

Demanda por saneamento básico no Brasil: uma análise com uso do modelo multinomial logit

Mário Jorge Cardoso de Mendonça[§]

Adolfo Sachsida[□]

Paulo Roberto Amorim Loureiro[‡]

RESUMO

Os serviços de saneamento, que englobam coleta de lixo, água tratada e esgotamento sanitário, exercem forte impacto sobre a saúde da população e o meio ambiente. No Brasil, verifica-se um elevado déficit desses serviços, causando diminuição da produtividade do trabalho e perda de produto da economia. Este trabalho tem como objetivo estimar um modelo de demanda por saneamento no Brasil. Em relação a isso, foram efetuados dois exercícios complementares. O primeiro diz respeito à estimação de um modelo para demanda por regressão logit, onde os três componentes do saneamento são estudados em conjunto. O outro visa à análise de cada um desses componentes separadamente. Neste sentido, foi utilizado um modelo de utilidade aleatória com base na análise multinomial logit para lixo e esgotamento sanitário. Apesar da limitação dos dados, em ambos os modelos foi possível observar que a demanda é sensível a fatores como escolaridade, herança familiar, raça, localização etc. Observou-se ainda que o impacto da renda sobre a demanda é sensível à faixa onde se situa a renda familiar, sobretudo nas camadas de renda menos favorecidas.

Palavras-chave: saneamento básico, meio ambiente, utilidade aleatória, modelo multinomial logit.

ABSTRACT

Basic sanitary services, that include waste disposal, treated water supplies and sewage services, have a strong effect on human health and on the environment. Brazil stands as having a large deficit in these services which results in a decrease in the productivity of labor and losses in GDP. This study aims at estimating an exploratory sanitary service demand model for Brazil. Two complementary approaches are used. The first one is a demand model estimated by a logit type regression where the three sanitary services are jointly analyzed. The second model takes each component individually for analytical purposes. In this sense, a random utility model based on the multinomial logit is used for waste disposal and sewage. In spite of data limitation, in both models it was possible to observe that demand is sensitive to factors such as education, family background, race, location, etc. Also relevant as a determining factor is the effect of household income, being especially true for poorer households.

Key words: sanitary services, environment, random utility, multinomial logit model.

JEL classification: H4, I0, I1, Q2.

§ Pesquisador do IPEA. E-mail: mjorge@ipea.gov.br

□ Professor da Universidade Católica de Brasília - UCB. E-mail: sachsida@pos.ucb.br

‡ Professor da Universidade Católica de Brasília - UCB. E-mail: paulo@pos.ucb.br

Recebido em março de 2003. Aceito em janeiro de 2004.

1 Introdução

Os serviços de saneamento básico são essenciais à vida, com fortes impactos sobre a saúde da população e o meio ambiente. Se entendermos a demanda por saneamento básico como uma demanda por insumos que melhoram a qualidade de vida do indivíduo, teremos uma ampla literatura que trabalha com esse tema. (Leibowitz e Friedman, 1979; Nocera e Zweifel, 1998; Dow, 1999, *inter alia*). Mais recentemente, Persson (2002) analisou as implicações de bem-estar geradas por mudanças nos preços de insumos relacionados a condições de saneamento. Para o Brasil, Carrera-Fernandez e Menezes (2002) fizeram uso do método de avaliação contingente para estimar uma função demanda por esgotamento sanitário.

Boa parte da população brasileira reside em locais em que as condições de saneamento são precárias. Devido à falta de saneamento e condições mínimas de higiene, a população está sujeita a diversos tipos de enfermidades. No Brasil são verificados altos índices de internações hospitalares pela carência de saneamento básico, em particular nas regiões Norte e Nordeste. Estes fatores nocivos repercutem sobre a saúde e certamente causam a diminuição da produtividade do trabalho, o que, por sua vez, gera perda no produto da economia. Além disso, observa-se que a parcela da população sujeita à falta de saneamento reside em locais impróprios para habitação, como nas encostas dos morros e nas margens dos rios. Isso traz como consequência, em princípio, a deterioração das áreas de floresta urbana e a poluição dos rios.

Define-se, nessa pesquisa, saneamento básico como sendo o conjunto de três elementos distintos que deveriam estar presentes num mesmo domicílio: abastecimento de água, esgotamento sanitário e serviço de coleta de lixo. De modo a assegurar uma gerência eficiente de recursos, é importante investigar que fatores determinam a demanda por saneamento básico, pois não é suficiente ampliar a oferta de serviços a menos que os agentes apreciem os benefícios relacionados. Além disso, caso não se tenha uma idéia correta acerca de tais determinantes, a atuação do governo com relação aos resultados desejados pode ficar comprometida. Por exemplo, caso o nível de escolaridade seja uma variável importante na determinação da demanda do indivíduo, e o governo decida pela ampliação da oferta de serviços de saneamento em regiões onde o nível de escolaridade seja baixo, é possível que os agentes, devido a fatores como escolaridade ou herança familiar, mantenham o hábito de utilizar práticas precárias, em razão da pouca utilidade que, para ele, tal bem deriva. Diferentemente, no caso da renda ser a variável fundamental, políticas

como distribuição de renda, ou ainda o simples aumento dos serviços, poderiam alcançar os objetivos desejados.

Boa parte dos estudos que tratam da questão do saneamento no Brasil está alicerçada apenas na mera ilustração de informações que ressaltam o déficit desse bem em relação aos domicílios brasileiros. Contudo, deve-se ter em mente que os dados refletem substancialmente o resultado da interação da oferta e demanda por esse serviço. Assim, conhecendo os determinantes da demanda, é possível minimizar o custo de expansão dos serviços de saneamento.

Este trabalho tem como objetivo estimar um modelo exploratório de demanda por saneamento no Brasil. Além desta introdução, a seção 2 apresenta os fatos estilizados sobre saneamento básico no Brasil. A seção 3 explicita um modelo de variável dicotômica para analisar a demanda por saneamento básico, que diz respeito à estimação de um modelo geral para demanda por saneamento por regressão logit; no qual os três componentes do saneamento são estudados em conjunto. A seção 4 tem como meta analisar cada um desses componentes separadamente. Neste sentido, foi utilizado um modelo de utilidade aleatória com base na análise multinomial logit para os casos do lixo e esgotamento sanitário e que mostra o impacto que diversas variáveis exercem sobre a escolha dos indivíduos entre as diversas alternativas. Por fim, na seção 5 são apresentadas as conclusões da pesquisa.

