

CEA – USP – RAE 15P02

**RELATÓRIO DE ANÁLISE ESTATÍSTICA DO PROJETO “O GLP NO SETOR
RESIDENCIAL: DETERMINANTES DO CONSUMO”**

Cláudia Monteiro Peixoto

Gilberto Alvarenga Paula

Simone Bega Harnik

São Paulo, 5 de julho de 2015

CENTRO DE ESTATÍSTICA APLICADA – CEA – USP

CENTRO DE ESTATÍSTICA APLICADA - CEA

RELATÓRIO DE ANÁLISE ESTATÍSTICA 15P02

TÍTULO: RELATÓRIO DE ANÁLISE ESTATÍSTICA DO PROJETO “O GLP NO SETOR RESIDENCIAL: DETERMINANTES DO CONSUMO”

PESQUISADOR: Felipe de Albuquerque Sgarbi

ORIENTADOR: Edmilson Moutinho dos Santos

INSTITUIÇÃO: Instituto de Energia e Ambiente (IEE) da Universidade de São Paulo (USP)

FINALIDADE: Doutorado

RESPONSÁVEIS PELA ANÁLISE: Cláudia Monteiro Peixoto

Gilberto Alvarenga Paula

Simone Bega Harnik

FICHA TÉCNICA

REFERÊNCIA DESTE TRABALHO: HARNIK, S.B., PAULA, G.A. e PEIXOTO, C.M. **Relatório de análise estatística do projeto: “O GLP no setor residencial: determinantes do consumo”.** São Paulo, IME – USP, 2015. (RAE – CEA – 15P02)

REFERÊNCIAS

BUSSAB, W.O. e MORETTIN, P.A. (2013). **Estatística Básica.** 8ª Edição. São Paulo. Editora Saraiva. 568p.

LUCON, O., COELHO, S.T. and GOLDEMBERG, J. (2004). LPG in Brazil: lessons and challenges. **Energy for Sustainable Development**, 8, n. 3, 82-90.

EMPRESA DE PESQUISA ENERGÉTICA. (2013). **Balanço Energético Nacional - Ano base 2012.** Rio de Janeiro. Disponível em: <https://ben.epe.gov.br/downloads/Relatorio_Final_BEN_2013.pdf>. Acesso em: 15.abr.2015.

GREENE, W. H. (2012). **Econometric analysis.** 7ª Edição. Pearson Education.

IBGE. (2010). **Pesquisa de Orçamentos Familiares 2008-2009 – Despesas, Rendimentos e condições de vida.** Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/pof/2008_2009/POFpublicacao.pdf>. Acesso em: 23.abr.2015.

JANNUZZI, G.M. and SANGA, G.A. (2004). LPG subsidies in Brazil: an estimate. **Energy for Sustainable Development**, 8, n. 3, 127-129.

PEREIRA, I.M. and DE ASSIS, E.S.(2013). Urban energy consumption mapping for energy management. **Energy Policy**, 59, 257-269.

PRADO, M. (2003). Preço alto faz consumidor trocar gás de cozinha por lenha. **Folha de São Paulo.** Disponível em: <<http://www1.folha.uol.com.br/folha/dinheiro/ult91u67693.shtml>>. Acesso em: 15.abr.2015.

SGARBI, F. DE A. (2013). **Modelos de transição energética residencial e o acesso a serviços energéticos limpos : uma análise a partir de dois estudos de caso.** São Paulo. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/86/86131/tde-11092013-124241/pt-br.php>>. Acesso em: 15.abr.2015.

PROGRAMAS COMPUTACIONAIS UTILIZADOS

Microsoft Excel 2010 para Windows

Microsoft Word 2010 para Windows

R para Windows (versão 3.2.0)

RStudio para Windows (versão 0.98.501)

TÉCNICAS ESTATÍSTICAS UTILIZADAS

Análise Descritiva Unidimensional (03:010)

Análise de Regressão Clássica (07:0200)

Outros (06:990)

ÁREA DE APLICAÇÃO

Econometria (14:070)

Outros (14:990)

RESUMO

O gás liquefeito de petróleo (GLP) está presente em mais de 96% dos domicílios brasileiros e respondeu por pouco mais de 25% do consumo de energia primária do setor residencial em 2012. É utilizado prioritariamente para a cocção, que representa, segundo dados do Balanço Energético Nacional, aproximadamente 60% do consumo final de energia nos domicílios do país. No entanto, há poucas informações sobre o perfil de consumo por região e por estrato geográfico (áreas urbanas ou rurais). Este trabalho pretende construir o perfil de consumo de GLP no país a partir de dados da Pesquisa de Orçamentos Familiares (POF), de 2008-2009, do Instituto Brasileiro de Geografia e Estatística (IBGE), levantamento que oferece detalhes sobre os produtos adquiridos pelas famílias brasileiras. Neste trabalho, serão apresentados modelos de regressão linear por região para a demanda de gás domiciliar, especificando as variáveis significativas, com impacto positivo ou negativo no consumo.

Palavras-chave: gás liquefeito de petróleo; GLP; energia; consumo residencial; Pesquisa de Orçamentos Familiares

SUMÁRIO

1. Introdução	1
1.1. Objetivos	4
2. Metodologia	5
2.1. Desenho amostral	6
2.2. Instrumentos de coleta	8
2.3. Variáveis escolhidas	9
3. Análise descritiva	12
4. Análise inferencial	33
4.1. Resultados para a região Centro-Oeste.....	36
4.2. Resultados para a região Nordeste.....	39
4.3. Resultados para a região Norte	42
4.4. Resultados para a região Sudeste	45
4.5. Resultados para a região Sul	48
4.6. Interpretação das estimativas	51
5. Considerações finais	58
Apêndice A – Gráficos da análise descritiva, segundo região	59
Região Centro-Oeste	59
Região Nordeste	66
Região Norte	73
Região Sudeste	80
Região Sul	87
Anexo 1 – Rotina em R para a extração da base de dados	94
Anexo 2 – Número de setores selecionados e domicílios esperados, selecionados e entrevistados, por Unidades da Federação – POF 2008-2009	106

1. Introdução

A produção de gás liquefeito de petróleo (GLP) no Brasil teve início em 1955, dois anos após a criação da Petrobras e 18 anos após o início do seu uso residencial (LUCON, COELHO and GOLDEMBERG, 2004). Atualmente, o GLP constitui uma importante fonte de energia: está presente em mais de 96% dos domicílios brasileiros (IBGE, 2010) e responde por pouco mais de 25% do consumo de energia primária do setor residencial em 2012 (EMPRESA DE PESQUISA ENERGÉTICA, 2013), sendo utilizado prioritariamente para a cocção – segundo dados do Balanço Energético Nacional, a cocção é responsável por aproximadamente 60% do consumo final de energia nos domicílios do país.

Jannuzzi e Sanga (2004) argumentam que os subsídios históricos direcionados ao GLP no Brasil são, em grande parte, responsáveis pela sua alta penetração residencial. No entanto, a partir de 2002, o GLP deixou de ser subsidiado, acarretando um aumento de cerca de 20% em seu preço no fim do mesmo ano (LUCON, COELHO and GOLDEMBERG, 2004). O novo mecanismo criado para facilitar o acesso da população ao combustível não acompanhou o constante aumento de preços (PRADO, 2003).

Como resultado, uma parcela da população, motivada por fatores econômicos, passou a utilizar a lenha em detrimento do GLP para atividades domésticas como cocção (LUCON, COELHO and GOLDEMBERG, 2004). Assim, o volume de GLP consumido em domicílios no Brasil chega, em 2012, ao mesmo patamar de antes da retirada dos subsídios ao produto, após leve recuperação na queda de consumo. A Figura 1 apresenta o consumo de GLP, lenha e eletricidade em toneladas equivalentes de petróleo (tep).

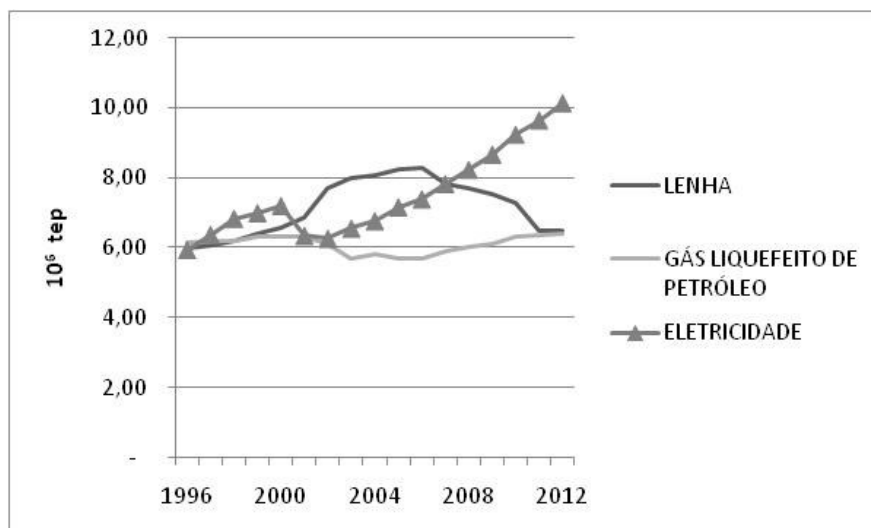


Figura 1 – Evolução do consumo residencial de lenha, GLP e eletricidade pelo setor residencial brasileiro

Fonte: Empresa de Pesquisa Energética, 2013

O ritmo de recuperação do consumo de GLP, considerado lento por especialistas do ramo, pode, segundo pesquisas, estar associado a mudanças no padrão de consumo energético do setor residencial brasileiro. Nos últimos dez anos, enquanto a demanda por eletricidade cresceu 54%, a demanda por GLP cresceu apenas 11% (EMPRESA DE PESQUISA ENERGÉTICA, 2013).

O consumo per capita de GLP acompanhou o de eletricidade até o início da década de 1990. Ao longo da década subsequente, seu ritmo de crescimento desacelerou. Já entre os anos 2000 e 2010, o consumo per capita de GLP diminuiu, ao passo que o de eletricidade continuou crescendo, conforme apresentado na Figura 2.

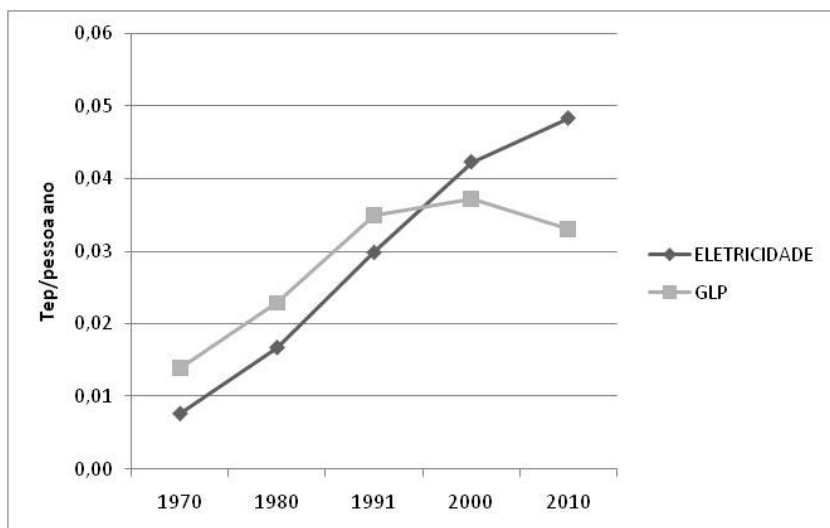


Figura 2 – Evolução do consumo anual per capita de eletricidade e GLP pelo setor residencial brasileiro

Fonte: EPE, 2013

De acordo com Jannuzzi e Sanga (2004), fatores não econômicos relacionados ao GLP são pouco compreendidos no país e podem influenciar o seu consumo ao longo do tempo. Pereira e de Assis (2013) desenvolveram um modelo de análise espacial do consumo energético de diferentes usos finais residenciais. Os autores argumentam que o aumento da eficiência dos fogões e a mudança nos hábitos de consumo energético residencial podem estar associados à redução do consumo per capita do GLP no Brasil.

Sgarbi (2013) apresenta informações sobre os combustíveis utilizados nos fogões de domicílios urbanos e rurais no país. Apesar de identificarem que o uso de múltiplos combustíveis é estratégia comum em áreas rurais, os dados não são capazes de esclarecer a proporção de consumo de cada um deles.

Nenhum dos trabalhos mencionados se dedica a discutir de que maneira o consumo energético residencial para a cocção, e, por consequência, a demanda deste setor por GLP, se comportará nos próximos anos. Por um lado, não há segurança sobre a quantidade total e proporcional de cada fonte de energia empregada neste uso, ou sobre as diferenças no padrão de consumo entre áreas rurais e urbanas ou ainda sobre o comportamento segundo diferentes faixas de

rendimento. Por outro, tendências de mudanças nos hábitos de consumo energético residencial, associadas ao aumento de renda da população e um maior acesso a equipamentos elétricos e ao gás natural em cidades, não foram analisadas.

O projeto busca diminuir lacunas de informação sobre o consumo energético em domicílios no Brasil. Para tanto, utilizará os dados secundários da Pesquisa de Orçamentos Familiares (POF) de 2008-2009, do Instituto Brasileiro de Geografia e Estatística (IBGE). Com o conjunto de dados, espera-se identificar o conjunto de variáveis relevantes para explicar o consumo de GLP no setor residencial.

1.1. Objetivos

Entre os objetivos principais do trabalho estão:

- Identificar os principais determinantes do consumo residencial de GLP a partir da POF 2008-2009 do IBGE.
- Prever o consumo residencial de GLP, dadas as variáveis explicativas relevantes.

Com base nestes dois objetivos, o pesquisador pretende realizar a discussão sobre estratégias para a indústria do GLP antecipar mudanças no mercado de combustíveis residenciais.

Ressalta-se ainda que um dos aspectos importantes desta análise é a construção de uma rotina em código aberto e livre para tratamento das informações coletadas pela POF. Até o momento, o IBGE ainda não disponibiliza rotinas em *R* para a exploração dos dados, e iniciativas esparsas têm tratado bancos de dados públicos.

2. Metodologia

Os dados utilizados na análise apresentada nos próximos capítulos pertencem à última edição da Pesquisa de Orçamentos Familiares (POF), conduzida pelo Instituto Brasileiro de Geografia e Estatística (IBGE) nos anos de 2008-2009. São, pois, dados secundários, trabalhados a fim de promoverem compreensão sobre o consumo doméstico do gás liquefeito de petróleo (GLP).

De acordo com a publicação que traz os resultados para a POF:

Os propósitos principais das Pesquisas de Orçamento Familiares – POFs realizadas pelo IBGE são disponibilizar informações sobre a composição orçamentária doméstica e sobre as condições de vida da população, incluindo a percepção subjetiva da qualidade de vida, bem como gerar bases de dados e estudos sobre o perfil nutricional da população. (IBGE, 2010, p.15)

O levantamento e a coleta de informações foram realizados em áreas urbanas e rurais de todo o território brasileiro, de 19 de maio de 2008 a 18 de maio de 2009. Os microdados – ou seja – os dados desagregados por domicílios – estão disponíveis na internet no *site* do IBGE e são abertos e públicos¹.

A POF é uma das poucas pesquisas públicas que detalham o consumo doméstico, permitindo traçar perfil das condições de vida e de consumo residencial. Assim, possui aplicações em diversas áreas e segmentos da economia, gerando também interesses variados para a academia. Seu principal e mais disseminado uso é a ponderação dos elementos da cesta de produtos que compõem índices de inflação:

A POF por investigar temas tão diversos possui múltiplas aplicações, tais como: contribuir com informações para subsidiar políticas públicas na área social para melhoria das condições de vida da população, incluídas as políticas públicas temáticas nos campos da nutrição, orientação alimentar, saúde, moradia, entre outras. Para o setor privado, a pesquisa pode ser útil na definição de estratégias de investimentos em que o

¹ Os microdados da POF 2008-2009 estão disponíveis em: <http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/pof/2008_2009/microdados.shtm>. Acesso em: 16.abr.2015.

conhecimento do perfil do consumidor e da demanda por bens e serviços seja determinante. Sob a ótica da qualidade das estatísticas públicas, destacam-se as atualizações das estruturas de ponderação das medidas de inflação, em particular do Sistema Nacional de Índices de Preços ao Consumidor produzido pelo IBGE, e da parcela de consumo das Contas Nacionais e Regionais do Brasil.

(IBGE, 2010, p. 18)

2.1. Desenho amostral

Segundo o relatório da POF 2008-2009, o desenho amostral permite a publicação de resultados nos níveis: Brasil, grandes regiões e recortes urbano-rural. Ainda de acordo com a publicação, é possível extrair resultados para as unidades da federação, mas apenas para os recortes urbano e total.

Como a pesquisa ocorreu durante 24 meses, os preços aplicados aos itens consumidos nos diferentes domicílios sofreram variações absolutas e relativas. Assim, os valores levantados foram valorados a preços de referência do dia 15 de janeiro de 2009. A data arbitrária está exposta nas notas técnicas da POF, mas sem uma justificativa para sua escolha.

A unidade amostral da pesquisa é o domicílio, definido como:

(...) a moradia estruturalmente separada e independente, constituída por um ou mais cômodos, sendo que as condições de separação e independência de acesso devem ser satisfeitas.

A condição de separação é atendida quando o local de moradia é limitado por paredes, muros, cercas e outros, quando é coberto por um teto e permite que seus moradores se isolem, arcando com parte ou todas as suas despesas de alimentação ou moradia.

A independência é atendida quando o local de moradia tem acesso direto, permitindo que seus moradores possam entrar e sair sem passar por local de moradia de outras pessoas.

(IBGE, 2010, p. 22)

Na POF, foram investigados apenas os domicílios particulares permanentes – os que servem total ou parcialmente de habitação para uma ou mais pessoas ligadas por laços de parentesco. A unidade de consumo é a família em questão que habita o domicílio. Cabe ressaltar que os domicílios podem ter mais de uma família. Por fim, pessoa moradora é aquela que não estava afastada do domicílio por período superior a 12 meses.

Para a seleção da amostra de domicílios, foram utilizados os setores censitários, unidades geográficas compostas por uma série de domicílios, já empregadas nos processos de produção dos Censos Demográficos. Foi utilizada a tecnologia de uma amostra mestra dos setores censitários. Ela consiste na seleção probabilística das áreas, de modo que seja possível selecionar subamostras para novas pesquisas ou temas específicos. A POF é, então, uma subamostra da amostra mestra.

A amostra mestra, que determinou a subamostra da POF, apresentou o seguinte esquema de estratificação:

- 1 - Divisão administrativa: municípios das capitais, regiões metropolitanas e Regiões Integradas de Desenvolvimento - RIDES;
 - 2 - Espacial/geográfica: áreas de ponderação, municípios;
 - 3 - Situação dos setores censitários: urbana ou rural; e
 - 4 - Estatística: a partir da variável renda do responsável, obtida no Censo Demográfico 2000.
- (IBGE, 2010, p. 35)

A publicação, então, classifica o tipo de procedimento amostral utilizado:

(...) para a POF 2008-2009, adotou-se um plano amostral denominado como conglomerado em dois estágios, com estratificações geográfica e estatística das unidades primárias de amostragem que correspondem aos setores da base geográfica do Censo Demográfico 2000, a partir da estrutura oferecida pela amostra mestra desenhada pelo IBGE para o Sistema Integrado de Pesquisas Domiciliares em construção. Os setores foram selecionados por amostragem com probabilidade proporcional ao número de domicílios existentes no setor, dentro de cada estrato final, compondo a amostra

mestra. A subamostra de setores para a POF 2008-2009 foi selecionada por amostragem aleatória simples em cada estrato. No plano adotado, as unidades secundárias de amostragem foram os domicílios particulares permanentes, que foram selecionados por amostragem aleatória simples sem reposição, dentro de cada um dos setores selecionados. Em seguida ao processo de seleção de setores e domicílios, os setores são distribuídos ao longo dos quatro trimestres da pesquisa, garantindo que em todos os trimestres, os estratos geográfico e socioeconômico estejam representados através dos domicílios selecionados. (IBGE, 2010, p. 36)

A quantidade de domicílios permanentes selecionados foi proporcional ao número total de domicílios permanentes no estrato. Cada estrato teve, no mínimo, três setores. O tamanho final da amostra foi de 4.696 setores com número esperado de domicílios com entrevista de 59.548. O número final de domicílios entrevistados foi de 55.970, distribuídos por unidades da federação conforme Anexo 2. Para nosso recorte específico, escolhemos dentre os 55.970 aqueles que possuíam algum consumo de gás – a saber, 49.835 domicílios, que compõem a amostra de interesse neste trabalho. Destes, foram eliminados 368 domicílios que possuíam consumo per capita de gás anual superior a 130 kg (dez botijões), o que foi considerado pelo pesquisador como incompatível com uso doméstico.

2.2. Instrumentos de coleta

Para a obtenção das respostas nos domicílios, o IBGE adota uma série de instrumentos de coleta. Neste trabalho são utilizados os seguintes:

- *POF 1 - Questionário de Características do Domicílio e dos Moradores:* informações gerais sobre o domicílio, como tipo, número de cômodos, forma de abastecimento de água, tipo de esgotamento sanitário, origem da energia elétrica, condição de ocupação, entre outros. Características dos moradores também são perguntadas.

- *POF 2 - Questionário de Aquisição Coletiva*: aquisições monetárias ou não de produtos de uso da família, além das despesas com energia, água, telefone, gás, aluguel, condomínio, entre outros. Também são perguntadas a aquisição de bens duráveis
- *POF 3 - Caderneta de Aquisição Coletiva*: produtos comprados frequentemente, como alimentos, bebidas, artigos de higiene.

2.3. Variáveis escolhidas

As variáveis utilizadas na análise foram selecionadas pelo pesquisador, que também implementou as rotinas em *R* para a construção do banco de dados específico para esta análise. Em parceria com o pesquisador, foi realizada a análise de consistência do banco de dados, com correções pontuais. As variáveis selecionadas foram:

- *Estrato geográfico*: diz respeito à área em que o domicílio se encontra. Categorias: urbana ou rural.
- *Peso amostral*: traz a informação sobre quantos domicílios cada observação (domicílio observado) representa. Valor numérico não necessariamente inteiro.
- *Quantidade de moradores*: número de moradores em cada domicílio. Valor inteiro.
- *Quantidade de famílias*: número de famílias no domicílio. Valor inteiro.
- *Tipo de domicílio*: identifica o tipo de domicílio. Categorias: casa, apartamento e cômodo.
- *Água canalizada*: identifica se existe água canalizada para pelo menos um cômodo do domicílio. Categorias: sim e não.

- *Rede geral de energia:* identifica se o domicílio possui energia elétrica proveniente de rede geral (com ou sem medidor ou relógio que registre o consumo) e cuja distribuição e comercialização são feitas por empresa concessionária de serviços públicos (permissionária). Categorias: sim e não.
- *Aquecimento de água por energia elétrica:* identifica se a energia elétrica é utilizada como fonte de aquecimento de água no domicílio. Exemplo: chuveiros elétricos. Categorias: sim e não.
- *Aquecimento de água por gás:* identifica se o gás, encanado ou de botijão, é utilizado como fonte de aquecimento de água no domicílio. Categorias: sim e não.
- *Aquecimento de água por energia solar:* identifica se a energia solar é utilizada como fonte de aquecimento de água no domicílio. Categorias: sim e não.
- *Aquecimento de água por lenha ou carvão:* identifica se lenha ou carvão são utilizados como fonte de aquecimento de água no domicílio. Categorias: sim e não.
- *Fogão a gás:* identifica se o gás, encanado ou de botijão, é utilizado como combustível no(s) fogão(ões) do domicílio. Categorias: sim e não.
- *Fogão a lenha:* identifica se a lenha é utilizada como combustível no(s) fogão(ões) do domicílio. Categorias: sim e não.
- *Fogão a carvão:* identifica se o carvão é utilizado como combustível no(s) fogão(ões) do domicílio. Categorias: sim e não.
- *Fogão a energia elétrica:* identifica se a energia elétrica é utilizada como combustível no(s) fogão(ões) do domicílio. Categorias: sim e não.
- *Forno elétrico:* identifica se o domicílio possui forno elétrico. Categorias: sim e não.

- *Grill*: identifica se o domicílio possui *grill*. Categorias: sim e não.
- *Micro-ondas*: identifica se o domicílio possui micro-ondas. Categorias: sim e não.
- *Quantidade de alimento anual*: construída pelos pesquisadores do IEE, identifica a quantidade de alimento (em kg), entre os 20 principais itens de cocção, consumida em um domicílio no período de um ano. Valor numérico.
- *Renda per capita mensal*: identifica a renda per capita (em reais) de cada morador do domicílio.
- *Região*: identifica a grande região a que pertence o domicílio. Texto.
- *Consumo de gás per capita*: identifica a quantidade de gás per capita (em kg) no domicílio. Valor numérico.
- *Alimento per capita*: identifica a quantidade per capita de alimento (em kg), entre os 20 principais itens de cocção, consumida em um ano. Valor numérico.

Pelas variáveis aqui elencadas, é possível observar que um primeiro grupo delas busca caracterizar o domicílio, quanto, por exemplo, à existência de energia elétrica ou existência de água canalizada. Um segundo grupo trata do aquecimento de água. O terceiro grupo trata do tipo de fogão utilizado para a cocção e o energético envolvido. Já um quarto grupo diz respeito a bens que podem estar associados a maior ou menor uso de gás, como existência de fogão, forno elétrico, *grill* ou micro-ondas. Por fim, são tratadas as variáveis como a quantidade de alimento ou a renda per capita.

3. Análise descritiva

Nesta etapa, construímos as estatísticas descritivas para o consumo de GLP do Brasil e para os consumos regionais, utilizando técnicas disponíveis em Bussab e Morettin (2013). Nessa análise exploratória, foram considerados os pesos amostrais relacionados a cada domicílio fornecidos pela Pesquisa de Orçamentos Familiares (POF) 2008-2009, do Instituto Brasileiro de Geografia e Estatística (IBGE). Assim, os gráficos se referem a estimativas para a população de domicílios do país.

Como o objetivo deste trabalho é construir e analisar o perfil de consumo residencial de gás liquefeito de petróleo (GLP), identificando variáveis relevantes que expliquem o comportamento do gasto domiciliar, foram retirados do banco de dados os domicílios cujos consumos per capita de gás excediam 130 kg por ano – valor equivalente a dez botijões de gás por pessoa no ano.

Adotou-se como variável resposta o consumo de gás anual per capita no domicílio, cujo histograma, já ponderado pelo peso relativo de cada domicílio, é apresentado na Figura 3.

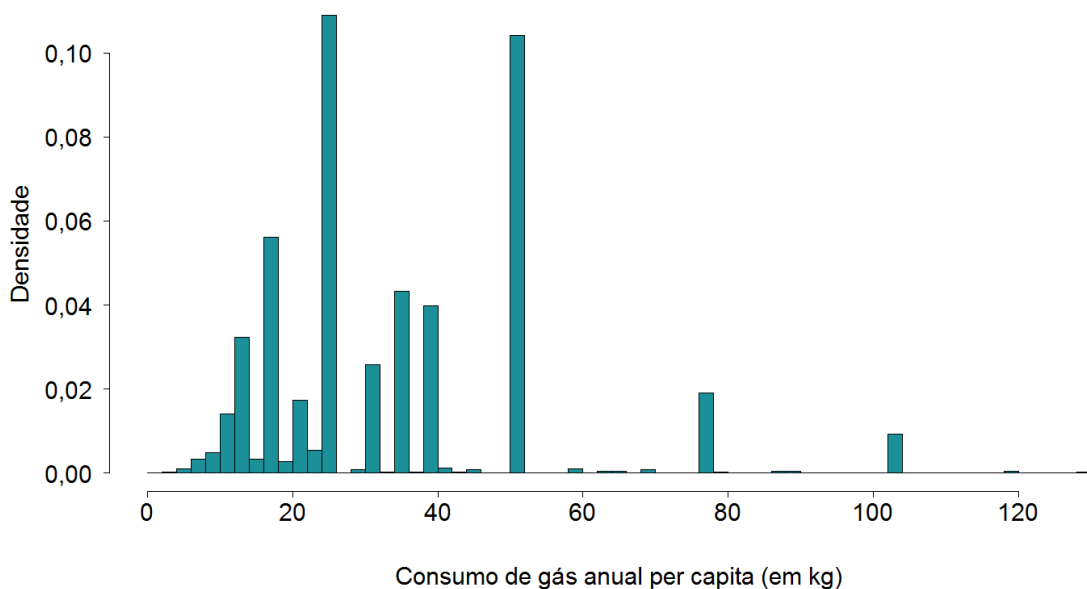


Figura 3 – Consumo de gás anual per capita (em kg) por domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

As medidas de posição e dispersão são apresentadas a seguir, na Tabela 1.

Tabela 1 – Medidas-resumo para a variável de consumo de gás anual per capita (em kg) por domicílio

Número de domicílios na amostra	49.467
Número de domicílios representados pela amostra	49.800.060
Mínimo (em kg)	1,6
Máximo (em kg)	130,0
Amplitude (em kg)	128,4
Mediana (em kg)	28,0
Média populacional estimada (em kg)	34,6
Variância populacional estimada (em kg²)	373,0
Desvio-padrão populacional estimado (em kg)	19,3
Coeficiente de variação	0,56

Fonte: POF 2008-2009, IBGE. Elaboração própria

O comportamento da variável resposta aparentemente não condiz com uma densidade específica bem comportada e conhecida. No entanto, o tamanho amostral – 49.467 domicílios – permite aproximações para a etapa de modelagem.

