

Estimação da proporção de usuários residenciais da COPANOR
potencialmente beneficiários da tarifa social

Bruno Vieira Andrade

Bruno Vieira Andrade
Matrícula: 2013673595

Estimação da proporção de usuários residenciais da COPANOR potencialmente beneficiários da tarifa social

Trabalho de conclusão de curso submetido
ao Departamento de Estatística da
Universidade Federal de Minas Gerais, sob
orientação da Professora Ilka Afonso Reis.

Belo Horizonte
23 de dezembro de 2015

AGRADECIMENTOS

Agradeço a todos os que contribuíram no decorrer deste projeto:

A Deus, a quem devo minha vida.

Em especial à orientadora Professora Ilka Afonso Reis, pela paciência, dedicação e seu fundamental papel na elaboração deste trabalho.

Aos meus colegas, pelo companheirismo e disponibilidade para me auxiliar em vários momentos.

RESUMO

A cobrança pelos serviços de saneamento é feita através de tarifas que são definidas, conforme determina a Lei Federal 11.445/2007, observando-se a capacidade de pagamento do cidadão e a utilização do serviço. A COPASA Serviços de Saneamento Integrado do Norte e Nordeste de Minas Gerais S/A (COPANOR), que atua em uma região carente do Estado de Minas Gerais, não possui tarifas que diferenciem, entre seus clientes residenciais, aqueles com capacidade de arcar com tarifas normais e aqueles que deveriam receber algum tipo de subsídio. Dessa forma, considerando que qualquer cidadão inscrito no Cadastro Único para Programas Sociais – CadÚnico faz jus a uma tarifa subsidiada, procurou-se estimar a proporção de usuários residenciais da empresa com potencial para serem beneficiados com uma eventual tarifa social, por meio de resultados amostrais e a posterior extrapolação para a população atendida pela COPANOR.

ABSTRACT

The charge for sanitation services is done through tariffs that are set, as required by Federal Law 11,445 / 2007, observing the payment capacity of citizens and intended use of the service. COPASA Northern Integrated Sanitation Services and Northeast of Minas Gerais S / A (COPANOR), which operates in a poor region of Minas Gerais, does not have tariffs that differentiate their residential customers among those with the capacity to pay for normal rates of those they should receive some kind of subsidy. Thus, considering that any citizen enrolled in the Single Registry for Social Programs - CadÚnico is entitled to a subsidized rate, we tried to estimate the proportion of residential users of the company with potential for benefit from any social tariff by means of sample results and the subsequent extrapolation to the population served by COPANOR.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1 – Mapa Área de abrangência de atuação da COPANOR..... | 9 |
| Figura 2 – Distribuição de frequências das unidades residenciais segundo a existência de CPF cadastrado na COPANOR..... | 22 |
| Figura 3 - Distribuição de frequências das unidades residenciais segundo a existência de CPF cadastrado na COPANOR e no CadÚnico..... | 23 |
| Figura 4 – Distribuição de frequência das 11.390 un. residenciais onde ocorreram entrevistas segundo a inclusão de membros familiares no CadÚnico..... | 24 |
| Figura 5 – Box Plot dos dados de consumo de água da população..... | 25 |
| Figura 6 – Box Plot dos dados de renda da população..... | 26 |
| Figura 7 – Gráfico de probabilidade normal e teste de Anderson-Darling para o consumo médio de água nas localidades da população amostrada..... | 32 |
| Figura 8 – Gráfico de probabilidade normal e teste de Anderson-Darling para o consumo médio de água nas localidades após transformação Box Cox..... | 33 |
| Figura 9 - Gráfico de probabilidade normal e teste de Anderson-Darling para a renda média nos municípios da população amostrada..... | 34 |

LISTAS DE TABELAS e QUADROS

| | |
|---|----|
| Tabela 1 – Informações operacionais da COPANOR..... | 10 |
| Tabela 2 - Distribuição de frequência das unidades residenciais consumidoras de energia elétrica no Brasil segundo a classificação da renda familiar (baixa ou não) em cada uma das cinco regiões geográficas do país em julho de 2015..... | 15 |
| Tabela 3 - Distribuição de frequência das un residenciais consumidoras de energia elétrica no Brasil segundo a classific. da renda familiar (baixa ou não) em cada um dos quatro estados da reg. sudeste do país em julho de 2015..... | 15 |
| Tabela 4 - Distribuição de frequência das unidades residenciais consumidoras de energia elétrica no Brasil segundo a classificação da renda familiar (baixa ou não) nas distribuidoras CEMIG e EMG em julho de 2015..... | 15 |
| Tabela 5 – Análise descritiva dos dados de consumo de água das localidades na COPANOR (N= 181)..... | 24 |
| Tabela 6 – Análise descritiva dos dados de renda dos Municípios na COPANOR (N=67)..... | 25 |
| Tabela 7 – Estimativa de erro de estimação do percentual de unidades que fazem jus à tarifa social na subpopulação sem CPF cadastrado na COPANOR..... | 27 |
| Tabela 8 – Caracterização da população atendida pela COPANOR e da amostra de residências visitadas segundo localidade, municípios e grupo de residências sem o CPF registrado no banco de dados da companhia..... | 29 |
| Tabela 9 – Estatísticas e parâmetros de ref. para os testes de hipóteses..... | 29 |
| | |
| Quadro 1 – Estatísticas utilizadas nos testes de hipóteses para as médias..... | 31 |
| Quadro 2 – Estatísticas e parâmetros de referência utilizados nos testes de hipóteses para a variância..... | 35 |

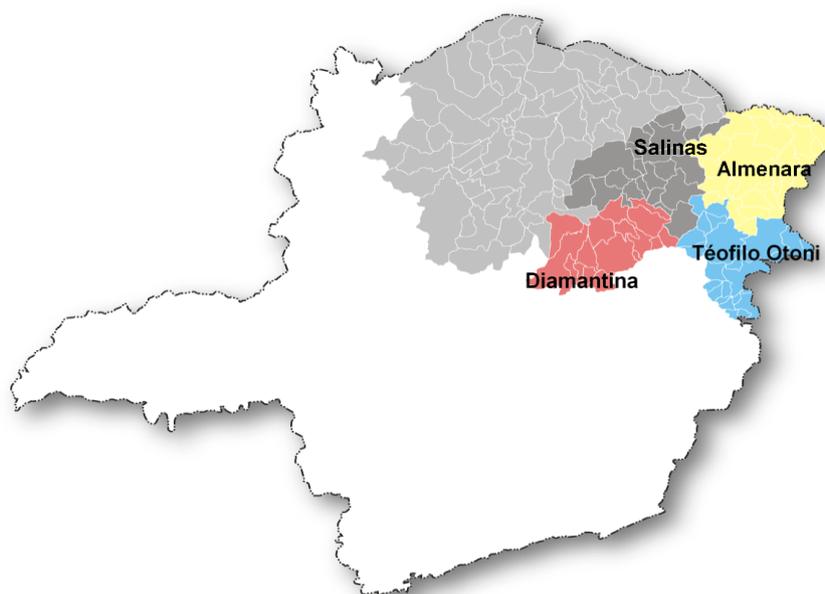
SUMÁRIO

| | |
|---|----|
| 1 – Introdução..... | 9 |
| 2 – A tarifa social no setor de energia elétrica..... | 14 |
| 3 – Materiais e Métodos..... | 16 |
| 3.1 – Os dados..... | 16 |
| 3.2 – Método..... | 17 |
| 3.2.1 – Avaliação da representatividade da amostra de unidades residenciais.. | 19 |
| 3.2.2 – O teste de hipóteses para uma média populacional..... | 20 |
| 3.2.3 – O teste de hipóteses para a variância populacional..... | 21 |
| 4 – Resultados..... | 22 |
| 4.1 – Análise descritiva dos dados..... | 22 |
| 4.2 – Estimativa do percentual de unidades que fazem jus à tarifa social..... | 27 |
| 4.3 – Validação da amostra..... | 28 |
| 4.3.1 – Teste de hipótese para a média..... | 29 |
| 4.3.2 – Teste de hipótese para a variância..... | 31 |
| 5 – Discussão..... | 36 |
| 6 – Conclusão..... | 37 |
| 7 – Bibliografia..... | 38 |

1. Introdução

A COPANOR (COPASA Serviços de Saneamento Integrado do Norte e Nordeste de Minas Gerais S/A) é uma empresa pública subsidiária da COPASA, criada pelo Governo de Minas, para atender as regiões Norte e Nordeste do Estado com os serviços de abastecimento de água, coleta e tratamento de esgotos sanitários e construção de módulos sanitários em todas as casas desprovidas dessas instalações, conforme a Lei Estadual 16.698/07. Na primeira fase, esse benefício chegará a mais de 450 localidades de 92 Municípios das Bacias Hidrográficas dos Rios Jequitinhonha, Mucuri, São Mateus, Buranhém, Itanhém e Jucuruçu, como mostra o mapa da Figura 1.

