

Estimativa do Limiar de Evasão de Clientes SABESP

Denisard Cneio de Oliveira Alves¹

Paula Carvalho Pereda²

Alexsandros Fraga³

Daniel Grimaldi³

Introdução

O objetivo deste estudo é estimar o limiar do fornecimento de água que, dada a atual estrutura tarifária da SABESP, leva os grandes clientes a buscar fontes alternativas de fornecimento de água.

Foi identificada pela SABESP a recorrência da busca por novas fontes de fornecimento de água. Entre as principais fontes alternativas de fornecimento de água destacam-se: poços (semi) artesianos; contratação de caminhões-pipa; e transporte de água de municípios vizinhos.

Algumas das evidências de aumento da concorrência com a SABESP, ou seja, do aumento de mercado concorrencial para o fornecimento de água para grandes clientes são: perda efetiva de clientes da Sabesp nos últimos anos (Base de dados da SABESP); e aumento do número de outorgas concedidas pelo DAEE para fontes alternativas (Site do DAEE).

Desta forma, acredita-se que os principais motivos que levam à evasão de clientes são: o diferencial entre tarifas SABESP e preço da água oferecido pelo mercado de fontes alternativas; a presença de condições necessárias para instalação de poços artesianos; e a estrutura tarifária da SABESP, que é progressiva com o consumo.

Com base nas evidências apontadas, este estudo pretende estimar os limiares de consumo que fazem com que os clientes da SABESP busquem outros tipos de

¹ Professor Titular do Departamento de Economia FEA-USP.

² Mestre em Economia pela FEA-USP.

³ Mestrando em Economia pela FEA-USP.

fornecimento de água, que não pelo sistema SABESP. Deste modo, este relatório divide-se em quatro seções. A primeira seção apresenta aspectos principais da base de dados disponibilizada pela SABESP; a maneira como esta foi organizada; a descrição das variáveis criadas; e as principais estatísticas descritivas dos dados. Na segunda seção discute-se o modelo utilizado para a estimação pretendida e alguns dos critérios definidos para a elaboração do modelo. A terceira seção apresenta os principais resultados das estimações e, por fim, a última seção é composta de considerações finais.

1 Base de Dados

1.1 Organização da Base de Dados

O trabalho se apóia em dados, disponibilizados diretamente pela SABESP, a respeito de **seus clientes comerciais e industriais**. Para cada um desses clientes foram fornecidas as seguintes informações: grupo de faturamento; ano e mês de referência da última conta faturada; número do RGI; código do município; código do atendimento comercial; tipo da ligação; tipo de cobrança; endereço; data da ligação de água; data da ligação de esgoto; status da ligação; tipo da codificação; ramo de atividade no fim do período de referência; quantidade de economia residencial, comercial, industrial e pública; tipo de rol; controle de faturamento de rol especial; código de leitura dos últimos 60 meses; consumo faturado mensal dos últimos 60 meses; consumo médio dos últimos 60 meses; e status do consumo médio dos últimos 60 meses. Dessas informações, algumas merecem ser comentadas.

O número do RGI (determinado pelo campo *rgi* no arquivo original) serviu para definir unicamente cada cliente, de tal forma que todas as outras informações estão a ele ligadas. Vale ressaltar que é possível que um determinado cliente mude de endereço e, com isso, passe a consumir do sistema SABESP através de outro RGI. No entanto, não é possível, dentro dessa base de dados, identificar estes casos. Apesar das limitações da metodologia, a suposição de que os clientes permaneceram no mesmo RGI durante o período analisado foi a opção viável.

O endereço disponível na base de dados tinha a seguinte peculiaridade: era informado pelos próprios clientes e de maneira não compulsória. Isto fez com que essa informação fosse pouco confiável e que se verificasse muita omissão. Por fim, o uso de tal informação foi inviabilizado.

Para a criação de uma variável capaz de controlar, nas estimações, possíveis efeitos de localidade, optou-se por utilizar o código de atendimento comercial. Por sua vez, o campo denominado *ramo*, que indicava o código do setor de atividade da ligação, foi utilizado para controlar efeitos específicos à atividade fim dos clientes, pois se espera que algumas atividades industriais induzam ao maior ou menor consumo de água.

O consumo faturado mensal para cada RGI foi disponibilizado para os últimos 60 meses até Agosto de 2008. Esse campo apresentava 420 caracteres, onde cada seqüência de 7 dígitos representava o consumo de um mês específico. Como a data de fechamento da fatura difere de acordo com o cliente, o último mês de consumo faturado também variou - sendo agosto para um grupo de clientes, e julho para outro. Tal fato exigiu um tratamento específico para evitar que efeitos de sazonalidade do consumo prejudicassem a estimação e a inferência estatística.

Com estas informações foi possível separar a amostra total, de 795.128 clientes, em três grupos. O primeiro, que totalizou 346.435 clientes, foi formado pelos clientes que tinham os dados de consumo faturado disponíveis para o período de agosto de 2003 até agosto de 2008. No segundo grupo, que totalizou 407.334 clientes, encontravam-se clientes para os quais a informação de consumo faturado mensal estava disponível para o período de julho de 2003 a julho de 2008. Por fim, foram identificados 41.299 clientes para os quais a informação de consumo faturado estava em branco, ou não se sabia o mês de referência da última conta faturada - e, portanto, não foi possível classificá-los em nenhum dos dois grupos anteriores. Esse último grupo foi desconsiderado para as estimações.

O período total de referência da pesquisa, portanto, englobou 61 meses entre julho de 2003 e agosto de 2008, contendo informações de RGI e consumo faturado para o período de referência. Ademais, mais uma base de dados foi fornecida. Nesta base foram listados todos os registros (RGIs) que apresentaram alterações no ramo de atividade ao longo do período de referência. A partir dessa base de dados foi possível definir como ramo de atividade inicial aquele informado na primeira atualização de

cada RGI. Se um RGI não aparecia em nenhum desses dois arquivos, isso significava que não havia ocorrido alteração de ramo para aquele RGI no período analisado. Neste caso, definiu-se que o *ramo de atividade final* (campo ramo das bases capital e outros) era igual ao *ramo inicial*. Tais informações foram transformadas em variáveis binárias.

1.2 Critérios para definição da Evasão dos Clientes SABESP

Com a base de dados pronta, o foco se tornou a criação de uma variável que fosse capaz de captar corretamente a evasão de grandes clientes do sistema SABESP, uma vez que o “ex-cliente” não informa formalmente à companhia sua saída. Técnicos da SABESP relatam que, na maior parte dos casos, somente por meio de inspeções de equipes de campo, motivadas por uma percepção de “queda incomum” do consumo, é possível identificar consumidores que se retiraram do sistema.

Portanto, seguindo essa prática da companhia e respeitando as limitações da base de dados, o comportamento do consumo de cada RGI traz a principal informação disponível para as estimações.

Primeiramente, foi preciso separar flutuações ocasionais de consumo de mudanças que indicassem uma tendência. Para tanto, optou-se por trabalhar sempre com as médias de consumo semestral ao invés do consumo mensal – já que esta última variável está muito mais sujeita a oscilações.

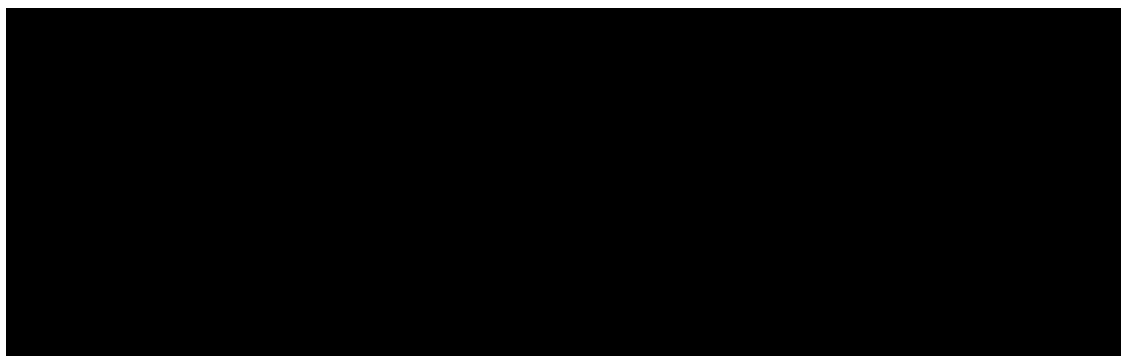
No entanto, ao se trabalhar com uma média semestral é preciso levar em consideração a possibilidade de oscilações sazonais comuns a diversas atividades econômicas - como ocorre na indústria de bebidas, por exemplo. Desta maneira, acredita-se que simplesmente comparar a média de semestres seguidos poderia levar a análises equivocadas, pois uma queda sazonal de demanda por água poderia ser interpretada como uma tendência.

Logo, para contornar esse problema, decidiu-se que a média de consumo semestral de cada semestre (seja Janeiro a Junho, ou Julho a Dezembro) seria comparada com a média de consumo do mesmo período no ano anterior.

Além disso, foi preciso determinar precisamente o ponto a partir do qual uma queda de consumo faturado deve ser classificada como “incomum”. Em trabalho anterior

contratado pela SABESP⁴, foram utilizadas as seguintes classificações para as tendências de consumo:

Quadro 1.1 – Critério de classificação de cliente a partir de comportamento de consumo



Essa classificação, no entanto, se mostrou muito rígida ao definir que somente com consumo zero um RGI poderia ser tido como “ex-cliente”. O mais comum entre os clientes que se retiram do sistema SABESP é uma queda drástica (já que passa a existir uma nova fonte de água), mas ainda com a presença de um consumo residual.

Os clientes industriais e comerciais utilizam uma grande porcentagem de água que não requer alta qualidade. É justamente essa parcela de consumo que é substituída. A água utilizada para fins mais nobres permanece sendo demandada junto à SABESP, justificando o consumo residual.