São expostos, então, os *boxplots* e tabelas com medidas-resumo para a variável consumo de gás anual per capita, segundo as variáveis explicativas, sempre levando em conta o peso domiciliar. A primeira delas é a região do Brasil (Figura 4).

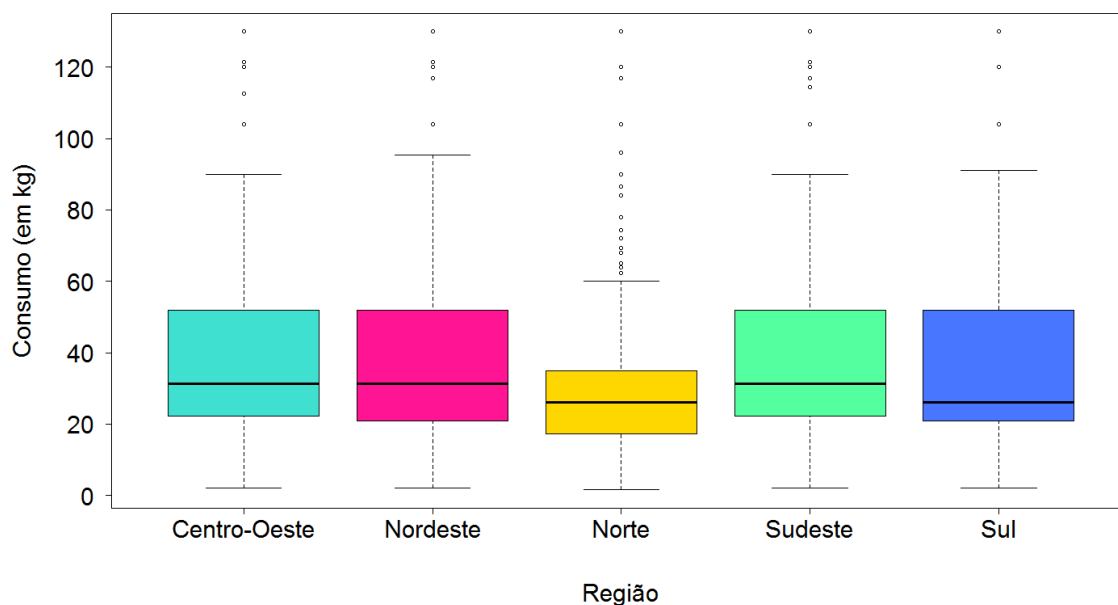


Figura 4 – *Boxplots* da variável consumo de gás anual domiciliar per capita, por região

Fonte: POF 2008-2009, IBGE. Elaboração própria

Verifica-se que as regiões Norte e Sul aparentam ter consumos medianos menores que os das demais regiões. No caso da Norte, notam-se pontos aberrantes a partir de consumos de cerca de 75 kg, devido ao maior número de consumos baixos per capita.

As medidas-resumo por região são apresentadas a seguir (Tabela 2):

Tabela 2 – Medidas-resumo do consumo de gás anual domiciliar per capita (em kg), segundo região

	Centro-Oeste	Nordeste	Norte	Sudeste	Sul
Número de domicílios na amostra	7.656	16.619	6.756	12.576	5.860
Número de domicílios representados pela amostra	3.998.909	12.779.995	3.513.664	21.844.944	7.662.549
Mínimo (em kg)	2	2	1,6	2	2
Máximo (em kg)	130,0	130,0	130,0	130,0	130,0
Amplitude (em kg)	128,0	128,0	128,4	128,0	128,0
Mediana (em kg)	31,2	31,2	26,0	31,2	26,0
Média (em kg)	35,7	34,4	29,5	35,2	35,0
Variância (emkg²)	380,4	384,3	326,3	359,9	395,1
Desvio-padrão (em kg)	19,5	19,6	18,1	19,0	19,9
Coeficiente de variação	0,55	0,57	0,61	0,54	0,57

Fonte: POF 2008-2009, IBGE. Elaboração própria

Um terceiro recorte de interesse é o estrato geográfico, isto é, se o domicílio pertence à área urbana ou rural. Na Figura 5, pode-se observar que o consumo nos domicílios de áreas rurais é mais assimétrico, com concentração em baixo consumo per capita.

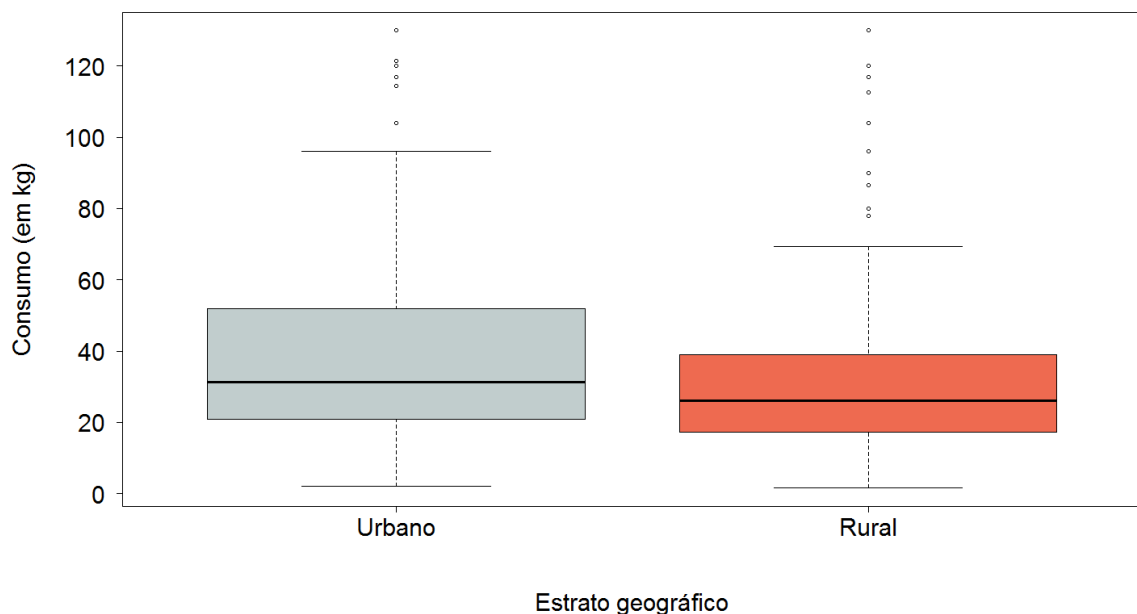


Figura 5 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo estrato geográfico

Fonte: POF 2008-2009, IBGE. Elaboração própria

As medidas-resumo evidenciam menores médias e medianas para os domicílios de área rural, conforme apresentado na Tabela 3.

Tabela 3 – Medidas-resumo do consumo de gás anual domiciliar per capita (em kg), segundo estrato geográfico

	Urbano	Rural
Número de domicílios na amostra	39.040	10.427
Número de domicílios representados pela amostra	42.809.836	6.990.225
Mínimo (em kg)	2,0	1,6
Máximo (em kg)	130,0	130,0
Amplitude (em kg)	128,0	128,4
Mediana (em kg)	31,2	26,0
Média (em kg)	35,1	31,7
Variância (em kg²)	367,0	400,0
Desvio-padrão (em kg)	19,2	20,0
Coeficiente de variação	0,55	0,63

Fonte: POF 2008-2009, IBGE. Elaboração própria

A fim de verificar se há diferenças entre o comportamento de consumo de gás anual domiciliar per capita entre as regiões, segundo estrato geográfico, foram construídos os *boxplots* apresentados na Figura 6.

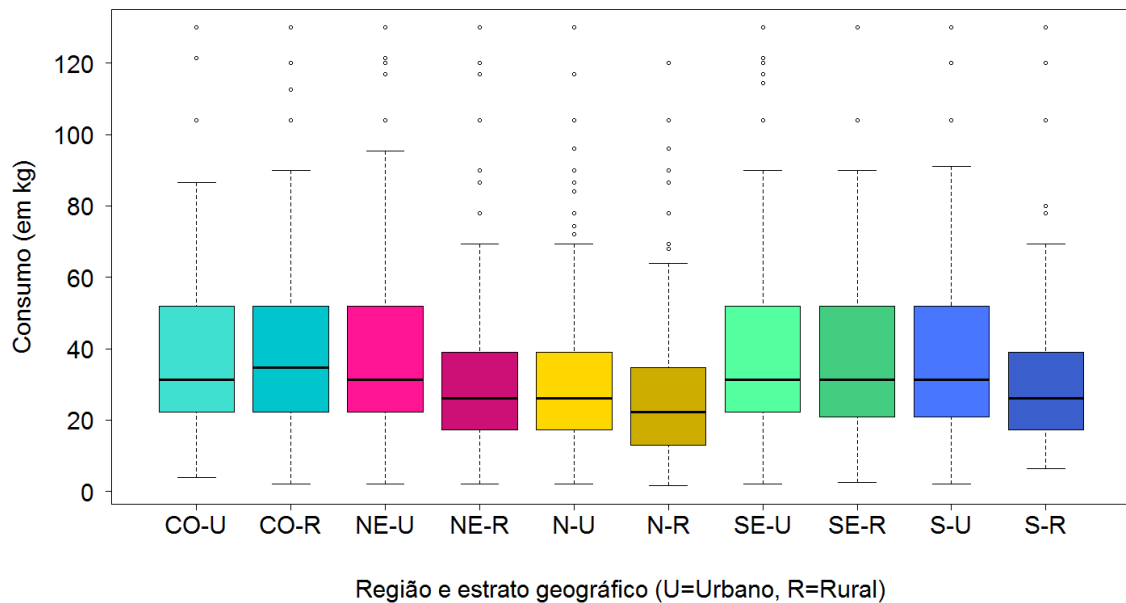


Figura 6 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo região e estrato geográfico

Fonte: POF 2008-2009, IBGE. Elaboração própria

A próxima variável explicativa disponível no banco de dados diz respeito à quantidade de famílias por domicílio. O número de famílias mais frequente é de apenas uma. No entanto, há até seis famílias em um mesmo domicílio (uma ocorrência).

Nesta análise, serão consideradas duas categorias de quantidade de famílias por domicílio: uma família e duas famílias ou mais. Da passagem da primeira categoria para a segunda, há diminuição do consumo de gás anual per capita, o que pode ser observado nos *boxplots* a seguir (Figura 7). Este fato indica uma economia de escala, ou seja, os domicílios com mais de uma família repartem o consumo entre mais pessoas. As medidas-resumo são apresentadas na Tabela 4.

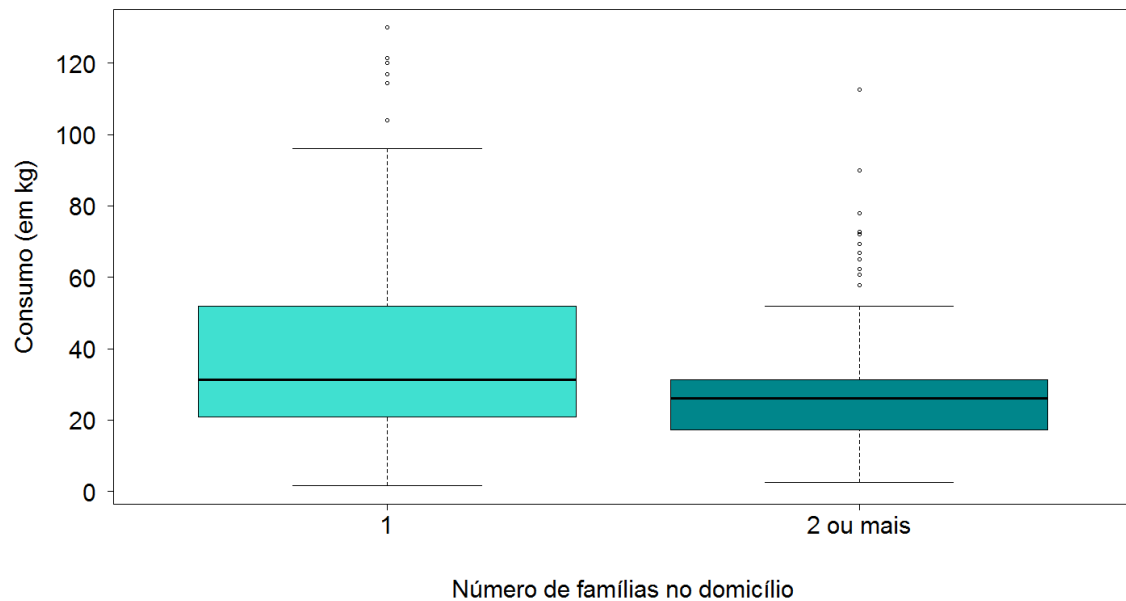


Figura 7 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo número de famílias por domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

Tabela 4 – Medidas-resumo do consumo de gás anual domiciliar per capita (em kg), segundo número de famílias por domicílio

	1 família	2 ou mais famílias
Número de domicílios na amostra	46.652	2.815
Número de domicílios representados pela amostra	46.989.000	2.811.060
Mínimo (em kg)	1,6	2,5
Máximo (em kg)	130,0	112,7
Amplitude (em kg)	128,4	110,2
Mediana (em kg)	31,2	26,0
Média (em kg)	35,2	24,5
Variância (em kg²)	380,9	133,3
Desvio-padrão (em kg)	19,5	11,5
Coefficiente de variação	0,55	0,47

Fonte: POF 2008-2009, IBGE. Elaboração própria

Quanto ao tipo de domicílio – se é casa, apartamento ou cômodo – parecem existir diferenças na mediana do consumo de gás per capita, principalmente no sentido de menor consumo mediano por apartamentos (Figura 8). A Tabela 5 oferece as medidas-resumo para esse cruzamento.

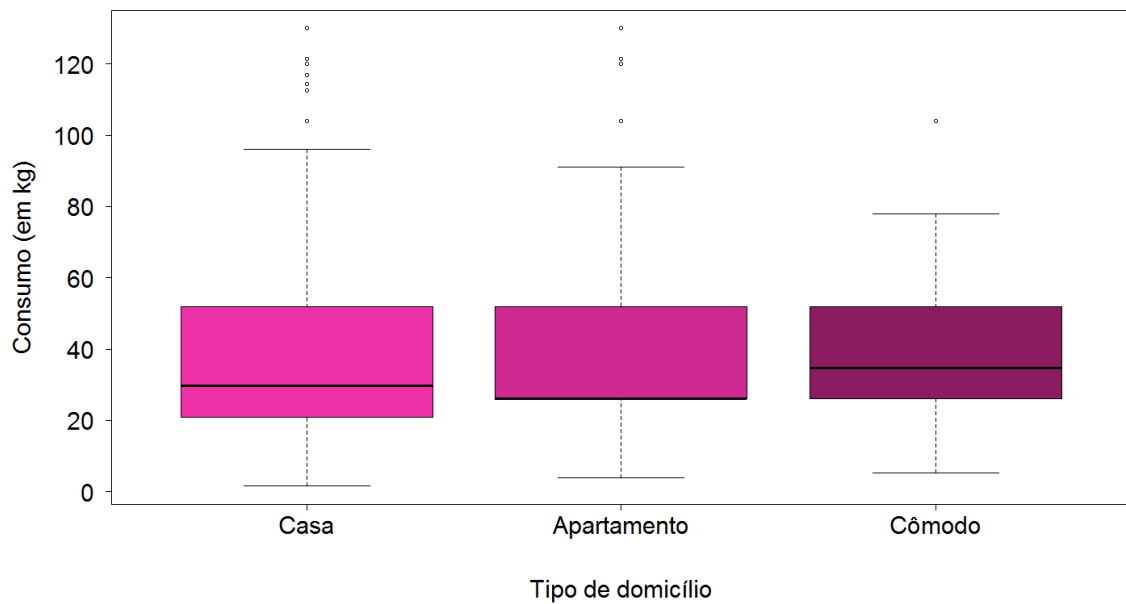


Figura 8 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo tipo de domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

Tabela 5 – Medidas-resumo do consumo de gás anual domiciliar per capita (em kg), segundo tipo de domicílio

	Casa	Apartamento	Cômodo
Número de domicílios na amostra	46.546	2.563	358
Número de domicílios representados pela amostra	46.077.974	3.409.353	312.733
Mínimo (em kg)	1,6	4,0	5,2
Máximo (em kg)	130,0	130,0	104,0
Amplitude (em kg)	128,4	126,0	98,8
Mediana (em kg)	29,7	26,0	34,7
Média (em kg)	34,5	36,1	39,4
Variância (em kg²)	371,3	386,1	424,0
Desvio-padrão (em kg)	19,3	19,6	20,6
Coeficiente de variação	0,56	0,54	0,52

Fonte: POF 2008-2009, IBGE. Elaboração própria

É apresentada, então, a variável que caracteriza a rede de água do domicílio. A existência de água canalizada no domicílio parece estar associada ao consumo de gás, ou seja, quando há água canalizada, há consumo mediano maior (Figura 9 e Tabela 6).

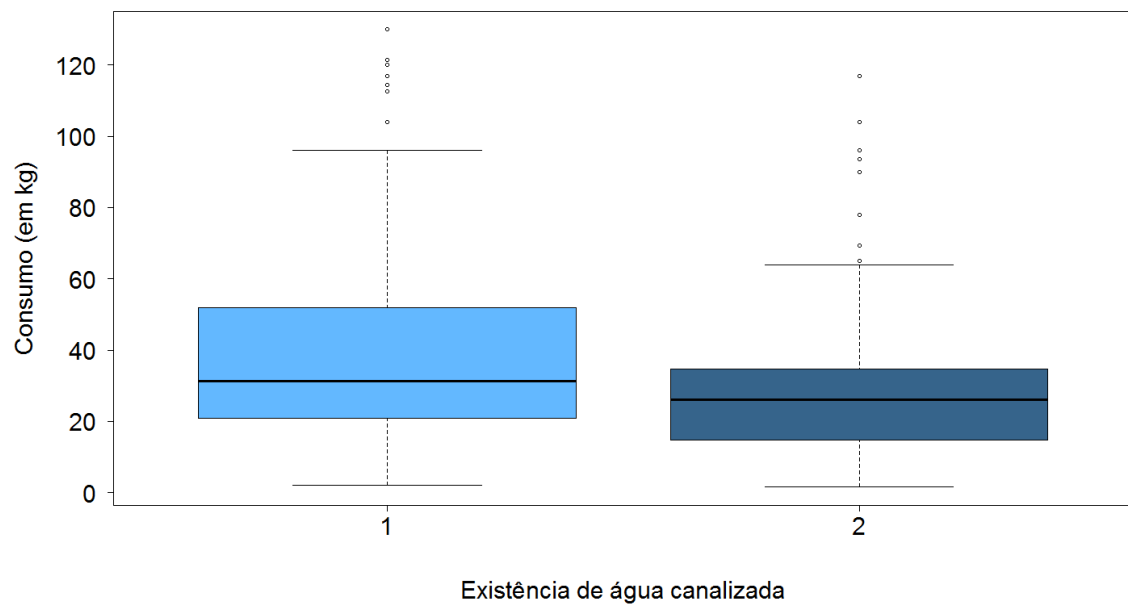


Figura 9 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de água canalizada no domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

Tabela 6 – Medidas-resumo do consumo de gás anual domiciliar per capita (em kg), segundo a existência de água canalizada no domicílio

	Tem água canalizada	Não tem água canalizada
Número de domicílios na amostra	45.682	3.785
Número de domicílios representados pela amostra	47.142.868	2.657.192
Mínimo (em kg)	2,0	1,6
Máximo (em kg)	130,0	117,0
Amplitude (em kg)	128,0	115,4
Mediana (em kg)	31,2	26,0
Média (em kg)	34,9	28,4
Variância (em kg²)	370,3	380,2
Desvio-padrão (em kg)	19,2	19,5
Coefficiente de variação	0,55	0,69

Fonte: POF 2008-2009, IBGE. Elaboração própria

Quanto à presença de rede de energia elétrica, a população de domicílios apresenta consumo per capita truncado mediano mais elevado entre os que são atendidos por esse serviço. Os domicílios sem rede de energia elétrica apresentam consumo mediano menor, como pode ser visualizado na Figura 10 e nas medidas-resumo da Tabela 7.

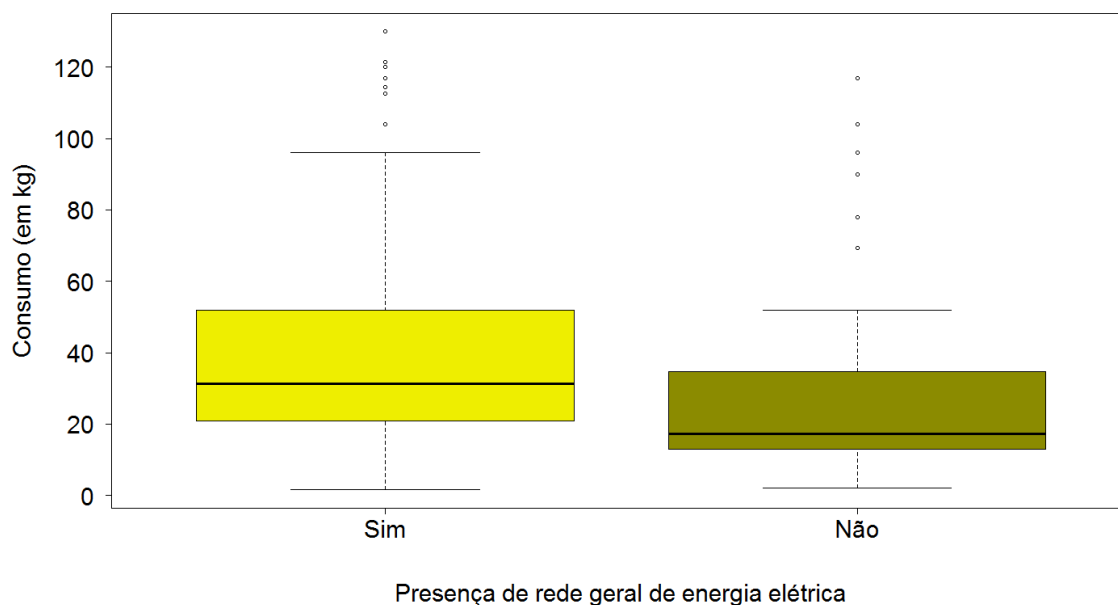


Figura 10 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de rede geral de energia elétrica no domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

Tabela 7 – Medidas-resumo do consumo de gás anual domiciliar per capita (em kg), segundo a existência de rede geral de energia elétrica no domicílio

	Tem rede geral de energia elétrica	Não tem rede geral de energia elétrica
Número de domicílios na amostra	48.591	876
Número de domicílios representados pela amostra	49.276.196	523.865
Mínimo (em kg)	1,6	2,0
Máximo (em kg)	130,0	117,0
Amplitude (em kg)	128,4	115,0
Mediana (em kg)	31,2	17,3
Média (em kg)	34,7	26,1
Variância (em kg²)	371,6	434,0
Desvio-padrão (em kg)	19,3	20,8
Coeficiente de variação	0,56	0,80

Fonte: POF 2008-2009, IBGE. Elaboração própria

Os próximos *boxplots* (Figura 11) dizem respeito ao tipo de aquecimento de água de que os domicílios dispõem. São tratadas a presença ou ausência de aquecimento de água por energia elétrica, gás, energia solar e lenha. Nos domicílios que possuem os três primeiros combustíveis para aquecimento de água, foi verificado consumo mediano maior de gás per capita. Já entre os que possuem aquecimento de água por lenha, há menor consumo mediano estimado.

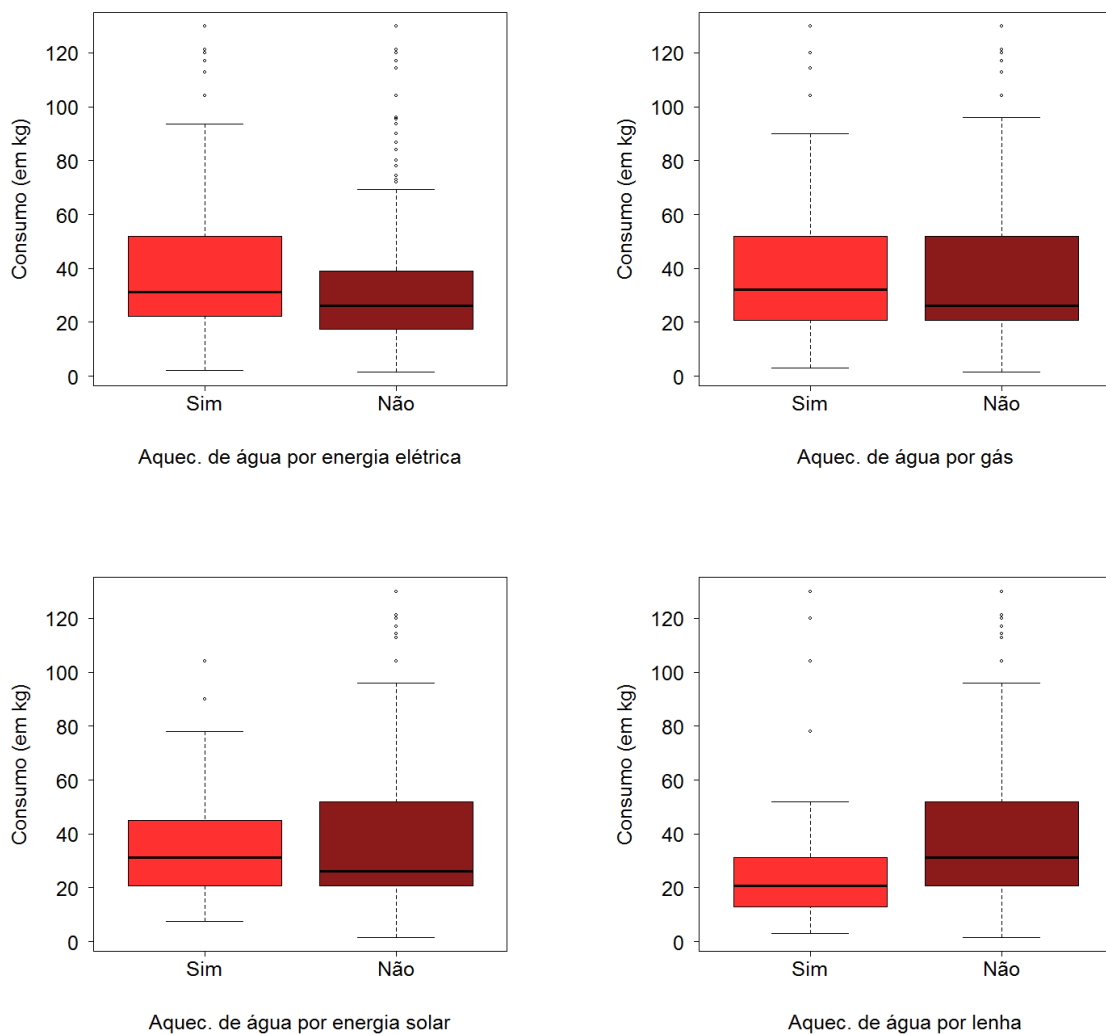


Figura 11 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de aquecimento de água por energia elétrica, gás, energia solar e lenha

Fonte: POF 2008-2009, IBGE. Elaboração própria

A Figura 12, por sua vez, traz os *boxplots* do consumo de gás anual per capita segundo a existência ou não dos diferentes tipos de fogão no domicílio: fogão

a gás, a lenha, a carvão ou a energia elétrica. Aparentemente, apenas os domicílios com fogão a gás apresentam consumo mediano de gás per capita mais elevado.