Figura 1 – Mapa Área de abrangência de atuação da COPANOR



Fonte: <http://www.copanor.com.br/wps/portal/copanor/a-copanor/area-de-atuacao>

A COPANOR tem por objetivo, portanto, planejar, projetar, executar, ampliar, remodelar, explorar e prestar serviços de abastecimento de água e de esgotamento sanitário no norte e nordeste do Estado de Minas. Apesar da Lei

Estadual 16.698 de 2007 autorizar atuação da empresa no norte de Minas Gerais, em junho de 2015, a COPANOR opera concessões apenas no nordeste do Estado (região verde escuro no mapa da Figura 1), com a performance operacional, em junho de 2015, descrita na Tabela 01.

Tabela 1 – Informações operacionais da COPANOR

| Situação do Município | Número de Municípios |
|------------------------------|-----------------------------|
| Com concessão | 49 |
| Em operação | 45 |

| Serviço Prestado | Unidades |
|-------------------------|-----------------|
| Água | 87.000 |
| Esgoto | 37.000 |

Fonte: <http://www.copasa.com.br/ri/>

Os serviços de saneamento, segundo a Lei Federal 11.445/2007, marco legal do setor, terão sua sustentabilidade assegurada mediante a cobrança de tarifas. As tarifas públicas, por sua vez, devem respeitar a capacidade de pagamento, tipo de uso dado ao recurso hídrico (consumidores residenciais, comerciais, industriais e públicos) e estimular o consumo consciente da água, dentre outros.

“Art. 29. Os serviços públicos de saneamento básico terão a sustentabilidade econômico-financeira assegurada, sempre que possível, mediante remuneração pela cobrança dos serviços:

I - de abastecimento de água e esgotamento sanitário: preferencialmente na forma de tarifas e outros preços públicos, que poderão ser estabelecidos para cada um dos serviços ou para ambos conjuntamente;

II - de limpeza urbana e manejo de resíduos sólidos urbanos: taxas ou tarifas e outros preços públicos, em conformidade com o regime de prestação do serviço ou de suas atividades;

III - de manejo de águas pluviais urbanas: na forma de tributos, inclusive taxas, em conformidade com o regime de prestação do serviço ou de suas atividades.

§ 1º Observado o disposto nos incisos I a III do caput deste artigo, a instituição das tarifas, preços públicos e taxas para os serviços de saneamento básico observará as seguintes diretrizes:

I - prioridade para atendimento das funções essenciais relacionadas à saúde pública;

II - ampliação do acesso dos cidadãos e localidades de baixa renda aos serviços;

III - geração dos recursos necessários para realização dos investimentos, objetivando o cumprimento das metas e objetivos do serviço;

IV - inibição do consumo supérfluo e do desperdício de recursos;

V - recuperação dos custos incorridos na prestação do serviço, em regime de eficiência;

VI - remuneração adequada do capital investido pelos prestadores dos serviços;

VII - estímulo ao uso de tecnologias modernas e eficientes, compatíveis com os níveis exigidos de qualidade, continuidade e segurança na prestação dos serviços;

VIII - incentivo à eficiência dos prestadores dos serviços.

§ 2º Poderão ser adotados subsídios tarifários e não tarifários para os usuários e localidades que não tenham capacidade de pagamento ou escala econômica suficiente para cobrir o custo integral dos serviços.

Art. 30. Observado o disposto no art. 29 desta Lei, a estrutura de remuneração e cobrança dos serviços públicos de saneamento básico poderá levar em consideração os seguintes fatores:

I - categorias de usuários, distribuídas por faixas ou quantidades crescentes de utilização ou de consumo;

II - padrões de uso ou de qualidade requeridos;

III - quantidade mínima de consumo ou de utilização do serviço, visando à garantia de objetivos sociais, como a preservação da saúde pública, o

adequado atendimento dos usuários de menor renda e a proteção do meio ambiente;

IV - custo mínimo necessário para disponibilidade do serviço em quantidade e qualidade adequadas;

V - ciclos significativos de aumento da demanda dos serviços, em períodos distintos; e

VI - capacidade de pagamento dos consumidores.

Art. 31. Os subsídios necessários ao atendimento de usuários e localidades de baixa renda serão, dependendo das características dos beneficiários e da origem dos recursos:

I - diretos, quando destinados a usuários determinados, ou indiretos, quando destinados ao prestador dos serviços;

II - tarifários, quando integrarem a estrutura tarifária, ou fiscais, quando decorrerem da alocação de recursos orçamentários, inclusive por meio de subvenções;

III - internos a cada titular ou entre localidades, nas hipóteses de gestão associada e de prestação regional”

(Artigos 29 a 31 da Lei Federal 11.445/2007)

A COPANOR, desde sua criação, possui uma tabela progressiva de tarifas com as seguintes categorias de usuários: (i) residencial, (ii) comercial, (iii) industrial e (iv) público, não havendo distinção, dentre os usuários residenciais, daqueles considerados de baixa renda, ou seja, aqueles enquadrados em uma provável categoria social. A ausência de uma categoria de tarifa social significa que não há distinção entre os usuários dos serviços, dessa forma não se separam os cidadãos de baixa renda daqueles com capacidade financeira para arcar com tarifas normais.

Uma das hipóteses plausíveis para não haver uma categoria social na tabela tarifária da empresa é que os habitantes da região de concessão seriam beneficiários de uma eventual tarifa social. Contudo, supor que, mesmo em uma

região carente, todos os usuários seriam de baixa renda é uma hipótese que merece ser testada.

Para estimar a proporção de usuários atendidos pela COPANOR que poderiam ser enquadrados como de baixa renda e, portanto, passíveis de enquadramento em uma futura tarifa social, foi necessário apurar a efetiva população de baixa renda na região.

Partindo do pressuposto que todo indivíduo inscrito no Cadastro Único para Programas Sociais – CadÚnico, é considerado de baixa renda, a família que possuir um membro inscrito será considerada potencial beneficiária da tarifa social.

“O Cadastro Único (CadÚnico) é um sistema que identifica e caracteriza as famílias de baixa renda, para que possam ter acesso aos programas Sociais do Governo Federal. Entende-se por família de baixa renda aquela que possui renda mensal de até meio salário mínimo per capita; ou renda mensal total de até três salários mínimos, ou seja, que a soma de todos os salários do grupo familiar seja de até três salários mínimos.”
Disponível em: <http://www.programadogoverno.org/cadastro-unico-do-governo-federal-cadunico>

Sendo assim, o principal objetivo deste trabalho é estimar a proporção de usuários residenciais na área de atendimento da COPANOR com potencial para serem beneficiários da tarifa social: a família residente em uma unidade usuária do serviço de saneamento com pelo menos um membro inscrito no CadÚnico.

Os objetivos específicos são:

1. Validar a amostra que permitirá inferir sobre a população amostrada.
2. Caracterizar as localidades e municípios atendidos pela COPANOR quanto o consumo de água e o nível de renda, medido pelo IDH-Renda.

2. A tarifa social no setor de energia elétrica

Em virtude da maturidade da regulação no setor elétrico brasileiro, uma análise quantitativa acerca da distribuição dos usuários desse setor, segundo seu enquadramento tarifário, servirá de comparativo à proporção estimada dos potenciais beneficiários dos serviços de saneamento na área de atuação da COPANOR.

A Agência Nacional de Energia Elétrica – ANEEL, no artigo 8º da *Resolução Normativa 414 de 9 de setembro de 2010*, estabelece as condições necessárias para que as unidades consumidoras possam ser classificadas como baixa renda.