2 Fatos estilizados do saneamento básico no Brasil

Os serviços de abastecimento básico, que incluem os serviços de água, esgotamento sanitário e coleta de lixo, estão ainda muito longe de atender, no Brasil, à totalidade da população. Apesar do aumento significativo verificado na oferta dos serviços nas últimas décadas, persiste uma demanda possivelmente não atendida, especialmente nos estratos sociais de renda mais baixa, localizados nas periferias de grandes cidades, nos menores municípios, nas pequenas localidades e na área rural. Na Tabela 1 são apresentados os números concernentes ao déficit de saneamento básico no Brasil.

Com relação ao abastecimento de água, o índice nacional desse serviço, por meio de ligações domiciliares às redes, alcançou, em 2000, 77,8%. No que se refere ao esgotamento sanitário, somente 47,2% dos domicílios estão ligados às redes coletoras, sejam elas exclusivas ou de drenagem de águas pluviais. Caso sejam consideradas as fossas sépticas, que em alguns casos podem se constituir em solução adequada, a cobertura em

esgotamento sanitário alcança 62,2%. Assim, o déficit dos serviços de abastecimento de água por redes públicas atinge cerca de 9,9 milhões de domicílios brasileiros, e 23,6 milhões não estão conectados às redes coletoras de esgotos. Se consideradas as fossas sépticas, o déficit em esgotamento sanitário atinge 16,9 milhões de domicílios.

No que tange às áreas rurais, o déficit é proporcionalmente muito superior àqueles verificados em áreas urbanas. Tomando por base o ano de censo mais recente (2000), o atendimento por redes de distribuição de água atinge somente cerca de 18,1% dos domicílios rurais. Em termos de esgotamento sanitário, apenas 3,3% dos domicílios rurais estão conectados às redes coletoras e somente 9,6% dispõem de fossas sépticas. Assim, de um total de 7,46 milhões de domicílios localizados em áreas rurais, apenas 1,35 milhões estão ligados às redes de abastecimento de água e 960 mil estão ligados às redes coletoras de esgotos ou dispõem de fossas sépticas.

Deve ser observado que o atendimento às populações dispersas em áreas de baixa densidade populacional pode ser considerado satisfatório com o uso de soluções individuais, inclusive para abastecimento de água para consumo humano. Por exemplo, nas áreas rurais, cerca de 4,3 milhões de domicílios se abastecem por meio de nascentes ou poços localizados na própria propriedade. Não há, todavia, dados que possam assegurar que estas fontes de água sejam seguras.¹ Por outro lado, ações de saúde pública em áreas rurais, como a simples desinfecção da água, poderiam assegurar uma qualidade mínima à água, tornando-a própria ao consumo humano.

No que se refere aos serviços de coleta de lixo, observa-se, de acordo com a Tabela 1, que ao longo da década de 1990 houve uma substancial melhora nos indicadores, principalmente no referente aos domicílios situados nas áreas rurais. Entretanto, a grande parcela desses domicílios continua sem ser atendida pelos serviços de coleta de lixo.

De modo a aprofundarmos a questão do saneamento básico no Brasil sob o prisma mais teórico, iremos adotar uma metodologia própria para, a partir dela, obter alguns fatos estilizados sobre o saneamento. Nesse sentido, a proposta aqui é primeiro definir uma

1 Para a OMS (Organização Mundial da Saúde), é atendido por serviço de abastecimento de água um domicílio urbano que se localize a, no máximo, 15 minutos de caminhada, ou 200 metros de distância de uma fonte de água segura como os chafarizes, e em esgotamento sanitário, aquele cujos dejetos humanos estejam afastados do contato direto com pessoas, animais e fontes de água, aceitando-se soluções do tipo latrina seca ou fossa rudimentar.

variável que possa exprimir um número que represente o nível de saneamento para cada domicílio. Essa variável pode ser construída a partir da combinação de três subconjuntos distintos de variáveis, cada qual relacionado a um dos elementos que definem o saneamento básico: abastecimento de água, esgotamento sanitário e coleta de lixo. No entanto, é necessário definir os critérios que o domicílio deve preencher para possuir cada um desses elementos.

Tabela 1
Evolução da Cobertura dos Serviços de Água e Esgotos no Brasil (em %)

Indicadores	1970	1980	1990	2000
Abastecimento de Água				
.domicílios urbanos - rede de distribuição	60,5	79,2	86,3	89,8
.domicílios rurais - rede de distribuição	2,6	5,0	9,3	18,1
Esgotamento Sanitário				
.domicílios urbanos - rede de coleta	22,2	37,0	47,9	56,0
.domicílios urbanos - fossas sépticas	25,3	22,9	20,9	16,0
.domicílios rurais - rede de coleta	0,45	1,4	3,7	3,3
.domicílios rurais - fossas sépticas	3,2	7,2	14,4	9,6
Coleta de Lixo*				
.domicílios urbanos - coleta direta	N.D.	N.D.	0,78	84,9
.domicílios urbanos - coleta indireta	N.D.	N.D.	0,06	0,08
.domicílios rurais - coleta direta	N.D.	N.D.	0,07	0,15
.domicílios rurais - coleta indireta	N.D.	N.D.	0,01	0,04

Fonte: IBGE, Censos Demográficos 1970, 1980, 1990 e 2000.

* N.D. = Dados não disponíveis.

Em relação ao abastecimento de água, o critério para que um domicílio possua esse serviço, num sentido pleno, é que tenha água encanada em pelo menos um cômodo e que ela seja proveniente de rede geral de distribuição. Para que um domicílio preencha o critério para esgotamento sanitário é necessário que exista banheiro ou sanitário nesse domicílio, sendo seu uso privativo ao domicílio, e que o escoadouro seja feito por rede coletora de esgoto ou fluvial, ou ainda fossa séptica. No caso do domicílio ter acesso à coleta de lixo, considerou-se que esse serviço é pleno no caso do lixo ser coletado direta ou indiretamente² por serviço de empresa de limpeza.