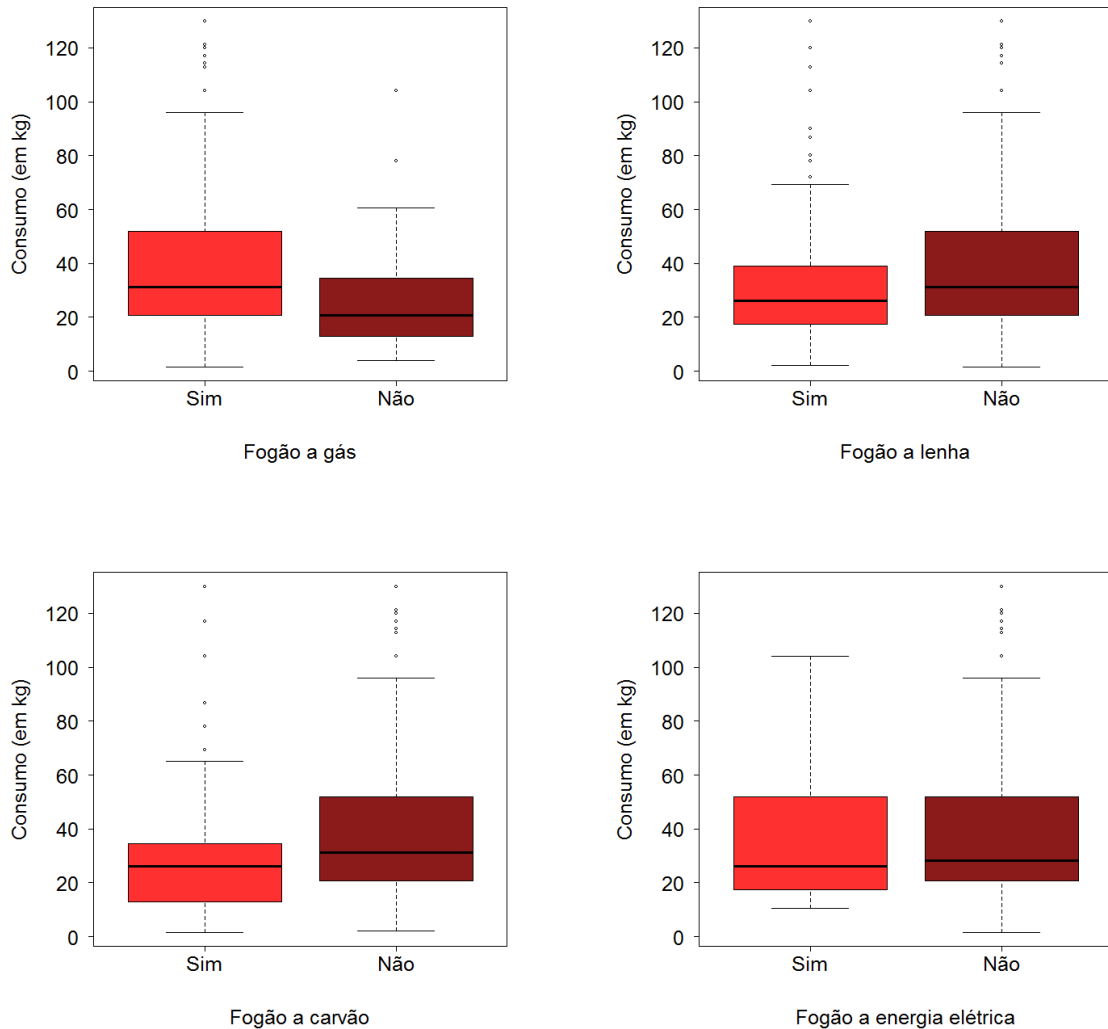


Figura 12 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de fogão a gás, a lenha, a carvão e a energia elétrica

Fonte: POF 2008-2009, IBGE. Elaboração própria

O grupo seguinte de variáveis diz respeito ao número de equipamentos de cocção alternativos disponíveis no domicílio: forno elétrico, *grill* e micro-ondas. Nos

três casos, o consumo mediano de gás per capita foi próximo entre os domicílios que dispunham do equipamento e os que não dispunham (Figura 13).

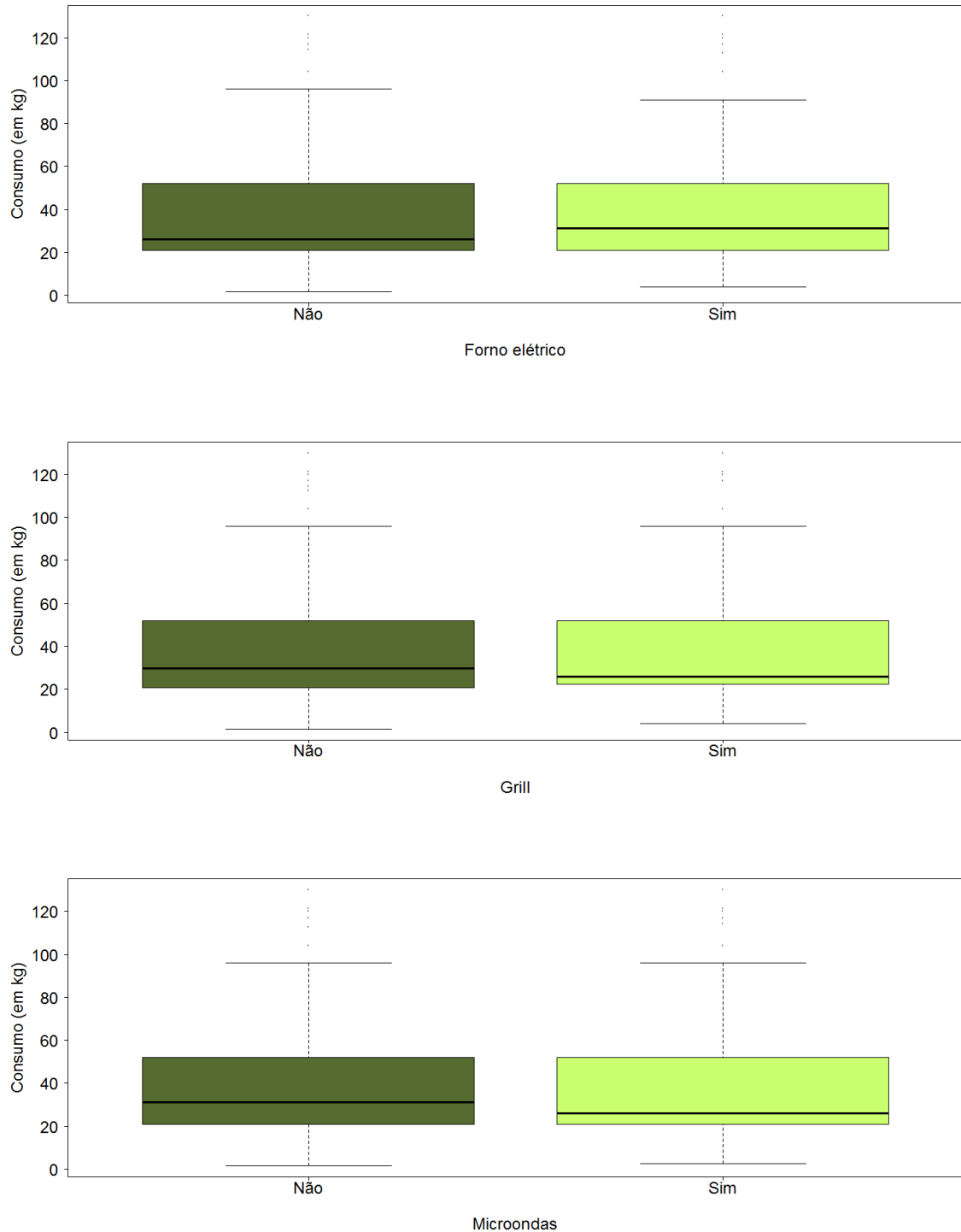


Figura 13 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de forno elétrico, *grill* e micro-ondas

Fonte: POF 2008-2009, IBGE. Elaboração própria

As próximas duas variáveis no banco de dados dizem respeito à renda per capita domiciliar mensal (Figura 14) e ao consumo anual de alimentos por pessoa – ambas com características contínuas. No entanto, dados pontos muito discrepantes, as duas receberão tratamento categorizado. O histograma para a variável renda per capita anual é apresentado a seguir. Como esperado para renda no Brasil, há alta concentração de observações em rendas por pessoa mais baixas.

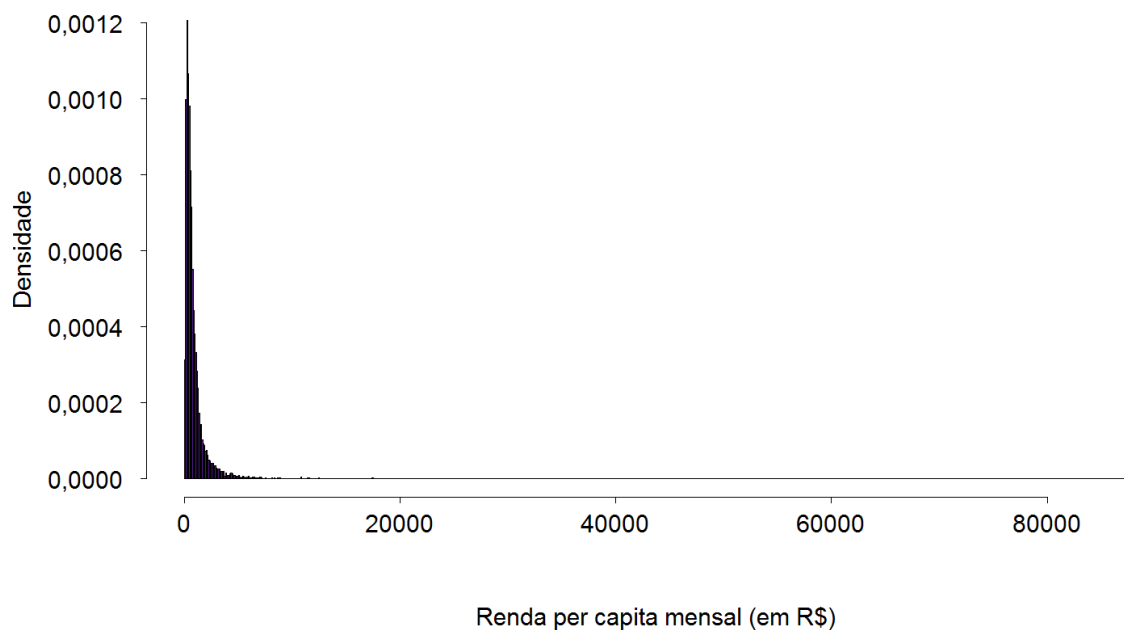


Figura 14 – Histograma da renda domiciliar per capita mensal (em R\$)

Fonte: POF 2008-2009, IBGE. Elaboração própria

Houve quatro domicílios com renda per capita mensal nula. Já o valor máximo observado foi de R\$ 87.430,75. As categorias utilizadas pelo IBGE para a renda no ano de 2009 são apresentadas na Tabela 8.

Tabela 8 – Categorias de renda utilizadas pelo IBGE na POF 2008-2009 para o rendimento familiar

Reais mensais (R\$)	Salários mínimos
Até 830,00 (inclui sem rendimento)	Até 2 (inclui sem rendimento)
Mais de 830,00 a 1.245,00	Mais de 2 a 3
Mais de 1.245,00 a 2.490,00	Mais de 3 a 6
Mais de 2.490,00 a 4.150,00	Mais de 6 a 10
Mais de 4.150,00 a 6.225,00	Mais de 10 a 15
Mais de 6.225,00 a 10.375,00	Mais de 15 a 25
Mais de 10.375,00	Mais de 25

Fonte: POF 2008-2009, IBGE

Para este trabalho, foram formuladas as seguintes categorias per capita, baseadas no padrão do IBGE (Tabela 9). O gráfico com o número de domicílios por categoria de renda per capita mensal pode ser observado na Figura 15.

Tabela 9 – Categorias de renda per capita mensal utilizadas neste trabalho

Reais mensais (R\$)	Salários mínimos
Até 207,50 (inclui sem rendimento)	Até 0,5
Mais de 207,50 a 311,25	Mais de 0,5, a 0,75
Mais de 311,25 a 622,50	Mais de 0,75 a 1,5
Mais de 622,50 a 1.037,50	Mais de 1,5 a 2,5
Mais de 1.037,50 a 1.556,25	Mais de 2,5 a 3,75
Mais de 1.556,25	Mais de 3,75

Fonte: Elaboração própria

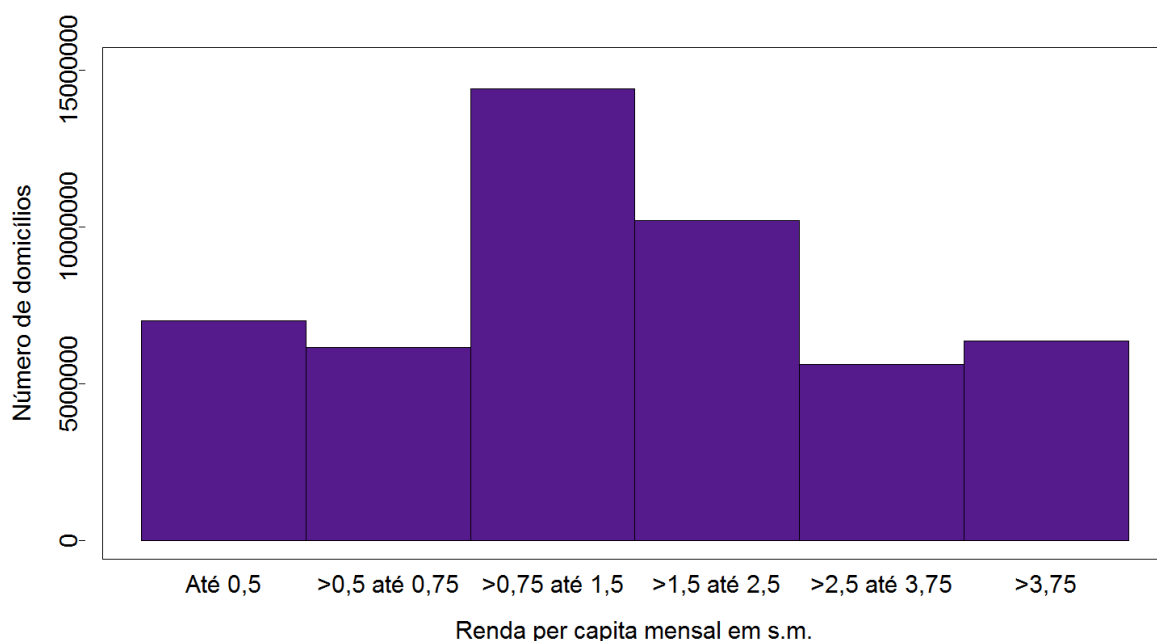


Figura 15 – Gráfico de barras das categorias de renda per capita mensal categorizadas em salários mínimos para a população estimada de domicílios

Fonte: POF 2008-2009, IBGE. Elaboração própria

A dispersão do consumo de gás anual per capita, segundo as categorias de renda, é apresentada no gráfico a seguir (Figura 16). Pode-se observar que o consumo mediano é crescente, conforme aumentam as categorias de renda, até uma estabilização, nas categorias mais elevadas. As medidas-resumo são apresentadas na Tabela 10.

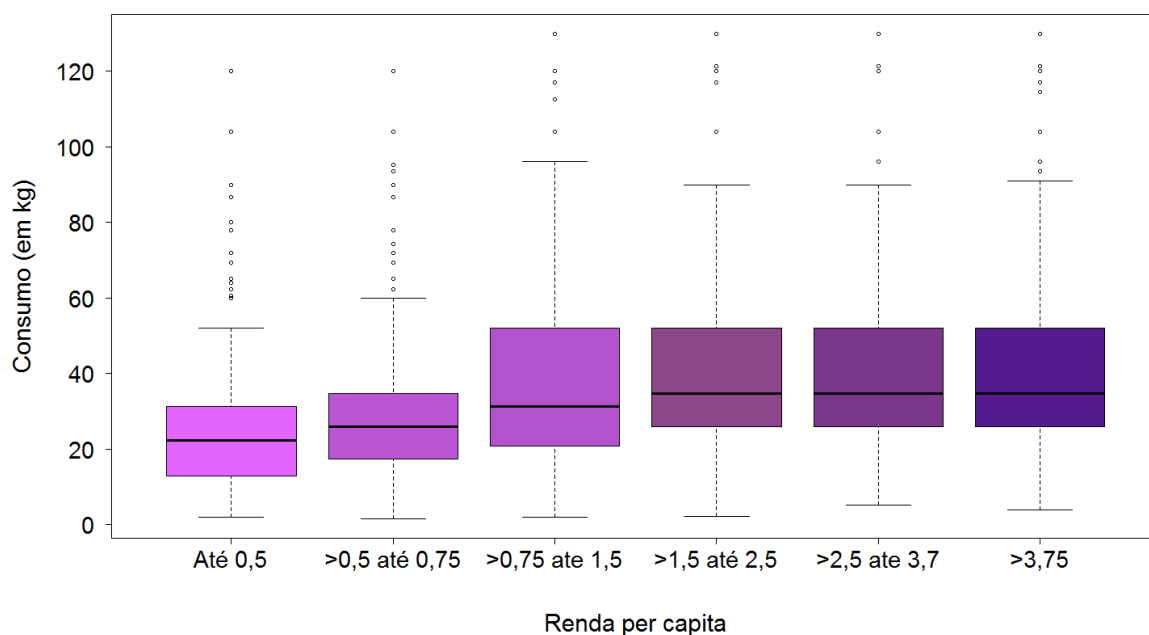


Figura 16 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de renda mensal per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

Tabela 10 – Medidas-resumo do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de renda per capita mensal

	Até 0,5	Mais de 0,5, a 0,75	Mais de 0,75 a 1,5	Mais de 1,5 a 2,5	Mais de 2,5 a 3,75	Mais de 3,75
Número de domicílios na amostra	8.799	6.940	15.137	9.198	4.527	4.866
Número de domicílios representados pela amostra	7.011.683	6.171.271	14.409.470	10.219.429	5.616.249	6.371.958
Mínimo (em kg)	2,0	1,6	2,0	2,2	5,2	4,0
Máximo (em kg)	120,0	120,0	130,0	130,0	130,0	130,0
Amplitude (em kg)	118,0	118,4	128,0	127,8	124,8	126,0
Mediana (em kg)	22,3	26,0	31,2	34,7	34,7	34,7
Média (em kg)	24,6	29,0	34,6	38,1	39,3	41,4
Variância (em kg²)	179,9	223,8	329,1	410,9	434,9	488,0
Desvio-padrão (em kg)	13,4	15,0	18,1	20,3	20,9	22,1
Coeficiente de variação	0,55	0,52	0,53	0,53	0,53	0,53

Fonte: POF 2008-2009, IBGE. Elaboração própria

A última variável do banco de dados a ser explorada trata da quantidade de alimento consumida. O pesquisador selecionou os 20 alimentos mais frequentes na POF que possuíam a característica de necessitar de cocção. O histograma da quantidade de alimentos per capita no ano é apresentado na Figura 17.

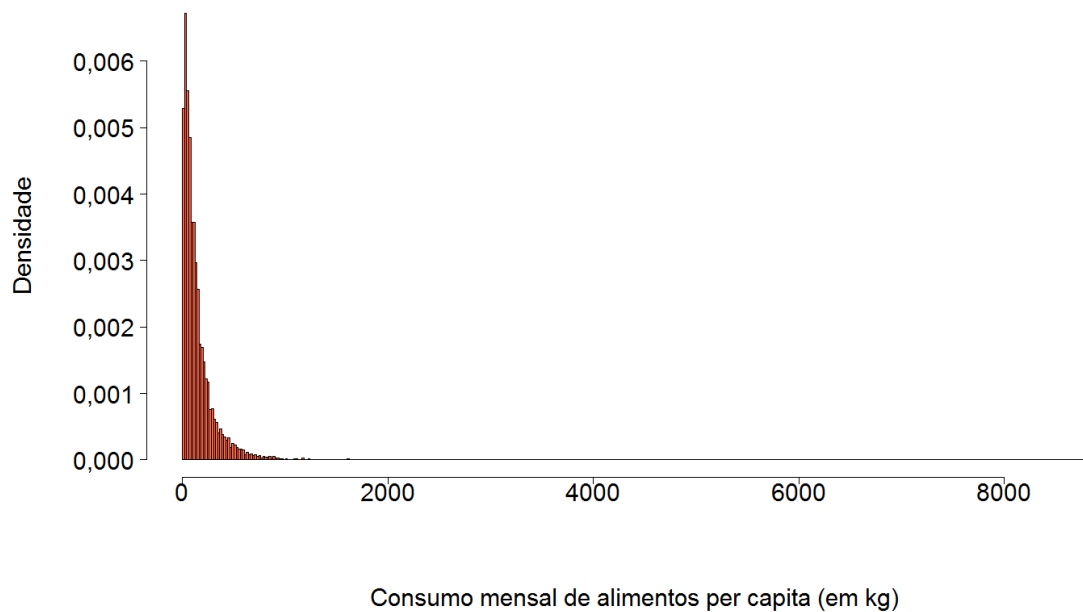


Figura 17 – Histograma do consumo de alimentos domiciliar per capita mensal (em kg)

Fonte: POF 2008-2009, IBGE. Elaboração própria

O consumo alimentar per capita é, assim como a renda, bastante assimétrica e com valores discrepantes de alto consumo. O valor mínimo de consumo anual encontrado foi de 0,18 kg por ano por pessoa, e o maior valor, de 8.805,8 kg por ano por pessoa. Optou-se por definir categorias de quantidade de consumo alimentar por pessoa (Tabela 11). Entre elas, foi incluída a categoria “Sem informação”, para que os dados de domicílios com informações faltantes nesta variável não fossem dispensados dos modelos de regressão propostos.

Tabela 11 – Categorias de consumo de alimento per capita mensal utilizadas neste trabalho

Quantidade de alimento (em kg)	Categoria
Até 50	Consumo baixo
Mais de 50 a 100	Consumo médio baixo
Mais de 100 a 200	Consumo médio alto
Mais 200	Consumo alto
NA	Sem informação

Fonte: Elaboração própria

Os *boxplots* do consumo de gás domiciliar per capita em relação às categorias de consumo de alimentos podem ser observados na Figura 18. Verifica-se consumo mediano de gás crescente, conforme aumenta o consumo de alimentos. As medidas-resumo são apresentadas na sequência (Tabela 12).

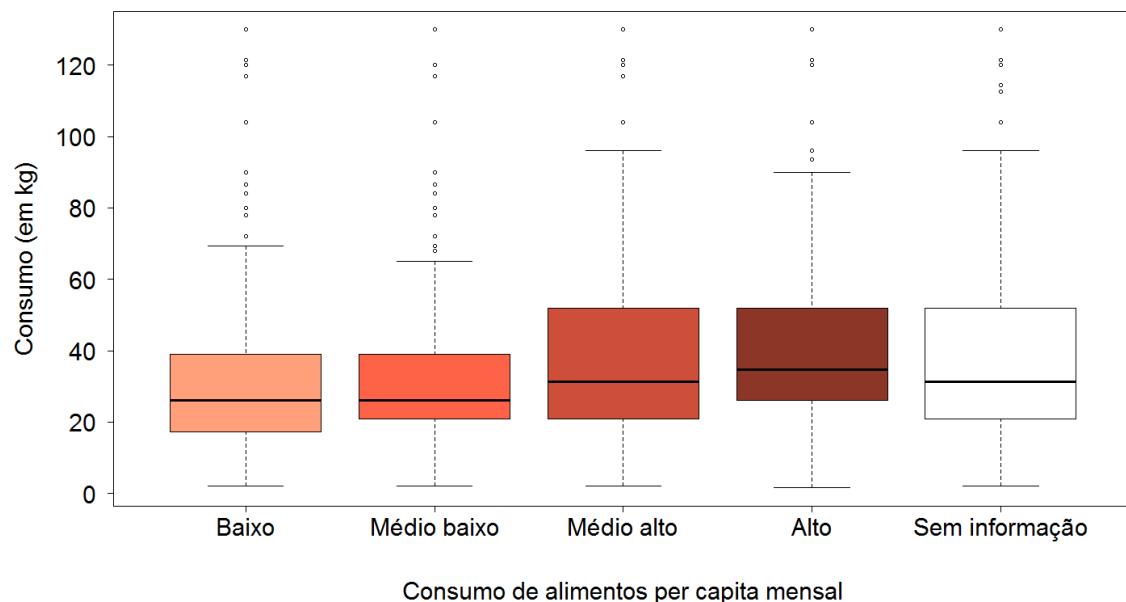


Figura 18 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de consumo de alimento per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

Tabela 12 – Medidas-resumo do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de consumo de alimento per capita mensal

	Baixo	Médio baixo	Médio alto	Alto	Sem informação
Número de domicílios na amostra	10.728	9.222	9.750	8.848	10.919
Número de domicílios representados pela amostra	10.966.759	9.202.429	9.725.142	8.875.676	11.030.054
Mínimo (em kg)	2,1	2,0	2,0	1,6	2,0
Máximo (em kg)	130,0	130,0	130,0	130,0	130,0
Amplitude (em kg)	127,9	128,0	128,0	128,4	128,0
Mediana (em kg)	26,0	26,0	32,0	36,0	32,0
Média (em kg)	29,9	33,0	34,9	40,0	36,1
Variância (em kg²)	271,4	305,8	368,2	477,0	402,3
Desvio-padrão (em kg)	16,5	17,5	19,2	21,8	20,1
Coeficiente de variação	0,55	0,53	0,55	0,55	0,56

Fonte: POF 2008-2009, IBGE. Elaboração própria

Pelo conjunto de gráficos e tabelas aqui apresentado, aparentemente, há variáveis relacionadas ao consumo de gás per capita nos domicílios brasileiros, como a renda ou o nível de consumo alimentar. Os modelos de regressão construídos na análise inferencial vão buscar verificar se essas variáveis são significativas.

Além disso, no Apêndice A estão disponíveis os *boxplots* do consumo de gás anual per capita domiciliar por região, o que permite analisar particularidades que serão alvos dos modelos de regressão linear deste trabalho.

4. Análise inferencial

Para a análise inferencial, optou-se por construir cinco modelos de regressão linear, um para cada grande região brasileira: Centro-Oeste, Nordeste, Norte, Sudeste e Sul. Como o interesse do pesquisador é verificar o comportamento e a significância das variáveis explicativas do consumo de gás per capita, e como a indústria do Gás Liquefeito de Petróleo (GLP) se organiza segundo grande região, a opção pareceu a mais adequada do ponto de vista interpretativo e técnico.

O modelo adotado foi o de regressão linear com log (a função matemática $\ln$) aplicado à variável resposta consumo de gás domiciliar per capita, o que caracteriza o que a econometria chama de modelo semilog (GREENE, 2012, p. 17).

Nos modelos lineares, algumas suposições precisam ser satisfeitas. Especificamente temos, segundo Greene:

- 1) Linearidade: há uma relação linear entre a variável resposta e as variáveis explicativas. Essa suposição é relaxada com a aplicação de log na variável resposta.
- 2) Esperança dos erros, dadas as covariáveis, igual a 0.
- 3) Homocedasticidade e erros não autocorrelacionados: cada erro precisa ter a mesma variância finita σ^2 .
- 4) Posto completo: nenhuma das covariáveis mantém relação linear com as demais, o que é necessário para estimar o modelo.

A suposição de distribuição normal dos erros é descartada por Greene para tamanhos amostrais grandes, que é exatamente o caso desta análise (2012, p.63). É provado que o estimador de mínimos quadrados tem distribuição assintótica normal, o que permite a realização de testes de hipóteses e a construção de intervalos de confiança.

O estimador utilizado é o de mínimos quadrados ponderado pelo peso domiciliar fornecido pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Este peso contabiliza a quantos domicílios cada observação se refere.

O modelo para cada região trabalha com categorias de referência e estima parâmetros para as demais categorias. Se, por exemplo, o estrato geográfico for

significativo para o modelo, a área urbana é tomada como referência e a rural tem seu parâmetro estimado.

As categorias de referência são apresentadas na Tabela 13.

Tabela 13 – Categorias de referência utilizadas na modelagem para cada variável

Variável	Categoria de referência
Estrato geográfico	Área urbana
Quantidade de famílias (em categorias)	Uma família
Tipo de domicílio	Casa
Existência de água canalizada	Sim
Presença de rede geral de energia elétrica	Sim
Presença de aquecimento de água por energia elétrica	Sim
Presença de aquecimento de água por gás	Sim
Presença de aquecimento de água por energia solar	Sim
Presença de aquecimento de água por lenha	Sim
Existência de fogão a gás	Sim
Existência de fogão a lenha	Sim
Existência de fogão a carvão	Sim
Existência de fogão a energia elétrica	Sim
Existência de forno elétrico	Não
Existência de <i>grill</i>	Não
Existência de micro-ondas	Não
Categorias de renda per capita	Até 0,5 s.m.
Categorias de alimento per capita	Baixo consumo per capita

Fonte: Elaboração própria

Por fim, o modelo ajustado tem a seguinte forma:

$$\log(y_i) = \beta_0 + \beta_1 \text{Rural} + \beta_2 \text{Apartamento} + \dots + w_i \text{Erro}_i$$

em que:

y_i corresponde ao consumo per capita de gás no domicílio i .

β_0 corresponde ao intercepto do modelo (consumo médio per capita para os domicílios nas categorias de referência da Tabela 13).

β_1 corresponde ao efeito de passar da zona urbana (referência) à rural no consumo de gás, quando esse efeito é significativo.

β_2 corresponde ao efeito de analisar um apartamento em vez de uma casa.

w_i corresponde ao peso do domicílio fornecido pelo IBGE.

Erro_i corresponde ao erro do modelo relacionado ao domicílio i .

Inicialmente, foram realizados ajustes com todas as variáveis explicativas disponíveis para cada uma das regiões, a saber:

- estrato geográfico,
- quantidade de famílias (em categorias),
- tipo de domicílio,
- existência de água canalizada,
- presença de rede geral de energia elétrica,
- presença de aquecimento de água por energia elétrica,
- presença de aquecimento de água por gás,
- presença de aquecimento de água por energia solar,
- presença de aquecimento de água por lenha,
- existência de fogão a gás,
- existência de fogão a lenha,
- existência de fogão a carvão,
- existência de fogão a energia elétrica,
- existência de forno elétrico,

- existência de *grill*,
- existência de micro-ondas,
- categorias de renda per capita,
- categorias de alimento per capita.