“Art. 8º: As unidades consumidoras serão classificadas nas Subclasses Residencial Baixa Renda, desde que sejam utilizadas por:

I - família inscrita no Cadastro Único para Programas Sociais do Governo Federal - Cadastro Único, com renda familiar mensal per capita menor ou igual a meio salário mínimo nacional;

II - quem receba o Benefício de Prestação Continuada da Assistência Social - BPC, nos termos dos arts. 20 e 21 da Lei no 8.742, de 7 de dezembro de 1993;

III - família inscrita no Cadastro Único com renda mensal de até 3 (três) salários mínimos, que tenha portador de doença ou patologia cujo tratamento ou procedimento médico requeira o uso continuado de aparelhos, equipamentos ou instrumentos que, para o seu funcionamento, demandem consumo de energia elétrica.”

As tabelas 2, 3 e 4 apresentam a distribuição das unidades residenciais consumidoras de energia elétrica no Brasil, na região Sudeste do país e em duas distribuidoras no estado de Minas Gerais, respectivamente, segundo a classificação da renda familiar (baixa ou não) em cada uma das cinco regiões geográficas do país (Tabela 2), em cada um dos cinco estados da região Sudeste (Tabela 3) e em duas companhias distribuidoras de energia elétrica de Minas Gerais em julho de 2015 (tabela 4).

Tabela 2 - Distribuição de frequência das unidades residenciais consumidoras de energia elétrica no Brasil segundo a classificação da renda familiar (baixa ou não) em cada uma das cinco regiões geográficas do país em julho de 2015.

| Região | Classificados como de "baixa renda" | | | | Total |
|--------------|-------------------------------------|---------------|-------------------|---------------|-------------------|
| | Sim | | Não | | |
| Centro-Oeste | 430.962 | 8,62% | 4.567.651 | 91,38% | 4.998.613 |
| Sudeste | 1.978.112 | 6,48% | 28.539.958 | 93,52% | 30.518.070 |
| Sul | 546.744 | 5,66% | 9.117.268 | 94,34% | 9.664.012 |
| Nordeste | 4.998.158 | 28,07% | 12.811.052 | 71,93% | 17.809.210 |
| Norte | 850.859 | 21,05% | 3.190.560 | 78,95% | 4.041.419 |
| Total | 8.804.835 | 13,14% | 58.226.489 | 86,86% | 67.031.324 |

Fonte: <http://www.aneel.gov.br/area.cfm?idArea=756&idPerfil=2&reltipo=presetReg>

Tabela 3 - Distribuição de frequência das unidades residenciais consumidoras de energia elétrica no Brasil segundo a classificação da renda familiar (baixa ou não) em cada um dos quatro estados da região sudeste do país em julho de 2015.

| Estados da Região Sudeste | Classificados como de "baixa renda" | | | | Total |
|---------------------------|-------------------------------------|--------------|-------------------|---------------|-------------------|
| | Sim | | Não | | |
| ES | 109.316 | 9,10% | 1.092.181 | 90,90% | 1.201.497 |
| MG | 751.722 | 10,96% | 6.107.776 | 89,04% | 6.859.498 |
| RJ | 312.640 | 5,36% | 5.520.803 | 94,64% | 5.833.443 |
| SP | 804.434 | 4,84% | 15.819.198 | 95,16% | 16.623.632 |
| Total | 1.978.112 | 6,48% | 28.539.958 | 93,52% | 30.518.070 |

Fonte: <http://www.aneel.gov.br/area.cfm?idArea=756&idPerfil=2&reltipo=presetReg>

Tabela 4 - Distribuição de frequência das unidades residenciais consumidoras de energia elétrica no Brasil segundo a classificação da renda familiar (baixa ou não) nas distribuidoras CEMIG e EMG em julho de 2015

| Distribuidoras | Classificadas como de "baixa renda" | | | | Total |
|----------------------|-------------------------------------|---------------|------------------|---------------|------------------|
| | Sim | | Não | | |
| EMG - Energisa Minas | 45.384 | 14,21% | 274.021 | 85,79% | 319.405 |
| CEMIG-D - CEMIG | 701.980 | 10,84% | 5.773.450 | 89,16% | 6.475.430 |
| Total | 747.364 | 11,00% | 6.047.471 | 89,00% | 6.794.835 |

Fonte: <http://www.aneel.gov.br/area.cfm?idArea=756&idPerfil=2&reltipo=presetEst>

Como se pode observar, o percentual de unidades consumidoras residenciais de baixa renda na região Sudeste é de 6,48%. Considerando o estado de Minas Gerais, esse percentual sobe para 10,96% e, dentro deste estado, a distribuidora

Energia Minas Gerais possui, entre seus clientes residenciais, 14,21% de unidades classificadas como de baixa renda.

A tarifa social permite aos cidadãos de baixa renda acessar os serviços públicos, respeitando sua capacidade de pagamento, além de distingui-los daqueles com capacidade financeira de arcar plenamente com os custos incorridos na prestação dos mesmos. Dessa forma, torna-se importante identificar os usuários passíveis do benefício, não apenas para dar a eles o acesso ao subsídio, mas determinar as tarifas justas às demais categorias de usuários.

Os resultados do setor elétrico demonstram que, mesmo em regiões mais carentes, nem todos os usuários dos serviços são enquadrados na categoria social. Neste sentido, a COPANOR buscou estimar aqueles passíveis de serem beneficiados com tais tarifas.

3. Materiais e Métodos

3.1. Os dados

O banco de dados comercial da COPANOR armazena os dados das unidades residenciais atendidas pela companhia, entre eles, a localidade onde está situada a residência e o consumo médio por unidade usuária (m³/un) por município e localidade. Um dos identificadores dessas unidades residenciais é o número do Cadastro de Pessoas Físicas (CPF) do usuário responsável pela unidade.

O CadÚnico é um banco de dados mantido pelo Governo Federal e, com base nele, é possível identificar quem teve atendimento prioritário em programas sociais promovidos pelo Governo, incluindo, por exemplo, os beneficiários do Bolsa Família. Os dados do CadÚnico não são de acesso público, visto a natureza dos dados nele disponíveis. Dessa forma, para este trabalho, os dados disponibilizados foram consolidados pela empresa, por localidade e município, segundo o número de famílias cadastradas no CadÚnico e identificadas como clientes da COPANOR (cruzamento realizado pela empresa) e total de

beneficiários inscritos no CadÚnico. Os dados disponíveis, portanto, são uma agregação dos dados de variáveis por localidade e Município, não havendo dado discricionário por usuário.

Os dados referentes à renda média nos municípios analisados foram obtidos no sítio eletrônico do Instituto de Pesquisa Econômica Aplicada – IPEA, disponível em: <http://www.ipeadata.gov.br>.

Adicionalmente, a COPANOR construiu um banco para armazenar os dados de um levantamento feito junto aos moradores das unidades residenciais com objetivo de identificar usuários que porventura estivessem cadastrados no CadÚnico. Esses dados também foram agregados por localidade e município.

3.2. Método

O banco da COPANOR contém os dados de todas as unidades residenciais atendidas pela empresa, que se configura a população deste estudo. Essa população pode ser dividida em duas subpopulações:

- 1) subpopulação 1: unidades com dado do CPF e/ou CNPJ (condomínio) do responsável cadastrado na base da COPANOR;
- 2) subpopulação 2: unidades sem dados de CPF e/ou CNPJ (condomínio) do responsável cadastrado na base da COPANOR.

Na subpopulação 1, foi realizado o cruzamento entre o banco de dados da COPANOR e o banco de dados do CadÚnico, com o objetivo de identificar as unidades residenciais atendidas pela COPANOR cujos responsáveis também eram responsáveis por famílias beneficiárias de um algum programa social. Essas unidades residenciais fariam jus à tarifa social.

No entanto, o número obtido por meio desse cruzamento poderia não representar o real número de unidades com potencial para serem beneficiárias da tarifa social, visto que o número do CPF do responsável não estava disponível para todas as unidades no banco de dados da COPANOR e que o responsável pela unidade residencial cujo CPF estava cadastrado no CadÚnico poderia não

ser o mesmo membro da família que tem o seu CPF cadastrado no banco da COPANOR.

Sendo assim, a COPANOR planejou um levantamento de dados junto às unidades residenciais pertencentes à subpopulação 2. Esse levantamento foi realizado pelos leituristas da companhia quando das suas visitas às unidades.