Resolveu-se, após opiniões de técnicos ligados à companhia, estabelecer como critério para classificação de ex-cliente uma queda de consumo médio semestral igual ou superior a 90%. Uma queda dessa ordem, no período analisado, com certeza não pode ser explicada por fatores conjunturais de mercado – como crises de demanda ou fatores sazonais. Tal queda de consumo só pode ser explicada pela retirada do cliente do sistema SABESP – seja pelo fechamento do negócio (comércio ou indústria) ou pela migração para uma nova fonte de abastecimento de água. Assim, as tendências de consumo foram classificadas como a seguir:

⁴ “Perdas de Grandes Clientes” (Agosto de 2008), este trabalho utiliza ferramentas de datamining para definição de critérios.

Quadro 1.2 - Critério 1 definido para o estudo



A partir dessa classificação foram construídas as *dummies* utilizadas como variáveis de interesse das regressões estimadas. Logo, se o consumo médio semestral associado a um determinado RGI apresenta redução superior a 90% (quando comparado com o consumo médio no mesmo semestre do ano anterior), a variável *cons_exclientes* a ele associada assume valor unitário. Caso contrário, assume valor zero.

1.3 Descrição dos critérios de divisão de amostras

A base de dados que foi usada no trabalho engloba toda população de clientes comerciais e industriais da Sabesp e contém 753.829 RGI's. De acordo com a média de consumo do primeiro semestre de 2004, mais de 95% desses RGI's são de clientes fora do sistema (consumo médio zero) ou de clientes de pequeno porte (consumo médio semestral abaixo de 50m³).

O objetivo deste trabalho é testar a hipótese de que os grandes clientes da Sabesp têm uma probabilidade maior de buscar fontes alternativas de água. Conclui-se que a população de interesse do estudo é uma pequena fração da população total de clientes industriais e comerciais da Sabesp. Diante desse quadro, optou-se por usar apenas a parcela da população que representava a população de interesse.

2 Metodologia

Dois modelos distintos foram utilizados para estimar o volume de fornecimento de água a partir do qual a probabilidade de evasão do cliente SABESP aumenta, sendo eles:

- 1) Modelo Probit com dados da comparação de 1 período;
- 2) Modelo Logit com efeitos fixos com dados da comparação de 7 períodos.

Ambos os modelos probit e logit são modelos de escolha binária, ou modelos de probabilidade. O princípio básico de modelos de probabilidade é o fato de a variável dependente assumir valor 1 para o caso de interesse, e 0 caso contrário. Em um modelo de regressão clássica, o interesse está no efeito marginal (que é o próprio coeficiente da regressão) de alguma variável x sobre a média condicional da variável dependente, y . Em modelos de probabilidade, ou modelos de escolha binária, o interesse se concentra em encontrar o efeito que uma variação em x provoca na probabilidade condicional de $y|x$ ⁵.

A seguir, ambos os métodos de estimação são apresentados em detalhes.

2.1 Modelo Probit

Uma vez detectada a evasão do cliente, com base nos critérios definidos na seção anterior e na comparação entre os dados de 2008 e 2004, é possível estimar o volume de fornecimento de água a partir do qual a probabilidade de evasão do cliente SABESP cresce, ou seja, a partir de que nível de consumo o cliente da SABESP tem maior probabilidade de procurar fornecedor alternativo.

O modelo probit foi utilizado devido à natureza da variável de interesse do estudo (evasão), que assume valor um, ou zero. Neste caso, o modelo permite que se explique o efeito do volume de consumo e de outras variáveis intrínsecas dos clientes na probabilidade de os clientes deixarem o sistema de abastecimento da SABESP e buscarem alternativas para o fornecimento de água. A constatação de efeito significativo do nível de consumo na probabilidade de evasão do sistema pode ser evidência de elasticidade preço da demanda diferente entre os tipos de clientes da SABESP (pequeno, médio, ou grande).

Varáveis disponíveis no cadastro de clientes da SABESP, tais como: localidade do RGI; e setores de atividade, permitiram qualificar melhor a estimativa da

⁵ Lê-se “*y* dado *x*”.

probabilidade de evasão, explicando possíveis diferenças de probabilidade de saída do sistema de acordo com o setor ao qual pertence o cliente e à região de instalação de suas atividades.

Logo, as variáveis dos ramos de atividades das empresas clientes da SABESP foram utilizadas como controles no modelo probit. A amostra utilizada na estimação é representativa para os municípios atendidos pela SABESP na Grande São Paulo. As ligações da SABESP pertencentes a municípios fora da Grande São Paulo foram excluídas da amostra por apresentarem poucas informações sobre os ramos de atividades.

O volume de água, dada a estrutura tarifária constante, ao longo dos cinco últimos anos, é a variável que permitirá ao cliente se ajustar no seu objetivo de minimizar custos. O volume médio faturado de água entrou na estimação no nível (em termos de m³) e ao quadrado, de maneira a captar a diferença de sensibilidade dos clientes à estrutura tarifária da SABESP, conforme aumenta o consumo médio dos mesmos.

Para a construção do modelo, definiu-se:

$$Y = \begin{cases} 1, & \text{se o cliente saiu do sistema SABESP} \\ 0, & \text{caso contrário} \end{cases}$$

A variável Y assume o valor 1 com base nos critérios definidos na Seção 1.2 deste estudo. Assim, as variáveis de controle são: *dummies* para ramos de atividades⁶; volume de água faturado mensalmente na SABESP; e volume ao quadrado de água faturado mensalmente na SABESP. Tais variáveis representadas pela matriz X, com número de colunas igual ao total de variáveis (nº de *dummies* + constante + variáveis de volume de água), com N linhas representando as N observações da amostra.

A partir da comparação entre o consumo médio faturado do primeiro semestre de 2008 e do primeiro semestre de 2004, calculou-se a variável Y. Assim, foi possível detectar se houve redução extraordinária de consumo que possa indicar que esses

⁶ *Dummies* são variáveis binárias que têm valor unitário quando ocorre o comportamento que se pretende medir e valor nulo caso contrário. Neste caso, as variáveis *dummies* para ramos de atividade tiveram valor unitário para cada setor de atividade, de acordo com os códigos de atividades disponibilizados pela SABESP.

clientes buscaram fontes alternativas de água. A probabilidade de evasão é dada por $p(Y = 1)$, que é uma curva na forma de S e expressa uma relação não linear entre Y e as variáveis indicadores.

A relação funcional representando esta curva é a função Probit. A função Probit está relacionada à função de distribuição da normal padronizada [Wooldridge, 2006]. Se Z for uma variável aleatória normal padronizada, a sua função densidade de probabilidade é dada por:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-0.5z^2}$$

A função Probit é, então, dada por:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-0.5z^2}$$

O modelo estatístico expressa a probabilidade do Y ser igual a 1, ou seja, o cliente saiu do sistema SABESP, é dado por:

$$p = P[Z \leq \Phi(\beta_1 + \beta_2 \text{setor} + \beta_3(m^3) + \beta_4(m^3)^2)$$

em que os betas são os parâmetros a serem estimados usando-se as N observações da amostra.

É importante ressaltar que, nestes tipos de modelos, o vetor de parâmetros β não tem a interpretação usual dos modelos de regressão tradicionais, ou seja, estes parâmetros não representam diretamente os efeitos marginais das variáveis explicativas sobre a variável dependente.

2.1.1 Efeitos Marginais

O impacto da variação de uma unidade em qualquer das variáveis explicativas, no caso em estudo: ramo de atividade; m^3 ; e $(m^3)^2$, sobre a probabilidade do cliente sair do sistema SABESP pode ser escrito da seguinte maneira:

$$\frac{\partial p}{\partial x} = \frac{\partial \Phi(t)}{\partial t} \frac{\partial t}{\partial x} = \phi(t) \beta_k, \text{ com } k = 2, 3, 4 \text{ e}$$

$$t = \beta_1 + \beta_2 \text{setor} + \beta_3 m^3 + \beta_4 (m^3)^2$$

Assim, para se obter a probabilidade marginal de uma variável – ou seja, o efeito de uma variação de uma variável explicativa na probabilidade de evasão – torna-se necessário avaliar a função densidade de probabilidade em algum ponto t estimado. Uma vez escolhidos os níveis das variáveis explicativas para t , multiplica-se a função densidade de probabilidade neste ponto pelo coeficiente estimado da variável (beta) que se deseja obter o efeito marginal.

Na medida em que se altera o setor, a probabilidade do cliente sair do sistema SABESP também se altera. O efeito marginal capta este impacto ao mesmo tempo em que testa se este aumento ou redução de probabilidade de evasão do sistema é significativamente diferente de zero.

No caso do volume de consumo em m^3 , a presença do termo quadrático (não-linear) altera o cálculo do efeito marginal do volume médio de consumo na probabilidade do cliente sair. Assim, tem-se:

$$\frac{\partial p}{\partial m^3} = \frac{\partial \Phi(t)}{\partial t} \frac{\partial t}{\partial m^3} = \phi(t)[\beta_3 + 2\beta_4 m^3],$$

em que $t = \beta_1 + \beta_2 \text{setor} + \beta_3 m^3 + \beta_4 (m^3)^2$

Uma vez obtidos os coeficientes estimados do Probit, calcula-se os efeitos marginais com base em alguma unidade das variáveis explicativas. É comum que sejam escolhidas medidas de posição, como média e mediana, para descrever os resultados dos efeitos marginais.

2.2 Modelo Logit com Efeitos Fixos

O modelo logit, assim como o probit, é um modelo de probabilidade, ou escolha binária. Para definir qual dos modelos é o mais adequado, Greene (2003) afirma que as duas distribuições (Normal e Logística) são muito parecidas, exceto nas caudas, onde a distribuição logística apresenta-se mais densa. Entretanto, para valores

intermediários de $x'\beta$ as duas distribuições tendem a gerar resultados semelhantes. A Distribuição Logística apresenta tendência a dar probabilidades maiores para $Y=0$ quando t é extremamente pequeno (e probabilidades menores para $Y=0$ quando t é grande) em relação à Distribuição Normal.