Após o ajuste do modelo de efeitos principais para cada uma das regiões, foi utilizado para a seleção de variáveis o algoritmo de *stepwise*, pelo Critério de Informação de Akaike (AIC), disponível pela função *stepAIC* no R, na biblioteca MASS. Quando permaneceram variáveis significativas a níveis de significância superiores a 10%, elas foram sendo retiradas, uma a uma, segundo ordem decrescente do valor P.

Os resultados por regiões, com as variáveis significativas ao nível de 5%, as estimativas, erros-padrões, intervalos de confiança para as estimativas e análise de resíduos para verificação das suposições, são expostos a seguir para cada um dos modelos ajustados.

4.1. Resultados para a região Centro-Oeste

O modelo selecionado para a região Centro-Oeste contém as seguintes covariáveis, todas significativas ao nível de 5%:

- estrato geográfico,
- quantidade de famílias (em categorias),
- tipo de domicílio,
- existência de aquecimento de água por energia elétrica,
- existência de fogão a lenha,
- existência de *grill*,
- existência de micro-ondas,
- categorias de renda per capita,
- categorias de alimento per capita.

As estimativas e erros padrões são apresentados na Tabela 14.

Tabela 14 – Estimativas, erros padrões, valor P, limites inferior e superior do intervalo de confiança (95%) para a região Centro-Oeste

	Estimativa	Erro padrão	Valor P	Limite inferior	Limite superior
Intercepto	2,928	0,031	<0,001	2,866	2,989
Área rural	0,082	0,020	<0,001	0,042	0,121
Duas ou mais famílias	-0,251	0,031	<0,001	-0,313	-0,189
Apartamento	-0,070	0,025	0,005	-0,119	-0,021
Cômodo	0,162	0,066	0,015	0,031	0,292
Ausência de aquecimento de água por energia elétrica	-0,072	0,017	<0,001	-0,105	-0,039
Ausência de fogão a lenha	0,139	0,025	<0,001	0,091	0,188
Presença de <i>grill</i>	-0,164	0,022	<0,001	-0,207	-0,122
Presença de micro-ondas	-0,095	0,016	<0,001	-0,126	-0,065
Renda maior que 0,5 até 0,75 s.m.	0,104	0,024	<0,001	0,057	0,151
Renda maior que 0,75 até 1,5 s.m.	0,296	0,020	<0,001	0,256	0,335
Renda maior que 1,5 até 2,5 s.m.	0,422	0,022	<0,001	0,379	0,465
Renda maior que 2,5 até 3,75 s.m.	0,464	0,025	<0,001	0,414	0,514
Renda maior que 3,75 s.m.	0,497	0,026	<0,001	0,446	0,548
Consumo de alimento médio baixo	0,102	0,020	<0,001	0,063	0,141
Consumo de alimento médio alto	0,160	0,019	<0,001	0,123	0,197
Consumo de alimento alto	0,222	0,019	<0,001	0,184	0,259
Sem informação sobre consumo de alimento	0,114	0,017	<0,001	0,081	0,148

Fonte: Elaboração própria

Os resíduos padronizados não apresentaram um padrão de distribuição (Figura 17), o que indica que a suposição de homocedasticidade foi satisfeita. Eles se distribuem ao redor da média (0,005), o que pode ser observado na Figura 18. A média da exponencial dos resíduos foi de 1,1.

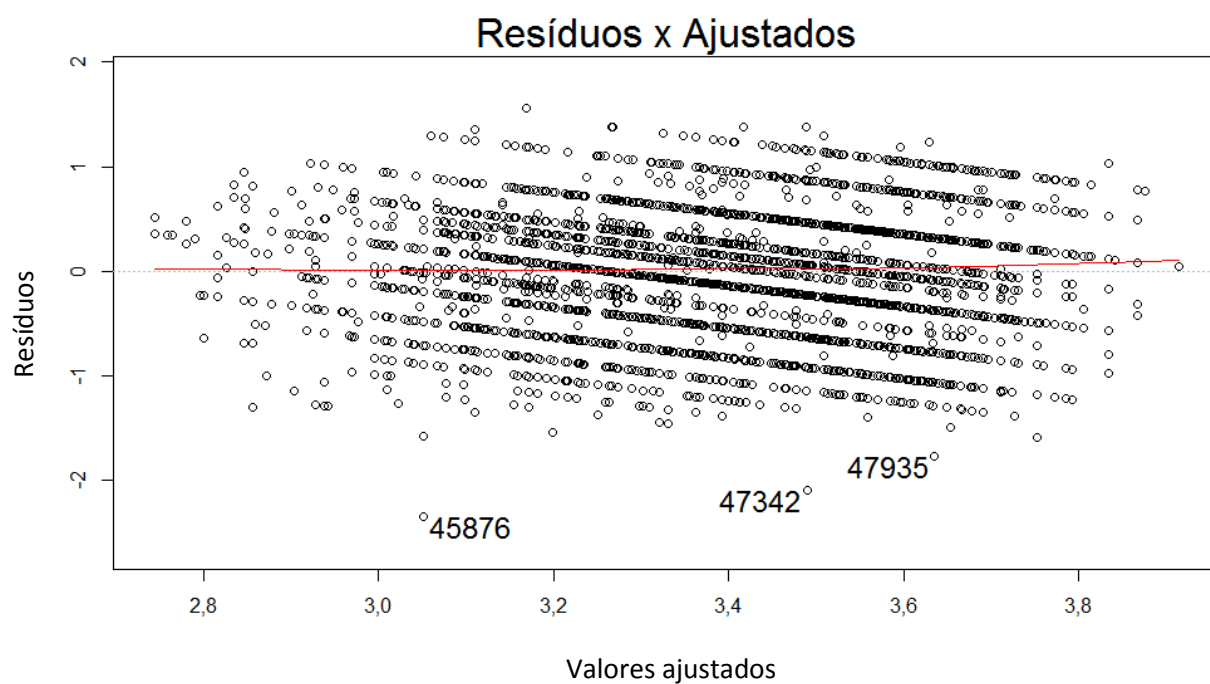


Figura 17 – Gráfico dos resíduos padronizados versus valores ajustados para o modelo da região Centro-Oeste

Fonte: Elaboração própria

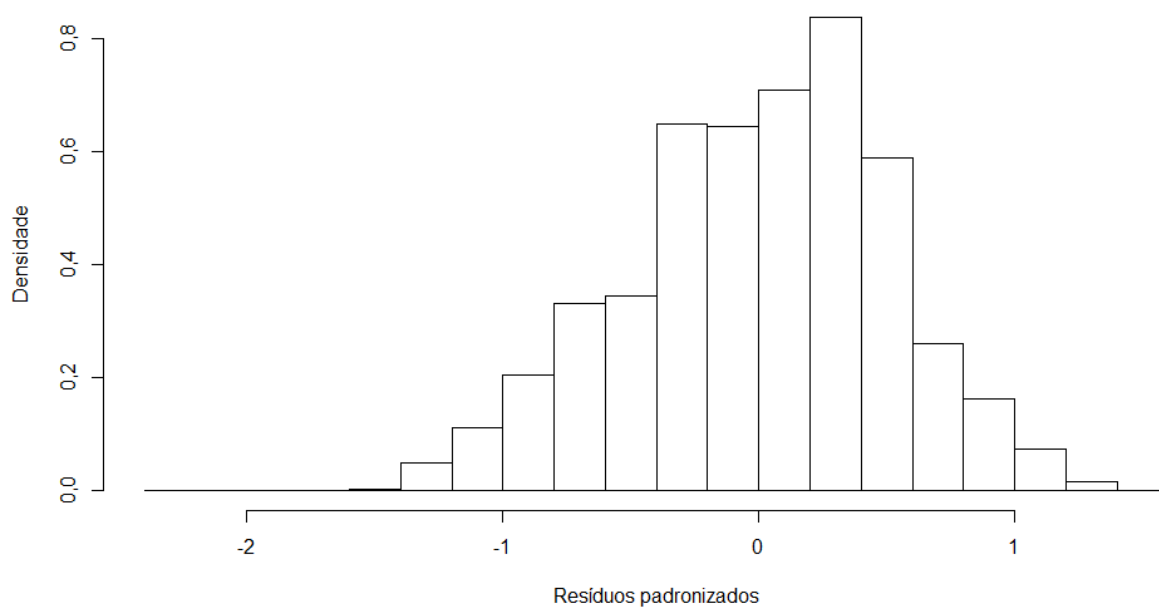


Figura 18 – Histograma dos resíduos padronizados para o modelo da região Centro-Oeste

Fonte: Elaboração própria

4.2. Resultados para a região Nordeste

O modelo selecionado para a região Nordeste contém as seguintes covariáveis, todas significativas ao nível de 5%:

- estrato geográfico,
- quantidade de famílias (em categorias),
- tipo de domicílio,
- presença de rede geral de energia elétrica,
- existência de aquecimento de água por energia elétrica,
- existência de aquecimento de água por lenha,
- existência de fogão a gás,
- existência de fogão a lenha,
- existência de fogão a carvão,
- existência de *grill*,
- existência de micro-ondas,
- categorias de renda per capita,
- categorias de alimento per capita.

A categoria cômodo, da variável tipo de domicílio, não apresentou diferença da categoria referência, que é casa. Ela foi mantida apenas para facilitar as comparações entre regiões. Caso contrário, cada região teria uma categorização diferente para esta variável. As estimativas e respectivos erros padrões são apresentados na Tabela 15.

Tabela 15 – Estimativas, erros padrões, valor P, limites inferior e superior do intervalo de confiança (95%) para a região Nordeste

	Estimativa	Erro padrão	Valor P	Limite inferior	Limite superior
Intercepto	2,607	0,040	<0,001	2,529	2,684
Área rural	-0,088	0,011	<0,001	-0,110	-0,065
Duas ou mais famílias	-0,236	0,015	<0,001	-0,266	-0,205
Apartamento	-0,117	0,019	<0,001	-0,154	-0,08
Cômodo	0,029	0,058	0,620	-0,086	0,144
Ausência de rede geral de energia elétrica	-0,136	0,030	<0,001	-0,195	-0,077
Ausência de aquecimento de água por energia elétrica	0,025	0,012	0,038	0,001	0,049
Ausência de aquecimento de água por lenha	0,176	0,035	<0,001	0,108	0,244
Ausência de fogão a gás	-0,195	0,036	<0,001	-0,265	-0,124
Ausência de fogão a lenha	0,190	0,013	<0,001	0,165	0,215
Ausência de fogão a carvão	0,117	0,013	<0,001	0,090	0,143
Presença de <i>grill</i>	-0,075	0,016	<0,001	-0,107	-0,044
Presença de micro-ondas	-0,136	0,015	<0,001	-0,166	-0,106
Renda maior que 0,5 até 0,75 s.m.	0,189	0,012	<0,001	0,165	0,212
Renda maior que 0,75 até 1,5 s.m.	0,362	0,011	<0,001	0,341	0,383
Renda maior que 1,5 até 2,5 s.m.	0,482	0,014	<0,001	0,455	0,509
Renda maior que 2,5 até 3,75 s.m.	0,475	0,020	<0,001	0,436	0,514
Renda maior que 3,75 s.m.	0,570	0,020	<0,001	0,531	0,609
Consumo de alimento médio baixo	0,073	0,011	<0,001	0,050	0,095
Consumo de alimento médio alto	0,143	0,011	<0,001	0,120	0,165
Consumo de alimento alto	0,287	0,013	<0,001	0,262	0,312
Sem informação sobre consumo de alimento	0,140	0,013	<0,001	0,115	0,165

Fonte: Elaboração própria

Os resíduos padronizados parecem ter satisfeito a suposição de homocedasticidade, por não apresentarem padrão de distribuição (Figura 19). Eles se distribuem ao redor da média, que é praticamente zero (0,016), como pode ser observado na Figura 20. A média da exponencial dos resíduos foi de 1,1.

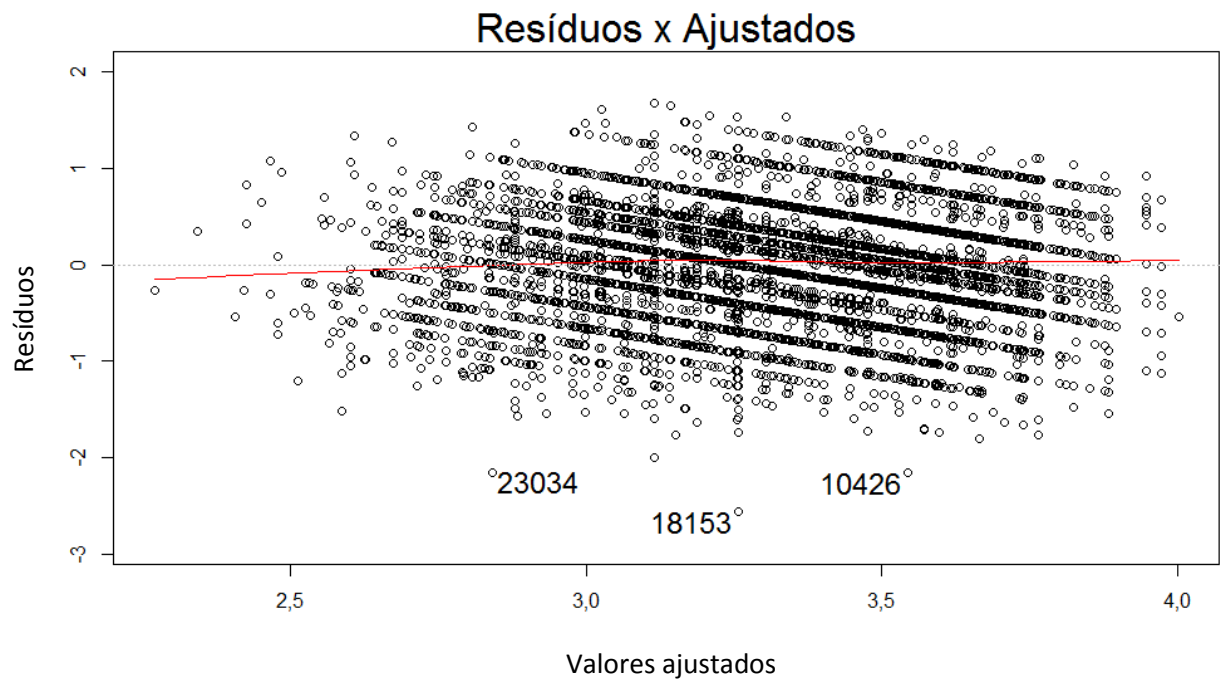


Figura 19 – Gráfico dos resíduos padronizados versus valores ajustados para o modelo da região Nordeste

Fonte: Elaboração própria

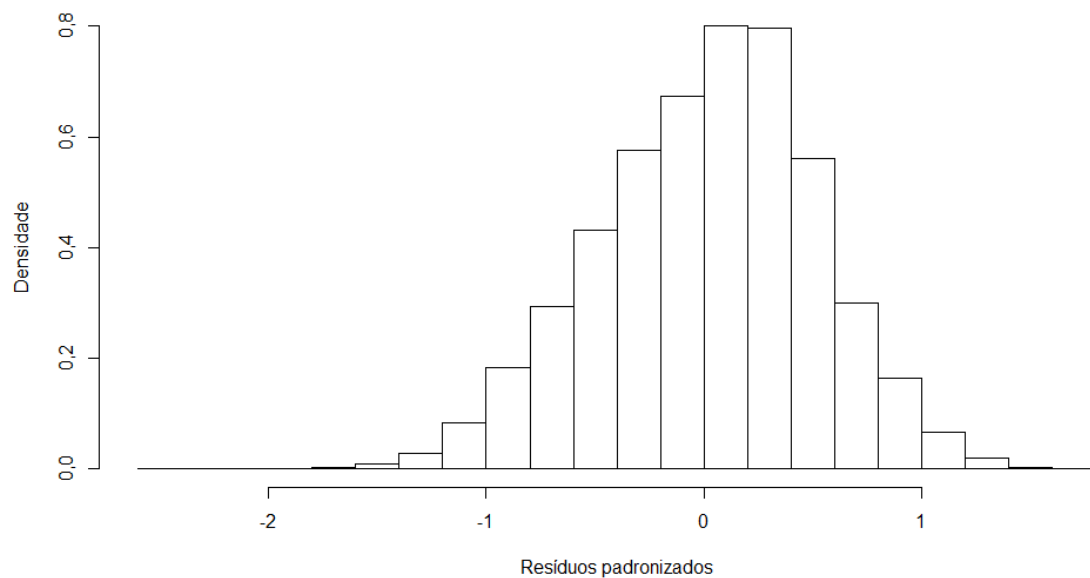


Figura 20 – Histograma dos resíduos padronizados para o modelo da região Nordeste

Fonte: Elaboração própria

4.3. Resultados para a região Norte

O modelo selecionado para a região Norte contém as seguintes covariáveis, todas significativas ao nível de 5%:

- quantidade de famílias (em categorias),
- tipo de domicílio,
- existência de água canalizada,
- presença de rede geral de energia elétrica,
- existência de fogão a gás,
- existência de fogão a lenha,
- existência de fogão a carvão,
- existência de *grill*,
- existência de micro-ondas,
- categorias de renda per capita,
- categorias de alimento per capita.

O consumo de gás per capita para a categoria apartamento, da variável tipo de domicílio, não diferiu da categoria de referência, que era casa. As estimativas e respectivos erros padrões são apresentados na Tabela 16.

Tabela 16 – Estimativas, erros padrões, valor P, limites inferior e superior do intervalo de confiança (95%) para a região Norte

	Estimativa	Erro padrão	Valor P	Limite inferior	Limite superior
Intercepto	2,518	0,036	<0,001	2,447	2,588
Duas ou mais famílias	-0,187	0,022	<0,001	-0,230	-0,145
Apartamento	0,024	0,033	0,468	-0,041	0,088
Cômodo	0,114	0,043	0,009	0,029	0,199
Ausência de água canalizada	-0,043	0,018	0,017	-0,078	-0,008
Ausência de rede geral de energia elétrica	-0,126	0,031	<0,001	-0,188	-0,065
Ausência de fogão a gás	-0,120	0,049	0,015	-0,216	-0,023
Ausência de fogão a lenha	0,191	0,023	<0,001	0,146	0,236
Ausência de fogão a carvão	0,142	0,025	<0,001	0,092	0,192
Presença de <i>grill</i>	-0,111	0,030	<0,001	-0,169	-0,053
Presença de micro-ondas	-0,096	0,024	<0,001	-0,142	-0,05
Renda maior que 0,5 até 0,75 s.m.	0,202	0,021	<0,001	0,162	0,243
Renda maior que 0,75 até 1,5 s.m.	0,395	0,018	<0,001	0,360	0,431
Renda maior que 1,5 até 2,5 s.m.	0,539	0,022	<0,001	0,497	0,582
Renda maior que 2,5 até 3,75 s.m.	0,563	0,027	<0,001	0,509	0,617
Renda maior que 3,75 s.m.	0,730	0,029	<0,001	0,674	0,786
Consumo de alimento médio baixo	0,042	0,019	0,027	0,005	0,078
Consumo de alimento médio alto	0,110	0,018	<0,001	0,074	0,145
Consumo de alimento alto	0,247	0,020	<0,001	0,207	0,287
Sem informação sobre consumo de alimento	0,180	0,023	<0,001	0,135	0,226

Fonte: Elaboração própria

Os resíduos padronizados não apresentaram um padrão de distribuição, o que indica que a suposição de homocedasticidade foi satisfeita (Figura 21). Eles se distribuem ao redor da média (0,005), como pode ser visto na Figura 22. A média da exponencial dos resíduos foi de 1,1.

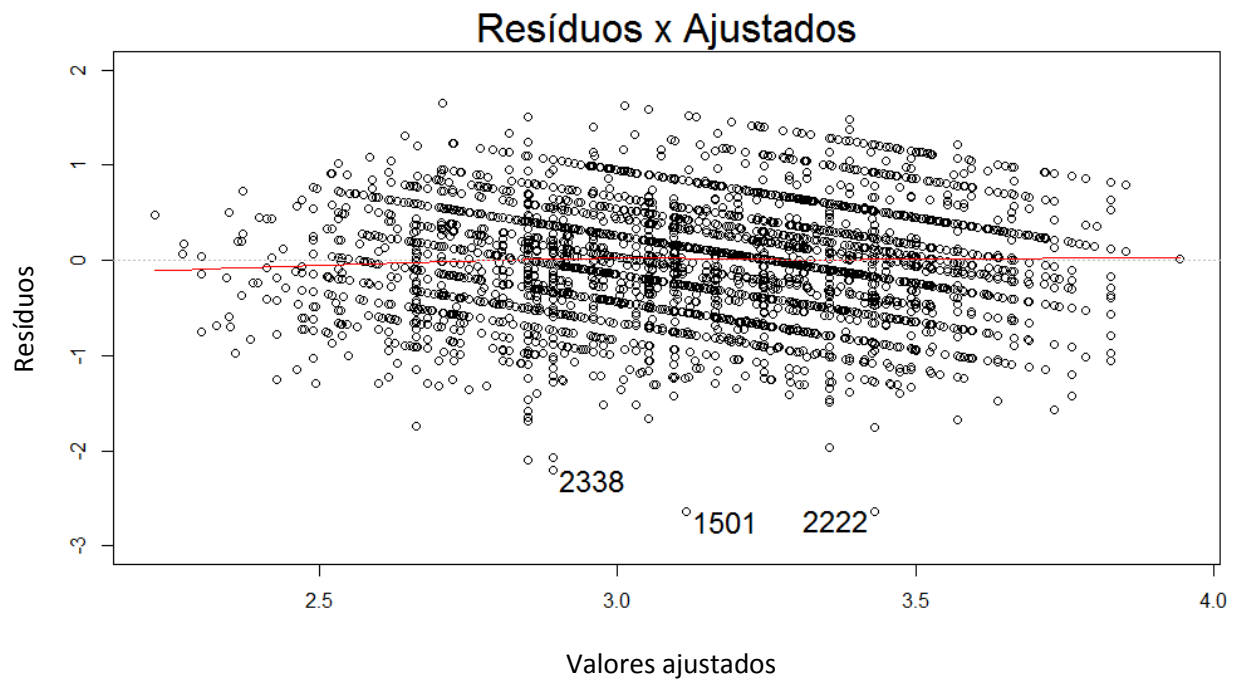


Figura 21 – Gráfico dos resíduos padronizados versus valores ajustados para o modelo da região Norte

Fonte: Elaboração própria

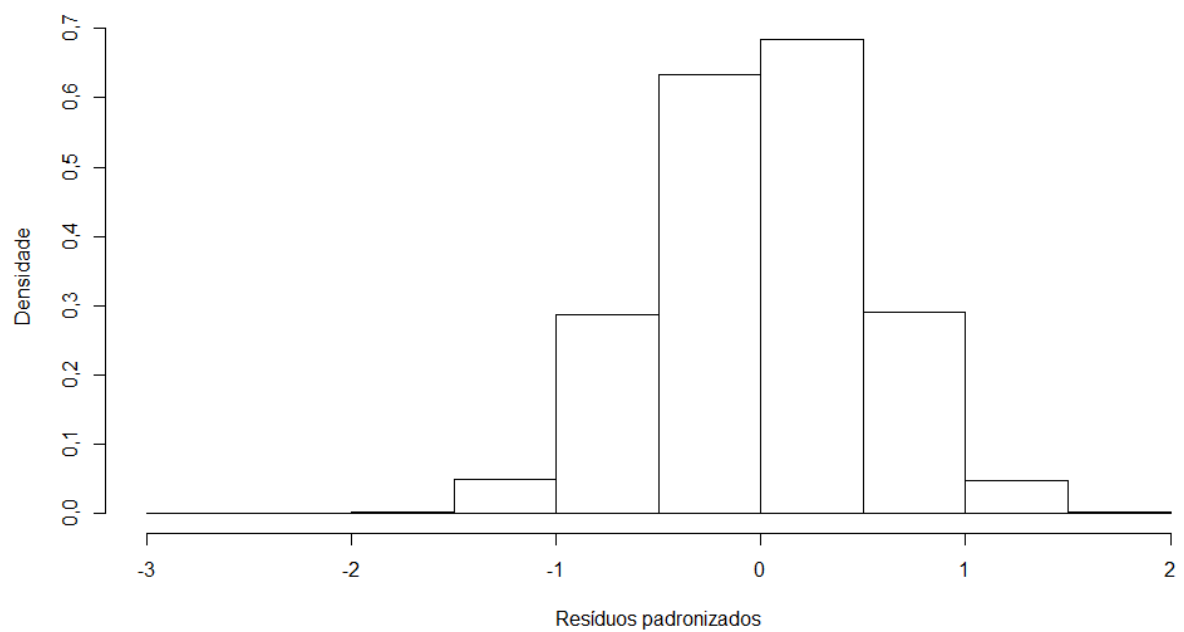


Figura 22 – Histograma dos resíduos padronizados para o modelo da região Norte

Fonte: Elaboração própria

4.4. Resultados para a região Sudeste

O modelo selecionado para a região Sudeste contém as seguintes covariáveis, todas significativas ao nível de 5%:

- estrato geográfico,
- quantidade de famílias (em categorias),
- tipo de domicílio,
- existência de aquecimento de água por lenha,
- existência de fogão a gás,
- existência de fogão a lenha,
- existência de *grill*,
- existência de micro-ondas,
- categorias de renda per capita,
- categorias de alimento per capita.

As estimativas e respectivos erros padrões são apresentados na Tabela 17.

Tabela 17 – Estimativas, erros padrões, valor P, limites inferior e superior do intervalo de confiança (95%) para a região Sudeste

	Estimativa	Erro padrão	Valor P	Limite inferior	Limite superior
Intercepto	2,733	0,045	<0,001	2,645	2,822
Área rural	0,062	0,018	<0,001	0,027	0,096
Duas ou mais famílias	-0,260	0,020	<0,001	-0,300	-0,22
Apartamento	-0,053	0,017	0,001	-0,086	-0,021
Cômodo	0,314	0,057	<0,001	0,202	0,425
Ausência de aquecimento de água por lenha	0,181	0,044	<0,001	0,096	0,266
Ausência de fogão a gás	-0,223	0,086	0,009	-0,390	-0,055
Ausência de fogão a lenha	0,160	0,022	<0,001	0,116	0,203
Presença de <i>grill</i>	-0,056	0,015	<0,001	-0,085	-0,027
Presença de micro-ondas	-0,160	0,010	<0,001	-0,180	-0,14
Renda maior que 0,5 até 0,75 s.m.	0,103	0,022	<0,001	0,060	0,146
Renda maior que 0,75 até 1,5 s.m.	0,288	0,019	<0,001	0,252	0,325
Renda maior que 1,5 até 2,5 s.m.	0,392	0,019	<0,001	0,355	0,43
Renda maior que 2,5 até 3,75 s.m.	0,456	0,021	<0,001	0,415	0,497
Renda maior que 3,75 s.m.	0,504	0,021	<0,001	0,462	0,545
Consumo de alimento médio baixo	0,107	0,015	<0,001	0,078	0,135
Consumo de alimento médio alto	0,112	0,014	<0,001	0,083	0,14
Consumo de alimento alto	0,184	0,015	<0,001	0,155	0,213
Sem informação sobre consumo de alimento	0,124	0,013	<0,001	0,099	0,15

Fonte: Elaboração própria

Os resíduos padronizados não apresentaram um padrão de distribuição, o que indica que a suposição de homocedasticidade foi satisfeita (Figura 23). Eles se distribuem ao redor da média (0,036), conforme histograma da Figura 24. A média da exponencial dos resíduos foi de 1,2.