2 Quando o lixo é depositado em caçamba, tanque ou depósito de serviço.

É razoável a assertiva de que a escolaridade, a renda, assim como a educação dos pais, daqueles que habitam o mesmo domicílio, estejam entre os principais fatores socioeconômicos que determinam a escolha por saneamento. Fatores ligados à raça e sexo podem também exercer influência na decisão devido à herança cultural. A análise exploratória que se segue toma por base os dados da PNAD de 1996. A Tabela 2 ilustra a dispersão, quanto ao nível de saneamento, dos domicílios no Brasil para a amostra da PNAD. Ela se associa ao nível de saneamento, e a referida variável é classificada com valores que vão de 0 a 3, sendo que (zero) representa o domicílio que não possui nenhum dos três itens de saneamento, 1 (um) o domicílio que possui um dos dois itens, 2 (dois) o que possui 2 dos itens de saneamento e 3 (três) representa o domicílio que possui todos as 3 características de saneamento. Conforme pode ser observado, aproximadamente 45% do total dos domicílios não atingem o saneamento pleno (nível 3).

No que diz respeito ao efeito da educação sobre o saneamento, o exercício seguinte ressalta quanto o nível de escolaridade é importante. No caso em questão, a amostra foi separada entre domicílios cuja pessoa de referência possui pelo menos o primeiro grau completo e os domicílios em que isto não ocorre. Para a amostra da PNAD de 1996, a Tabela 3 mostra que cerca de 60% dos domicílios possuem saneamento pleno em que a pessoa de referência tem pelo menos o nível elementar de escolaridade. Já para o outro grupo, observa-se que o saneamento pleno não alcança 40%.

Tabela 2
Nível de Saneamento Básico dos Domicílios no Brasil em 1996

Nível de Saneamento Básico	Porcentual	Porcentual Acumulado
0	7,14%	7,14%
1	7,32%	14,46%
2	30,57%	45,03%
3	54,97%	100,00%

Fonte: PNAD (1996).

Ocorre, contudo, que não apenas fatores de natureza socioeconômica são responsáveis pela grande disparidade existente no volume de saneamento básico entre os domicílios brasileiros. Como assinala Carrera-Fernandez e Menezes (2002), essa disparidade também é percebida caso seja feita uma comparação entre as regiões do Brasil. Assim, tomando-se por base os dados da PNAD-1996, e de acordo com a Tabela 4, observa-se que os Estados do Sudeste e do Sul possuem uma quantidade de domicílios plenamente

saneados superiores aos Estados das regiões Nordeste e Centro-Oeste do País, para não mencionarmos a região Norte, que não aparece na tabela. Enquanto no Sudeste 77% dos domicílios possuem saneamento pleno, esse número na região Nordeste é de apenas 34%, sendo que na região Centro-Oeste esse valor é de 41%.

Tabela 3
Nível de Saneamento Básico de acordo com o Nível Educacional

Nível de Saneamento Básico	Pelo Menos 4 anos de Estudo		Menos de 4 anos de Estudo	
	Porcentual	Acumulado	Porcentual	Acumulado
0	5,89%	5,89%	11,30%	11,30%
1	5,05%	10,93%	14,87%	26,16%
2	28,58%	39,52%	37,16%	63,32%
3	60,48%	100,00%	36,68%	100,00%

Fonte: PNAD (1996).

Isso parece ser conseqüência do fato de as regiões Sudeste e Sul serem as mais ricas do País e que, portanto, possuem melhor infra-estrutura para a população, facilitando o acesso aos serviços de saneamento. Também é sabido que as áreas urbanas possuem os domicílios com maior nível de saneamento, possivelmente devido ao menor custo marginal de se obter tal serviço nas áreas urbanas.

Tabela 4
Saneamento Básico por Região

Nível de Saneamento	Nordeste		Sudeste		Sul		Centro-Oeste	
	%	Acum.	%	Acum.	%	Acum.	%	Acum.
0	9,38	9,38	2,94	2,94	4,40	4,40	15,17	15,17
1	13,99	23,38	4,53	7,48	3,16	7,56	4,98	20,14
2	42,10	65,47	14,76	22,23	36,74	44,30	38,64	58,78
3	34,53	100,00	77,77	100,00	55,70	100,00	41,22	100,00

Fonte: PNAD (1996).

Como pode ser visto na Tabela 5, de acordo com os dados da PNAD-1996, 57% dos domicílios urbanos são dotados de saneamento pleno, enquanto que este número cai para 17% tomando por base os domicílios que se localizam em áreas rurais.

Tabela 5
Saneamento Básico: Área Urbana x Área Rural

Nível de Saneamento	Área Urbana		Área Rural	
	%	Acumulado	%	Acumulado
0	6,58	6,58	19,14	19,14
1	6,02	12,60	35,10	54,24
2	30,66	43,26	28,62	82,86
3	56,74	100,00	17,14	100,00

Fonte: PNAD (1996).

3 Modelo de demanda por saneamento básico

Na seção anterior foram apresentados alguns fatos estilizados relacionados ao saneamento básico no Brasil. Uma vez tendo introduzido esses pontos fundamentais, o objetivo agora é explicitar o modelo para a demanda por saneamento básico. Inicialmente, sabe-se que a utilidade advinda do saneamento é uma variável não observada, ou seja, observa-se apenas o nível de saneamento representado aqui por y .

Uma vez introduzidos os conceitos relacionados ao saneamento básico para um domicílio, a tarefa a seguir é desenvolver um modelo que tem como objetivo captar a demanda por esse bem. Neste sentido, iremos introduzir uma variável dicotômica y que traduz, de modo simplificado, se um domicílio possui ou não saneamento. Assim sendo, temos que

$y = 0$, se o domicílio possui menos que três elementos de saneamento;

$y = 1$, caso o domicílio possua os três elementos definidores de saneamento.

A razão para a utilização desse expediente é que ele permite, ou pelo menos facilita, inicialmente, a análise simplificada do impacto das variáveis explicativas do modelo.