Ainda de acordo com Greene (2003), embora existam razões práticas para favorecer o uso de um modelo em detrimento de outro por conveniências matemáticas, é muito difícil encontrar justificativas fundamentadas teoricamente para esta escolha. Ambos os modelos são substitutos e a utilização dos dois serve para corroborar os resultados encontrados, independentemente da distribuição de probabilidade assumida para os erros.

Entretanto, o modelo logit apresenta vantagens sobre o modelo probit quando se tem base de dados em painel e, assim, a presença de efeitos específicos, pois o primeiro não exige que sejam feitas hipóteses sobre a relação do efeito específico não observado e as variáveis independentes do modelo.

Antes de se entrar nos detalhes do modelo logit com efeitos fixos, para facilitar o entendimento do modelo, decidiu-se acrescentar uma introdução sobre o Método de Efeitos Fixos. Os textos a seguir foram escritos com base em Wooldridge (2001).

2.2.1 Modelo de Efeitos Fixos

A utilização do método de Efeitos Fixos, no caso deste estudo, deve-se à possibilidade de transformar a base de dados disponibilizada em um Painel de dados. Essa transformação oferece diversas vantagens na estimação, uma vez que elimina alguns problemas inerentes aos dados em *cross-section* e o problema da falta de controles na regressão que explica a evasão de clientes da SABESP. Dados em painel requerem o emprego de modelos econométricos adequados. Um modelo básico de dados em painel pode ser representado por:

$$y_{it} = x'_{it} \beta + z'_i \alpha + \varepsilon_{it}$$

O termo x'_{it} possui K regressores, incluindo o vetor de 1's (para a constante). Na equação acima, o efeito individual, ou heterogeneidade, é representado por $z'_i \alpha$, onde

z_i contém um conjunto de variáveis específicas para cada indivíduo, que podem ser observados ou não, e se mantêm constantes ao longo do tempo. O modelo a ser empregado depende do tipo de heterogeneidade assumida.

Se z_i é não observado, mas correlacionado com x_{it} , então o estimador de Mínimos Quadrados Ordinários é viesado e inconsistente⁷, como consequência de uma variável relevante omitida. Neste caso, o modelo pode ser formulado por:

$$y_{it} = x'_{it} \beta + \alpha_i + \varepsilon_{it}$$

Nesta equação, $\alpha_i = z'_i \alpha$ engloba os efeitos não observados, e constantes no tempo, que afetam y_{it} . Esta abordagem é chamada de efeito fixo exatamente por assumir que α_i é um termo constante específico de cada indivíduo que não varia ao longo do tempo.

Para cada i , pode-se tirar a média da equação: $\bar{y}_i = \bar{x}'_i \beta + \alpha_i + \bar{\varepsilon}_i$

Como o termo α_i é constante no tempo, ele aparecerá em ambas as equações. Subtraindo a equação de sua média, temos:

$$y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i) \text{ ou } \bar{y}_{it} = \bar{x}_{it} \beta + \bar{\varepsilon}_{it}$$

em que $\bar{y}_{it} = y_{it} - \bar{y}_i$; $\bar{x}_{it} = x_{it} - \bar{x}_i$ e $\bar{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$.

É importante notar que o efeito individual, α_i , desaparece, o que sugere que esta equação pode ser estimada por Mínimos Quadrados Ordinários. Os estimadores obtidos por este procedimento são chamados de estimadores de efeitos fixos ou estimadores *within*. O intuito da formulação por efeitos fixos é eliminar α_i , ou seja, todos os efeitos não observados que são constantes no tempo⁸.

⁷ Um estimador/parâmetro estimado é considerado viesado quando, em média, não acerta o valor verdadeiro.

⁸ Não se podem incluir variáveis constantes no tempo na estimação de Efeitos Fixos, pois estas seriam perfeitamente colineares à variável α_i , ou seja, não seria possível estimar o modelo (não atende à hipótese de multicolinearidade perfeita).

O modelo de EF perde N graus de liberdade para estimar, pois é preciso calcular as médias para cada observação no tempo, assim o total de graus de liberdade é $N.T - N = N(T-1)$ ⁹.

2.2.2 Modelo Logit com Efeitos Fixos

Como foi visto no início desta seção, o modelo logit apresenta vantagens sobre o modelo probit quando há efeitos específicos. Desta forma, utilizou-se o modelo logit na estimação do efeito do tamanho do cliente (volume) na sua probabilidade de sair do sistema de abastecimento da SABESP. De acordo com Wooldridge (2002), o modelo de resposta discreto de efeitos fixos pode ser definido por:

$$P(y_{it} = 1 / x_{it}, \alpha_i, 0 \leq n_i \leq 7) = \Lambda(x_{it}\beta + \alpha_i)$$

$$P(y_{it} = 0 / x_{it}, \alpha_i, 0 \leq n_i \leq 7) = 1 - \Lambda(x_{it}\beta + \alpha_i)$$

em que $\Lambda(\cdot)$ é a função logística, x_{it} é a matriz de controles, β é o vetor de coeficientes, α_i o efeito não observado e $n_i = \sum_{t=1}^7 y_{it}$.

O volume médio faturado de água entrou nesta estimação, assim como no modelo probit, no nível (em termos de m³) e ao quadrado, de maneira a captar a diferença de sensibilidade dos clientes à estrutura tarifária da SABESP, conforme aumenta o consumo médio dos mesmos. Outros controles como: localização do RGI; ramo de atividade do cliente; entre outras características da ligação, tiveram seus impactos capturados pelo efeito específico não observado, o chamado ‘efeito fixo’.

Para a construção deste modelo, definiu-se:

$$Y = \begin{cases} 1, & \text{se o cliente saiu do sistema SABESP} \\ 0, & \text{caso contrário} \end{cases}$$

A variável Y assume o valor 1 com base nos seguintes critérios:

- *Para semestres pares:*

⁹ Este método também é idêntico a se estimar a equação adicionando-se uma *dummy* para que cada indivíduo tenha um intercepto diferente (modelo LSDV).

$$Y = \begin{cases} 1, & \text{se Consumo_M\u00e9dio_Semestre}_i \leq k \cdot \text{Consumo_M\u00e9dio_Semestre}_{t-2} \\ 0, & \text{caso contr\u00e1rio} \end{cases}$$

- *Para semestres \u00edmpares:*

$$Y = \begin{cases} 1, & \text{se Consumo_M\u00e9dio_Semestre}_i \leq k \cdot \text{Consumo_M\u00e9dio_Semestre}_{t-2} \\ 0, & \text{caso contr\u00e1rio} \end{cases}$$

em que $k=0,5$, ou $0,75$, ou $0,90$; dependendo do crit\u00e9rio utilizado para a defini\u00e7\u00e3o de ex-clientes.

Os dados foram organizados em painel da seguinte maneira:

- Definiu-se, em cada semestre, a vari\u00e1vel dependente para o cliente, que mensurou a sa\u00edda ou perman\u00eancia do cliente na SABESP;
- A vari\u00e1vel de consumo m\u00e9dio (vari\u00e1vel independente) considerada para cada per\u00edodo \u00e9 igual ao consumo no semestre de compara\u00e7\u00e3o anterior, ou seja, se Y foi definido a partir da compara\u00e7\u00e3o do 1\u00b0 semestre de 2005 com o 1\u00b0 semestre de 2004, a m\u00e9dia do consumo faturado utilizada como controle foi a do 1\u00b0 semestre de 2004.

Para os valores extremos de n , 0 e 7, a distribui\u00e7\u00e3o condicional \u00e9 dita n\u00e3o informativa para os par\u00e2metros do modelo (betas), sendo estas observa\u00e7\u00f5es exclu\u00eddas da modelagem. Isto ocorre devido ao fato de estes valores determinarem o resultado da vari\u00e1vel dependente.

O modelo logit com efeitos fixos depende da hip\u00f3tese de exogeneidade estrita condicional a α_i para a consist\u00eancia de seus estimadores, logo,

$$P(y_{it} = 1 / x_i, \alpha_i) = P(y_{it} = 1 / x_{it}, \alpha_i) = \Lambda(x_{it} \beta + \alpha_i).$$

Neste modelo, o que se faz \u00e9 encontrar a distribui\u00e7\u00e3o conjunta de $\mathbf{y}_i \equiv (y_{i1}, \dots, y_{i7})'$ condicional a $\mathbf{x}_i, \alpha_i, n_i$. Esta distribui\u00e7\u00e3o condicional n\u00e3o depender\u00e1 de α_i , sendo a distribui\u00e7\u00e3o de $\mathbf{y}_i \equiv (y_{i1}, \dots, y_{i7})'$ condicional a \mathbf{x}_i, n_i .

$$P(y_{i1} = y_1, \dots, y_{i7} = y_7 | x_i, \alpha_i, n_i = n) = P(y_{i1} = y_1, \dots, y_{i7} = y_7 | x_i, \alpha_i) / P(n_i = n | x_i, \alpha_i)$$

O método utilizado para estimar os coeficientes (β) é o de Máxima Verossimilhança Condicional, que gera estimadores logit de efeitos fixos.

$$l_i(\beta) = \log \left\{ \exp \left(\sum_{t=1}^T y_{it} x_{it} \beta \right) \left[\sum_{\theta \in R_i} \exp \left(\sum_{t=1}^T \theta_t x_{it} \beta \right) \right] \right\}$$

em que $\left\{ \theta_t \in R^T : \theta_t \in [0,1], \sum_{t=1}^T \theta_t = n_i \right\}$

A intuição deste modelo e a interpretação do efeito não observado são as mesmas do método de efeitos fixos. Entretanto, para encontrar os efeitos parciais nas probabilidades de resposta, é preciso supor valores para o alfa. A maneira de derivar os efeitos marginais é equivalente ao modelo probit, especificado na subseção anterior, com exceção à suposição que deve ser feita para o efeito fixo. No logit com efeitos fixos, para que se encontre o t estimado (combinação linear das variáveis explicativas + efeito fixo), que entra na função densidade de probabilidade para o cálculo dos efeitos marginais, deve-se supor algum valor para o efeito fixo ou considera-lo zero. Esta segunda alternativa é mais prudente de acordo com Wooldridge (2002) e menos arbitrária. Logo, os efeitos marginais calculados neste trabalho consideram nulo o efeito fixo, sendo a equação para os efeitos marginais idêntica à apresentada para o modelo probit.