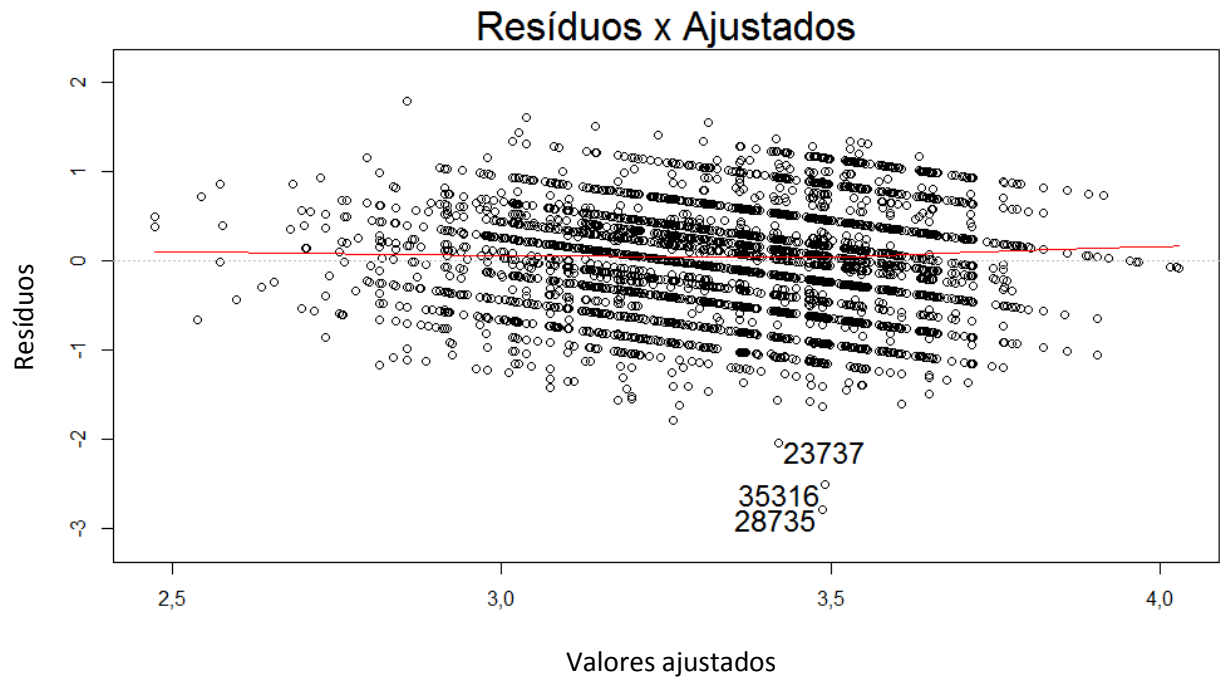


Figura 23 – Gráfico dos resíduos padronizados versus valores ajustados para o modelo da região Sudeste

Fonte: Elaboração própria

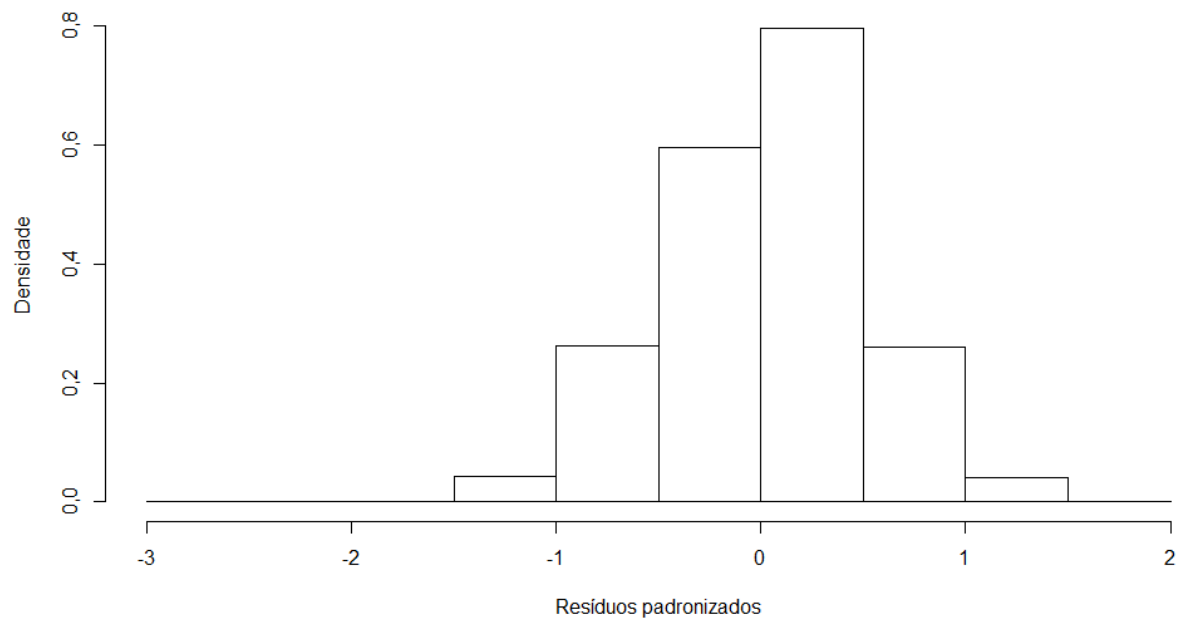


Figura 24 – Histograma dos resíduos padronizados para o modelo da região Sudeste

Fonte: Elaboração própria

4.5. Resultados para a região Sul

O modelo final para a região Sul contém as seguintes covariáveis, todas significativas ao nível de 5%:

- estrato geográfico,
- quantidade de famílias (em categorias),
- existência de aquecimento de água por gás,
- existência de aquecimento de água por energia solar,
- existência de aquecimento de água por lenha,
- existência de fogão a lenha,
- existência de forno elétrico,
- existência de *grill*,
- existência de micro-ondas,
- categorias de renda per capita,
- categorias de alimento per capita.

As estimativas e respectivos erros padrões são apresentados na Tabela 18.

Tabela 18 – Estimativas, erros padrões, valor P, limites inferior e superior do intervalo de confiança (95%) para a região Sul

	Estimativa	Erro padrão	Valor P	Limite inferior	Limite superior
Intercepto	2,930	0,128	<0,001	2,679	3,180
Área rural	-0,061	0,022	0,005	-0,103	-0,018
Duas ou mais famílias	-0,264	0,034	<0,001	-0,33	-0,197
Ausência de aquecimento de água por gás	-0,312	0,037	<0,001	-0,384	-0,240
Ausência de aquecimento de água por energia solar	0,188	0,094	0,046	0,003	0,373
Ausência de aquecimento de água por lenha	0,210	0,081	0,009	0,052	0,369
Ausência de fogão a lenha	0,146	0,019	<0,001	0,110	0,183
Presença de forno elétrico	-0,034	0,015	0,027	-0,065	-0,004
Presença de <i>grill</i>	-0,067	0,019	<0,001	-0,105	-0,030
Presença de micro-ondas	-0,115	0,015	<0,001	-0,145	-0,084
Renda maior que 0,5 até 0,75 s.m.	0,111	0,034	0,001	0,044	0,178
Renda maior que 0,75 até 1,5 s.m.	0,209	0,028	<0,001	0,153	0,264
Renda maior que 1,5 até 2,5 s.m.	0,300	0,029	<0,001	0,243	0,357
Renda maior que 2,5 até 3,75 s.m.	0,385	0,032	<0,001	0,322	0,447
Renda maior que 3,75 s.m.	0,504	0,032	<0,001	0,441	0,567
Consumo de alimento médio baixo	0,055	0,022	0,013	0,012	0,099
Consumo de alimento médio alto	0,081	0,022	<0,001	0,037	0,124
Consumo de alimento alto	0,150	0,021	<0,001	0,108	0,192
Sem informação sobre consumo de alimento	0,075	0,022	0,001	0,032	0,117

Fonte: Elaboração própria

Os resíduos padronizados não apresentaram um padrão de distribuição, o que indica que a suposição de homocedasticidade foi satisfeita (Figura 25). Eles se distribuem ao redor da média, que é praticamente zero (-0,009), como pode ser observado na Figura 26. A média da exponencial dos resíduos foi de 1,1.

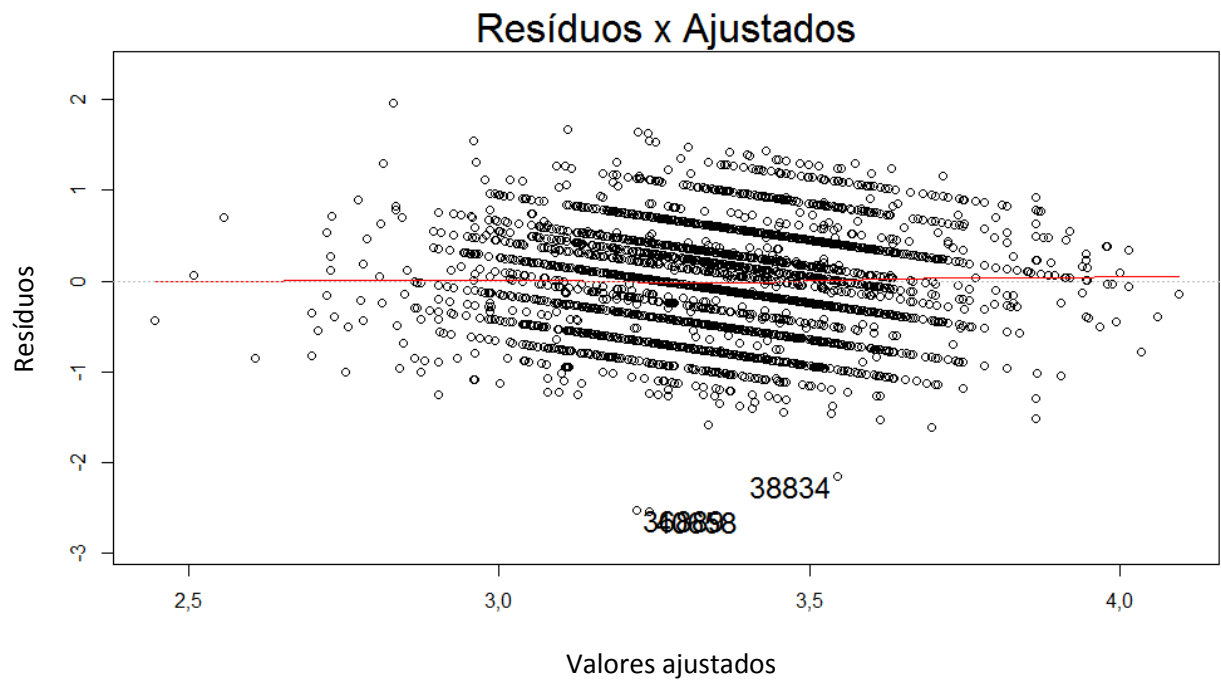


Figura 25 – Gráfico dos resíduos padronizados versus valores ajustados para o modelo da região Sul

Fonte: Elaboração própria

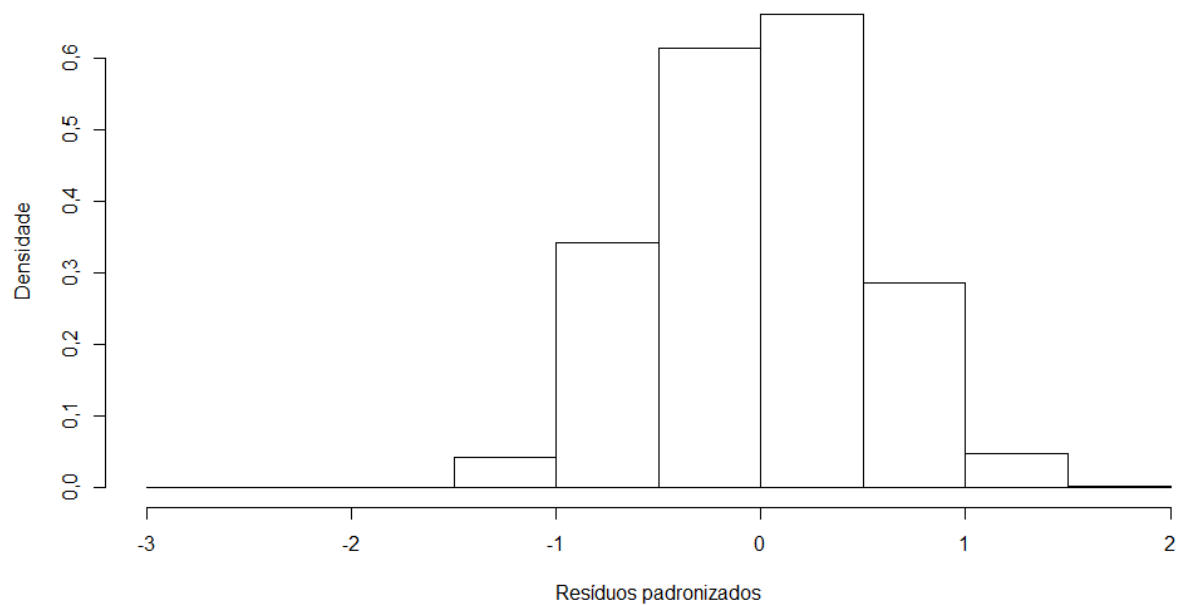


Figura 26 – Histograma dos resíduos padronizados para o modelo da região Sul

Fonte: Elaboração própria

4.6. Interpretação das estimativas

Como foram obtidas médias de aproximadamente 1 para a exponencial dos resíduos nos cinco modelos, é possível interpretar as estimativas de maneira intuitiva. A estimativa β associada a cada categoria, tudo o mais mantido constante, implica variação de e^β no consumo de gás per capita domiciliar.

A seguir, são oferecidas as variações estimadas no consumo de gás bem como os respectivos intervalos de confiança (Tabelas 19 a 23). Uma variação hipotética de 1,10, por exemplo, ao mudar de categoria, passando da referência à que teve a estimativa calculada, significa que há variação média de mais 10% no consumo de gás. Outra variação hipotética, agora de 0,90, significa que a mudança de categoria, tudo o mais mantido constante, representa redução média do consumo gás de 10% em relação à categoria de referência.

Para a leitura das próximas tabelas, é possível utilizar como exemplo a variação no consumo de gás que ocorre ao passar da área urbana para a área rural na região Centro-Oeste (Tabela 19). Na área rural, segundo o modelo proposto, há aumento no consumo de gás domiciliar per capita que varia de 4,3% a 12,9%, com relação à área urbana, tudo o mais mantido constante.

Tabela 19 – Variação estimada e os limites inferior e superior para seu intervalo de confiança (95%), para o modelo da região Centro-Oeste

	Variação	Limite inferior	Limite superior	Variação em %	Limite inferior em %	Limite superior em %
Área rural	1,09	1,04	1,13	8,5	4,3	12,9
Duas ou mais famílias	0,78	0,73	0,83	-22,2	-26,9	-17,2
Apartamento	0,93	0,89	0,98	-6,8	-11,2	-2,1
Cômodo	1,18	1,03	1,34	17,6	3,1	33,9
Ausência de aquecimento de água por energia elétrica	0,93	0,90	0,96	-6,9	-10,0	-3,8
Ausência de fogão a lenha	1,15	1,10	1,21	14,9	9,5	20,7
Presença de <i>grill</i>	0,85	0,81	0,89	-15,1	-18,7	-11,5
Presença de micro-ondas	0,91	0,88	0,94	-9,1	-11,8	-6,3
Renda maior que 0,5 até 0,75 s.m.	1,11	1,06	1,16	11,0	5,9	16,3
Renda maior que 0,75 até 1,5 s.m.	1,34	1,29	1,40	34,4	29,2	39,8
Renda maior que 1,5 até 2,5 s.m.	1,53	1,46	1,59	52,5	46,1	59,2
Renda maior que 2,5 até 3,75 s.m.	1,59	1,51	1,67	59,0	51,3	67,2
Renda maior que 3,75 s.m.	1,64	1,56	1,73	64,4	56,2	73,0
Consumo de alimento médio baixo	1,11	1,07	1,15	10,7	6,5	15,1
Consumo de alimento médio alto	1,17	1,13	1,22	17,4	13,1	21,8
Consumo de alimento alto	1,25	1,20	1,30	24,9	20,2	29,6
Sem informação sobre consumo de alimento	1,12	1,08	1,16	12,1	8,4	16,0

Fonte: Elaboração própria

Tabela 20 – Variação estimada e os limites inferior e superior para seu intervalo de confiança (95%), para o modelo da região Nordeste

	Variação	Limite inferior	Limite superior	Variação em %	Limite inferior em %	Limite superior em %
Área rural	0,92	0,90	0,94	-8,4	-10,4	-6,3
Duas ou mais famílias	0,79	0,77	0,81	-21,0	-23,4	-18,5
Apartamento	0,89	0,86	0,92	-11,0	-14,3	-7,7
Ausência de rede geral de energia elétrica	0,87	0,82	0,93	-12,7	-17,7	-7,4
Ausência de aquecimento de água por energia elétrica	1,03	1,00	1,05	2,5	0,1	5,0
Ausência de aquecimento de água por lenha	1,19	1,11	1,28	19,2	11,4	27,6
Ausência de fogão a gás	0,82	0,77	0,88	-17,7	-23,3	-11,7
Ausência de fogão a lenha	1,21	1,18	1,24	20,9	17,9	24,0
Ausência de fogão a carvão	1,12	1,09	1,15	12,4	9,4	15,4
Presença de grill	0,93	0,90	0,96	-7,2	-10,1	-4,3
Presença de micro-ondas	0,87	0,85	0,90	-12,7	-15,3	-10,1
Renda maior que 0,5 até 0,75 s.m.	1,21	1,18	1,24	20,8	17,9	23,6
Renda maior que 0,75 até 1,5 s.m.	1,44	1,41	1,47	43,6	40,6	46,7
Renda maior que 1,5 até 2,5 s.m.	1,62	1,58	1,66	61,9	57,6	66,4
Renda maior que 2,5 até 3,75 s.m.	1,61	1,55	1,67	60,8	54,7	67,2
Renda maior que 3,75 s.m.	1,77	1,70	1,84	76,8	70,1	83,9
Consumo de alimento médio baixo	1,08	1,05	1,10	7,6	5,1	10,0
Consumo de alimento médio alto	1,15	1,13	1,18	15,4	12,7	17,9
Consumo de alimento alto	1,33	1,30	1,37	33,2	30,0	36,6
Sem informação sobre consumo de alimento	1,15	1,12	1,18	15,0	12,2	17,9

Fonte: Elaboração própria

Tabela 21 – Variação estimada e os limites inferior e superior para seu intervalo de confiança (95%), para o modelo da região Norte

	Variação	Limite inferior	Limite superior	Variação em %	Limite inferior em %	Limite superior em %
Duas ou mais famílias	0,83	0,79	0,87	-17,1	-20,5	-13,5
Cômodo	1,12	1,03	1,22	12,1	2,9	22,0
Ausência de água canalizada	0,96	0,92	0,99	-4,2	-7,5	-0,8
Ausência de rede geral de energia elétrica	0,88	0,83	0,94	-11,8	-17,1	-6,3
Ausência de fogão a gás	0,89	0,81	0,98	-11,3	-19,4	-2,3
Ausência de fogão a lenha	1,21	1,16	1,27	21,0	15,7	26,6
Ausência de fogão a carvão	1,15	1,10	1,21	15,3	9,6	21,2
Presença de <i>grill</i>	0,89	0,84	0,95	-10,5	-15,5	-5,2
Presença de micro-ondas	0,91	0,87	0,95	-9,2	-13,2	-4,9
Renda maior que 0,5 até 0,75 s.m.	1,22	1,18	1,28	22,4	17,6	27,5
Renda maior que 0,75 até 1,5 s.m.	1,48	1,43	1,54	48,4	43,3	53,9
Renda maior que 1,5 até 2,5 s.m.	1,71	1,64	1,79	71,4	64,4	79,0
Renda maior que 2,5 até 3,75 s.m.	1,76	1,66	1,85	75,6	66,4	85,3
Renda maior que 3,75 s.m.	2,08	1,96	2,19	107,5	96,2	119,5
Consumo de alimento médio baixo	1,04	1,01	1,08	4,3	0,5	8,1
Consumo de alimento médio alto	1,12	1,08	1,16	11,6	7,7	15,6
Consumo de alimento alto	1,28	1,23	1,33	28,0	23,0	33,2
Sem informação sobre consumo de alimento	1,20	1,14	1,25	19,7	14,5	25,4

Fonte: Elaboração própria

Tabela 22 – Variação estimada e os limites inferior e superior para seu intervalo de confiança (95%), para o modelo da região Sudeste

	Variação	Limite inferior	Limite superior	Variação em %	Limite inferior em %	Limite superior em %
Área rural	1,06	1,03	1,10	6,4	2,7	10,1
Duas ou mais famílias	0,77	0,74	0,80	-22,9	-25,9	-19,7
Apartamento	0,95	0,92	0,98	-5,2	-8,2	-2,1
Cômodo	1,37	1,22	1,53	36,9	22,4	53,0
Ausência de aquecimento de água por lenha	1,20	1,10	1,30	19,8	10,1	30,5
Ausência de fogão a gás	0,80	0,68	0,95	-20,0	-32,3	-5,4
Ausência de fogão a lenha	1,17	1,12	1,23	17,4	12,3	22,5
Presença de <i>grill</i>	0,95	0,92	0,97	-5,4	-8,1	-2,7
Presença de micro-ondas	0,85	0,84	0,87	-14,8	-16,5	-13,1
Renda maior que 0,5 até 0,75 s.m.	1,11	1,06	1,16	10,8	6,2	15,7
Renda maior que 0,75 até 1,5 s.m.	1,33	1,29	1,38	33,4	28,7	38,4
Renda maior que 1,5 até 2,5 s.m.	1,48	1,43	1,54	48,0	42,6	53,7
Renda maior que 2,5 até 3,75 s.m.	1,58	1,51	1,64	57,8	51,4	64,4
Renda maior que 3,75 s.m.	1,66	1,59	1,72	65,5	58,7	72,5
Consumo de alimento médio baixo	1,11	1,08	1,14	11,3	8,1	14,5
Consumo de alimento médio alto	1,12	1,09	1,15	11,9	8,7	15,0
Consumo de alimento alto	1,20	1,17	1,24	20,2	16,8	23,7
Sem informação sobre consumo de alimento	1,13	1,10	1,16	13,2	10,4	16,2

Fonte: Elaboração própria

Tabela 23 – Variação estimada e os limites inferior e superior para seu intervalo de confiança (95%), para o modelo da região Sul

	Variação	Limite inferior	Limite superior	Variação em %	Limite inferior em %	Limite superior em %
Área rural	0,94	0,90	0,98	-5,9	-9,8	-1,8
Duas ou mais famílias	0,77	0,72	0,82	-23,2	-28,1	-17,9
Ausência de rede geral de energia elétrica	0,73	0,68	0,79	-26,8	-31,9	-21,3
Ausência de aquecimento de água por gás	1,21	1,00	1,45	20,7	0,3	45,2
Ausência de aquecimento de água por energia solar	1,23	1,05	1,45	23,4	5,3	44,6
Ausência de aquecimento de água por lenha	1,16	1,12	1,20	15,7	11,6	20,1
Ausência de fogão a lenha	0,97	0,94	1,00	-3,4	-6,3	-0,4
Presença de forno elétrico	0,93	0,90	0,97	-6,5	-10,0	-3,0
Presença de <i>grill</i>	0,89	0,87	0,92	-10,8	-13,5	-8,1
Presença de micro-ondas	1,12	1,04	1,19	11,7	4,5	19,5
Renda maior que 0,5 até 0,75 s.m.	1,23	1,17	1,30	23,2	16,5	30,2
Renda maior que 0,75 até 1,5 s.m.	1,35	1,28	1,43	35,0	27,5	42,9
Renda maior que 1,5 até 2,5 s.m.	1,47	1,38	1,56	46,9	38,0	56,4
Renda maior que 2,5 até 3,75 s.m.	1,66	1,55	1,76	65,5	55,4	76,3
Renda maior que 3,75 s.m.	1,06	1,01	1,10	5,7	1,2	10,4
Consumo de alimento médio baixo	1,08	1,04	1,13	8,4	3,8	13,2
Consumo de alimento médio alto	1,16	1,11	1,21	16,2	11,4	21,2
Consumo de alimento alto	1,08	1,03	1,12	7,7	3,3	12,4
Sem informação sobre consumo de alimento	0,94	0,90	0,98	-5,9	-9,8	-1,8

Fonte: Elaboração própria

Por fim, a título de documentação, a Tabela 24 apresenta as informações relativas aos modelos ajustados, como o R quadrado múltiplo e o R quadrado ajustado, que são medidas da qualidade e do poder de explicação do modelo. Já era esperado que os valores seriam baixos, uma vez que outras variáveis que não foram consideradas no modelo poderiam ter poder de explicação sobre o consumo de gás domiciliar, mas não estavam disponíveis. Além disso, o modelo só considerou variáveis categóricas, que também têm limitações em seu poder de explicação.

Tabela 24 – Informações selecionadas sobre o ajuste dos modelos: R quadrado múltiplo e ajustado, estatística F, graus de liberdade, Valor P, total de observações na base de dados, observações descartadas devido a ausência de informações

	Centro- Oeste	Nordeste	Norte	Sudeste	Sul
R quadrado múltiplo	0,1274	0,2277	0,2566	0,1196	0,1125
R quadrado ajustado	0,1254	0,2267	0,2545	0,1184	0,1098
Estatística F	65,2	231,9	121,3	94,4	41,0
Graus de liberdade	17 e 7.597	21 e 16.522	19 e 6.677	18 e 12.509	18 e 5.822
Valor P	<0,001	<0,001	<0,001	<0,001	<0,001
Total de observações na base de dados	7.656	16.619	6.756	12.576	5.860
Observações não utilizadas por dados faltantes	41	75	59	48	19

Fonte: Elaboração própria

5. Considerações finais

Pelos modelos aqui ajustados, o pesquisador tem em mãos dados para avaliar as variáveis que impactam significativamente no consumo de gás per capita em cada região do país. Cabe salientar que os modelos dependem das variáveis disponíveis na base de dados da POF. Outras variáveis poderiam levar a interpretações distintas.

É necessário pontuar também que o fato de uma variável não aparecer como significativa no modelo não significa que ela não esteja de fato associada a maior ou menor consumo de gás. O que ocorre é que as técnicas de seleção de variáveis explicativas buscam as com maior poder de explicação no conjunto das disponíveis.

Futuros trabalhos podem ser realizados no sentido de construir a série histórica de dados da POF e verificar mudanças nas duas pesquisas disponíveis (2002-2003 e 2008-2009, esta última utilizada nesse trabalho).

Outra possibilidade de análise dos dados aqui não realizada é verificar se há diferenças no perfil dos domicílios consumidores de gás de cozinha e dos não consumidores. Uma técnica a ser empregada seria a construção de um modelo de regressão logística, tendo como resposta o consumo ou não de Gás Liquefeito de Petróleo (GLP).

Também como indicação de possíveis estudos futuros estão as eventuais interações entre variáveis explicativas. A inclusão das interações nos modelos, além de dar complexidade, possibilita detalhamentos que os modelos ajustados apenas com efeitos principais não são capazes de captar. Para tanto, cabe ao pesquisador determinar as interações de interesse, de maneira parcimoniosa, a fim de realizar novas análises com o mesmo conjunto de dados.