Os leituristas fizeram duas perguntas: (i) *alguém que more aqui (casa ou apartamento) está cadastrada no CadÚnico?* (ii) *alguém que more aqui (casa ou apartamento) recebe bolsa-família ou algum benefício do Governo?*

O objetivo das duas perguntas era identificar se algum morador estava cadastrado no CadÚnico, mas, como o nome do cadastro poderia não ser familiar para o respondente, procedeu-se ao segundo questionamento. Isto porque todos os beneficiários do programa bolsa-família devem estar cadastrados no CadÚnico e a maioria dos programas de subvenção o tem como base cadastral. A resposta afirmativa para qualquer das perguntas implicou em considerar aquela unidade usuária como potencial beneficiária da tarifa social.

O levantamento deveria acontecer em todos os municípios atendidos pela COPANOR. Com este procedimento, esperava-se contemplar todos os extratos socioeconômicos.

Embora o objetivo fosse realizar as visitas à quase totalidade das unidades residenciais da subpopulação 2, os dados puderam ser coletados em apenas uma amostra dessas unidades.

Segundo a empresa, não houve sorteio para a realização da pesquisa, foi solicitado ao leiturista que, ao realizar a aferição de consumo da residência, procedesse à pesquisa, nas residências da subpopulação 2.

As informações obtidas pela contagem dos resultados demonstram que nem todas as residências da subpopulação 2 participaram do levantamento, visto que, em alguns casos, não havia morador presente e, em outros, os residentes se recusaram a responder ao questionário.

Com base no exposto acima, pode-se notar que, para estimar do número de unidades residenciais atendidas pela COPANOR que fazem jus à tarifa social, há dois tipos de inferência a serem feitas neste trabalho: uma inferência exata no caso da subpopulação 1 (unidades com CPF) e uma inferência estatística no caso da subpopulação 2 (unidades sem CPF).

A estimativa para o percentual de unidades residenciais atendidas pela COPANOR que fazem jus à tarifa social será uma composição dos resultados dos dois tipos de inferência.

3.2.1. Avaliação da representatividade da amostra de unidades residenciais

Apesar do tamanho da amostra ser considerável, mais importante é obter uma amostra que seja uma representação honesta da população e que conduza à estimação das características da população com grande precisão e pouco ou nenhum vício.

Para estender os resultados obtidos na população que respondeu a pesquisa (11.390 residências) para a parte que não respondeu (14.716 residências), devemos mostrar que essas duas partes da população não são diferentes com respeito às características que possam ter relação com o fato de terem ou não CPF no CadÚnico.

Neste sentido, para dar confiabilidade à extrapolação do resultado da pesquisa para a população, testes de hipóteses¹ deveriam ser realizados, visando validar a amostra por meio da comparação das duas subpopulações com respeito a características que possam ter relação com o benefício da tarifa social. No caso deste estudo, apenas os dados sobre o consumo médio de água das localidades e o IDHM-Renda do município estavam disponíveis.

O teste de hipótese é, segundo Montgomery & Runger (2012), *um procedimento levando a uma decisão acerca de uma hipótese particular.*

¹ “Hipótese é uma alegação, ou afirmação, sobre uma propriedade de uma população.” Triola, Mario F., página 171. Introdução à Estatística.

Dessa forma, em Estatística, uma hipótese é uma suposição formulada a respeito dos parâmetros de uma distribuição de probabilidade de uma ou mais populações. Esta hipótese é testada com base em resultados amostrais, sendo rejeitada ou não. Essa hipótese somente será rejeitada se o resultado da amostra for claramente improvável de ocorrer quando a hipótese for verdadeira.

Consideremos H_0 (a hipótese nula) a ser testada e H_1 (a hipótese alternativa) a hipótese complementar a H_0 . O teste de hipóteses pode levar à rejeição ou não de H_0 . Em caso de rejeição de H_0 , a hipótese a ser considerada será a alternativa (H_1).

Os parâmetros testados foram média e variância, tanto para o consumo médio de água quanto para o IDHM-Renda.

3.2.2 O teste de hipóteses para uma média populacional

Como dito anteriormente, o teste de hipótese é uma técnica de inferência estatística que permite julgar se a característica observada numa amostra da população é compatível com o que se supõe para esta característica na população.

O teste de hipótese para uma média populacional (μ) compara o valor da média amostral observada (\bar{x}) com o valor esperado para \bar{x} caso a média μ assuma o valor estabelecido para ela sob a hipótese nula do teste, ou seja, μ_0 . Essa comparação é feita via a estatística de teste Z, que é uma diferença padronizada entre os valores da média amostral e de μ_0 , como apresentado na equação 1, que mostra a estatística de teste para a média, supondo uma população de dados que segue a distribuição Normal

$$Z_{\text{exp}} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}, \quad \text{equação 1}$$

onde μ é o valor médio populacional, \bar{x} é o valor médio da amostra, σ é o desvio-padrão populacional e n é o tamanho da amostra.

Se as hipóteses são $H_0: \mu = \mu_0$ e $H_1: \mu \neq \mu_0$, então a estatística de teste é calculada como $Z_{\text{exp}} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$. Considerando um nível de significância de 5%, se $-1.96 \leq Z_{\text{exp}} \leq 1.96$, a hipótese nula não é rejeitada. Caso contrário, a hipótese nula é rejeitada em favor da hipótese alternativa.

3.2.3 – O teste de hipóteses para a variância populacional

No caso das medidas de dispersão, as variâncias foram analisadas por meio de um teste de hipóteses de homogeneidade das variâncias, assumindo que a população tem uma distribuição Normal. Este teste utiliza a razão das variâncias como estatística de teste, como mostra a equação 2

$$X^2 = \frac{(n - 1)S^2}{\sigma^2}, \quad \text{equação 2}$$

onde σ^2 é a variância populacional, S^2 é a variância amostral e n é o tamanho da amostra.

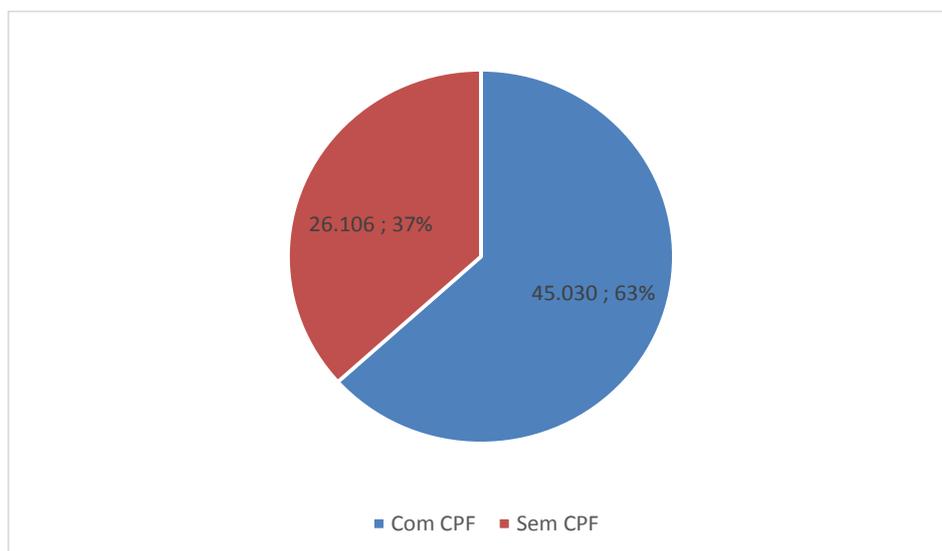
Se as hipóteses são $H_0: \sigma = \sigma_0$ e $H_1: \sigma \neq \sigma_0$, então a estatística de teste é calculada como $X_{\text{exp}} = \frac{(n-1)S^2}{\sigma_0^2}$. Considerando um nível de significância de 5%, a hipótese nula não é rejeitada se $LI < X_{\text{exp}} < LS$, onde LI e LS são, respectivamente os percentis da distribuição qui-quadrado com $(n-1)$ graus de liberdade que deixam 2,5% e 97,5% abaixo deles. Caso contrário, a hipótese nula é rejeitada em favor da hipótese alternativa.