3 Resultados das Estimações

3.1 Resultados do Probit

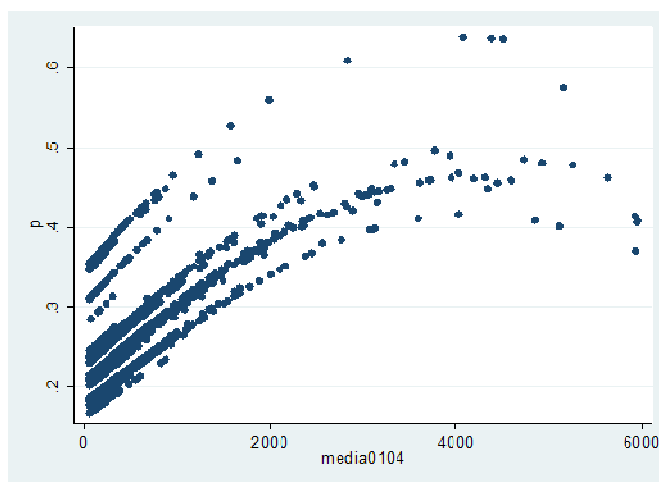
Como foi visto na seção anterior, as estimações do modelo probit utilizaram dados em *cross-section* das ligações (clientes) da SABESP para a Grande São Paulo.

Seguem abaixo as especificações amostrais do modelo estimado:

A amostra consiste dos clientes cujo consumo faturado médio do 1º semestre de 2004 foi acima de 50m³.

Os clientes que apresentaram redução maior de 90% entre os primeiros semestres de 2004 e 2008 foram considerados como ex-clientes do sistema SABESP, i.e. fazem parte da amostra dos que se evadiu do sistema SABESP¹⁰.

Gráfico 3.1: probabilidade de evasão por nível de consumo (critério 1)



3.1.1- Interpretação dos resultados

O gráfico indica que a probabilidade de evasão da SABESP é crescente conforme aumenta o nível de consumo dos RGIs¹¹; observou-se uma redução na probabilidade de evasão a partir do consumo médio 3000-4000m³, o que corrobora a importância do programa **Sabesp Soluções Ambientais**, que garante contratos com estrutura tarifária diferenciada para clientes cujo consumo seja igual ou superior a 3.000m³/mês.

As diferenças de patamar verificadas nos gráficos das probabilidades devem-se à inclusão de *dummies* para os ramos de atividade de cada cliente. Este resultado mostra que a probabilidade de saída da SABESP varia significativamente entre os ramos de atividades.

¹⁰ Optou-se por reportar apenas os resultados para amostra com clientes com consumo superior a 50m³ e apenas para ex clientes do tipo 1 devido a baixa significância estatística dos para os dados de cross-section. Veja o Apêndice A.

¹¹ Esse resultado correu independentemente do critério utilizado para definição de "ex-cliente", pois outros dois critérios foram usados para definir um ex cliente: redução de 50% e de 75% no consumo de água entre os semestres.

Os ramos de Atividades que apresentaram maiores probabilidade de evasão, em média, foram: Construção Civil; P&D; Eletricidade e Saneamento; Extrativismo; Telecom; e Transporte. Em contrapartida, os setores de atividade com menores probabilidades médias de evasão foram: Educação e Saúde; Atividades Imobiliárias; Financeiras; e Petroquímica.

3.2 Resultados do Modelo Logit com Efeitos Fixos

Como foi visto na seção anterior, as estimações do modelo probit utilizaram dados em *cross-section* das ligações (clientes) da SABESP para a Grande São Paulo. Foram estimados 9 modelos, utilizando os diversos critérios de seleção de amostra (cortes no consumo médio)¹² e as classificações para a definição de "ex-cliente", ou cliente que saiu do sistema de abastecimento da SABESP¹³. A descrição dos resultados estimados para ex cliente Tipo 1(mais de 90% de redução no consumo de água), os gráficos com os resultados estimados e a interpretação dos resultados estão resumidos nas subseções abaixo¹⁴.

3.2.1 Amostra: Consumo médio acima de 50m³

3.2.1.1 - Ex-cliente tipo 1

Critério de definição de amostra e de ex-clientes e da amostra:

- *Amostra 1*: RGIs cujo consumo faturado médio do 1º semestre de 2004 foi acima de 50m³;

¹² Três amostras foram utilizadas nas estimações: Amostra 1, para clientes com consumo mensal médio superior a 50m³ no 1º semestre de 2004; Amostra 2, para clientes com consumo mensal médio superior a 300m³ no 1º semestre de 2004; Amostra 3, para clientes com consumo mensal médio superior a 1000m³ no 1º semestre de 2004. Neste trabalho são reportados os resultados da amostra definida para clientes com consumo mensal médio superior a 50m³ no primeiro semestre de 2004. Os resultados para as demais definições de clientes estão No Apêndice D, em saída do Stata, mas os gráficos e discussão dos resultados estão no Relatório FIPE-SABESP (2008)

¹³ Foram definidas três classificações para "ex-cliente": Tipo 1: Clientes que apresentaram redução maior que 90% entre os semestres pares/impares; Tipo 2: Clientes que apresentaram redução maior que 75% entre os semestres pares/impares; Tipo 3: Clientes que apresentaram redução maior que 50% entre os semestres pares/impares. .

¹⁴ Detalhes das estimações podem ser vistos no Anexo D deste trabalho.

- *Ex-cliente tipo 1*: RGIs que apresentaram redução maior que 90% entre os semestres pares/ímpares.

Gráfico 3.10: Efeito marginal do consumo (critério 1)

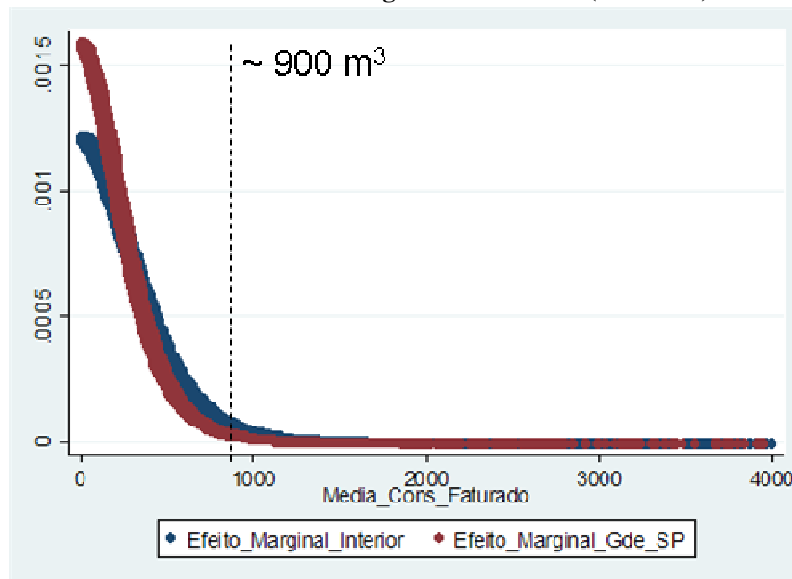
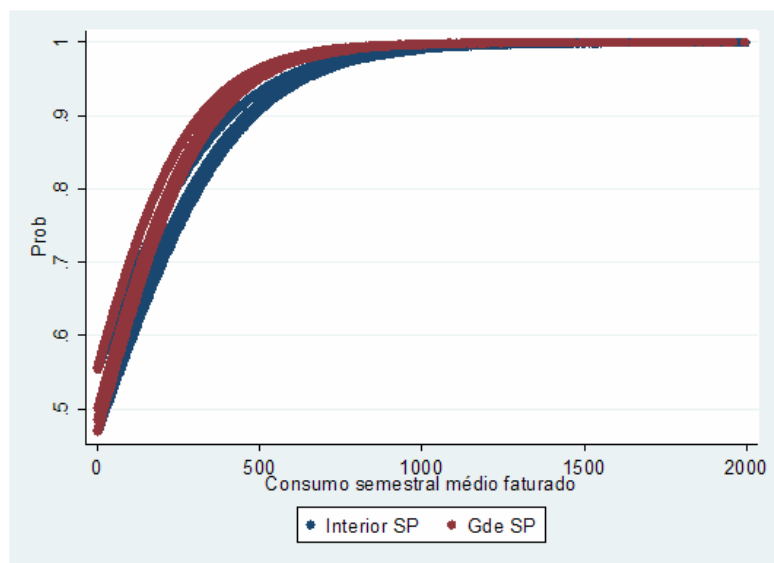


Gráfico 3.11: probabilidade de evasão por nível de consumo (critério 1)



3.2.1.2 - Ex-cliente tipo 2

Critério de definição de amostra e de ex-clientes e da amostra:

- *Amostra 1*: RGIs cujo consumo faturado médio do 1º semestre de 2004 foi acima de 50m³;

- *Ex-cliente tipo 2*: RGIs que apresentaram redução maior que 75% entre os semestres pares/ímpares.