Apêndice A – Gráficos da análise descritiva, segundo região

Região Centro-Oeste

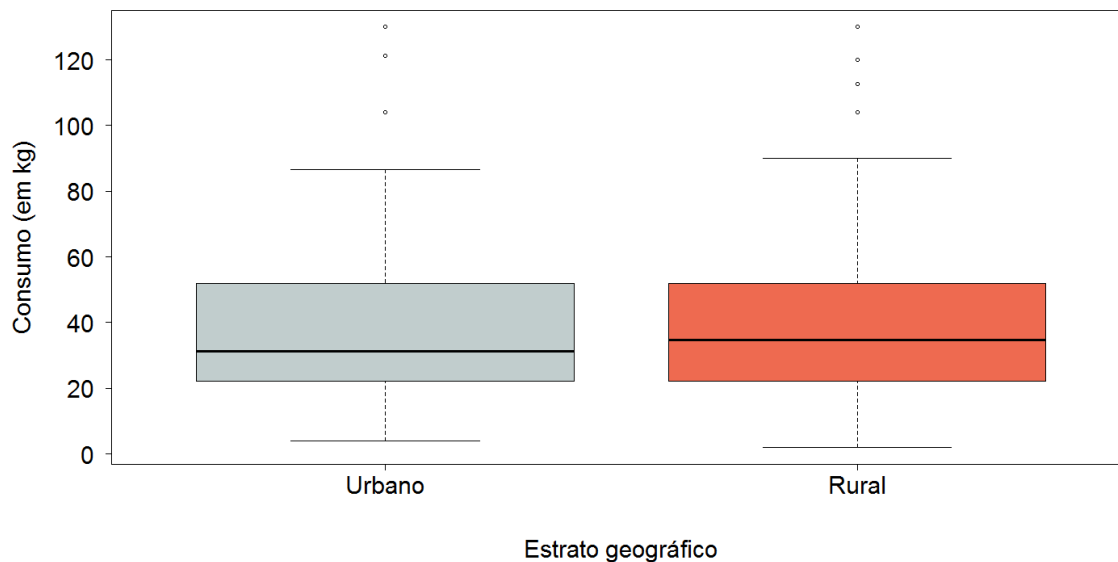


Figura A1 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo estrato geográfico

Fonte: POF 2008-2009, IBGE. Elaboração própria

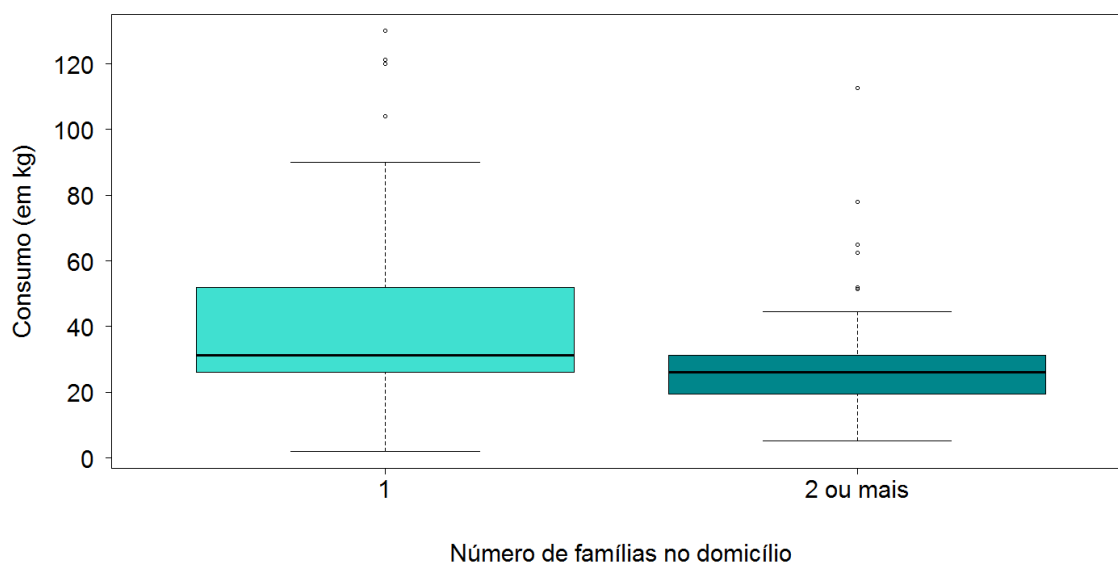


Figura A2 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo número de famílias por domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

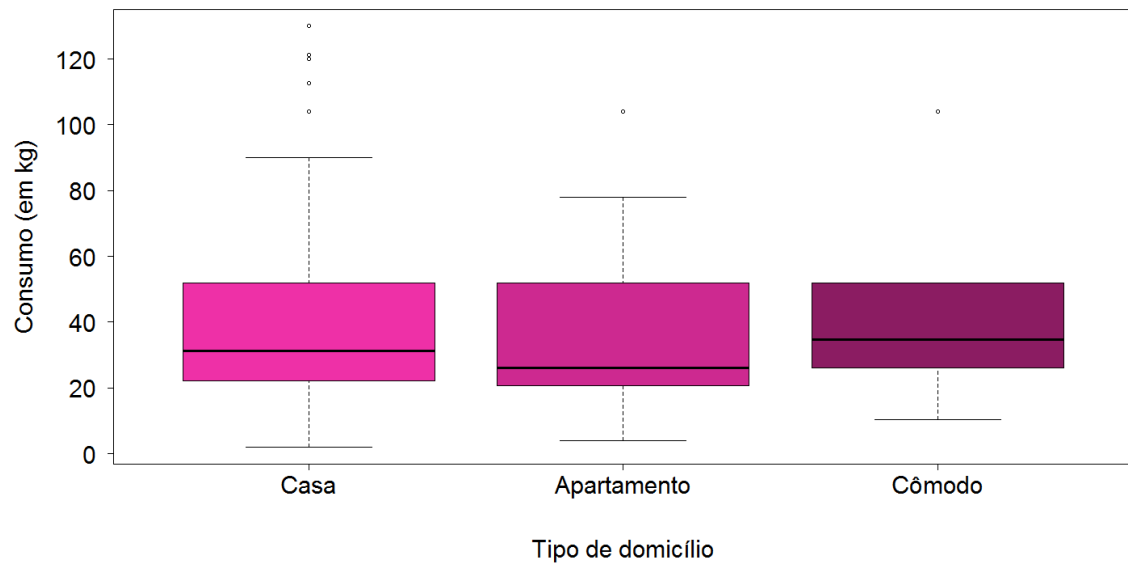


Figura A3 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo tipo de domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

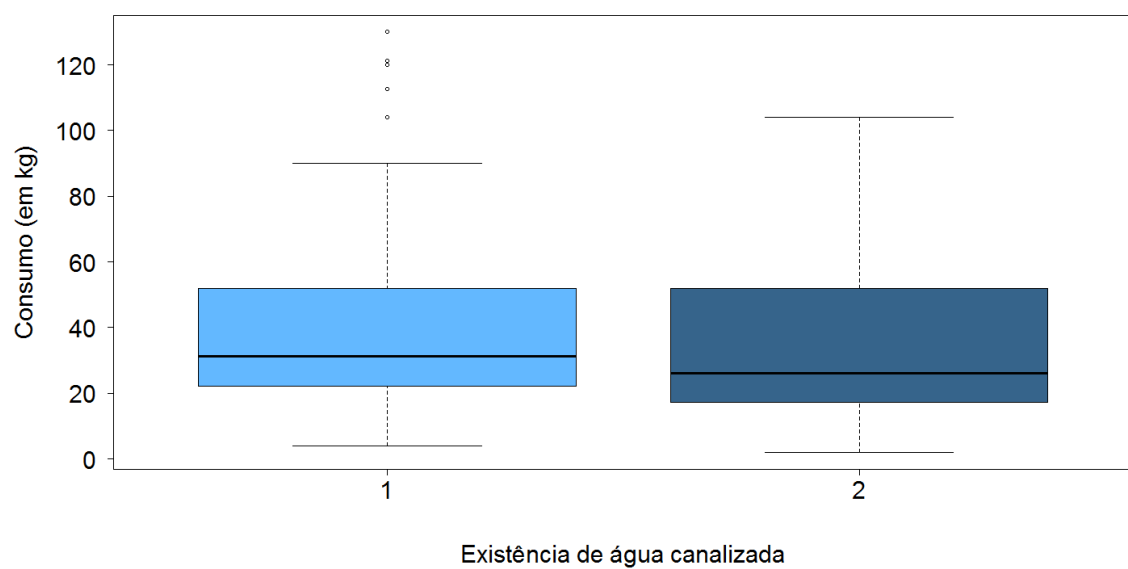


Figura A4 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de água canalizada no domicílio

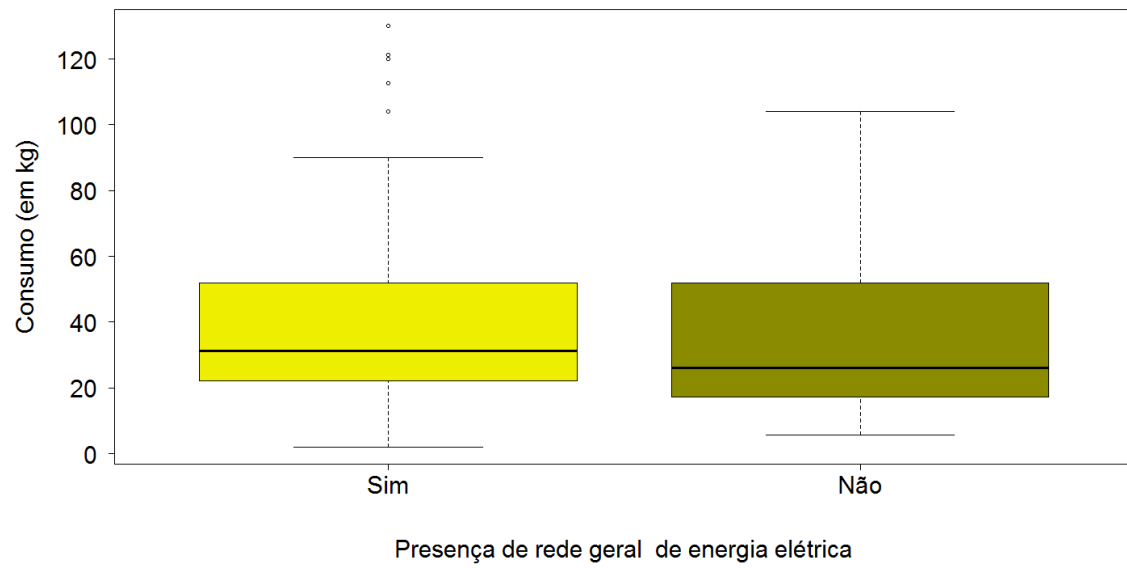


Figura A5 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de rede geral de energia elétrica no domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

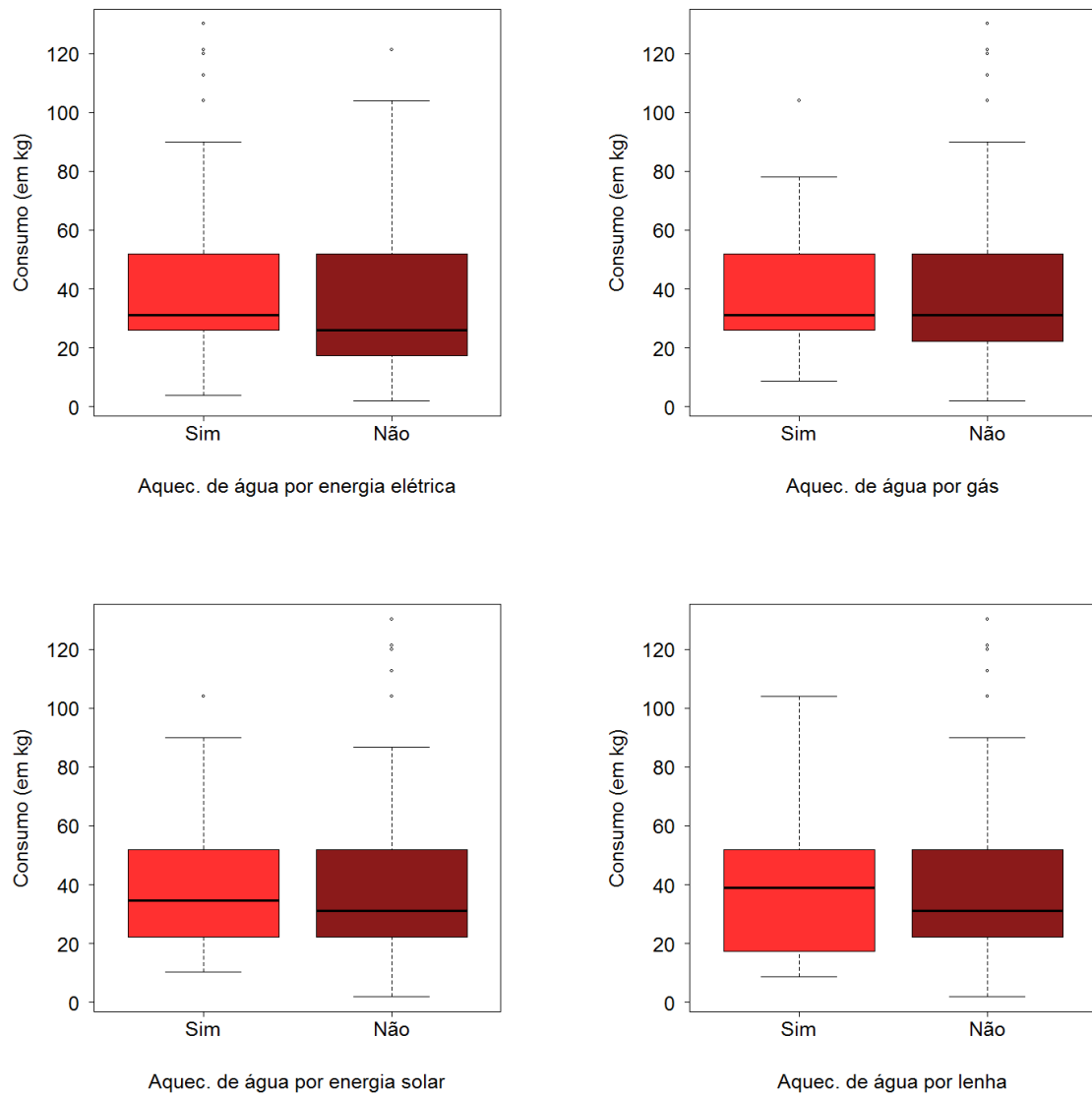


Figura A6 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de aquecimento de água por energia elétrica, gás, energia solar e lenha

Fonte: POF 2008-2009, IBGE. Elaboração própria

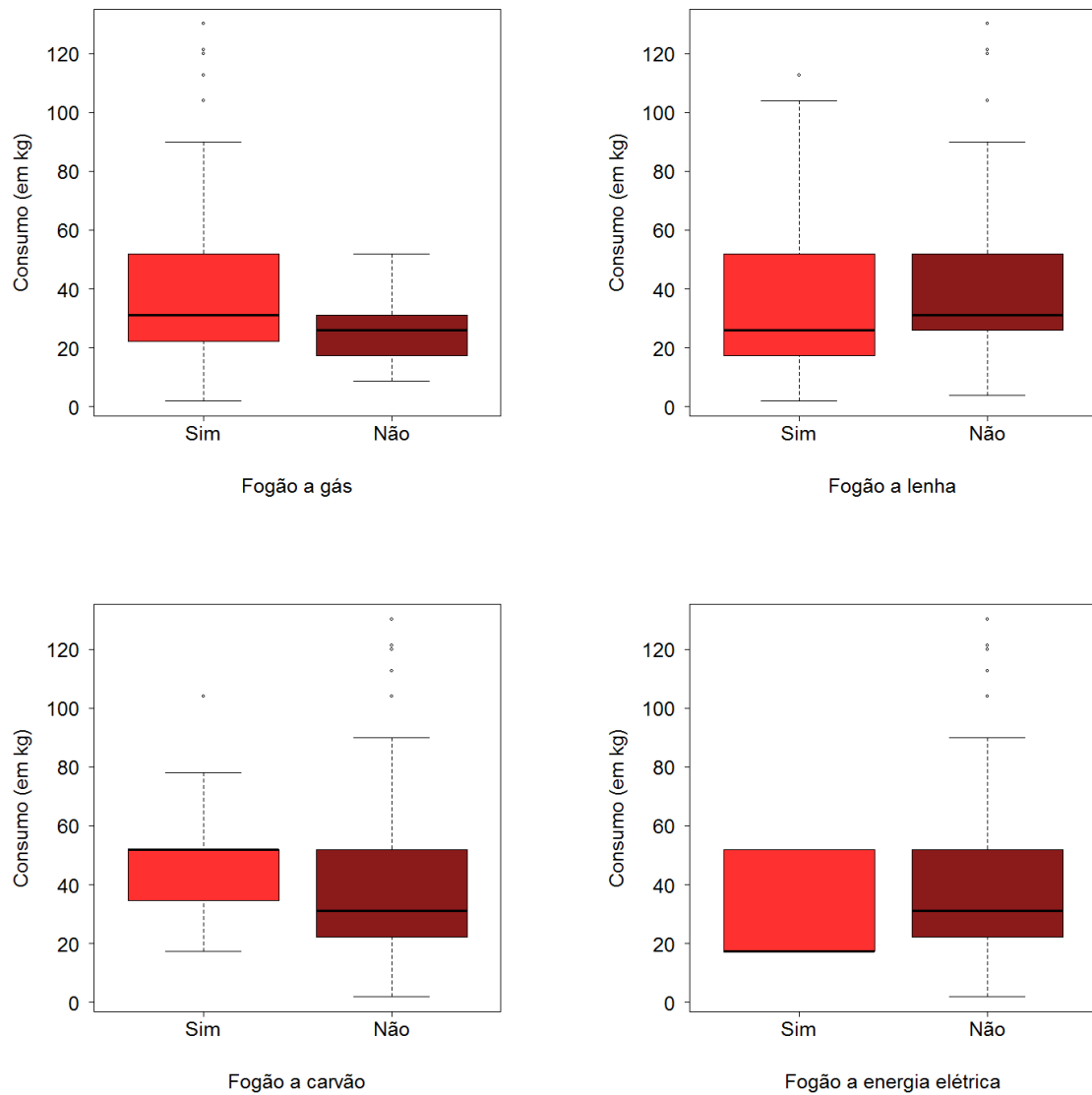


Figura A7 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de fogão a gás, a lenha, a carvão e a energia elétrica

Fonte: POF 2008-2009, IBGE. Elaboração própria

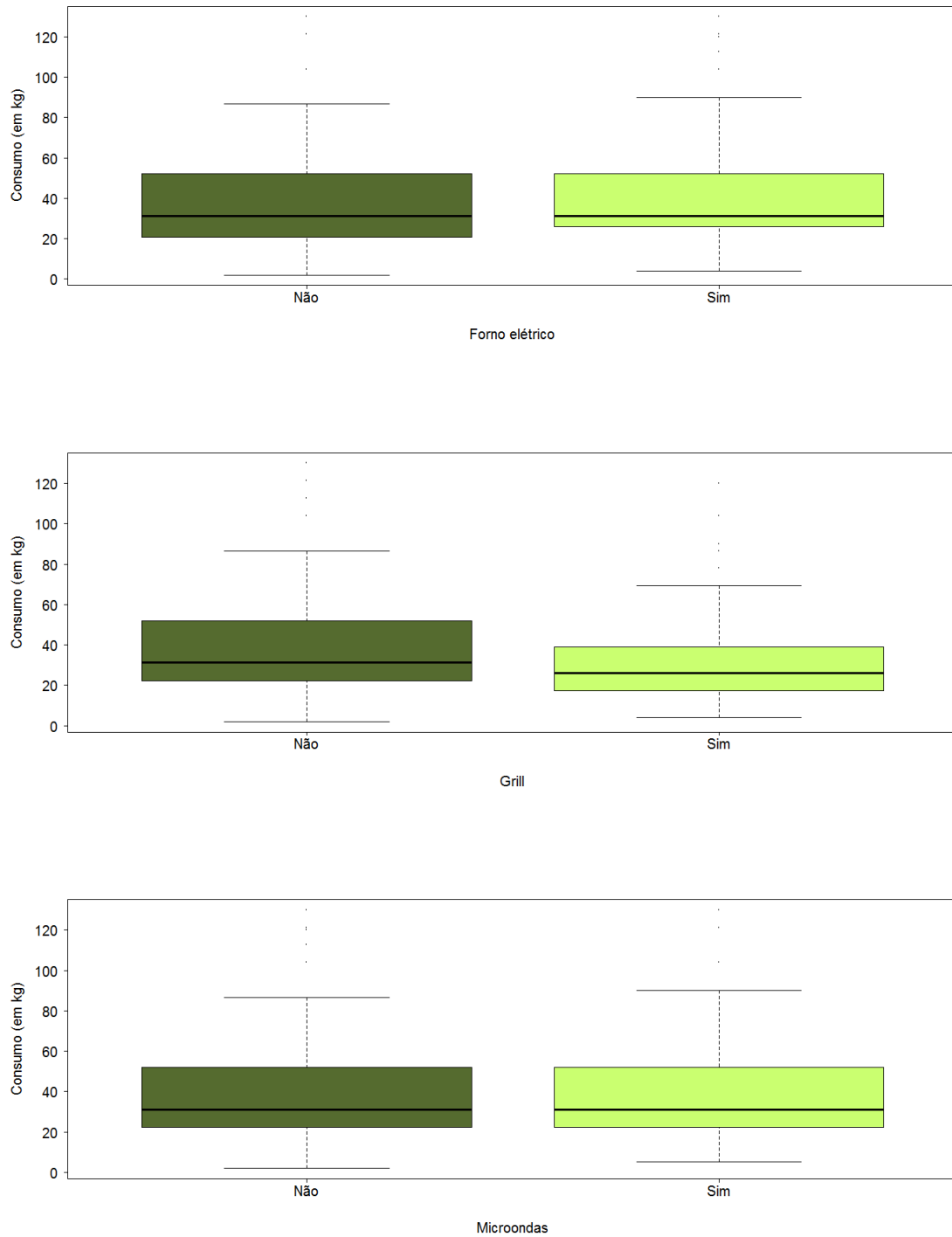


Figura A8 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de forno elétrico, *grill* e micro-ondas

Fonte: POF 2008-2009, IBGE. Elaboração própria

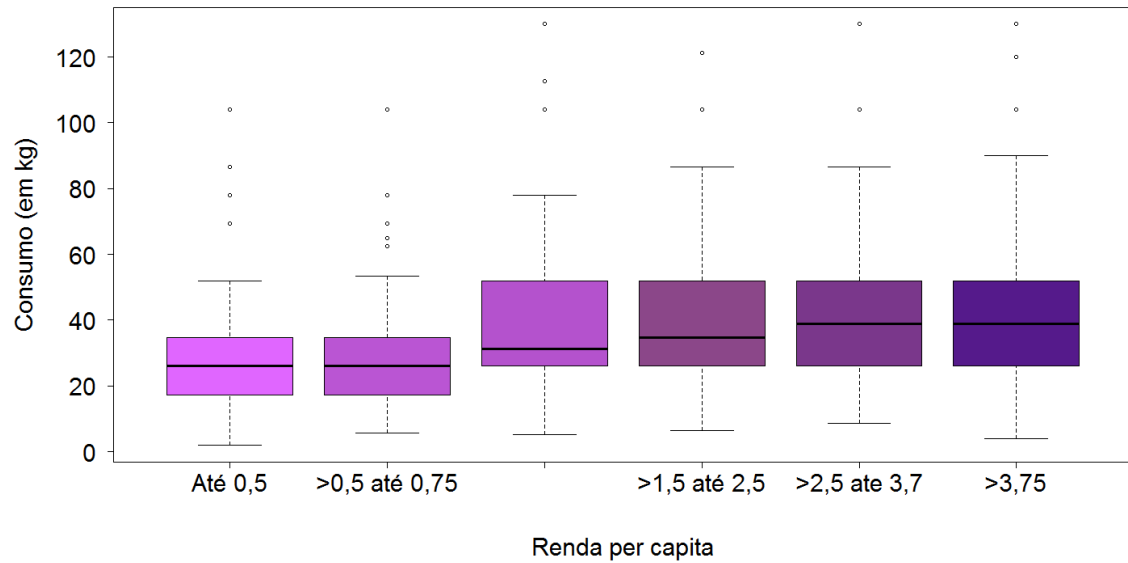


Figura A9 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de renda mensal per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

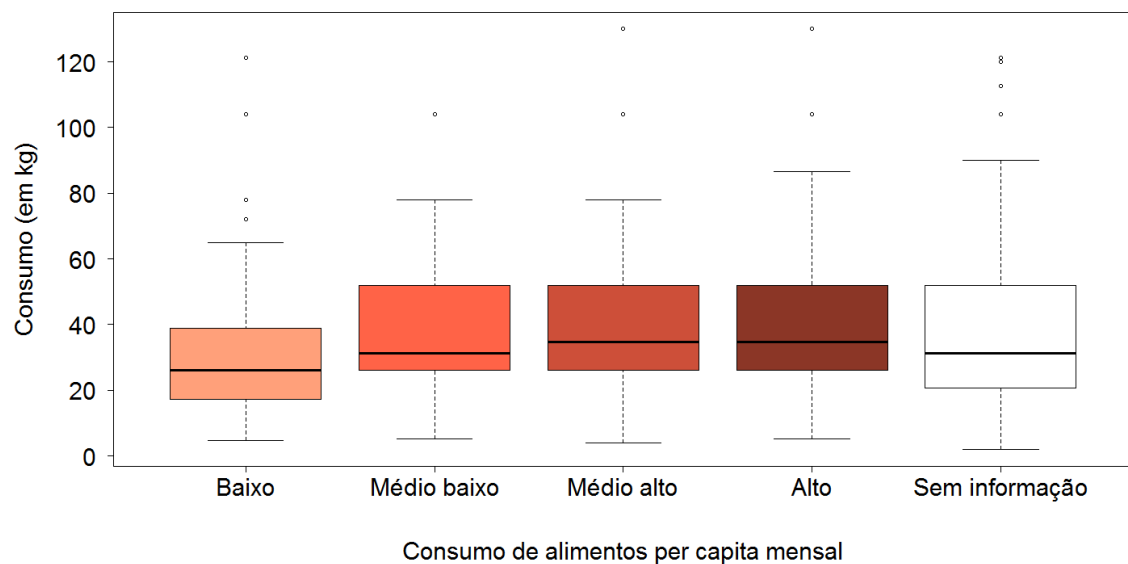


Figura A10 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de consumo de alimento per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

Região Nordeste

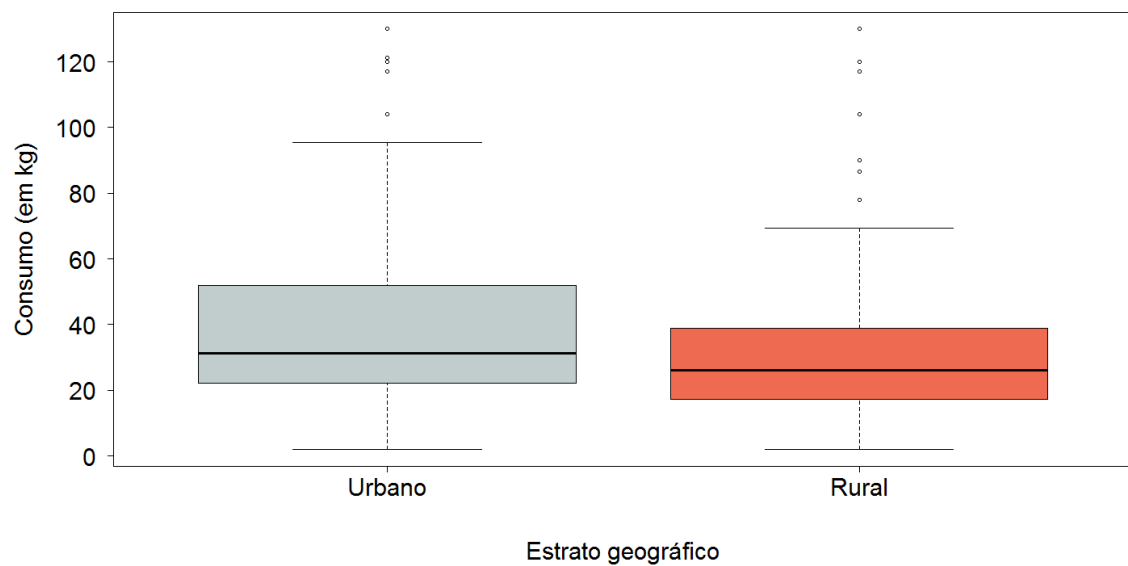


Figura A11 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo estrato geográfico

Fonte: POF 2008-2009, IBGE. Elaboração própria

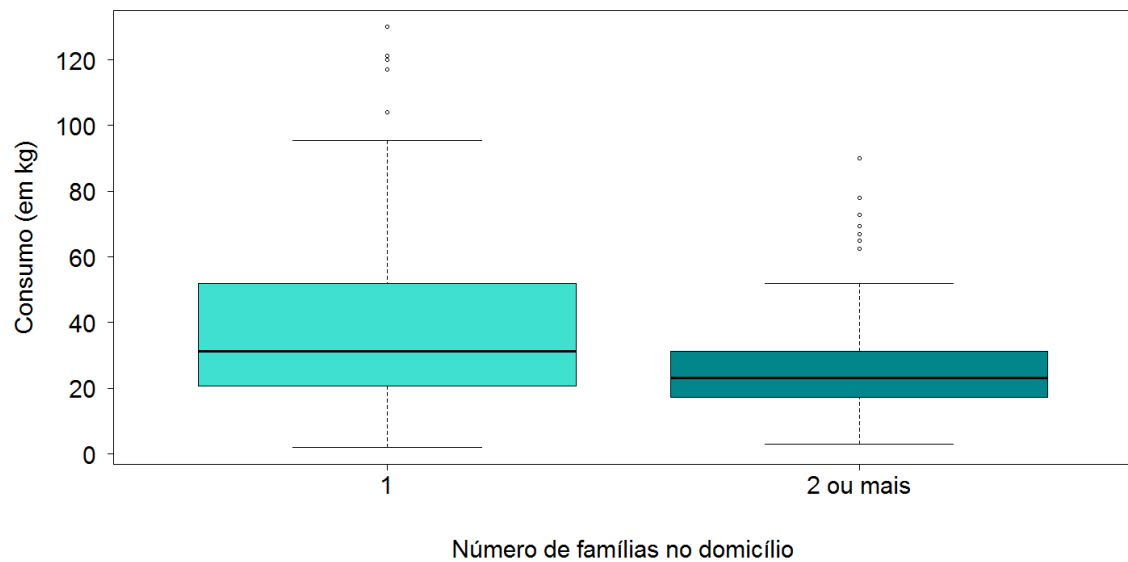


Figura A12 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo número de famílias por domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

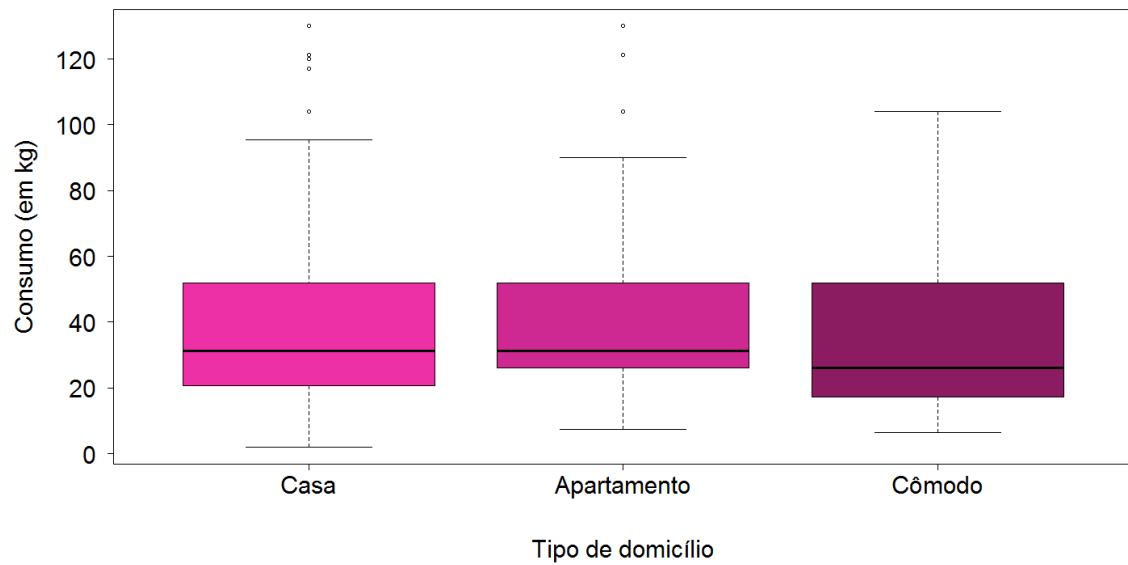


Figura A13 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo tipo de domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

