem impacto positivo e negativo para o primeiro e o segundo componente, respectivamente.

Por sua vez, ao restringir-se a amostra para os indivíduos que têm plano de saúde a renda e a escolaridade têm impacto positivo sobre a demanda de consultas médicas para qualquer subpopulação.

A partir destes resultados podem-se concluir alguns fatos. Em primeiro lugar, os indivíduos que utilizam o sistema privado de saúde possuem renda e escolaridade suficiente para adquirir um seguro-saúde. Aqueles que possuem plano de saúde têm uma renda média de R\$ 1.058,65 e 10,6 anos de estudo, enquanto os que utilizam o sistema público de saúde têm uma renda média de R\$ 533,84 e 7,2 anos de estudo. Em segundo lugar, para quem utiliza o sistema público a renda não representa um fator importante, uma vez que o SUS é universal. Assim, nota-se que existe desigualdade no acesso aos serviços de saúde (TRAVASSOS et al., 2000; URIBE, 2008).

#### 4. Considerações finais

As transformações recentes do setor de saúde – reorganização dos modelos de atenção e incorporação de novas tecnologias – têm alterado os modelos de utilização dos serviços médicos, impactando sobre o nível de bem-estar e a produtividade dos indivíduos. Um fenômeno resultante de tais mudanças é o aumento expressivo dos gastos *per capita* em saúde e da fração do PIB destinado ao setor de saúde nos últimos anos em vários países. O objetivo deste trabalho residiu no estudo dos determinantes da demanda por cuidados médicos, visto os novos paradigmas dos sistemas de saúde, de forma a alocar os recursos de maneira mais eficiente e organizar a infraestrutura dos serviços. A partir dos microdados oriundos da PNAD/2008, procurou-se investigar como o estado de saúde, a qualidade da saúde, os atributos natos, os atributos adquiridos, as variáveis relacionadas à morbidade e as variáveis relacionadas à moradia afetam a demanda individual por saúde. Adotaram-se como estratégia econométrica os modelos de dados contáveis.

A partir da estimação de vários modelos, a parametrização especificada pelo modelo *Finite Mixture* mostrou-se a mais adequada considerando-se o número médio previsto de consultas como critério de escolha. Os resultados indicaram duas subpopulações com comportamentos distintos: o grupo “*low-users*” e o grupo “*heavy-users*”. O primeiro grupo com um peso de 0,7724 obteve um valor previsto médio de 1,97 consultas ao ano, enquanto o segundo, com 0,2276, obteve um valor previsto médio de 5,67 consultas ao ano. Este resultado indica que a dicotomia “usuários” e “não-usuários” não é válida como proposto pelo modelo *Hurdle*.

A renda contribui de forma positiva nos dois grupos, apesar de seu baixo valor em módulo. As mulheres, os negros e os indígenas são os maiores demandantes de cuidados médicos, assim como os indivíduos que possuem plano de saúde. A escolaridade apresentou sinal positivo em ambas as populações para a *dummy* que indica 15 anos ou mais de estudo.

Verificou-se, também, a existência de desigualdades sociais no acesso aos serviços de saúde. Os resultados demonstraram que os indivíduos mais pobres têm maiores dificuldades de fazer o provimento privado de assistência a saúde, sendo necessário recorrer ao sistema público de saúde. Assim, em razão do menor nível educacional e da renda reduzida dos indivíduos que não possuem plano de saúde a adoção de políticas públicas é fundamental. Neste sentido, o provimento gratuito de planos ou seguros de saúde e de um maior fornecimento de medicamentos para esses grupos se apresenta como uma opção.

Um dos possíveis desdobramentos deste trabalho é o estudo intertemporal dos determinantes da utilização de serviços médicos através de pseudo-painéis.

## 5. Referências Bibliográficas

ANDRADE et al. Determinantes dos gastos das famílias com saúde no Brasil. **Economia**, 7(3): 485–508, 2006.

ANDRADE, et al. **Análise da estrutura do setor saúde e sua inserção na economia brasileira utilizando as matrizes de insumo-produto de 2000 e 2005**. UFMG/Cedeplar, 2011. 38p. (Texto para discussão; 424).

BRITO, R. A. **Demanda por serviços de saúde: o caso SABESPREV**. 2005. 140 f. Dissertação (Mestrado em Economia), UFMG / CEDEPLAR, Belo Horizonte, 2005.

CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics: methods and applications**. Cambridge University Press, New York, 2005.

GERDTHAM, U. G.; JONSSON, B. International comparisons of health expenditure. In CULYER, A. J. & NEWHOUSE, J. P., editors, **Handbook of Health Economics**. Elsevier, Amsterdam. 1988.

GROSSMAN, Michael. On the concept of health and the demand for health. **Journal of Political Economy**, 80(2): 223-255, 1972.

HITIRIS, T.; POSNETT, J. The determinants and effects of health expenditure in developed countries. **Journal of Health Economics**, 11(2):173–181, 1992.

KASSOUF, A. L. Acesso aos serviços de saúde nas áreas urbana e rural do Brasil. **Revista de Economia e Sociologia Rural**, 43(1): 29-44, 2005.

LONG, J. S.; FREESE, J. **Regression models for categorical dependent variables using STATA**. College Station: Stata Press Publication, 2001.

NORONHA, Kênia V. M. S.; ANDRADE, Mônica V. Desigualdades sociais em saúde e na utilização na utilização entre os idosos da América Latina. **Pan American Journal of Public Health**. Washington, D. C., 2005.

PALERMO, P; PORTUGAL, M; SOUZA, F. Um estudo sobre desigualdade no acesso à saúde na região Sul. In: X ENCONTRO REGIONAL DE ECONOMIA, 2005, Fortaleza. **Anais ...** Fortaleza, CE, 2005.

PESQUISA NACIONAL POR AMOSTRA DE DOMICÍLIOS (PNAD). Microdados, IBGE, 2008.

RIBEIRO, Mirian Martins. **Utilização de serviços de saúde no Brasil: uma investigação do padrão etário por sexo e cobertura por plano de saúde**. 2005. 100 f. Dissertação (Mestrado em Economia), UFMG / CEDEPLAR, Belo Horizonte, 2005.

RODRIGUES, et al. **Análise de taxas de utilização de serviços de saúde temporalmente: elementos para a formulação de um estudo**. UFMG/Cedeplar, 2010. 13p. (Texto para discussão; 404).

TRAVASSOS et al. Social and geographical inequalities in health services utilization in Brazil. **Ciência & Saúde Coletiva**, 5(1): 133-149, 2000.

URIBE, Paloma V. **Utilização de modelos de contagem na estimação da demanda por consultas médicas**. 2008. 106 f. Dissertação (Mestrado em Economia), Universidade de São Paulo, São Paulo, 2008.

WORLD HEALTH ORGANIZATION, **Data and statistics**, 2012.