4. Resultados

4.1. Análise descritiva dos dados

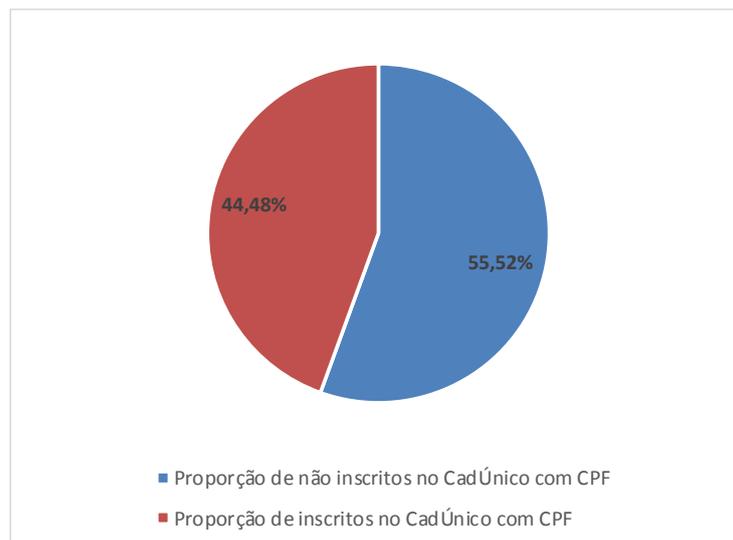
Da população total de 71.136 de unidades residenciais cadastradas no banco de dados da COPANOR, 63,30% (45.030) possuía CPF cadastrado no banco de dados da empresa, como pode ser visto na Figura 2.

Figura 2 – Distribuição de frequências das unidades residenciais segundo a existência de CPF cadastrado na COPANOR



Dentre as 45.030 unidades residenciais que possuíam CPF cadastrado, 44,48% (20.030) foram identificadas como beneficiárias de algum programa social associado ao CadÚnico (Figura 3). Desse modo, inicialmente, o percentual de unidades residenciais atendidas pela COPANOR que fazem jus à tarifa social foi estimado em 44,48%.

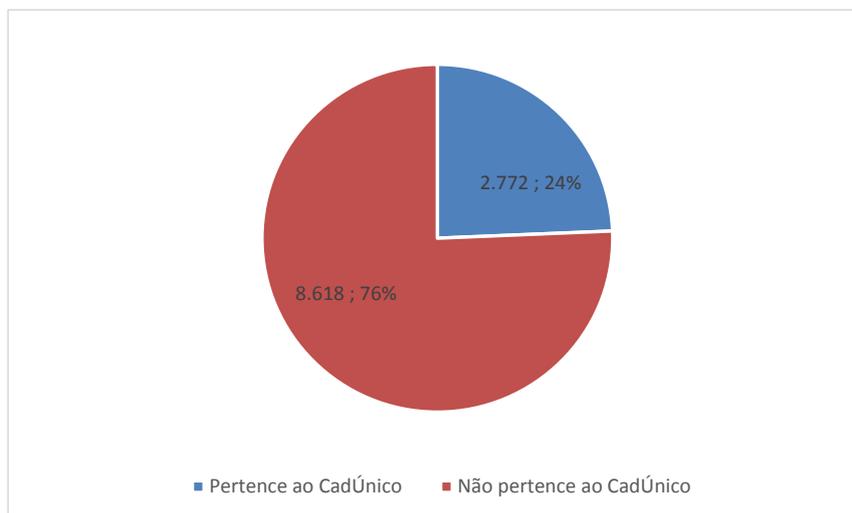
Figura 3 - Distribuição de frequências das unidades residenciais segundo a existência de CPF cadastrado na COPANOR e no CadÚnico



No grupo de 26.106 a serem visitadas (unidades residenciais sem CPF cadastrado), foram realizadas entrevistas em 11.390 (43,63%) residências. Segundo informações dos responsáveis pelas visitas, as causas para a não realização das entrevistas em mais da metade das unidades a serem visitadas foram as seguintes: (i) recusa em responder ao leitorista, (ii) ausência do morador, (iii) perda das folhas de resposta, dentre outras explicações.

Como se pode observar na Figura 4, das 11.390 visitas com entrevistas válidas, em 2.772 (24,32%) residências foi identificado que algum membro da família residente no imóvel pertencia ao CadÚnico.

Figura 4 – Distribuição de frequências das 11.390 unidades residenciais onde ocorreram entrevistas segundo a inclusão de membros familiares no CadÚnico



Os hábitos de consumo de água nas localidades e a renda média dos municípios caracterizados através das medidas resumo apresentadas nas Tabelas 5 e 6.

Tabela 5 – Análise descritiva dos dados de consumo de água das localidades na COPANOR (N= 181)

| Medidas Resumo (m ³ /ec) | |
|-------------------------------------|--------|
| Média | 7,4050 |
| Mediana | 7,3565 |
| Mínimo | 5,0681 |
| Máximo | 9,8490 |
| Desvio Padrão | 1,2308 |
| Coeficiente de Variação | 16,62% |

Tabela 6 – Análise descritiva dos dados de renda dos Municípios na COPANOR (N=67)

| Medidas Resumo IDH - Renda | |
|----------------------------|--------|
| Média | 0,5877 |
| Mediana | 0,5820 |
| Mínimo | 0,5230 |
| Máximo | 0,6970 |
| Desvio Padrão | 0,0360 |
| Coeficiente de Variação | 6,13% |

As figuras 5 e 6 mostram os box-plots dos dados de consumo de água das localidades e da renda (IDHM-Renda) dos municípios, respectivamente.

Figura 5 – Box Plot dos dados de consumo de água da população

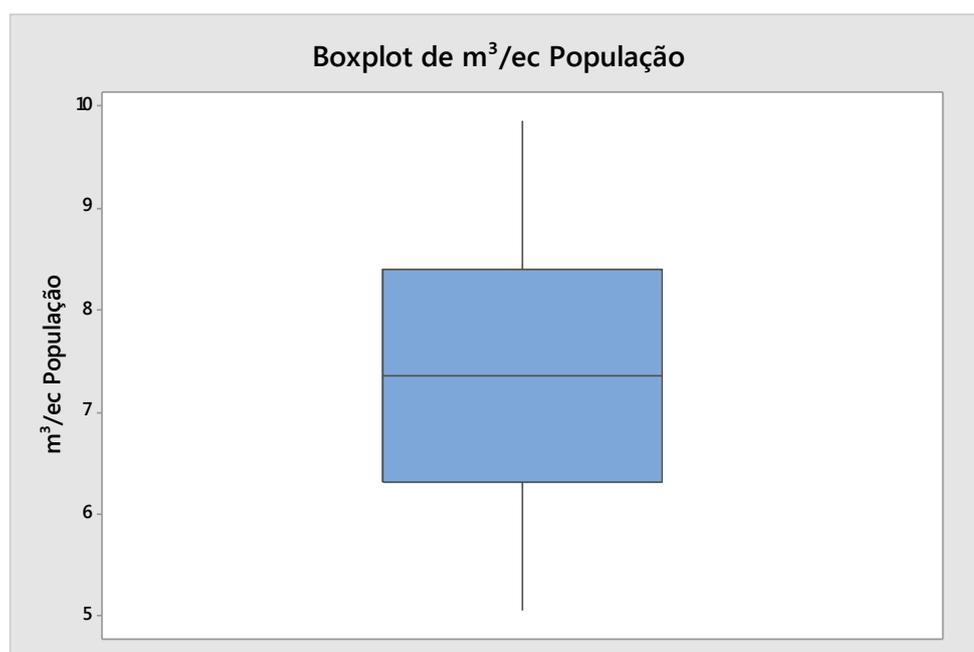
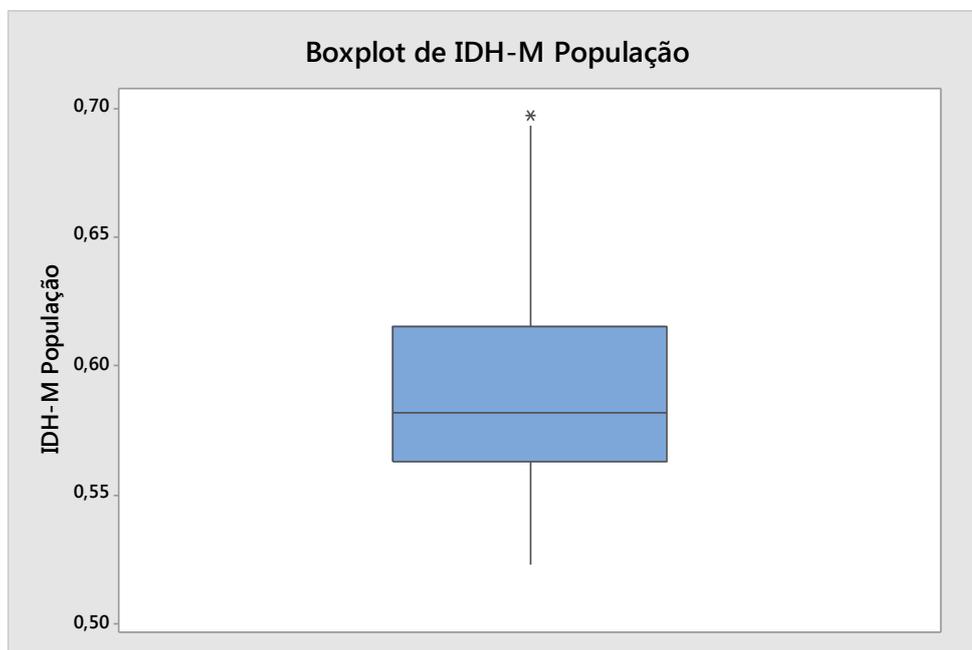