Iremos assumir agora que a utilidade do saneamento, y^* , é expressa por uma função linear, tal que

$$y_i^* = \beta' x_i + \varepsilon_i \quad i = 1, \dots, N \quad (1)$$

onde x é o vetor de variáveis explicativas do modelo, β é o vetor de parâmetros e ε é termo aleatório que assume distribuição normal padrão. Aqui i representa o i -ésimo domicílio. Como não é observado, tem-se então que

$$y_i = 0, \quad \text{se } y_i^* \leq u_i \quad (2)$$

$$y_i = 1, \quad \text{c. c.}$$

onde u representa o nível de corte para a utilidade ante a escolha de y . Desde que y é uma variável qualitativa, o modelo usual de mínimos quadrados ordinário (MQO) apresenta alguns problemas, dentre os quais podem ser destacados os erros heteroscedásticos³ e a possibilidade de se obter estimativas de probabilidade fora do intervalo [0,1]. Isto significa que, no caso em questão, o estimador MQO é ineficiente, gerando ainda previsões imprecisas.

O referencial econométrico utilizado para analisar a escolha dos consumidores por serviço de esgotamento sanitário pleno deve partir, então, de um modelo de variável qualitativa dicotômica. (Greene, 1993). Neste caso, o modelo pode ser utilizado para analisar o impacto de diferentes fatores (ou variáveis explicativas) sobre a probabilidade de o domicílio possuir/conectar saneamento pleno. Assim, temos que:

$$\Pr(y_i^* > u) = \Pr(y_i = 1) = F(\beta' x_i) = \frac{e^{\beta' x_i}}{1 + e^{\beta' x_i}} \quad (3)$$

onde F é uma função de distribuição. Como se pode notar, assume-se que F é do tipo logit.

Deve-se ainda ter em mente que o modelo para estimar a demanda por saneamento guarda relação com o modelo de demanda por habitação - na medida em que entre os atributos que os indivíduos tomam por base para a escolha de uma habitação certamente se encontram os serviços de saneamento. Nesse sentido, uma estratégia razoável para avançar na investigação é testar, no modelo de saneamento, algumas variáveis explicativas que também aparecem no modelo de habitação. Garcia e Rebelo (2002) utilizam uma ampla gama de variáveis - tais como renda domiciliar, idade do chefe da família, número

3 Pode-se demonstrar que a variância do erro depende das probabilidades, o que significa que o erro aleatório é heteroscedástico. Porém, esse é um problema superável, porquanto existem procedimentos econométricos que podem ser utilizados para corrigir a heteroscedasticidade.

de moradores, Estado da federação onde se localiza o domicílio, se o domicílio está situado em área urbana ou rural - para se estudar a demanda por habitação.

3.1 Modelo logit para saneamento básico pleno

Conforme foi visto na seção anterior, o modelo assume que a variável dependente y é dicotômica. Nesse caso, temos uma situação simplificada, embora menos realista, onde um domicílio atende à condição de saneamento somente quando todos os elementos de saneamento - água, esgotamento sanitário e coleta de lixo - estão presentes.

As variáveis explicativas que serão testadas no modelo são as seguintes: número de moradores (*moradores*), logaritmo da renda média do domicílio (*lrendmed*), escolaridade da pessoa de referência (*escol*), idade da pessoa de referência (*idade*); escolaridade da mãe da pessoa de referência (*instm*). As variáveis de controle utilizadas neste constituem as variáveis *dummies*, que definem a região geográfica do município, Sudeste, Sul, Centro-Oeste e Nordeste (a região Norte foi usada como referência), e uma *dummy*, que indica se o domicílio se localiza em área urbana ou rural (*urban*). Há ainda uma variável *dummy* para raça (*branco*) e outra para o sexo (homem) do indivíduo. A descrição completa das variáveis do modelo encontra-se no Anexo 1.

O uso da transformação para logaritmo da renda média domiciliar deveu-se ao fato desta variável apresentar uma grande dispersão. A aplicação do log permite suavizar a série. A introdução da escolaridade da mãe da pessoa de referência visa captar alguma espécie de herança familiar ou aspecto cultural herdado.

Embora haja pouca referência na literatura acerca desse aspecto, presume-se que os sinais esperados para as variáveis renda, escolaridade e idade sejam positivos. Com relação às variáveis representativas de cor e sexo, não se definirá nenhuma proposição *a priori*, bem como para o número de moradores. Quanto às variáveis de controle para região e área urbana, espera-se que haja uma significativa dispersão indicando, por exemplo, que os domicílios localizados nas regiões Sudeste e Sul, e em áreas urbanas, tenham uma probabilidade maior de possuir saneamento. Por fim, os dados utilizados foram obtidos a partir da PNAD (Pesquisa Nacional por Amostra de Domicílio) de 1996. Os resultados para o modelo probit são apresentados na Tabela 6.

Tabela 6
Modelo Logit para Demanda por Saneamento

Variáveis Independentes	OLS	Logit	Logit	Logit	Efeito Marginal
	(1)	Amostra Plena (2)	Renda <= R\$ 1000 (3)	Renda <= R\$ 500 (4)	Renda <=R\$ 500 (5)
MORADS	0,002 (0,042)	0,009 (0,201)	0,048 (0,000)	0,051 (0,000)	0,018 (0,000)
LRENDMED	0,005 (0,000)	0,021 (0,000)	0,302 (0,000)	0,333 (0,000)	0,119 (0,000)
ESCOL	0,024 (0,000)	0,089 (0,000)	0,058 (0,000)	0,060 (0,000)	0,021 (0,000)
INSTM	0,0101 (0,000)	0,037 (0,000)	0,028 (0,000)	0,030 (0,000)	0,011 (0,000)
IDADE	0,003 (0,000)	0,012 (0,000)	0,007 (0,000)	0,008 (0,000)	0,002 (0,000)
BRANCO	0,043 (0,000)	0,156 (0,000)	0,110 (0,000)	0,117 (0,000)	0,042 (0,000)
HOMEM	-0,023 (0,000)	-0,092 (0,000)	-0,140 (0,000)	-0,130 (0,000)	-0,047 (0,000)
URBAN	0,359 (0,000)	1,757 (0,000)	1,676 (0,000)	1,683 (0,000)	0,414 (0,000)
SUDESTE	0,566 (0,000)	1,918 (0,000)	1,907 (0,000)	1,896 (0,000)	0,650 (0,000)
SUL	0,353 (0,000)	1,236 (0,000)	1,202 (0,000)	1,200 (0,000)	0,450 (0,000)
CENTRO	0,254 (0,000)	0,923 (0,000)	0,879 (0,000)	0,859 (0,000)	0,330 (0,000)
NORDESTE	0,220 (0,000)	0,736 (0,000)	0,814 (0,000)	0,826 (0,000)	0,303 (0,000)
CONSTANTE	-0,554 (0,000)	-4,283 (0,000)	-5,427 (0,000)	-5,535 (0,000)	- (0,000)
VIF	4,72	-	-	-	-
1/VIF	0,21	-	-	-	-
R ²	0,397	-	-	-	-
PSEUDO R ²	-	0,334	0,337	0,351	0,351
Observações	64.859	64.859	59.490	53.747	53.747