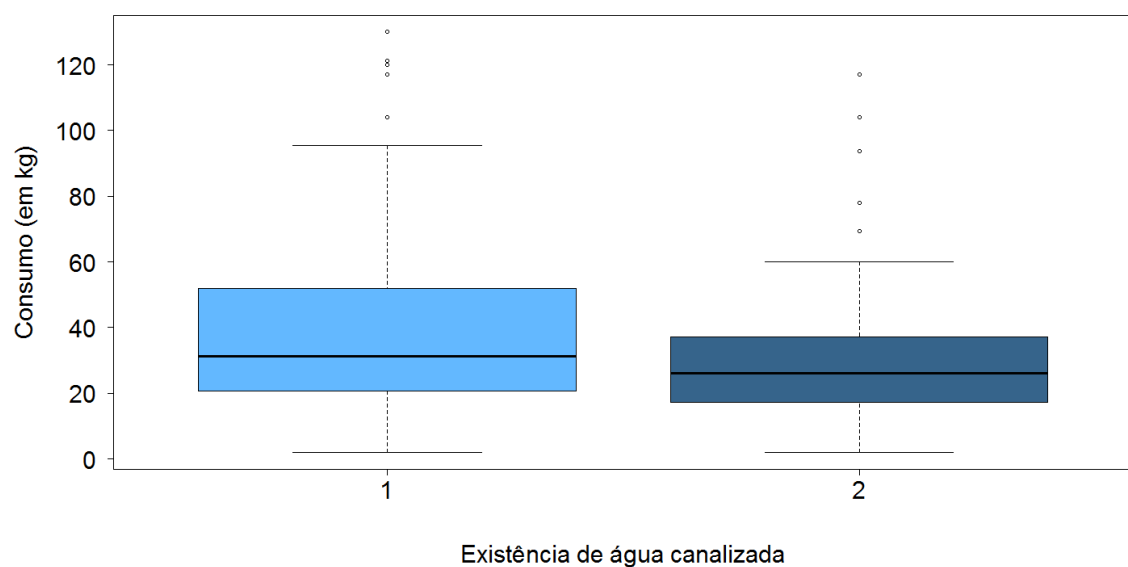


Figura A14 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de água canalizada no domicílio

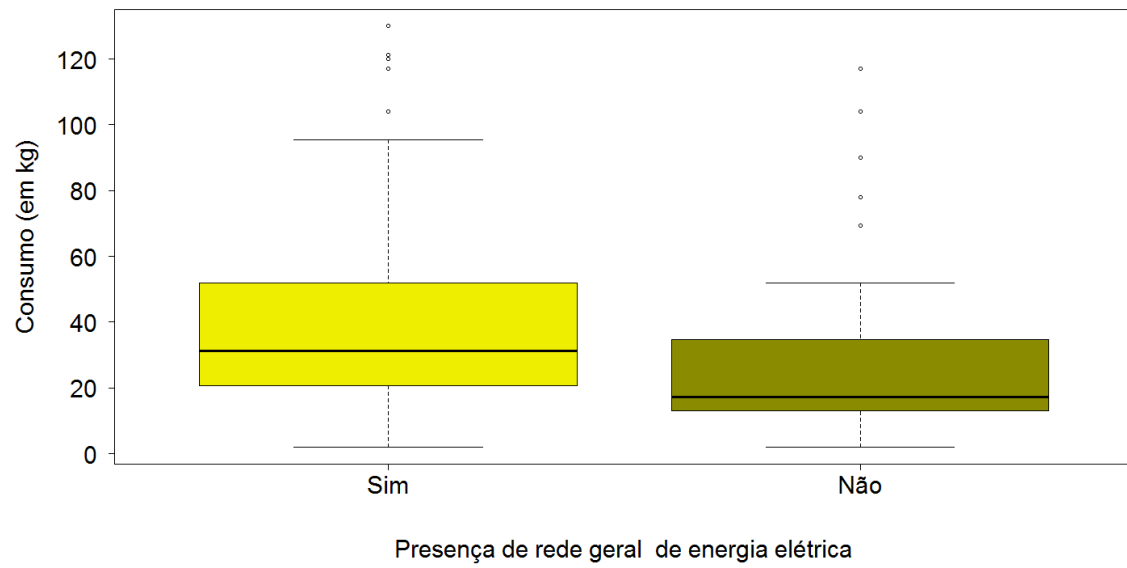


Figura A15 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de rede geral de energia elétrica no domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

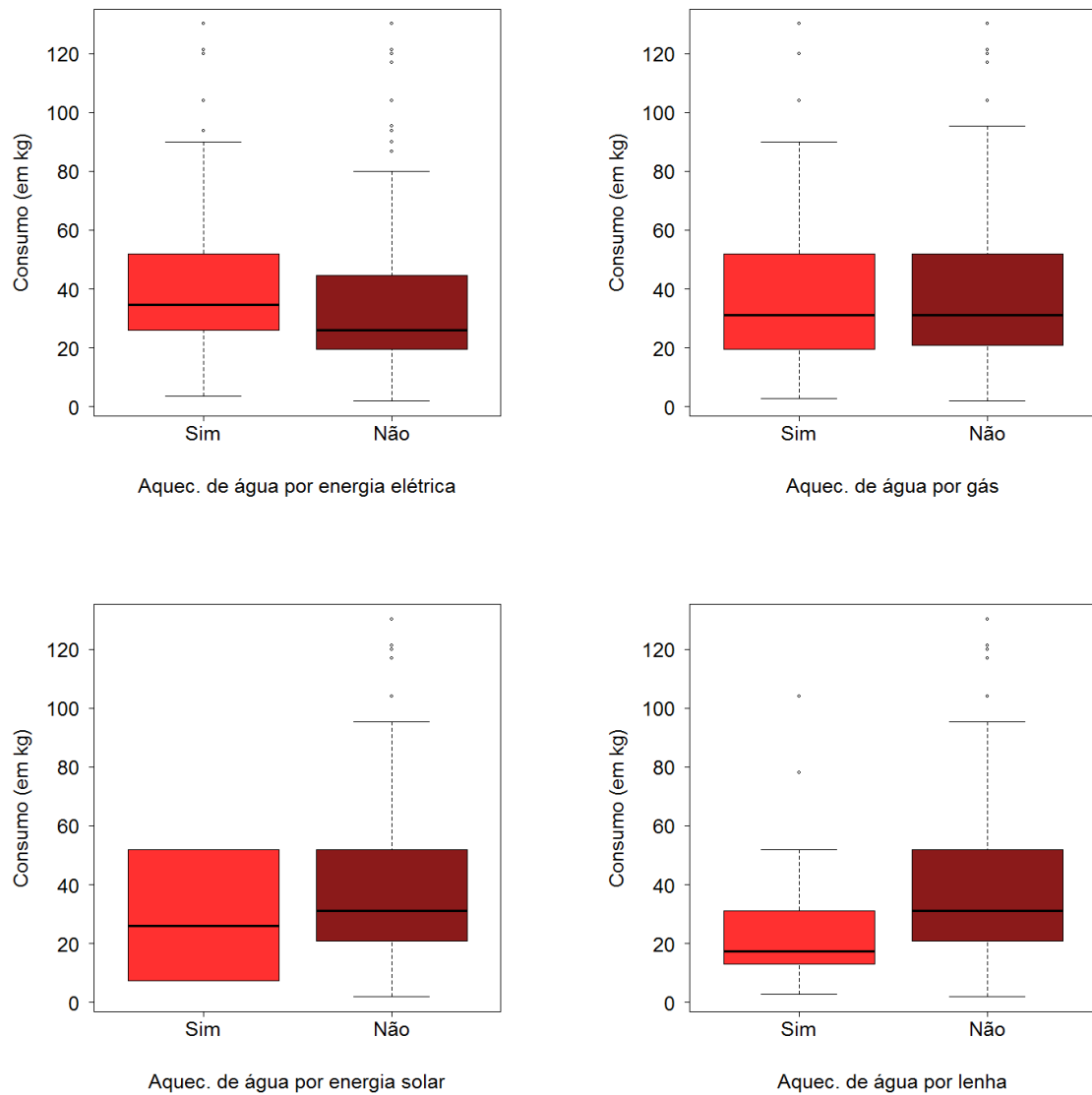


Figura A16 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de aquecimento de água por energia elétrica, gás, energia solar e lenha

Fonte: POF 2008-2009, IBGE. Elaboração própria

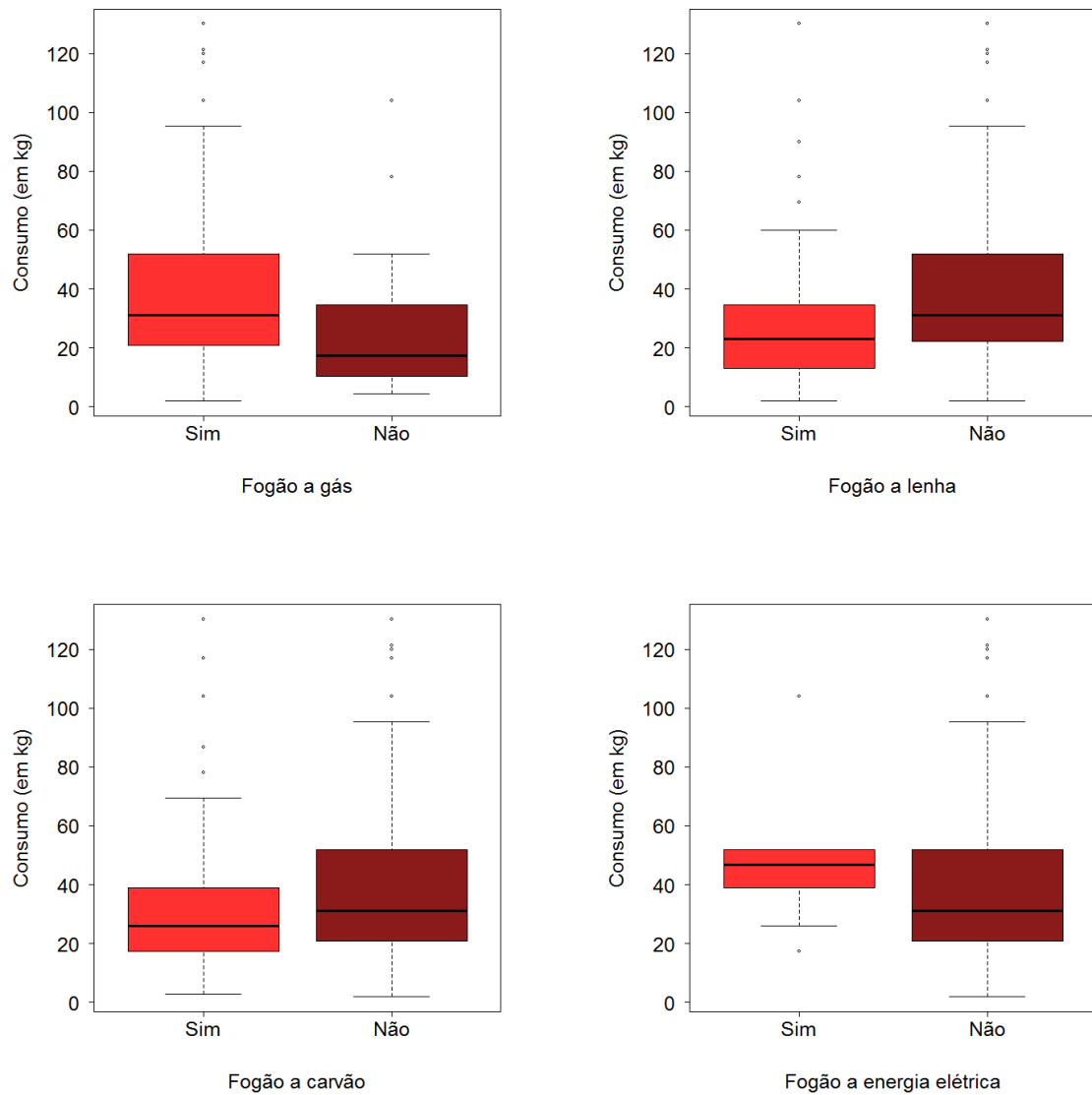


Figura A17 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de fogão a gás, a lenha, a carvão e a energia elétrica

Fonte: POF 2008-2009, IBGE. Elaboração própria

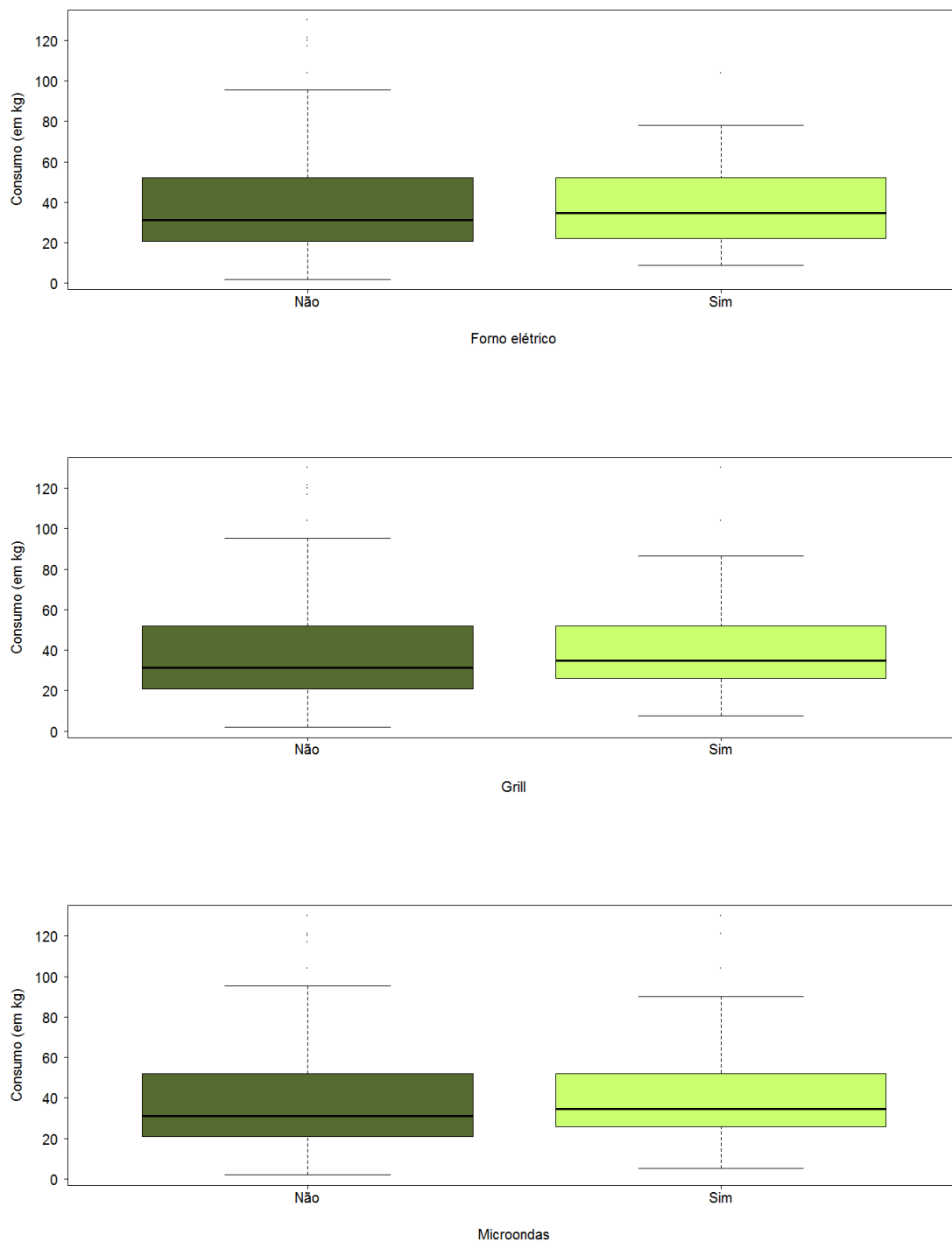


Figura A18 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de forno elétrico, *grill* e micro-ondas

Fonte: POF 2008-2009, IBGE. Elaboração própria

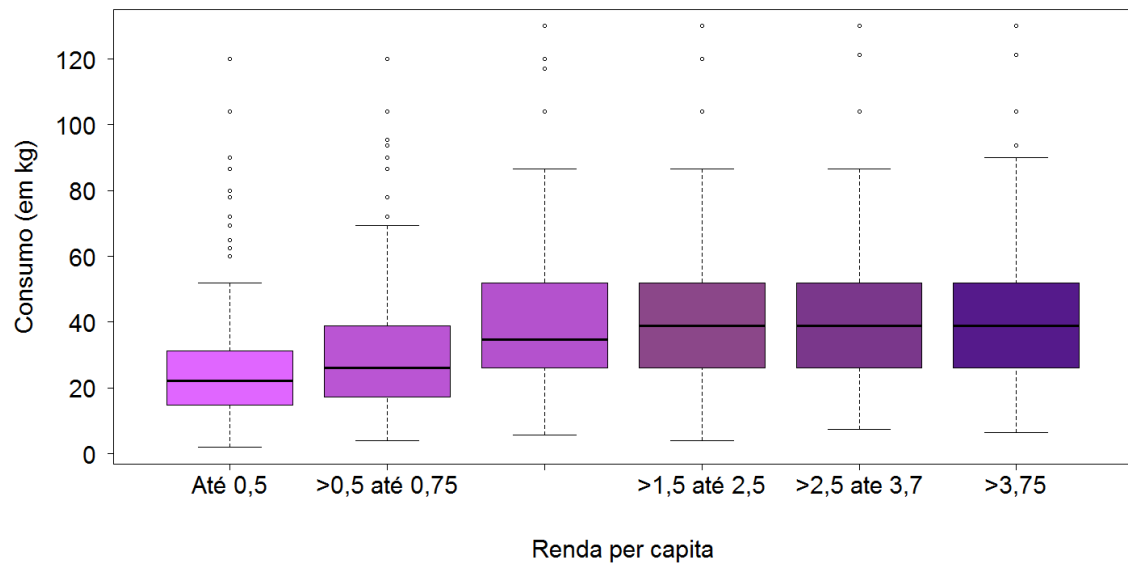


Figura A19 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de renda mensal per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

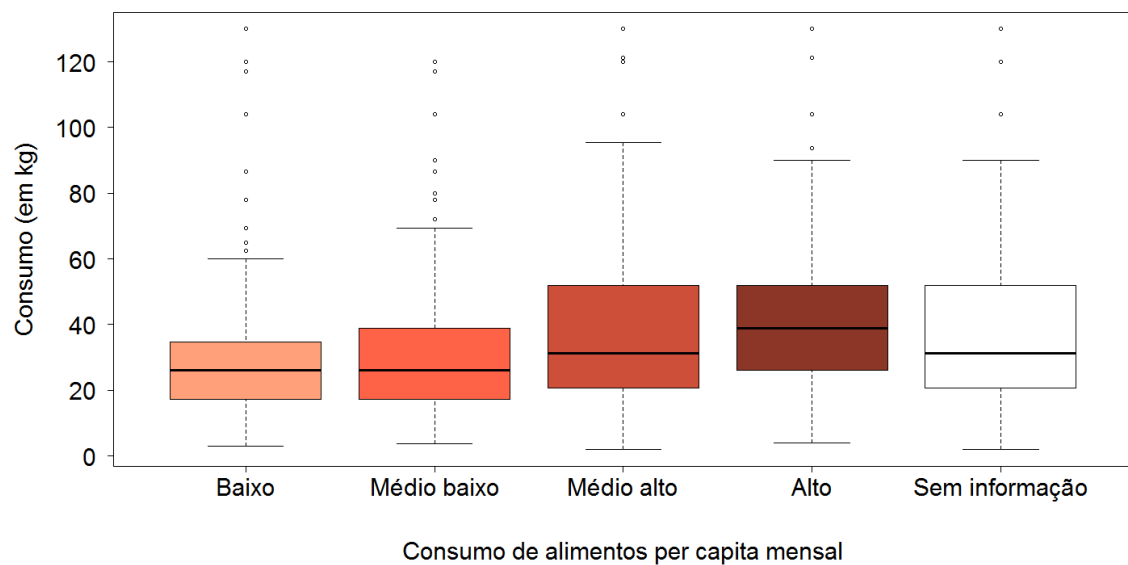


Figura A20 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de consumo de alimento per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

Região Norte

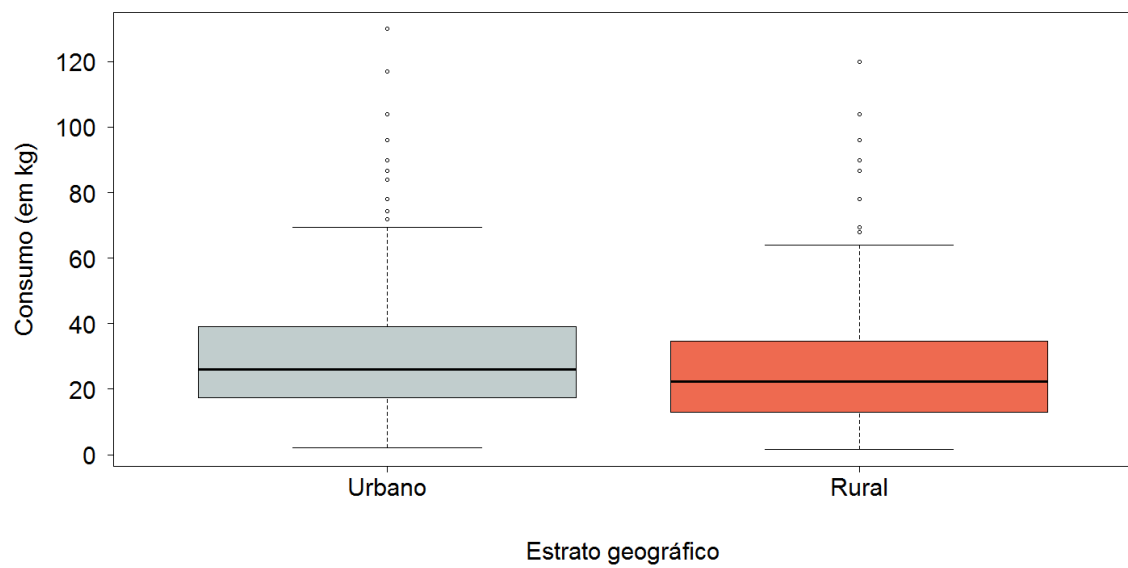


Figura A21 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo estrato geográfico

Fonte: POF 2008-2009, IBGE. Elaboração própria

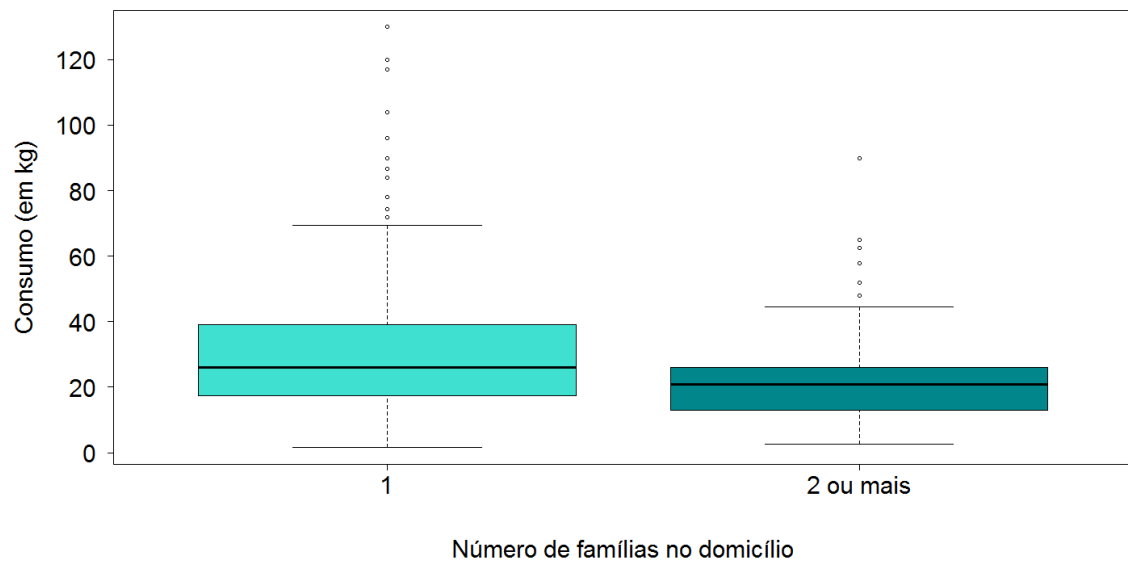


Figura A22 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo número de famílias por domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

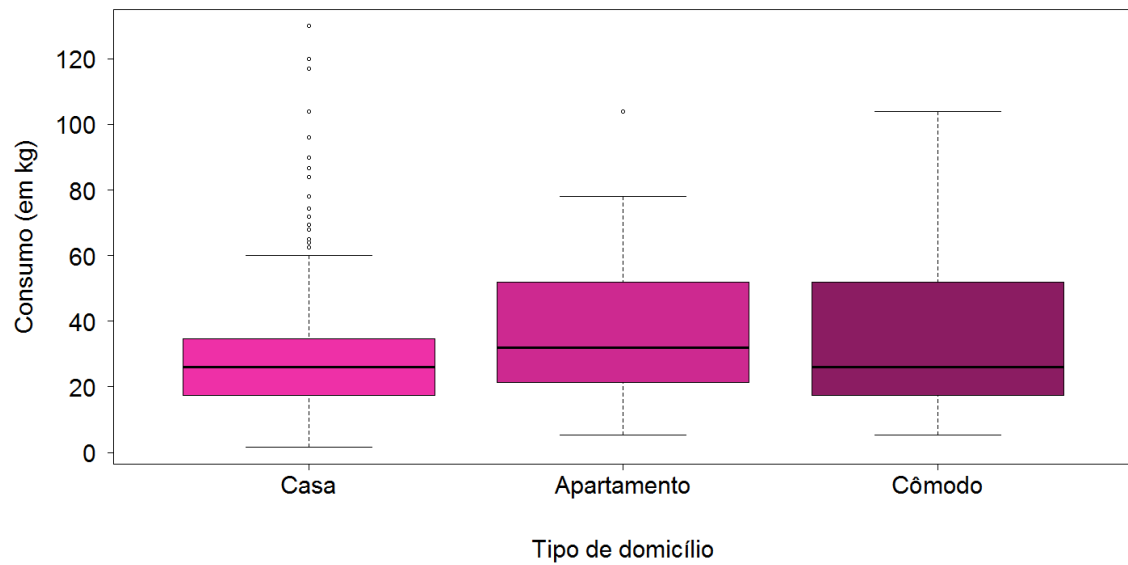


Figura A23 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo tipo de domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

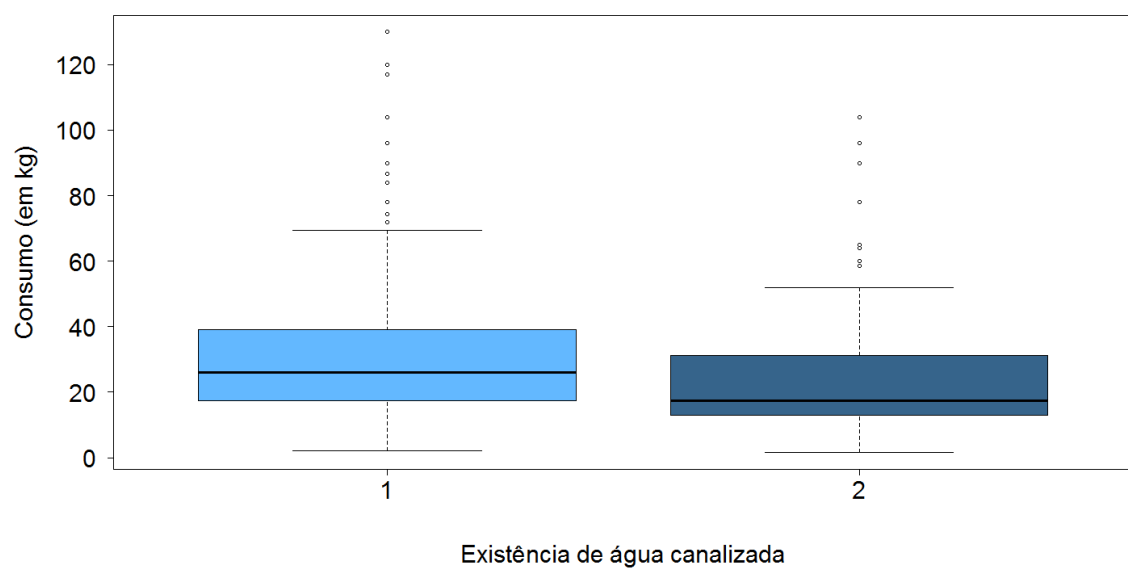


Figura A24 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de água canalizada no domicílio