Figura 6 – Box Plot distribuição dos dados de renda da população



O consumo médio de água na região é inferior à média dos consumidores atendidos pela concessionária regional dos serviços de abastecimento de água e esgotamento sanitário do Estado de Minas Gerais (COPASA –MG), que é de 11,7 m³/ec, segundo dados de junho de 2014, disponíveis em www.copasa.com.br/ri. O consumo máximo das localidades analisadas ainda é inferior à média dos usuários da COPASA – MG, sendo 9,84 m³/ec. Observa-se, também que há baixa dispersão de consumo em torno da média, visto que os coeficientes de variação são 16,62% e 6,13% respectivamente.

O IDH – Renda no Estado de Minas Gerais, para o mesmo período de apuração dos municípios da COPANOR é 0,798, demonstrando que se trata de uma região, em média, mais carente. Esta análise é reforçada ao se observar que o IDH-renda mais alto dos municípios analisados é 0,697, abaixo do IDH-renda do Estado e trata-se de um outlier. Observa-se, ainda que há baixa dispersão em torno da média, com variância e desvio-padrão 0,0013 e 0,0360, respectivamente.

4.2. Estimativa do percentual de unidades que fazem jus à tarifa social

O percentual de unidades residenciais que faz jus à tarifa social na subpopulação de usuários sem CPF foi estimado em 24,32%, com erro de estimação apurado em 0,59% (equação 3) com os dados apresentados na Tabela 5.

$$e = \sqrt{\frac{(pqz^2)(N - n)}{n(N - 1)}}$$

Equação 3

Tabela 7 – Estimativa de erro de estimação do percentual de unidades que fazem jus à tarifa social na subpopulação sem CPF cadastrado na COPANOR

| Total amostra | | |
|---|---|--------|
| Tamanho da População | N | 26.106 |
| Proporção de casos positivos | p | 24,34% |
| Proporção de casos negativos | q | 75,66% |
| Percentil da distribuição gaussiana, nível de confiança 95% | z | 1,96 |
| Tamanho da Amostra | n | 11.390 |
| Erro | e | 0,0059 |

Sendo assim, a estimativa para o total de unidades residenciais que fazem jus à tarifa social na subpopulação de residências sem cadastro de CPF na COPANOR (subpopulação 2) foi calculada em 6.354 unidades (26.106 x 24,34%).

Considerando o número de unidades que fazem jus à tarifa social apurados nas duas subpopulações, a companhia estimou uma quantidade de potenciais beneficiários da tarifa social como a soma dos clientes pré-identificados dentre aqueles com CPF cadastrados (20.030) e com a estimativa feita na população de unidades sem o registro do CPF na empresa (6.354), resultando em um valor estimado de 26.384 unidades residenciais potenciais beneficiárias da tarifa social.

Assim, a estimativa da porcentagem de unidades residenciais potenciais de receber a tarifa social é de 37,09% do total dos 71.136 clientes residenciais da COPANOR.

4.3. Validação da amostra

Para poder extrapolar o resultado obtido na amostra de visitas domiciliares para a população de unidades residenciais sem CPF no cadastro da COPANOR, foi necessário saber, a partir de testes estatísticos, se a amostra é representativa da população.

No entanto, os dados foram agregados por localidade e não houve registros sobre quais residências foram visitadas. Desse modo, não foi possível calcular estatísticas sobre o consumo de água, por exemplo, para residências visitadas e residências não-visitadas. Além disso, o modo como os dados foram disponibilizados impossibilitou o cálculo até mesmo de estatísticas para os grupos de residências sem o CPF registrado na COPANOR (subpopulação 2).

Considerando que os dados estão agregados por localidade (no caso do consumo de água) e por município (no caso do IDH-Renda), a comparação dos parâmetros média e variância da população de onde foi retirada a amostra de residências visitadas foi feita com os parâmetros da população da região de atuação da COPANOR, incluindo residências com e sem CPF registrado na companhia. A Tabela 6 mostra a caracterização da população total e da amostra de residências visitadas segundo a quantidade de localidades, municípios e imóveis sem registro de CPF no banco de dados da COPANOR. Como se pode observar, das informações quantitativas dessa tabela, a amostra obtida da população é relativamente expressiva, o que possibilitará uma inferência mais consistente acerca da população.

Tabela 8 – Caracterização da população atendida pela COPANOR e da amostra de residências visitadas segundo localidade, municípios e grupo de residências sem o CPF registrado no banco de dados da companhia

| | População | Amostra | Tamanho |
|---|-----------|---------|---------|
| Localidades | 181 | 80 | 44% |
| Municípios | 67 | 41 | 61% |
| Imóveis sem registro de CPF no Banco de Dados COPANOR | 26.106 | 11.390 | 44% |

Para avaliar a amostra obtida nas visitas domiciliares, foram realizados testes de hipóteses destinados a comparar as medidas descritivas de posição (média) e dispersão (variância) da população e a amostra.

4.3.1. Teste de hipótese para a média

O IDHM Renda considera a renda municipal per capita, ou seja, a renda média mensal dos indivíduos residentes em determinado município

As estatísticas necessárias para a realização dos testes de hipóteses são detalhadas na Tabela 9.

Tabela 9 – Estatísticas e parâmetros de referência para os testes de hipóteses

| | Variável | |
|---|-------------------------------|----------------------------|
| | Consumo de água na localidade | IDH-M (Renda) no município |
| Média na População de Referência (μ_0) | 7,4050 | 0,5877 |
| Variância na População de Referência (σ_0^2) | 1,5066 | 0,0013 |
| Tamanho da População de Referência (N) | 181 | 67 |

| | | |
|------------------------------|--------|--------|
| Média amostral (\bar{x}) | 7,5673 | 0,5863 |
| Variância amostral (S^2) | 1,3547 | 0,0015 |
| Tamanho da Amostra (n) | 80 | 41 |

Teste de hipóteses para o consumo médio de água nas localidades

O conjunto de hipóteses utilizado foi $H_0: \mu = 7,40$ e $H_1: \mu \neq 7,40$, onde μ é a média de consumo de água nas localidades com residências amostradas. Conforme critérios anteriormente explanados, não se rejeita H_0 se $-1,96 < Z_{exp} < 1,96$, onde Z_{exp} é valor da estatística do teste e o nível de significância utilizado foi 5%.

A estatística de teste foi calculada como descrito na equação 1, com os dados da tabela 9, sendo:

$$Z_{exp} = \frac{7,5673 - 7,4050}{\frac{\sqrt{1,5066}}{\sqrt{80}}} = 1,18$$

Como $Z_{exp}=1,18$, não se rejeita a hipótese de que o consumo médio de água nas localidades com residências amostradas seja igual ao consumo médio de água na população total no nível de 5% de significância.