Nota: 1. Os valores entre parênteses representam o p-valor.

Conforme pode ser visto na Tabela 6, as estimativas foram obtidas para três subconjuntos amostrais. As colunas (1) e (2) mostram os resultados para o total da amostra, a coluna (3) aponta os resultados cujo domicílio possui renda média igual ou menor que R\$ 1.000, enquanto a que coluna (4) mostra os resultados para aqueles cuja renda média se situa abaixo de R\$ 500,00. Por fim, a coluna (5) apresenta os efeitos marginais para esse último modelo, ou seja, o impacto que cada variação marginal da variável tem sobre a escolha de uma família morar num domicílio com saneamento pleno.

Na coluna (1) são mostrados os resultados gerados por mínimos quadrados ordinários (MQO). A estimação por MQO é feita apenas para observar se existe problema de multicolinearidade no modelo. A partir do emprego do instrumento VIF (*Variance Inflation Factor*), calcula-se o impacto sobre a variância de cada variável decorrente das correlações advindas da presença dos outros regressores. Conforme pode ser observado, o valor desta estatística indica que não existe problema de multicolinearidade, já que o valor da estatística VIF é 4,72. A literatura assinala que para que exista indicação de multicolinearidade o valor de VIF deve estar acima de 5. (Judge *et alii*, 1982). O outro valor para essa estatística indica o maior percentual para uma determinada variável, dentre todas as variáveis explicativas analisadas, relativo à maior influência que as outras variáveis exercem sobre a sua variância. Assim, o valor dessa estatística aponta que, para a variável que sofre maior efeito da presença dos outros regressores, 21% (1/VIF) da sua variância não é explicado pela presença das correlações existentes entre ela e as outras variáveis, o que é bastante razoável.

A coluna (2) apresenta os resultados para o modelo logit. Entretanto, diferentemente do modelo linear, numa estimação logit os coeficientes não traduzem diretamente o efeito de uma variação marginal da variável decorrente da característica não-linear desse modelo. (Greene, 1993). As colunas (3) e (4) estimam o modelo para duas faixas de rendas. Por fim, na coluna (5) são apresentados os resultados que mostram os efeitos marginais. Conforme pode ser visto, exceto *morads* para a amostra plena, todas as variáveis são significativas.

No caso das *dummies* geográficas, pode-se observar que elas também apresentam os sinais esperados, o que indica que domicílios localizados nas regiões Sudeste e Sul demonstram possuir uma probabilidade bem maior de serem saneados do que os localizados em outras regiões. Da mesma forma, um domicílio localizado em área urbana mostra igual característica. Os resultados também indicam que a cor e o sexo da pessoa de referência exercem influência sobre a decisão de escolha de morar em um domicílio plenamente saneado. Assim, o fato da pessoa de referência ser do sexo feminino e de cor branca tem impacto positivo sobre a escolha por saneamento.

Uma vez apresentado o modelo básico de demanda por saneamento, a tarefa agora é estudar os determinantes que fazem com que os agentes optem por escolher determinada opção por saneamento. Para entender o que se está querendo dizer, deve-se ter em mente que, de acordo com a metodologia da PNAD, existem para cada um dos elementos de saneamento diversas categorias nas quais a opção do agente pode ser enquadrada. Por exemplo, no caso do esgotamento sanitário, as categorias podem ser enquadradas em seis tipos distintos, que vão desde coleta direta pela rede pública até o lixo que é jogado em rio, lago ou mar. Assim, devido à ocorrência de múltiplas opções, não faz sentido o emprego de variável dicotômica para a formulação de um modelo ilustrativo da escolha de determinada categoria para cada um dos elementos, sendo necessária a utilização de técnicas aplicáveis a esse tipo de questão. Este é o propósito da próxima seção.

4 Análise multinomial para o modelo de saneamento

4.1 Modelo de utilidade aleatória

Conforme assinalado ao final da seção anterior, cada componente de saneamento - água, esgotamento sanitário e lixo - possui categorias distintas de classificação que representam tipos diferentes de opção para o agente. Nesse sentido, a escolha j , $j = 1, \dots, J$, para o indivíduo i , $i = 1, \dots, I$, visa maximizar o nível de utilidade U_{ij} . Tendo em vista que a informação acerca dos determinantes de cada escolha é incompleta, pode-se definir U_{ij} da seguinte forma:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad j = 1, \dots, J, i = 1, \dots, I \quad (4)$$

onde V_{ij} representa sua parte determinística e ε_{ij} , o componente aleatório.