Gráfico 3.12: Efeito marginal do consumo (critério 2)

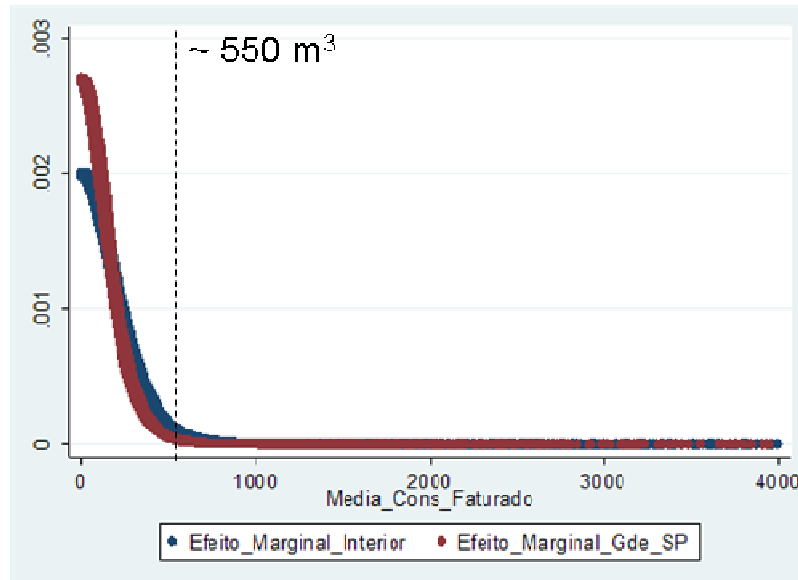
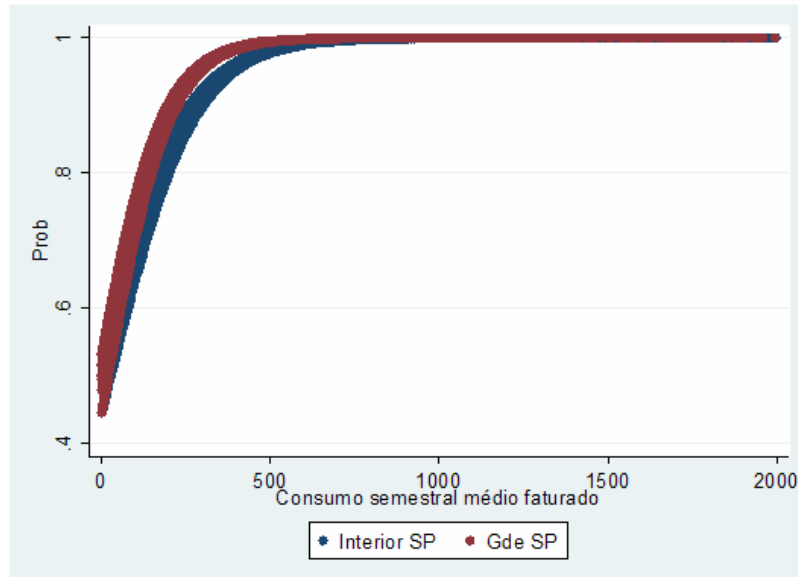


Gráfico 3.13: Probabilidade de evasão por nível de consumo (critério 2)



3.2.1.3 - Ex-cliente tipo 3

Critério de definição de amostra e de ex-clientes e da amostra:

- *Amostra 1*: RGIs cujo consumo faturado médio do 1º semestre de 2004 foi acima de 50m³;

- *Ex-cliente tipo 3*: RGIs que apresentaram redução maior que 50% entre os semestres pares/ímpares.

Gráfico 3.14 Efeito marginal do consumo (critério 3)

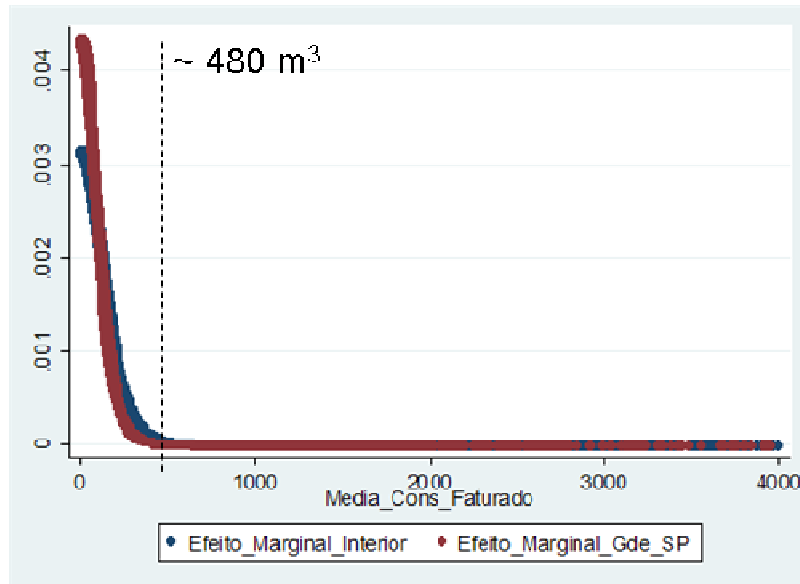
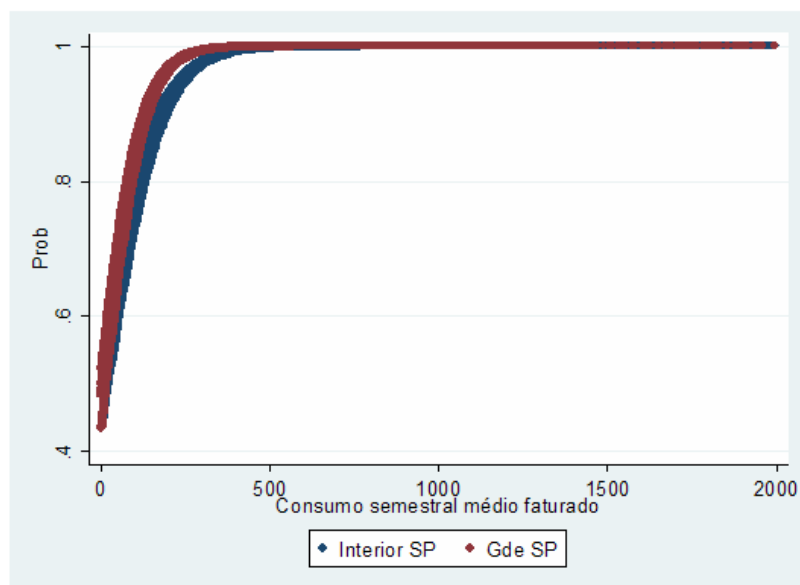


Gráfico 3.15: probabilidade de evasão por nível de consumo (critério 3)



3.2.1.4 - Interpretação dos resultados

Independente do critério utilizado para a definição da variável dependente, é possível observar que as curvas de efeito marginal são não-crescentes para todas as faixas de consumo estimadas. A contrapartida direta desse fato é que a probabilidade de evasão

apresenta comportamento crescente com respeito ao nível de consumo faturado - corroborando a hipótese de que grandes clientes apresentam maior probabilidade de buscar fontes de água alternativas.

No entanto, o nível de consumo médio semestral a partir do qual a probabilidade estimada de evasão deixa de aumentar, ou seja, torna-se estável, se mostrou bastante sensível ao critério utilizado para definição da variável dependente. Pode-se notar que o efeito marginal converge para valores muito próximos de zero a partir de: **900m³** de consumo médio semestral faturado, pelo critério 1; **500m³**, pelo critério 2; e **480m³**, pelo critério 3.

Deve-se ressaltar também que o efeito de localidade não se mostrou muito relevante, pois, independente do critério utilizado, as curvas de probabilidade de evasão estimadas para clientes da grande São Paulo estiveram sempre muito próximas daquelas estimadas para clientes do interior do estado.

4. Considerações Finais

O objetivo deste estudo foi estimar o limiar do fornecimento de água, dada a atual estrutura tarifária. A hipótese que se pretendia testar era a de que a atual estrutura tarifária da SABESP leva grandes clientes a buscar fontes alternativas de fornecimento de água. Os resultados dos modelos estimados, probit e logit, corroboraram a hipótese inicial. As principais conclusões que podem ser tiradas deste trabalho são:

- Há evidências a favor da crença de que empresas de distintos setores de atividade possuem diferentes tendências de sair ou não do sistema de abastecimento da SABESP;
- Há poucas evidências de grandes diferenças entre as sensibilidades de grandes clientes da Grande São Paulo e do Interior de São Paulo;
- A probabilidade de evasão dos clientes SABESP aumenta conforme cresce o nível de consumo médio dos mesmos;

- Os níveis de consumo que estabilizam a probabilidade de evasão, ou seja, níveis em que a probabilidade chega próxima ao seu máximo, variaram entre 500m³ e 1200m³, indo ao encontro de evidências encontradas pela SABESP.

Com base neste estudo, pode-se concluir que o limite mínimo imposto pelo programa *Sabesp Soluções Ambientais*, que garante contratos com estrutura tarifária diferenciada para clientes cujo consumo seja igual ou superior a 3.000m³/mês, deveria ser revisto, uma vez que clientes de grande porte com consumos médios abaixo deste patamar estão saindo do sistema de abastecimento da SABESP em busca de reduções de custos no mercado concorrencial. A revisão do limite mínimo deste programa acirraria a concorrência entre os fornecedores de água, trazendo benefícios aos utilitários deste insumo.

BIBLIOGRAFIA

Greene, W. H. (2003), *Econometric Analysis*. New Jersey: Prentice Hall, 5th ed.

Superintendência de Marketing da SABESP, “Perda de grandes clientes”: agosto de 2008. Estudo disponibilizado pela SABESP.

Wooldridge, J. M. (2001), *Econometrics Analysis of Cross Section and Panel Data*. The Massachusetts Institute of Technology Press.

Wooldridge, J. M. (2006), *Introdução à Econometria: Uma abordagem moderna*. Thomson Learning, São Paulo, 2006.