Teste de hipóteses para a renda média (IDH-M) nos municípios

O conjunto de hipóteses utilizado foi $H_0: \mu = 0,5877$ e $H_1: \mu \neq 0,5877$, onde μ é a média do IDH-M nos municípios com residências amostradas. Conforme critérios anteriormente explanados, não se rejeita H_0 se $-1,96 < Z_{exp} < 1,96$, onde Z_{exp} é valor da estatística do teste e o nível de significância utilizado foi 5%.

A estatística de teste foi calculada como descrito na equação 1, com os dados da tabela 9, sendo:

$$Z_{exp} = \frac{0,5863 - 0,5877}{\frac{\sqrt{0,0013}}{\sqrt{41}}} = -0,24$$

Como $Z_{exp} = -0,24$, não se rejeita a hipótese de que o renda média nos municípios com residências amostradas seja igual à renda média na população total no nível de 5% de significância.

Quadro 1 – Estatísticas utilizadas nos testes de hipóteses para as médias

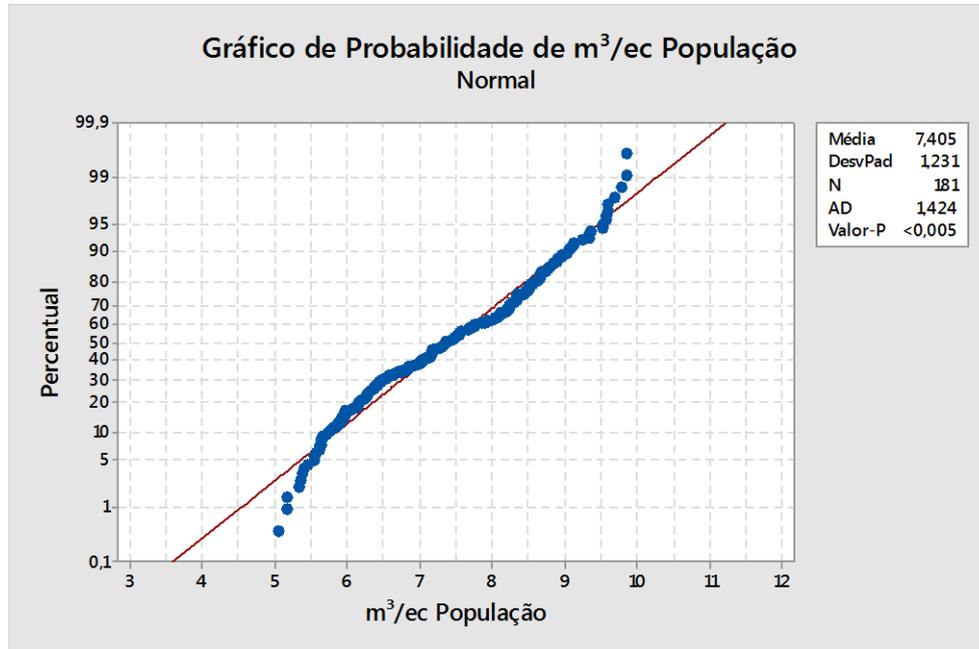
| Variável | μ | X | σ/\sqrt{n} | Z_{exp} |
|--------------------------------------|--------|--------|-------------------|-----------|
| Consumo médio das localidades (n=80) | 7,40 | 7,57 | 0,14 | 1,18 |
| IDH (renda) dos municípios (n=41) | 0,5877 | 0,5863 | 0,0056 | -0,24 |

Dessa forma, ao nível de significância de 5%, não há evidências para se rejeitar a hipótese de que média de consumo de água e renda média da amostra não sejam as mesmas da população.

4.3.2. Teste de hipótese para a variância

Antes de se testar a hipótese de que a dispersão do consumo médio de água na população de localidades amostradas é igual à da população total, é necessário assumir que a distribuição do consumo de água nas localidades tem distribuição Normal. Dessa forma, observando o gráfico de probabilidade normal da Figura 7, elaborado com o auxílio do software Minitab 17, se pode inferir que parece haver desvios da normalidade nas caudas da distribuição, o que também é captado pelo teste de Anderson-Darling (valor-p < 0,05, rejeitando a hipótese de normalidade dos dados de consumo médio de água nas localidades).

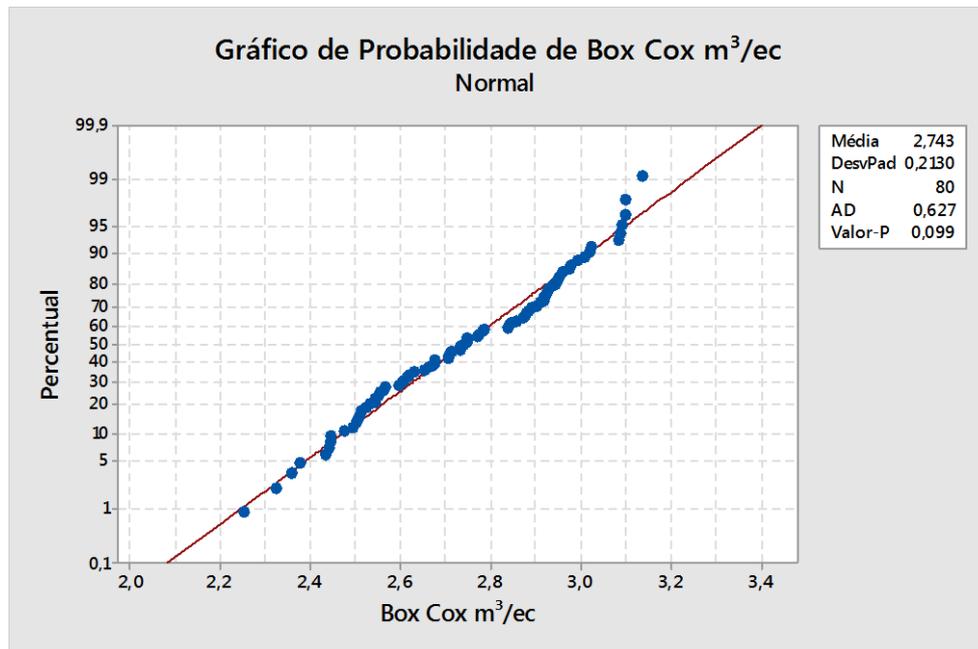
Figura 7 – Gráfico de probabilidade normal e teste de Anderson-Darling para o consumo médio de água nas localidades da população amostrada.



Sendo assim, foi utilizado o método de Box-Cox para se conseguir uma transformação dos dados de consumo de água que satisfizesse a suposição de normalidade para a distribuição destes dados. A transformação apontou para a transformação “raiz quadrada”, visto que o valor estimado de lambda é 0,57 com intervalo de 95% de confiança indo de -0,44 a 1,47.

A Figura 8 mostra o gráfico de probabilidade normal para os dados de consumo na escala raiz quadrada. Podemos ver que a transformação resolveu a hipótese de normalidade da distribuição.

Figura 8 – Gráfico de probabilidade normal e teste de Anderson-Darling para o consumo médio de água nas localidades após transformação Box Cox



O teste de variância foi realizado com os dados de consumo de água na escala raiz quadrada.

O conjunto de hipóteses utilizado foi $H_0: \sigma = 0,0514$ e $H_1: \sigma \neq 0,0514$, onde σ é a dispersão da raiz quadrada do consumo de água na população de localidades com residências amostradas. Conforme critérios anteriormente explanados, não se rejeita H_0 no nível de 5% de significância se $56,309 < X_{esp} < 105,473$, onde 56,309 e 105,473 são, respectivamente, os percentis da distribuição qui-quadrado com 79 graus de liberdade que deixam 2,5% e 97,5% abaixo deles. Caso contrário, a hipótese nula é rejeitada em favor da hipótese alternativa.