A probabilidade P_{ij} de que o indivíduo i escolha certa alternativa é igual à probabilidade que U_{ij} seja a maior utilidade entre U_{i1}, \dots, U_{iJ} . Denotando por $x_i \in \{1, \dots, J\}$ a escolha feita pelo indivíduo i , temos então que:

$$P_{ij} = \Pr(x_i = j) = \Pr(U_{ij} > U_{ik}, \forall k = 1, \dots, J : k \neq j) = \Pr(\varepsilon_{ik} - \varepsilon_{ij} \leq V_{ij} - V_{ik}, \forall k = 1, \dots, J : k \neq j) \quad (5)$$

Dados os componentes determinísticos das funções de utilidade, V_{i1}, \dots, V_{iJ} , essa probabilidade irá depender das suposições acerca das distribuições (ou das diferenças) dos termos estocásticos $\varepsilon_{i1}, \dots, \varepsilon_{ij}$. O componente determinístico V_{ij} é afetado por diferentes tipos de determinantes, podendo ser definido da seguinte forma:

$$V_{ij} = \alpha_j + x_{ij}' \beta + z_i' \gamma_j \quad j = 1, \dots, J, i = 1, \dots, I \quad (6)$$

onde z_i representa o vetor de variáveis específicas do indivíduo, x_{ij} é o vetor de variáveis que varia tanto em relação às alternativas como aos indivíduos e, por fim, as constantes próprias das alternativas α_j . Nesta pesquisa, em razão da limitação da base de dados, não há disponibilidade de qualquer variável do tipo x_{ij} . Neste caso, temos então que:

$$V_{ij} = \alpha_j + z_i' \gamma_j \quad j = 1, \dots, J, i = 1, \dots, I \quad (7)$$

Tendo em vista as características desse modelo, o modo mais apropriado de estimar os parâmetros é a partir da aplicação de um modelo multinomial logit, onde:

$$\Pr(Y_i = j) = \frac{e^{\beta_j' z_i}}{\sum_{k=1}^J e^{\beta_k' z_i}} \quad (8)$$

Nesse modelo, as equações estimadas geram um conjunto de probabilidades para $J+1$ escolhas para o indivíduo i . Um modo de remover essa indeterminação é introduzir uma normalização para a alternativa de referência como, por exemplo, fazendo o vetor $\beta_1 = 0$, Assim temos que:

$$\Pr(Y_i = 0) = \frac{1}{1 + \sum_{k=1}^J e^{\beta_k' z_i}}, \text{ e}$$

$$\Pr(Y_i = j) = \frac{e^{\beta_j' z_i}}{1 + \sum_{k=2}^J e^{\beta_k' z_i}} \quad \text{para } j = 2, \dots, J - 1 \quad (9)$$

A partir do emprego de método de máxima verossimilhança, e também com o uso de otimização não-linear, é possível obter as estimativas para os coeficientes que aparecem em (9).

4.2 Aplicação empírica: esgotamento sanitário e lixo

Tendo apresentado o aparato teórico para tratar a questão da escolha, onde aparecem diversas alternativas não ordenadas, a questão agora recai sobre como aplicar o modelo multinomial no contexto de saneamento, observando como as variáveis explicativas

afetam a decisão tomada em cada uma das escolhas. Neste sentido, o que se irá propor é o estudo dos fatores que determinam a probabilidade de se efetuar cada uma das alternativas disponíveis para o conjunto dos elementos que formam o saneamento básico. Aqui serão tratados os casos do lixo e esgotamento sanitário. De acordo com a metodologia usada pela PNAD, são diversas as alternativas que um domicílio pode apresentar para cada um desses elementos. Todas essas alternativas são explicitadas, com detalhes, ao final do trabalho no Anexo 1.

O Anexo 1 mostra que os elementos que compõem o saneamento estudado neste trabalho - lixo e esgotamento sanitário - possuem cada qual seis alternativas, sendo que as escolhas dessas alternativas são naturalmente determinadas por fatores socioeconômicos e geográficos, tal como aparece na análise feita na seção anterior. Tendo em vista que um número tão elevado de escolhas se traduz num obstáculo para uma análise mais clara do modelo multinomial, um artifício que pode ser usado é efetuar a aglutinação de escolhas semelhantes num único grupo. Por exemplo, em ambos os casos - lixo e esgotamento sanitário - as duas primeiras escolhas são bastante semelhantes no sentido de que ambas, por hipótese, parecem estar ligadas a grupos homogêneos de indivíduos no que se refere às preferências e a dotações. Da mesma forma, as escolhas JT⁴ e JMR,⁵ para o lixo, e VL⁶ e JRM,⁷ para esgotamento sanitário, podem ser agrupadas assumindo a mesma justificativa. Diferentemente, as demais escolhas são tomadas individualmente.

Um ponto que merece ser ressaltado é que o estudo dos fatores que determinam tais escolhas é importante para o meio ambiente e o bem-estar da sociedade, pois como está descrito no anexo, essas alternativas para o lixo e esgotamento são as que mais agridem o meio ambiente e propiciam o surgimento de problemas para a saúde.

A Tabela 7 apresenta os resultados do modelo multinomial para esgotamento e lixo. Antes de efetuar a análise desses resultados é necessário tecer alguns comentários. Conforme foi assinalado no parágrafo anterior, algumas escolhas para esses dois elementos mostram muita similaridade. Assim sendo, uma opção foi elaborar um novo arranjo para essas alternativas considerando as definições que aparecem no Anexo 1, de modo a enquadrar as alternativas similares dentro de um mesmo grupo. Naturalmente, o estabelecimento de opções similares deve carregar um certo grau de subjetividade.

4 Jogar em terreno baldio ou logradouro.

5 Jogado diretamente em mar ou rio.

6 Vala.

7 Rio ou mar.

Tendo em vista essas colocações, a nova ordenação foi feita da seguinte forma: para o lixo, temos que [0] = CD (coletado diretamente) ou CI (coletado indiretamente), [1] = QE (queimado), [2] = JT ou JMR, e [3] = OT (outros), enquanto que para esgotamento sanitário, temos [0] = RC (rede coletora de esgoto) ou FSR (fossa séptica ligada à rede), [1] = FSNR (fossa séptica não ligada à rede), [2] = FR (fossa rudimentar) e [3] = VL (vala) ou JRM (jogado em rio ou mar). Observa-se que cada modelo mostra duas regressões distintas. O modelo restrito é estimado excluindo-se as variáveis não significativas do modelo onde não existe restrição. No que diz respeito às variáveis explicativas, seu número foi reduzido em relação à regressão logit levada a cabo na seção 3 para facilitar a análise, permanecendo as alternativas julgadas as mais importantes.