Anexo A: Resultados das estimações em STATA – Modelo Probit

Tabela A.1 - Base de Dados 50m³ e Critério 1

Regressão Probit	Numero de obs	=	21572
	LR chi2(21)	=	190.42
	Prob > chi2	=	0.0000
Log-verossimil.= -7304.411	Pseudo R2	=	0.0129

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf]	
media0104	.0005762	.0001064	5.41	0.000	.0003676	.0007848
mediasqr_1	-6.93e-08	3.23e-08	-2.14	0.032	-1.33e-07	-5.94e-09
ativ_aoi~c	-.5255802	.0795765	-6.60	0.000	-.6815472	-.3696132
comercio_i~c	-.4051634	.0800826	-5.06	0.000	-.5621224	-.2482045
constr_civ~c	.1675649	.1360726	1.23	0.218	-.0991325	.4342623
educ_saude~c	-.7069916	.0883922	-8.00	0.000	-.880237	-.5337461
eletric_sa~c	.198283	.3305241	0.60	0.549	-.4495324	.8460984
extrat_inic	.196175	.3621641	0.54	0.588	-.5136536	.9060037
fabr_maq_i~c	-.4988205	.117737	-4.24	0.000	-.7295809	-.2680602
financ_inic	-.7328927	.1173767	-6.24	0.000	-.9629469	-.5028386
imobil_inic	-.6804313	.1293418	-5.26	0.000	-.9339366	-.4269261
ind_final~c	-.5192976	.0918861	-5.65	0.000	-.6993911	-.3392041
metalurg_i~c	-.5763596	.1054003	-5.47	0.000	-.7829404	-.3697788
outros_inic	-.5438058	.0822251	-6.61	0.000	-.7049639	-.3826476
p_e_d_inic	-.1287348	.6488551	-0.20	0.843	-1.400468	1.142998
petroq_inic	-.5198065	.1102122	-4.72	0.000	-.7358183	-.3037946
resid_inic	-.3606135	.0818007	-4.41	0.000	-.52094	-.200287
serv_graf~c	-.3251819	.1510933	-2.15	0.031	-.6213194	-.0290444
servicos_i~c	-.4999653	.0845788	-5.91	0.000	-.6657368	-.3341939
telecom_inic	-.8043303	.2973539	-2.70	0.007	-1.387133	-.2215274
transp_inic	-.3712755	.097474	-3.81	0.000	-.562321	-.1802299
_cons	-.8263276	.0760948	-10.86	0.000	-.9754708	-.6771845

Tabela A.2 - Base de Dados 50 m³ e Critério 2

Regressão Probit	Numero de obs	=	21572
	LR chi2(21)	=	213.82
	Prob > chi2	=	0.0000
Log-verossimil.= -9639.373	Pseudo R2	=	0.0110

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf]	
media0104	.0003916	.0000977	4.01	0.000	.0002001	.0005831
mediasqr_1	-2.67e-08	2.90e-08	-0.92	0.356	-8.36e-08	3.01e-08
ativ_aoi~c	-.5647649	.0738171	-7.65	0.000	-.7094437	-.4200861
comercio_i~c	-.4238418	.0742801	-5.71	0.000	-.569428	-.2782556
constr_civ~c	.1254069	.1292168	0.97	0.332	-.1278533	.3786671
educ_saude~c	-.7657133	.0811583	-9.43	0.000	-.9247806	-.606646
eletric_sa~c	-.1019267	.3290781	-0.31	0.757	-.746908	.5430545
extrat_inic	-.110551	.3614023	-0.31	0.760	-.8188864	.5977844
fabr_maq_i~c	-.5564837	.1073711	-5.18	0.000	-.7669272	-.3460401
financ_inic	-.7496792	.1044352	-7.18	0.000	-.9543685	-.5449899
imobil_inic	-.701305	.1152555	-6.08	0.000	-.9272016	-.4754085
ind_final~c	-.5692772	.0844362	-6.74	0.000	-.7347692	-.4037853
metalurg_i~c	-.5371155	.0943642	-5.69	0.000	-.722066	-.352165
outros_inic	-.5733337	.0759701	-7.55	0.000	-.7222323	-.424435
p_e_d_inic	-.4202834	.6465978	-0.65	0.516	-1.687592	.847025
petroq_inic	-.4659745	.09857	-4.73	0.000	-.6591682	-.2727808
resid_inic	-.4236974	.0759837	-5.58	0.000	-.5726227	-.2747721
serv_graf~c	-.3051674	.1363979	-2.24	0.025	-.5725024	-.0378323
servicos_i~c	-.5320095	.0780714	-6.81	0.000	-.6850265	-.3789924
telecom_inic	-.614588	.2358344	-2.61	0.009	-1.076815	-.1523611
transp_inic	-.3650056	.0892937	-4.09	0.000	-.5400181	-.1899932
_cons	-.4993641	.0709033	-7.04	0.000	-.6383319	-.3603962

Tabela A.3 - Base de Dados 50 m³ e Critério 3

Regressão Probit	Numero de obs	=	21572
	LR chi2(21)	=	192.75
	Prob > chi2	=	0.0000
Log-verossimil.= -12709.743	Pseudo R2	=	0.0075

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf]	
media0104	.000065	.0001013	0.64	0.521	-.0001336	.0002636
mediasqr_1	6.00e-08	4.11e-08	1.46	0.144	-2.05e-08	1.40e-07
ativ_aoi~c	-.5823844	.0706091	-8.25	0.000	-.7207758	-.443993
comercio_i~c	-.4359039	.0710945	-6.13	0.000	-.5752465	-.2965613
constr_civ~c	-.0298182	.1262022	-0.24	0.813	-.2771699	.2175335
educ_saude~c	-.7144289	.0759589	-9.41	0.000	-.8633055	-.5655522
eletric_sa~c	-.3099597	.3195403	-0.97	0.332	-.9362472	.3163279
extrat_inic	.251077	.3435106	0.73	0.465	-.4221915	.9243455
fabr_maq_i~c	-.4678201	.0979487	-4.78	0.000	-.6597961	-.2758441
financ_inic	-.5357792	.092502	-5.79	0.000	-.7170798	-.3544785
imobil_inic	-.5951221	.1029269	-5.78	0.000	-.7968551	-.3933891
ind_final_~c	-.5668429	.0793451	-7.14	0.000	-.7223565	-.4113293
metalurg_i~c	-.5598874	.0880655	-6.36	0.000	-.7324926	-.3872823
outros_inic	-.5733895	.0723097	-7.93	0.000	-.7151139	-.4316651
p_e_d_inic	-.187515	.5715515	-0.33	0.743	-1.307735	.9327054
petroq_inic	-.4647968	.0919894	-5.05	0.000	-.6450928	-.2845008
resid_inic	-.4232469	.072558	-5.83	0.000	-.565458	-.2810358
serv_graf_~c	-.3095289	.1274531	-2.43	0.015	-.5593325	-.0597254
servicos_i~c	-.5273829	.0741358	-7.11	0.000	-.6726865	-.3820794
telecom_inic	-.3457469	.2004957	-1.72	0.085	-.7387113	.0472175
transp_inic	-.3473916	.0843303	-4.12	0.000	-.512676	-.1821072
_cons	-.0804923	.0683234	-1.18	0.239	-.2144037	.0534192

Tabela A.4 - Base de Dados 300 m³ e Critério 1

Regressão Probit	Numero de obs	=	1078
	LR chi2(19)	=	57.09
	Prob > chi2	=	0.0000
Log-verossimil.= -409.96144	Pseudo R2	=	0.0651

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf]	
media0104	.0006251	.0002229	2.80	0.005	.0001882	.001062
mediasqr_1	-8.58e-08	5.34e-08	-1.61	0.108	-1.90e-07	1.88e-08
ativ_aoi~c	-1.603147	.3726822	-4.30	0.000	-2.33359	-.872703
comercio_i~c	-1.297084	.3937742	-3.29	0.001	-2.068868	-.525301
constr_civ~c	-.8153843	.5517685	-1.48	0.139	-1.896831	.2660621
educ_saude~c	-1.566228	.3951621	-3.96	0.000	-2.340731	-.7917242
eletric_sa~c	-.6867848	.8395786	-0.82	0.413	-2.332329	.958759
fabr_maq_i~c	-.8889714	.4622183	-1.92	0.054	-1.794903	.0169598
financ_inic	-2.10899	.5626172	-3.75	0.000	-3.2117	-1.006281
imobil_inic	-1.654486	.4725207	-3.50	0.000	-2.580609	-.728362
ind_final_~c	-1.278371	.4417637	-2.89	0.004	-2.144212	-.4125305
metalurg_i~c	-1.388459	.4993448	-2.78	0.005	-2.367157	-.4097611
outros_inic	-1.560415	.4207823	-3.71	0.000	-2.385133	-.7356966
petroq_inic	-.8169233	.4317471	-1.89	0.058	-1.663132	.0292854
resid_inic	-1.280235	.3787418	-3.38	0.001	-2.022555	-.5379144
serv_graf_~c	-1.3173	.7310613	-1.80	0.072	-2.750154	.1155534
servicos_i~c	-1.373097	.3949568	-3.48	0.001	-2.147198	-.5989959
telecom_inic	-1.373005	.6893584	-1.99	0.046	-2.724123	-.0218877
transp_inic	-1.186242	.4153662	-2.86	0.004	-2.000344	-.372139
_cons	-.0049324	.3670656	-0.01	0.989	-.7243677	.7145029

Tabela A.5 - Base de Dados 300 m³ e Critério 2

Regressão Probit	Numero de obs	=	1078
	LR chi2(19)	=	71.30
	Prob > chi2	=	0.0000
Log-verossimil.= -493.07895	Pseudo R2	=	0.0674