A estatística de teste foi calculada como descrito na equação 2, sendo:

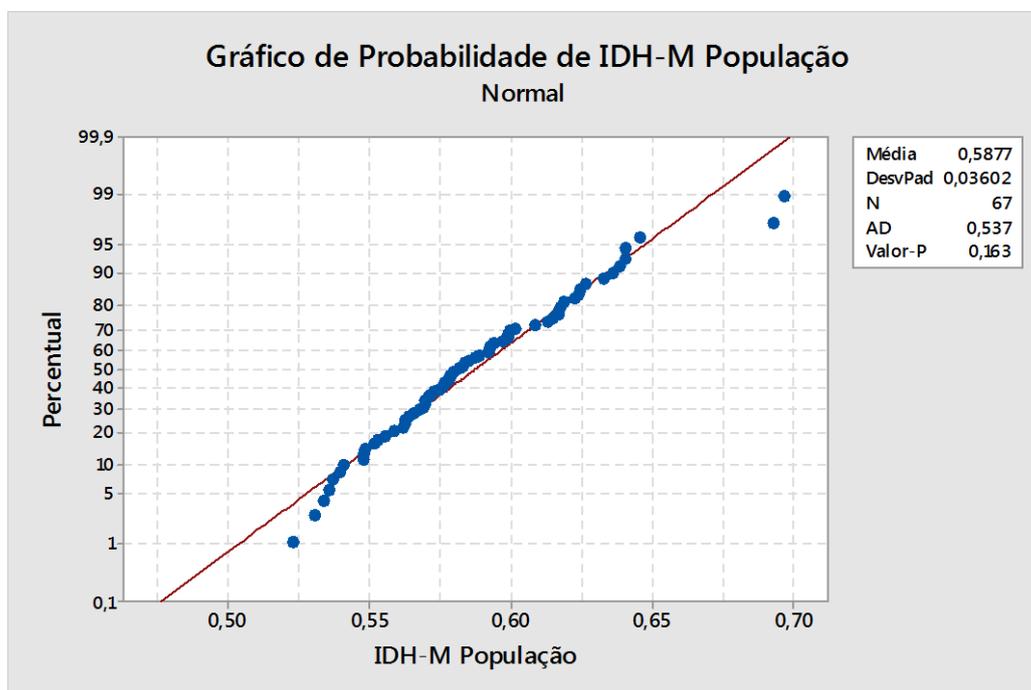
$$X^2 = \frac{(80 - 1)0,0454}{0,0514} = 69,7063$$

Como $X_{exp} = 69,7063$, não se rejeita a hipótese de que a dispersão do consumo de água nas localidades com residências amostradas seja igual à dispersão na população total no nível de 5% de significância.

Teste de hipóteses para a variância da renda (IDH-M) nos municípios

Antes de se testar a hipótese de que a dispersão da renda média nos municípios da população amostrada é igual à da população total, é necessário assumir que a distribuição da IDH-Renda nos municípios da população amostrada é normal. Dessa forma, observando o gráfico de probabilidade normal da Figura 5, pode-se inferir que não há evidências contra a suposição de normalidade, o que também é captado pelo teste de Anderson-Darling (valor-p > 0,05, não rejeitando a hipótese de normalidade dos dados de IDH-Renda nos municípios).

Figura 9 - Gráfico de probabilidade normal e teste de Anderson-Darling para a renda média nos municípios da população amostrada.



Com a hipótese de normalidade dos dados de IDH-Renda satisfeita, procedeu-se ao teste de hipóteses para a dispersão dessa variável. O conjunto de

hipóteses utilizado foi $H_0: \sigma = 0,0013$ e $H_1: \sigma \neq 0,0013$, onde σ é a dispersão da renda (IDH-M) nos municípios com residências amostradas. Conforme critérios anteriormente explanados, não se rejeita H_0 se $24,4330 < X_{\text{esp}} < 59,3417$, onde 24,4330 e 59,3417 são, respectivamente os percentis da distribuição qui-quadrado com 40 graus de liberdade que deixam 2,5% e 97,5% abaixo deles. Caso contrário, a hipótese nula é rejeitada em favor da hipótese alternativa.

A estatística de teste foi calculada como descrito na equação 2, com os dados da tabela 9, sendo:

$$X^2 = \frac{(41 - 1)0,0015}{0,0013} = 47,3353$$

Como $X_{\text{exp}} = 47,3353$, não se rejeita a hipótese de que a dispersão da renda nos municípios com residências amostradas seja igual à dispersão na população total no nível de 5% de significância.

O Quadro 2 apresenta um resumo dos testes de hipóteses para as variâncias.

Quadro 2 – Estatísticas e parâmetros de referência utilizados nos testes de hipóteses para a variância

| Variável | n | G.L | S ² | σ ² |
|-------------------------------|----|----------------|------------------------------------|--------------------------------------|
| Consumo médio das localidades | 80 | 79 | 0,0454 | 0,0514 |
| IDH (renda) dos municípios | 41 | 40 | 0,0015 | 0,0013 |
| | | | | |
| Variável | α | X ² | X ² _{α/2, n-1} | X ² _{1-α/2, n-1} |
| Consumo médio das localidades | 5% | 69,7063 | 56,309 | 105,473 |
| IDH (renda) dos municípios | 5% | 47,3353 | 24,4330 | 59,3417 |

Os resultados obtidos nos cenários descritos demonstram que os resultados da amostra das visitas domiciliares são representativos da COPANOR.

5. Discussão

Este trabalho teve o objetivo principal de estimar a proporção de usuários dos serviços de saneamento da COPANOR que eventualmente seriam beneficiários de uma tarifa social, considerando que estaria apta toda família com pelo menos um membro inscrito no CadÚnico. Neste sentido, uma análise acerca dos indivíduos pré-identificados através do cadastro da empresa e uma extrapolação de dados amostrais, após teste de validação desta amostra, foram necessários para se estimar a provável proporção de usuários potenciais.

A análise dos dados sócio econômicos da população residente na área de atuação da COPANOR demonstra que, apesar de se tratar de uma região carente, nem todos os habitantes podem ser considerados carentes, tomando-se a inscrição no CadÚnico como referencial de carência.

Os resultados da amostragem realizada pela empresa foram considerados consistentes e significativos em relação à população investigada, visto que os testes de hipóteses a respeito das características sócio econômicas elencadas (consumo de água e renda) demonstraram que não há evidências para se acreditar que a população de onde foi retirada a amostra de residências visitadas seja diferente da população total de residências atendidas pela COPANOR.

O teste de hipótese para as médias não necessitou pressupor que a distribuição o consumo de água (m^3/ec) e da variável *proxy* da renda (IDH-renda) fossem normais, supondo-se que o desvio-padrão para essas variáveis na população de residências a serem visitadas (subpopulação 2) fosse o mesmo da população total.

Quanto ao teste de hipótese para a variância, a suposição de normalidade é necessária para a realização do teste qui-quadrado e, no caso do consumo de água nas localidades, ela não foi considerada válida. No entanto, por meio do uso das transformações de Box-Cox, foi possível encontrar uma transformação que conseguiu normalizar a distribuição dos dados. O teste de variância foi realizado com os dados transformados.

Há, entretanto, de se registrar que a falta de treinamento dos leituristas da companhia para realizar as entrevistas e registrar os dados de forma individual e a falta de dados individuais para a população de unidades residenciais sem o cadastro de CPF no banco de dados da COPANOR, implicaram em limitações para uma análise mais acurada.

6. Conclusão

Partindo do princípio de que basta um residente da unidade usuária estar inscrito no CadÚnico para que essa unidade seja beneficiada com a tarifa social, o potencial de unidades residenciais a serem beneficiadas com essa categoria tarifária na área de atuação da COPANOR foi estimado em cerca de 37,09%, com margem de erro de 0,0059, para 95% de confiança.

7. Bibliografia

- Aneel. Disponível em: <www.aneel.gov.br>. Acesso em: 15.ago.2015
- Copanor. Disponível em: <www.copanor.com.br> e <www.copasa.com.br/ri>. Acesso em: 20.set.2015
- COPASA. Disponível em:<www.copasa.com.br/ri> Acesso em: 29.out.2015
- Diretrizes Nacional para o Saneamento Básico. **Lei Federal 11.445/2007**
- Montgomery, Douglas C. & Runger, George C.; **Estatística Aplicada e Probabilidade para Engenheiros**. Rio de Janeiro: LTC, 2012
- Programas Sociais do Governo Federal. Disponível em: <<http://www.programadogoverno.org/cadastro-unico-do-governo-federal-cadunico>>. Acesso em: 20.set.2015
- Real Statistics Using Excel. Disponível em: <<http://www.real-statistics.com/chi-square-and-f-distributions/one-sample-hypothesis-testing-variance/>>. Acesso em: 20.set.2015
- Triola, Mario F., **Introdução à Estatística**. Rio de Janeiro: LTC, 1999.