Tendo em vista que estamos tratando de um modelo de escolha, é preciso ressaltar que somente foram levados em consideração na amostra aqueles domicílios nos quais todas as alternativas apresentadas para esgotamento sanitário e lixo encontram-se disponíveis, pois, caso contrário, a análise estaria sendo aplicada indevidamente. Vale a pena lembrar que a utilização de qualquer opção de fossa (ligada ou não à rede) ou opção de enterrar o lixo, principalmente em locais de baixa urbanização, é prática tão saudável quanto o uso de esgoto ligado à rede ou à coleta direta de lixo.

Em relação aos resultados, alguns pontos merecem ser destacados. Conforme se pode observar a partir da coluna (2) para ambos os modelos de lixo e esgotamento sanitário, a variável *lrendmed*, excetuando-se *urban*, é aquela cujo coeficiente tem o maior impacto sobre a probabilidade de se fazer opção pela alternativa [0], que é a mais salutar à saúde do indivíduo e ao meio ambiente. A escolaridade também mostra ser um fator importante para esta alternativa. Vale a pena estar atento ao comportamento dos determinantes que influenciam as escolhas [2] para o modelo de lixo e [3] para esgotamento, que são as que mais danos causam ao meio ambiente e à saúde. Em ambos os casos, a renda não tem impacto sobre a probabilidade de escolha para essas alternativas. Provavelmente isso se deve ao fato de que as famílias que habitam domicílios nos quais prevalecem esgotamento e lixo com tais características possuem renda que somente as qualificam a morar em tais domicílios.

Tabela 7
Modelo Multinomial para Saneamento: Lixo e Esgotamento

Variáveis Independentes	Lixo		Esgotamento Sanitário	
	Irrestrito (1)	Restrito (2)	Irrestrito (1)	Restrito (2)
Escolha =	= [0]	= [0]	= [0]	= [0]
LRENDMED	0,757 (0,000)	0,786 (0,000)	0,635 (0,000)	0,632 (0,000)
ESCOL	0,132 (0,000)	0,112 (0,000)	0,099 (0,000)	0,101 (0,000)
IDADE	-0,007 (0,015)	-0,008 (0,002)	0,124 (0,000)	0,124 (0,000)
MORADS	0,036 (0,050)	0,050 (0,000)	0,024 (0,000)	0,020 (0,000)
BRANCO	0,261 (0,003)	0,333 (0,000)	0,523 (0,000)	0,523 (0,000)
URBAN	4,811 (0,000)	4,829 (0,000)	2,674 (0,000)	2,673 (0,000)
CTE	-2,641 (0,000)	-2,731 (0,000)	-6,375 (0,000)	-6,356 (0,000)
Escolha =	= [1]	= [1]	= [1]	= [1]
LRENDMED	0,159 (0,004)	0,189 (0,000)	0,298 (0,000)	0,297 (0,000)
ESCOL	0,021 (0,244)	-	0,071 (0,000)	0,072 (0,000)
IDADE	-0,122 (0,000)	-0,136 (0,000)	0,004 (0,000)	0,005 (0,000)
MORADS	-0,015 (0,453)	-	0,042 (0,000)	0,041 (0,000)
BRANCO	-0,074 (0,390)	-	0,198 (0,000)	0,198 (0,000)
URBAN	1,061 (0,000)	1,081 (0,000)	0,551 (0,000)	0,549 (0,000)
CTE	2,943 (0,000)	2,398 (0,000)	-3,229 (0,000)	-3,211 (0,000)
Escolha =	= [2]	= [2]	= [3]	= [3]
LRENDMED	-0,076 (0,174)	-	0,015 (0,520)	-
ESCOL	-0,034 (0,000)	-0,054 (0,000)	-0,025 (0,000)	-0,029 (0,000)
IDADE	-0,012 (0,000)	-0,032 (0,000)	-0,002 (0,050)	-0,002 (0,050)
MORADS	-0,039 (0,050)	-0,045 (0,000)	0,012 (0,216)	-
BRANCO	-0,576 (0,000)	-0,656 (0,000)	0,137 (0,000)	0,134 (0,000)
URBAN	1,732 (0,000)	1,751 (0,000)	0,008 (0,838)	-
CTE	3,629 (0,000)	3,543 (0,000)	-1,616 (0,000)	-1,500 (0,000)
Grupo de Referência	[3]	[3]	[2]	[2]
Teste F para Restrição	-	2,761 (0,543)	-	1,661 (0,645)
Pseudo R ²	0,323	0,323	0,126	0,126
OBS	80493	80493	72994	72994

Nota: Os valores entre parênteses representam o p-valor.

Um ponto que merece destaque é o fato de a escolaridade também reduzir a probabilidade de se efetuar as escolhas [2] e [3]. No que tange às variáveis de controle, podem ser destacados os seguintes fatos: 1) a variável *dummy urban*, para áreas urbanizadas, mostra que os domicílios pertencentes a essas áreas têm maior probabilidade de estarem associados à alternativa [0]. Naturalmente isso se deve a fatores ligados à maior disponibilidade de oferta de serviços de saneamento básico em tais regiões; 2) observa-se também que indivíduos de cor branca têm maior probabilidade de morar num domicílio onde o lixo e esgotamento sanitário estão relacionados à alternativa [0] (que é a mais adequada à saúde do indivíduo e ao meio ambiente).

5 Comentários finais

Este estudo teve como objetivo efetuar uma análise acerca dos determinantes do saneamento básico no Brasil. Embora boa parte das pesquisas aponte para o fato de que o déficit é gerado fundamentalmente por fatores ligados à oferta, como a escassez dos serviços, em virtude da restrição de recursos é fundamental saber também que fatores atuam pelo lado da demanda por tais serviços.