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf]	
media0104	.0007276	.0002117	3.44	0.001	.0003127	.0011424
mediasqr_l	-1.05e-07	5.07e-08	-2.07	0.039	-2.04e-07	-5.45e-09
ativ_aoi~c	-1.605386	.3821092	-4.20	0.000	-2.354306	-.8564657
comercio_i~c	-1.156369	.3986198	-2.90	0.004	-1.937649	-.3750882
constr_civ~c	-.7682672	.5491085	-1.40	0.162	-1.8445	.3079656
educ_saude~c	-1.447723	.3983775	-3.63	0.000	-2.228529	-.6669177
eletric_sa~c	-.9063159	.846418	-1.07	0.284	-2.565265	.7526329
fabr_maq_i~c	-.9803486	.4674643	-2.10	0.036	-1.896562	-.0641354
financ_inic	-2.325265	.5703667	-4.08	0.000	-3.443163	-1.207367
imobil_inic	-1.884685	.4821543	-3.91	0.000	-2.82969	-.9396802
ind_final_~c	-1.028792	.4356569	-2.36	0.018	-1.882664	-.1749203
metalurg_i~c	-1.424459	.4936768	-2.89	0.004	-2.392048	-.4568708
outros_inic	-1.414932	.416574	-3.40	0.001	-2.231402	-.5984616
petroq_inic	-.9549747	.4402582	-2.17	0.030	-1.817865	-.0920844
resid_inic	-1.482363	.3901486	-3.80	0.000	-2.24704	-.7176858
serv_graf_~c	-1.551116	.7410329	-2.09	0.036	-3.003514	-.0987182
servicos_i~c	-1.382707	.4014223	-3.44	0.001	-2.169481	-.5959341
telecom_inic	-.7065238	.6038034	-1.17	0.242	-1.889957	.4769091
transp_inic	-1.196925	.4200136	-2.85	0.004	-2.020137	-.3737138
_cons	.1728615	.3781557	0.46	0.648	-.5683101	.914033

Tabela A.6 - Base de Dados 300 m³ e Critério 3

Regressão Probit	Numero de obs	=	1078
	LR chi2(19)	=	67.03
	Prob > chi2	=	0.0000
Log-verossimil.= -619.48211	Pseudo R2	=	0.0513

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf]	
media0104	.0003186	.0002021	1.58	0.115	-.0000775	.0007147
mediasqr_l	-1.33e-08	4.89e-08	-0.27	0.786	-1.09e-07	8.26e-08
ativ_aoi~c	-1.762257	.4363593	-4.04	0.000	-2.617505	-.9070084
comercio_i~c	-1.269245	.4498735	-2.82	0.005	-2.15098	-.3875087
constr_civ~c	-.9966777	.584654	-1.70	0.088	-2.142578	.1492231
educ_saude~c	-1.556274	.4489858	-3.47	0.001	-2.43627	-.6762782
eletric_sa~c	-.5395549	.8647234	-0.62	0.533	-2.234382	1.155272
fabr_maq_i~c	-1.206962	.5095669	-2.37	0.018	-2.205695	-.2082293
financ_inic	-1.665063	.4908777	-3.39	0.001	-2.627166	-.7029605
imobil_inic	-2.046035	.5049942	-4.05	0.000	-3.035805	-1.056264
ind_final_~c	-1.103996	.4821142	-2.29	0.022	-2.048922	-.1590696
metalurg_i~c	-1.897585	.5357806	-3.54	0.000	-2.947696	-.8474742
outros_inic	-1.481301	.4608398	-3.21	0.001	-2.38453	-.5780714
petroq_inic	-1.164247	.4864509	-2.39	0.017	-2.117673	-.2108206
resid_inic	-1.754885	.4433672	-3.96	0.000	-2.623869	-.8859015
serv_graf_~c	-1.441132	.685859	-2.10	0.036	-2.785391	-.0968736
servicos_i~c	-1.510291	.4510484	-3.35	0.001	-2.394329	-.6262519
telecom_inic	-.8243479	.6411988	-1.29	0.199	-2.081074	.4323786
transp_inic	-1.343291	.4661148	-2.88	0.004	-2.25686	-.4297232
_cons	.8392576	.4355071	1.93	0.054	-.0143207	1.692836

Tabela A.7 - Base de Dados 1000 m³ e Critério 1

Regressão Probit	Numero de obs	=	64
	LR chi2(12)	=	18.08
	Prob > chi2	=	0.1133
Log-verossimil.= -30.709161	Pseudo R2	=	0.2274

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf]	
media0104	.000501	.0009983	0.50	0.616	-.0014557	.0024576
mediasqr_1	-9.32e-08	1.87e-07	-0.50	0.618	-4.59e-07	2.73e-07
ativ_aoi~c	-7.458265	1.091233	-6.83	0.000	-9.597043	-5.319487
comercio_i~c	-6.583606	1.409499	-4.67	0.000	-9.346174	-3.821039
educ_saude~c	-7.805393	1.210442	-6.45	0.000	-10.17781	-5.432971
fabr_maq_i~c	-6.555233	1.463591	-4.48	0.000	-9.423818	-3.686647
ind_final~c	-6.155488	1.437697	-4.28	0.000	-8.973322	-3.337654
metalurg_i~c	-6.548695	1.437161	-4.56	0.000	-9.36548	-3.731911
petroq_inic	-6.549494	1.324953	-4.94	0.000	-9.146354	-3.952634
resid_inic	-7.561656	1.19541	-6.33	0.000	-9.904617	-5.218696
servicos_i~c	-7.360286	1.232704	-5.97	0.000	-9.776342	-4.944231
transp_inic	-6.090127	1.29594	-4.70	0.000	-8.630122	-3.550132
_cons	5.996619

Tabela A.8 - Base de Dados 1000 m³ e Critério 2

Regressão Probit	Numero de obs	=	67
	LR chi2(13)	=	17.21
	Prob > chi2	=	0.1897
Log-verossimil.= -35.102328	Pseudo R2	=	0.1969

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf]	
media0104	.0004912	.0006993	0.70	0.482	-.0008794	.0018618
mediasqr_1	-6.65e-08	1.07e-07	-0.62	0.535	-2.76e-07	1.44e-07
ativ_aoi~c	-7.138563	.8627378	-8.27	0.000	-8.829498	-5.447628
comercio_i~c	-5.730572	1.243904	-4.61	0.000	-8.168581	-3.292564
educ_saude~c	-6.874385	.9666532	-7.11	0.000	-8.76899	-4.979779
fabr_maq_i~c	-6.323092	1.27651	-4.95	0.000	-8.825005	-3.821178
ind_final~c	-6.046703	1.27084	-4.76	0.000	-8.537505	-3.555902
metalurg_i~c	-6.246879	1.237114	-5.05	0.000	-8.671579	-3.82218
outros_inic	-6.617663	1.072044	-6.17	0.000	-8.718832	-4.516495
petroq_inic	-6.315274	1.13193	-5.58	0.000	-8.533816	-4.096733
resid_inic	-7.265882	.9541855	-7.61	0.000	-9.136051	-5.395712
servicos_i~c	-7.046421	1.011529	-6.97	0.000	-9.028982	-5.06386
transp_inic	-5.774595	1.085248	-5.32	0.000	-7.901643	-3.647548
_cons	5.639093

Tabela A.9 - Base de Dados 1000m³ e Critério 3

Regressão Probit		Numero de obs = 67				
Log-verossimil.= -36.181175		LR chi2(13) = 17.98				
		Prob > chi2 = 0.1582				
		Pseudo R2 = 0.1990				
cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf]	
media0104	-.0002	.0008173	-0.24	0.807	-.001802	.0014019
mediasqr_1	8.05e-08	1.41e-07	0.57	0.569	-1.97e-07	3.58e-07
ativ_aoi~c	-7.062332	.9432441	-7.49	0.000	-8.911057	-5.213608
comercio_i~c	-5.628319	1.306677	-4.31	0.000	-8.189359	-3.067279
educ_saude~c	-6.473285	1.031197	-6.28	0.000	-8.494394	-4.452177
fabr_maq_i~c	-6.224797	1.343442	-4.63	0.000	-8.857896	-3.591699
ind_final~c	-5.988145	1.326915	-4.51	0.000	-8.588851	-3.38744
metalurg_i~c	-6.106186	1.302092	-4.69	0.000	-8.65824	-3.554133
outros_inic	-6.519514	1.146098	-5.69	0.000	-8.765824	-4.273204
petroq_inic	-6.219595	1.196355	-5.20	0.000	-8.564407	-3.874784
resid_inic	-7.135726	1.032504	-6.91	0.000	-9.159397	-5.112054
servicos_i~c	-6.35058	1.077408	-5.89	0.000	-8.46226	-4.2389
transp_inic	-5.661933	1.16601	-4.86	0.000	-7.947271	-3.376596
_cons	6.211214

Anexo B: Resultados das estimações em STATA – Modelo Logit

Tabela B.1 - Base de Dados 50m³ e Critério 1

Regressão Logística Condicional efeitos-fixos		Numero de obs = 35289				
Variável Grupo: rgi		Numero de grupos = 5043				
		Obs por grupo: min = 3				
		avg = 7.0				
		max = 7				
Log-verossim. = -12747.088		LR chi2(10) = 1337.08				
		Prob > chi2 = 0.0000				
cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf.]	
media_	.0048436	.0003975	12.19	0.000	.0040645	.0056226
mediasqr_	-7.39e-08	6.25e-09	-11.81	0.000	-8.61e-08	-6.16e-08
media_sp_	.0015162	.000449	3.38	0.001	.0006362	.0023962
mediasqr_sp_	-3.17e-07	1.93e-08	-16.43	0.000	-3.55e-07	-2.79e-07
D1	-.0139732	.049249	-0.28	0.777	-.1104994	.0825531
D2	-.0601879	.0480965	-1.25	0.211	-.1544553	.0340796
D3	-.1195541	.0481359	-2.48	0.013	-.2138987	-.0252095
D4	.2191234	.0457706	4.79	0.000	.1294146	.3088321
D5	.2279944	.046042	4.95	0.000	.1377538	.318235
D7	-.1226152	.049387	-2.48	0.013	-.219412	-.0258184

Tabela B.2 - Base de Dados 50 m³ e Critério 2

Regressão Logística Condicional efeitos-fixos		Numero de obs = 50485				
Variável Grupo: rgi		Numero de grupos = 7215				
		Obs por grupo: min = 3				
		avg = 7.0				
		max = 7				
		LR chi2(10) = 3334.98				
Log-verossim. = -17898.338		Prob > chi2 = 0.0000				
cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf.]	
media_	.0080342	.0003996	20.11	0.000	.007251	.0088174
mediasqr_	-1.23e-07	6.25e-09	-19.66	0.000	-1.35e-07	-1.11e-07
media_sp_	.0028111	.0004567	6.15	0.000	.0019159	.0037063
mediasqr_sp_	-5.52e-07	2.00e-08	-27.55	0.000	-5.91e-07	-5.13e-07
D1	-.1270276	.0413581	-3.07	0.002	-.2080879	-.0459673
D2	-.2193538	.0406636	-5.39	0.000	-.2990529	-.1396547
D3	-.0247804	.0393941	-0.63	0.529	-.1019914	.0524306
D4	.0598966	.0388158	1.54	0.123	-.016181	.1359741
D5	.1259359	.0388233	3.24	0.001	.0498437	.2020282
D7	-.0867898	.0407307	-2.13	0.033	-.1666204	-.0069591

Tabela B.3 - Base de Dados 50 m³ e Critério 3

Regressão Logística Condicional efeitos-fixos		Numero de obs = 83134				
Variável Grupo: rgi		Numero de grupos = 11881				
		Obs por grupo: min = 3				
		avg = 7.0				
		max = 7				
		LR chi2(10) = 8134.62				
Log-verossim. = -28889.482		Prob > chi2 = 0.0000				
cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf.]	
media_	.012579	.000381	33.01	0.000	.0118322	.0133259
mediasqr_	-1.93e-07	5.92e-09	-32.58	0.000	-2.05e-07	-1.81e-07
media_sp_	.0047973	.0004543	10.56	0.000	.003907	.0056877
mediasqr_sp_	-9.03e-07	2.05e-08	-43.97	0.000	-9.43e-07	-8.63e-07
D1	-.2447307	.0321149	-7.62	0.000	-.3076747	-.1817868
D2	-.2614784	.0314401	-8.32	0.000	-.3230998	-.199857
D3	-.0696208	.0305337	-2.28	0.023	-.1294658	-.0097758
D4	-.0502843	.0303422	-1.66	0.097	-.109754	.0091853
D5	.0919294	.0300434	3.06	0.002	.0330454	.1508134
D7	-.0327269	.0309711	-1.06	0.291	-.0934292	.0279754

Tabela B.4 - Base de Dados 300 m³ e Critério 1

Regressão Logística Condicional efeitos-fixos		Numero de obs	=	2644		
Variável Grupo: rgi		Numero de grupos	=	379		
		Obs por grupo: min	=	3		
		avg	=	7.0		
		max	=	7		
		LR chi2(10)	=	339.95		
Log-verossim. = -842.54997		Prob > chi2	=	0.0000		
cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf.]	
media_	.001476	.0003001	4.92	0.000	.0008877	.0020642
mediasqr_	-2.22e-08	4.99e-09	-4.44	0.000	-3.20e-08	-1.24e-08
media_sp_	.0030531	.0004773	6.40	0.000	.0021176	.0039885
mediasqr_sp_	-8.87e-07	1.62e-07	-5.46	0.000	-1.20e-06	-5.68e-07
D1	-.4537494	.2360337	-1.92	0.055	-.9163668	.0088681
D2	-.5522581	.2290772	-2.41	0.016	-1.001241	-.1032751
D3	-.251396	.2183385	-1.15	0.250	-.6793316	.1765395
D4	1.166553	.1919158	6.08	0.000	.7904053	1.542701
D5	1.194282	.1892777	6.31	0.000	.8233041	1.565259
D6	.1578653	.2168033	0.73	0.467	-.2670613	.5827919

Tabela B.5 - Base de Dados 300 m³ e Critério 2

Regressão Logística Condicional efeitos-fixos		Numero de obs	=	3455		
Variável Grupo: rgi		Numero de grupos	=	496		
		Obs por grupo: min	=	3		
		avg	=	7.0		
		max	=	7		
		LR chi2(10)	=	498.85		
Log-verossim. = -1164.0304		Prob > chi2	=	0.0000		
cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf.]	
media_	.0019024	.0002776	6.85	0.000	.0013584	.0024465
mediasqr_	-2.88e-08	4.57e-09	-6.31	0.000	-3.78e-08	-1.99e-08
media_sp_	.0037104	.0004313	8.60	0.000	.0028651	.0045558
mediasqr_sp_	-9.25e-07	1.03e-07	-9.02	0.000	-1.13e-06	-7.24e-07
D1	-.8035462	.1947543	-4.13	0.000	-1.185258	-.4218348
D2	-.6653131	.1844884	-3.61	0.000	-1.026904	-.3037224
D3	.3264545	.1660386	1.97	0.049	.0010248	.6518842
D4	.7487475	.1577534	4.75	0.000	.4395565	1.057938
D5	.8435921	.1556869	5.42	0.000	.5384513	1.148733
D6	-.0096443	.1759279	-0.05	0.956	-.3544566	.3351681

Tabela B.6 - Base de Dados 300 m³ e Critério 3

Regressão Logística Condicional efeitos-fixos		Numero de obs = 5150				
Variável Grupo: rgi		Numero de grupos = 740				
		Obs por grupo: min = 3				
		avg = 7.0				
		max = 7				
Log-verossim. = -1719.9758		LR chi2(10) = 834.57				
		Prob > chi2 = 0.0000				
cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf.]	
media_	.0027225	.0002647	10.29	0.000	.0022038	.0032413
mediasqr_	-4.17e-08	4.19e-09	-9.95	0.000	-4.99e-08	-3.35e-08
media_sp_	.0045604	.0004209	10.84	0.000	.0037355	.0053854
mediasqr_sp_	-1.31e-06	1.05e-07	-12.52	0.000	-1.52e-06	-1.11e-06
D1	-1.009033	.1547519	-6.52	0.000	-1.312341	-.705725
D2	-.5474943	.1425914	-3.84	0.000	-.8269684	-.2680202
D3	.3737985	.1307116	2.86	0.004	.1176085	.6299884
D4	.5166303	.1266895	4.08	0.000	.2683235	.7649371
D5	.7282818	.1251036	5.82	0.000	.4830833	.9734804
D6	.1641564	.1348874	1.22	0.224	-.1002179	.4285308

Tabela B.7 - Base de Dados 1000 m³ e Critério 1

Regressão Logística Condicional efeitos-fixos		Numero de obs = 320				
Variável Grupo: rgi		Numero de grupos = 47				
		Obs por grupo: min = 3				
		avg = 6.8				
		max = 7				
Log-verossim. = -95.256508		LR chi2(10) = 51.31				
		Prob > chi2 = 0.0000				
cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf.]	
media_	.0008096	.000264	3.07	0.002	.0002922	.0013271
mediasqr_	-1.14e-08	4.87e-09	-2.35	0.019	-2.10e-08	-1.88e-09
media_sp_	.0022934	.0006723	3.41	0.001	.0009757	.0036112
mediasqr_sp_	-4.71e-07	1.60e-07	-2.94	0.003	-7.84e-07	-1.57e-07
D2	.3102323	.5349931	0.58	0.562	-.7383349	1.3588
D3	.9425779	.5851656	1.61	0.107	-.2043255	2.089481
D4	1.213108	.6146515	1.97	0.048	.0084137	2.417803
D5	1.770749	.6511332	2.72	0.007	.4945513	3.046947
D6	.8282266	.7470734	1.11	0.268	-.6360103	2.292464
D7	1.649146	.7025122	2.35	0.019	.2722476	3.026045

Tabela B.8 - Base de Dados 1000 m³ e Critério 2

Regressão Logística Condicional efeitos-fixos		Numero de obs	=	410
Variável Grupo: rgi		Numero de grupos	=	61
		Obs por grupo: min	=	3
		avg	=	6.7
		max	=	7
		LR chi2(10)	=	79.98
Log-verossim.	= -123.82305	Prob > chi2	=	0.0000

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf.]	
media_	.0009623	.0002496	3.86	0.000	.0004731	.0014515
mediasqr_	-1.38e-08	4.58e-09	-3.02	0.003	-2.28e-08	-4.84e-09
media_sp_	.0023548	.0006428	3.66	0.000	.001095	.0036146
mediasqr_sp_	-3.94e-07	1.58e-07	-2.49	0.013	-7.05e-07	-8.36e-08
D2	.342255	.5087617	0.67	0.501	-.6548997	1.33941
D3	1.753714	.5419354	3.24	0.001	.6915398	2.815888
D4	1.313905	.5705178	2.30	0.021	.1957106	2.432099
D5	1.860438	.5954193	3.12	0.002	.6934376	3.027438
D6	1.113432	.6513705	1.71	0.087	-.1632305	2.390095
D7	1.84469	.6291652	2.93	0.003	.6115489	3.077831

Tabela B.9 - Base de Dados 1000m³ e Critério 3

Regressão Logística Condicional efeitos-fixos		Numero de obs	=	642
Variável Grupo: rgi		Numero de grupos	=	96
		Obs por grupo: min	=	3
		avg	=	6.7
		max	=	7
		LR chi2(10)	=	91.30
Log-verossim.	= -209.35787	Prob > chi2	=	0.0000

cons_exclientes	Coef.	Erro Pad.	z	P> z	[95% Intervalo Conf.]	
media_	.0010596	.0002073	5.11	0.000	.0006533	.001466
mediasqr_	-1.58e-08	3.40e-09	-4.64	0.000	-2.25e-08	-9.12e-09
media_sp_	.0019387	.0005691	3.41	0.001	.0008233	.0030541
mediasqr_sp_	-4.38e-07	1.25e-07	-3.50	0.000	-6.84e-07	-1.93e-07
D2	.4221398	.397366	1.06	0.288	-.3566833	1.200963
D3	1.294893	.3948387	3.28	0.001	.5210237	2.068763
D4	1.246982	.4087374	3.05	0.002	.4458709	2.048092
D5	1.24639	.4252338	2.93	0.003	.4129473	2.079833
D6	.9655774	.4545812	2.12	0.034	.0746146	1.85654
D7	1.200605	.4532036	2.65	0.008	.3123424	2.088868