Para perceber isso, basta tomar como exemplo o fato de que mesmo que uma comunidade esteja privada dos serviços públicos de coleta de lixo ou esgotamento sanitário bem como de água tratada, é possível, por iniciativa própria, encontrar formas alternativas de substituir a atuação do Estado, obtendo o mesmo grau de qualidade e bem-estar. Assim, numa comunidade onde o lixo não possa ser coletado diretamente, uma alternativa seria queimar ou enterrar esse lixo. De modo semelhante, em locais onde não existe rede coletora de esgoto, uma alternativa seria a introdução de fossas sépticas, mesmo que estas não sejam ligadas à rede. Em todos esses exemplos, a educação, mais do que a renda dos indivíduos, aparece como o fator decisivo para impulsionar essas iniciativas, na medida em que faz com que o agente tome consciência do valor que tais coisas representam. As iniciativas próprias permitem também que o Estado diminua o custo de expansão desses serviços.

Conforme se pôde constatar, algumas variáveis socioeconômicas desempenham papel importante na demanda por saneamento. Entre as que merecem destaque estão a renda e a escolaridade. No caso da primeira, observa-se que o impacto dessa sobre a demanda por saneamento é muito sensível, sofrendo descolamentos estruturais no seu coeficiente para diferentes faixas de renda. Observa-se ainda que a importância desta variável é significativamente mais acentuada para as faixas mais baixas de renda. Quando o exercício é ampliado de modo a incluir as alternativas para cada os casos do lixo e esgotamento sanitário, surgem alguns pontos interessantes. Em ambos os casos, a escolaridade exerce boa influência sobre a decisão do agente optar pelas alternativas que causam maior impacto negativo sobre o meio ambiente, e que em ambos os casos significa poluir rio ou mar.

Bibliografia

- Andrade, T. A.; Brandão, A. S. P.; Lobão, Waldir J. A.; Silva, S. L. Q. da. Saneamento urbano: a demanda residencial por água. *Pesquisa e Planejamento Econômico*, v. 25, n. 3, p. 427-448, 1995.
- Dow, W. H. Flexible discrete choice demand models consistent with utility maximization: an application to health care demand. *The American Agricultural Economics Association Annual Meeting*, New York, 1999.
- Fernandez, J. C.; Menezes, W. F. *A avaliação contingente e a demanda por serviço público de esgotamento sanitário: uma análise a partir da região do Alto Subaé*. Mimeografado.
- Garcia, F.; Rebelo, A. M. Déficit habitacional e desigualdade de renda no Brasil. *Economia Aplicada*, v. 6, n. 2, p. 239-264, abr./jun. 2002.
- Greene, W. *Econometric analysis*. Prentice Hall, 1993.
- Goldberg, A. *A course in econometrics*. Harvard University Press, 1991
- Hensher, D. A.; Greene, W. H. *Specification and estimation of nested logit model: alternative normalizations*. New York University, 2000. Mimeografado.
- Judge, G.; Hill, C.; Griffiths, W.; Lee, T.; Lütkepohl, H. *Introduction to the theory and practice of econometrics*. New York: Wiley, 1982.
- Leibwitz, A.; Friedman, B. Family bequests and the derived demand for health inputs. *Economic Inquiry*, v. 17, n. 3, p. 419-434, 1979.
- Maddala, G. S. *Limited dependent and qualitative variables in econometrics*. Cambridge University Press, 1998.
- McFadden, D. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (ed.), *Frontiers in econometrics*. Academic Press, 1974, p. 105-142.
- Nocera, S.; Zweifel, P. The demand for health: an empirical test of the Grossman model. In: Zweifel, P. (editor), *Medical profession and regulation*. Kluwer, 1998.
- Persson, T. H. Welfare calculations in models of the demand for sanitation. *Applied Economics*, v. 34, n. 12, p. 1509-1518, 2002.

_____. Demand for water and sanitation in Bangladesch. *Working Paper* 2001-03, Dept. of Economics, Lund University, Sweden.

_____. Household choice of drinking water source in Phillipines. *Working Paper* 2000-13, Dept. of Economics, Lund University, Sweden.

STATA. "Stata User's Guide, Release 8". College Station, Texas: Stata Press.

Anexo 1 - Descrição das variáveis

Descrição das variáveis da PNAD (1996) utilizadas na pesquisa

- MORADS = número de moradores do domicílio;
 - SEXO = sexo da pessoa de referência do domicílio;
 - IDADE = idade da pessoa de referência do domicílio quando foi realizada a pesquisa;
 - RENDOMC = renda média das pessoas residentes no domicílio;
 - ESCOL = nível de escolaridade da pessoa de referência do domicílio em número de anos completos de estudo;
 - INSTM = nível de escolaridade da mãe da pessoa de referência em número de anos completos de estudo;
 - BRANCO = variável *dummy* que assume o valor 1 se a pessoa de referência do domicílio é branca e 0 se ela é negra ou parda;
 - NEGRO = variável *dummy* que assume o valor 1 se a pessoa de referência do domicílio é negro e 0 se ela é branca ou parda;
 - PARDO = variável *dummy* que assume o valor 1 se a pessoa de referência do domicílio é parda e 0 se ela é branca ou negra;
 - URBAN = variável *dummy* que assume o valor 1 se o domicílio está localizado em área urbana; 0 para área rural;
 - SUDESTE = variável *dummy* que assume o valor 1 se o domicílio está localizado na região Sudeste, 0 caso contrário;
 - SUL = variável *dummy* que assume o valor 1 se o domicílio está localizado na região Sul, 0 caso contrário;
 - CENTRO = variável *dummy* que assume o valor 1 se o domicílio está localizado na região Centro-Oeste, 0 caso contrário;
 - NORDESTE = variável *dummy* que assume o valor 1 se o domicílio está localizado na região Nordeste e, 0 caso contrário.
-

Classificação do escoadouro do banheiro de uso dos moradores do domicílio

- RC = Rede coletora de esgoto ou fluvial;
 - FSR = Fossa séptica ligada à rede;
 - FSNR = Fossa séptica não ligada à rede;
 - FR = Fossa rudimentar;
 - VL = Vala; e
 - JMR = Jogado diretamente no mar ou rio.
-

Classificação do destino do lixo proveniente dos domicílios

- CD = Coletado diretamente;
 - CI = Coletado Indiretamente;
 - QE = Queimado ou enterrado na propriedade;
 - JT = Jogado em terreno baldio ou logradouro;
 - JMR = Jogado diretamente no mar ou rio; e
 - OT = Outros.
-