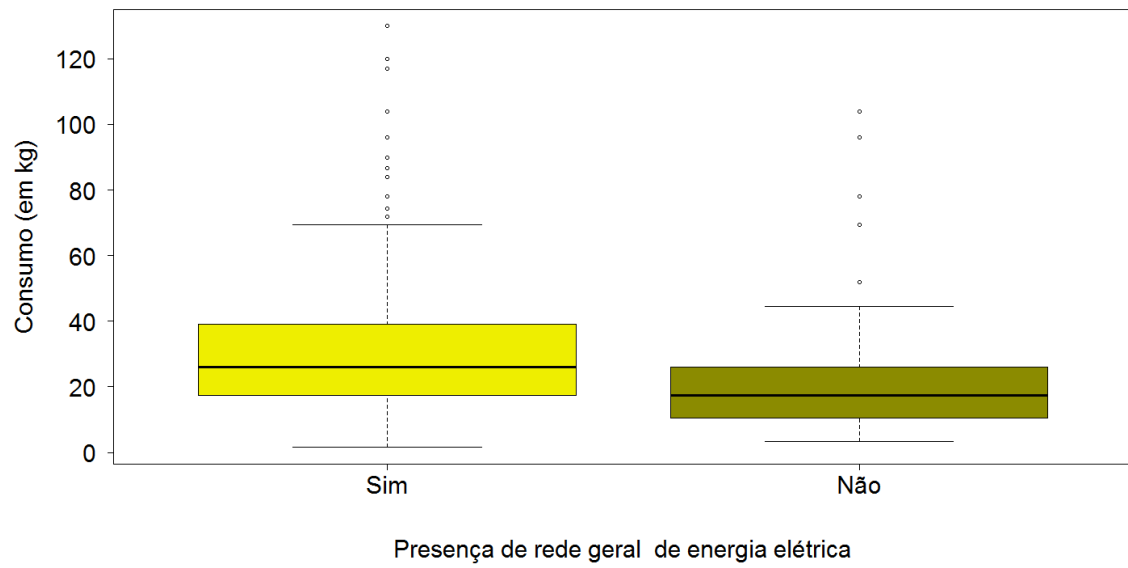


Figura A25 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de rede geral de energia elétrica no domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

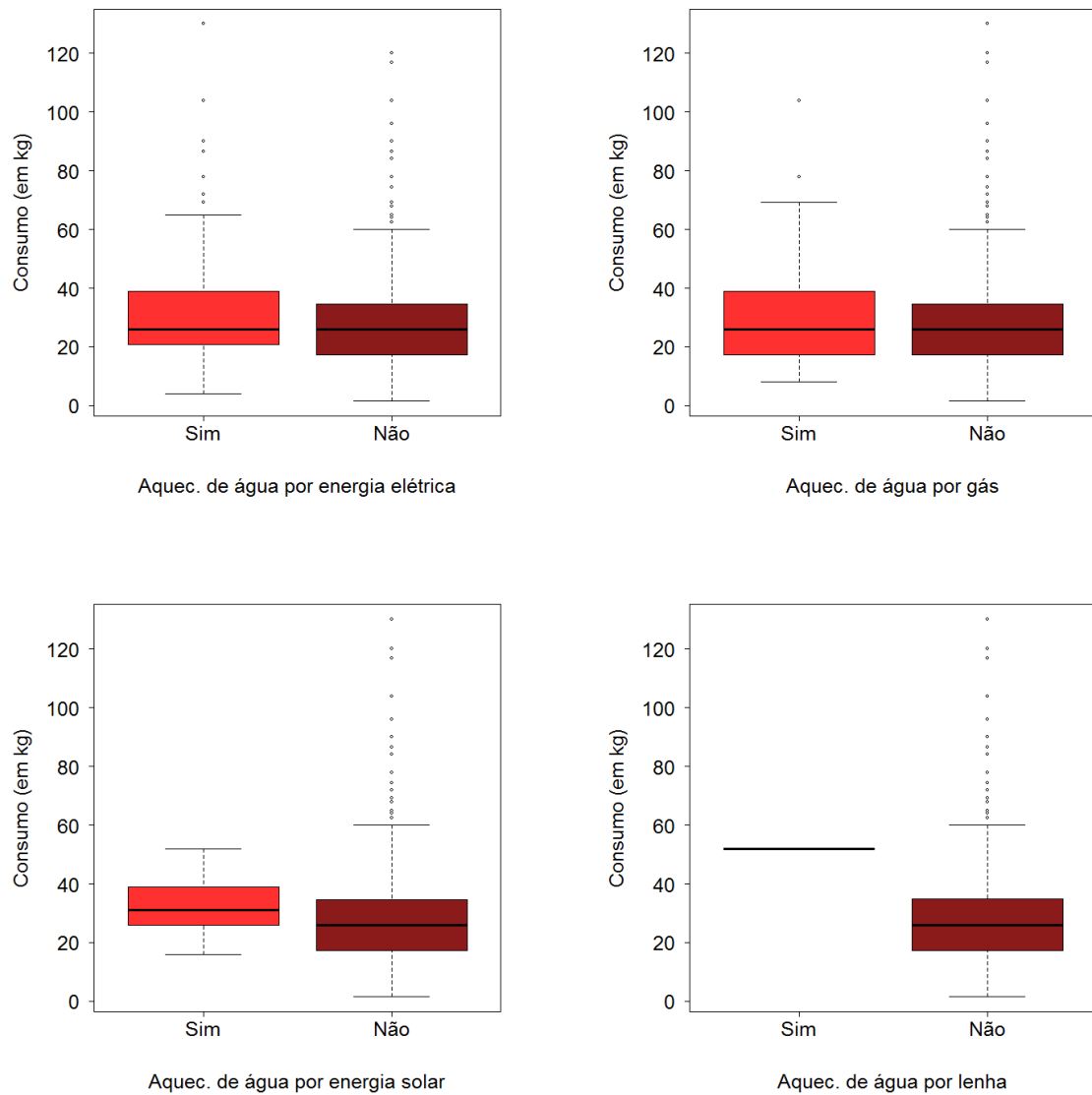


Figura A26 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de aquecimento de água por energia elétrica, gás, energia solar e lenha

Fonte: POF 2008-2009, IBGE. Elaboração própria

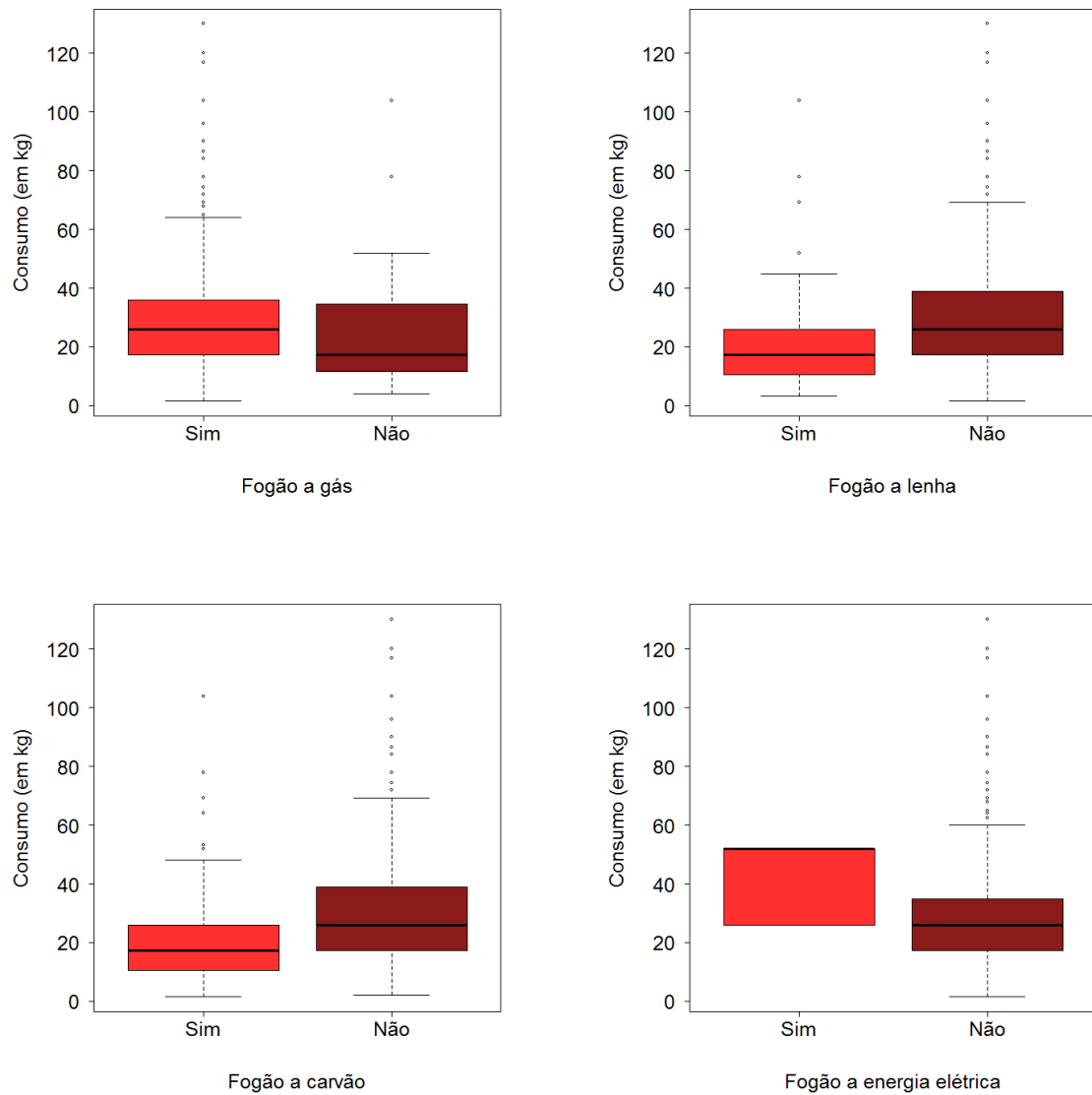


Figura A27 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de fogão a gás, a lenha, a carvão e a energia elétrica

Fonte: POF 2008-2009, IBGE. Elaboração própria

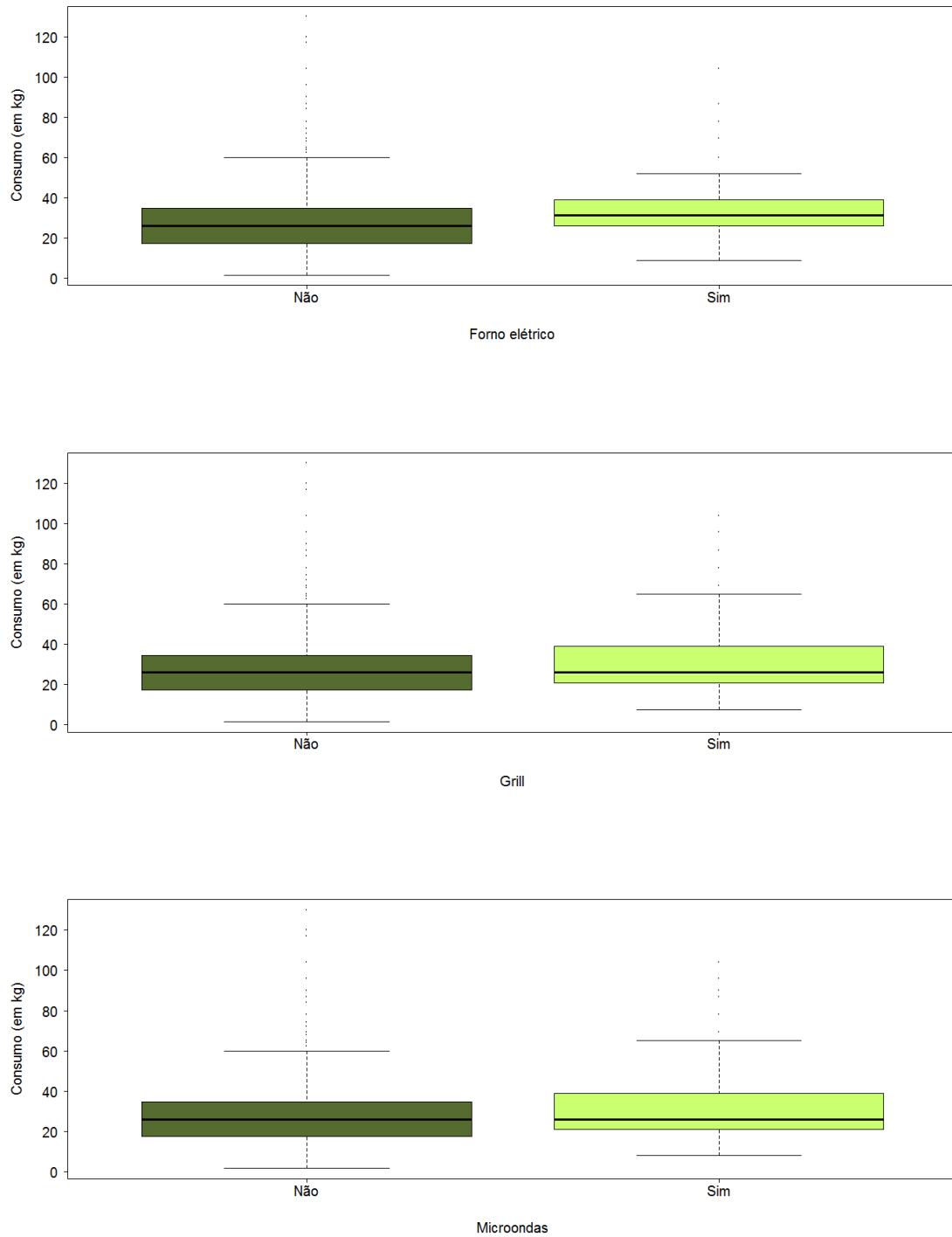


Figura A28 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de forno elétrico, *grill* e micro-ondas

Fonte: POF 2008-2009, IBGE. Elaboração própria

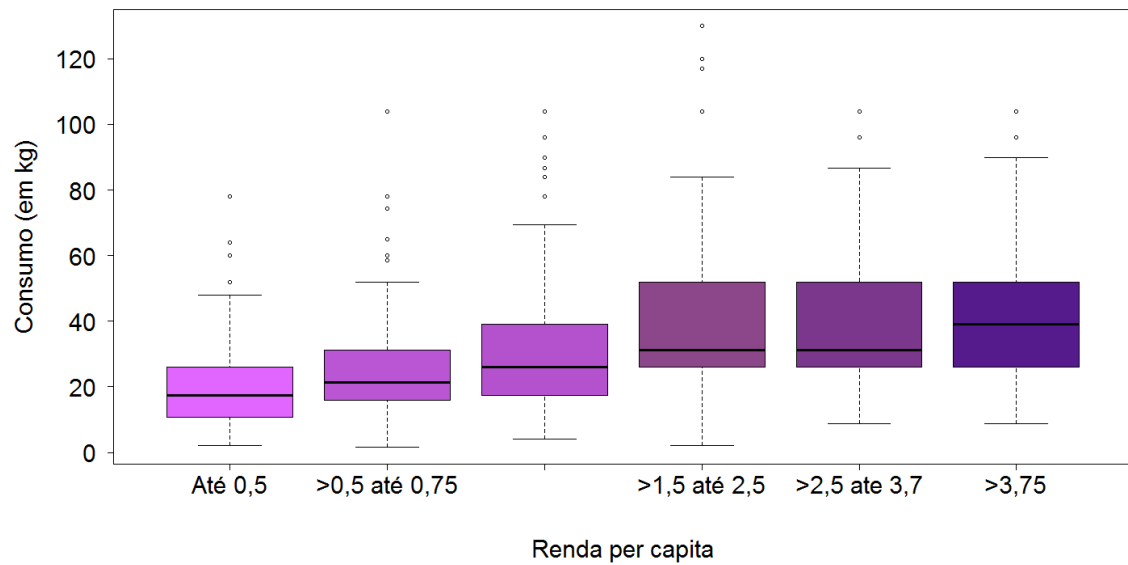


Figura A29 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de renda mensal per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

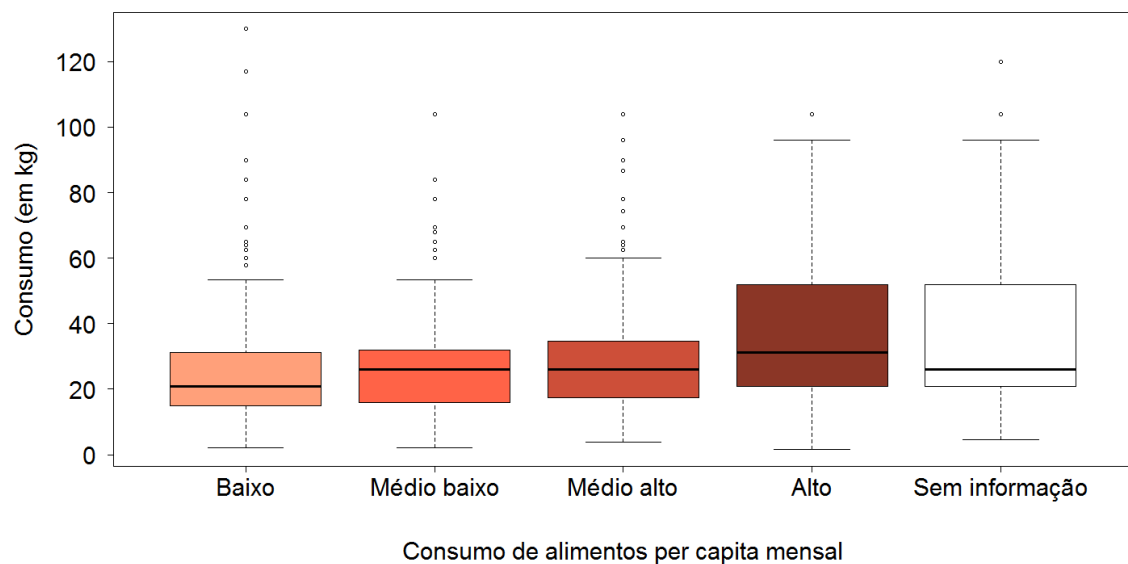


Figura A30 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de consumo de alimento per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

Região Sudeste

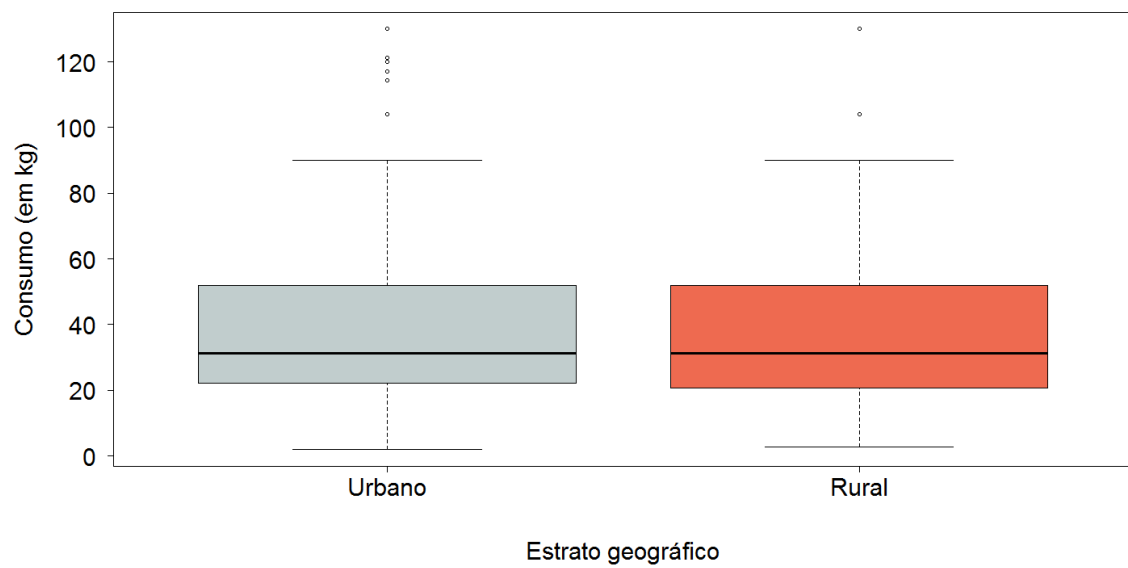


Figura A31 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo estrato geográfico

Fonte: POF 2008-2009, IBGE. Elaboração própria

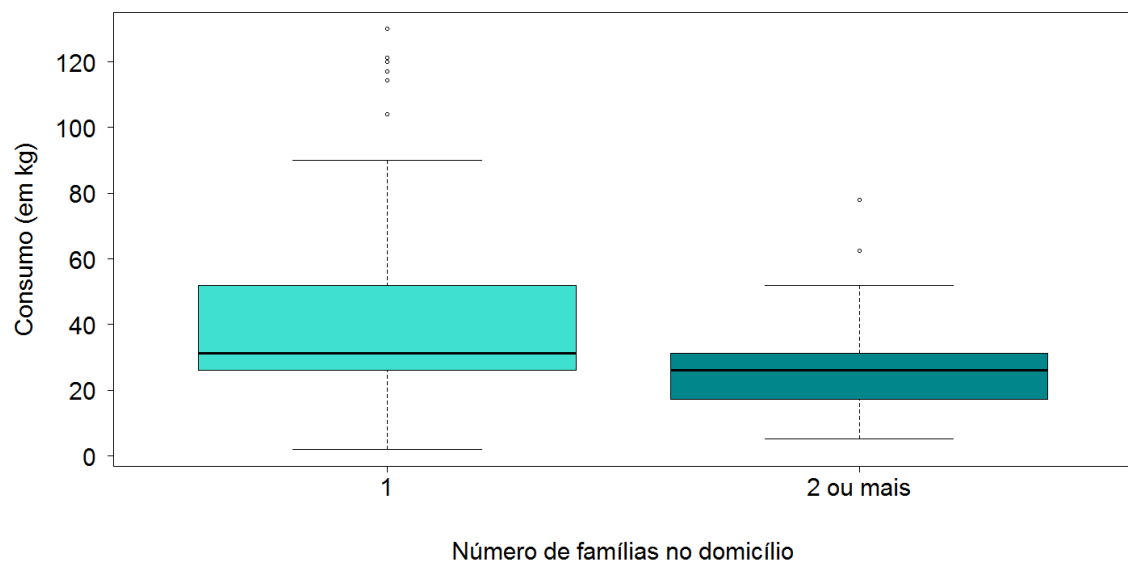


Figura A32 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo número de famílias por domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

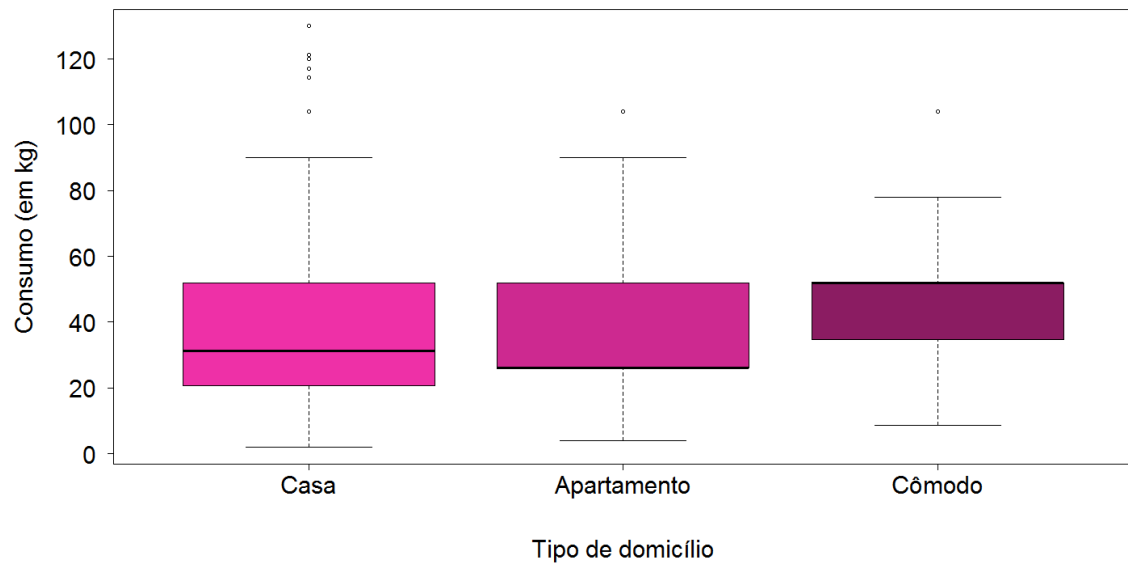


Figura A33 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo tipo de domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

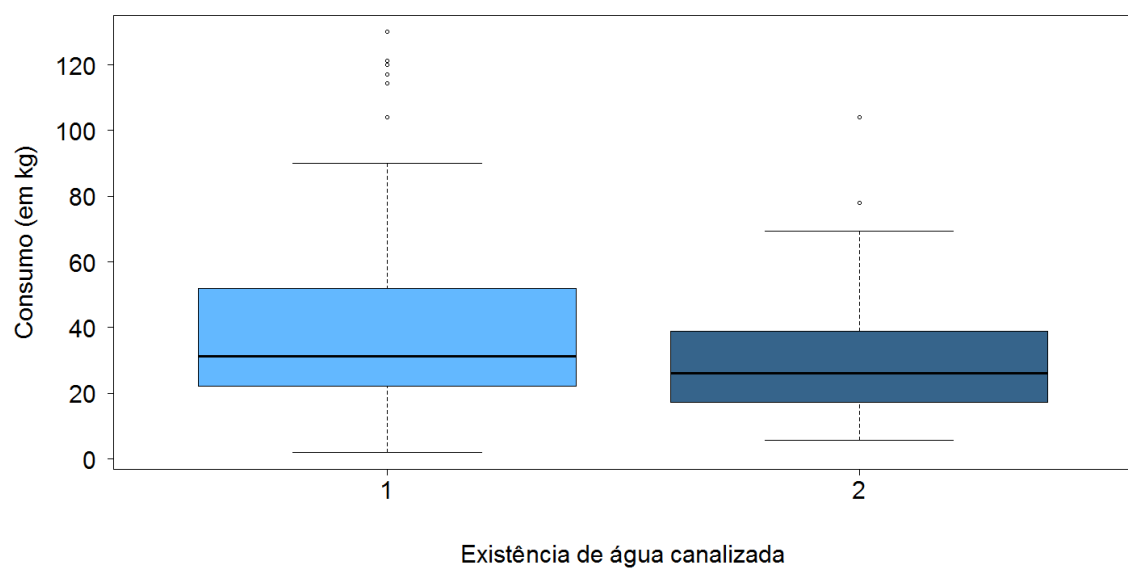


Figura A34 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de água canalizada no domicílio

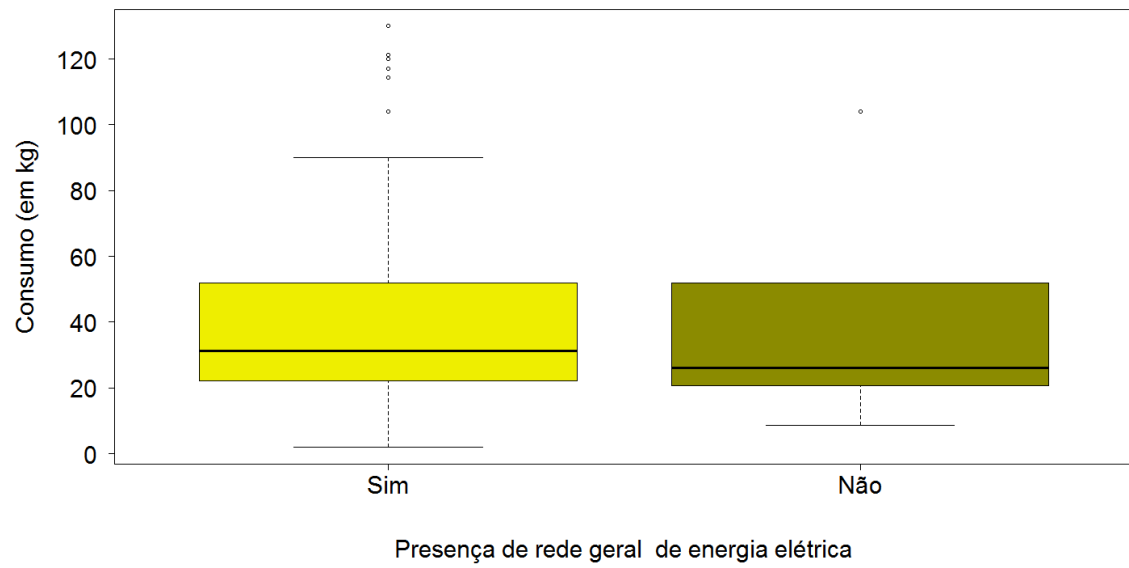


Figura A35 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de rede geral de energia elétrica no domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

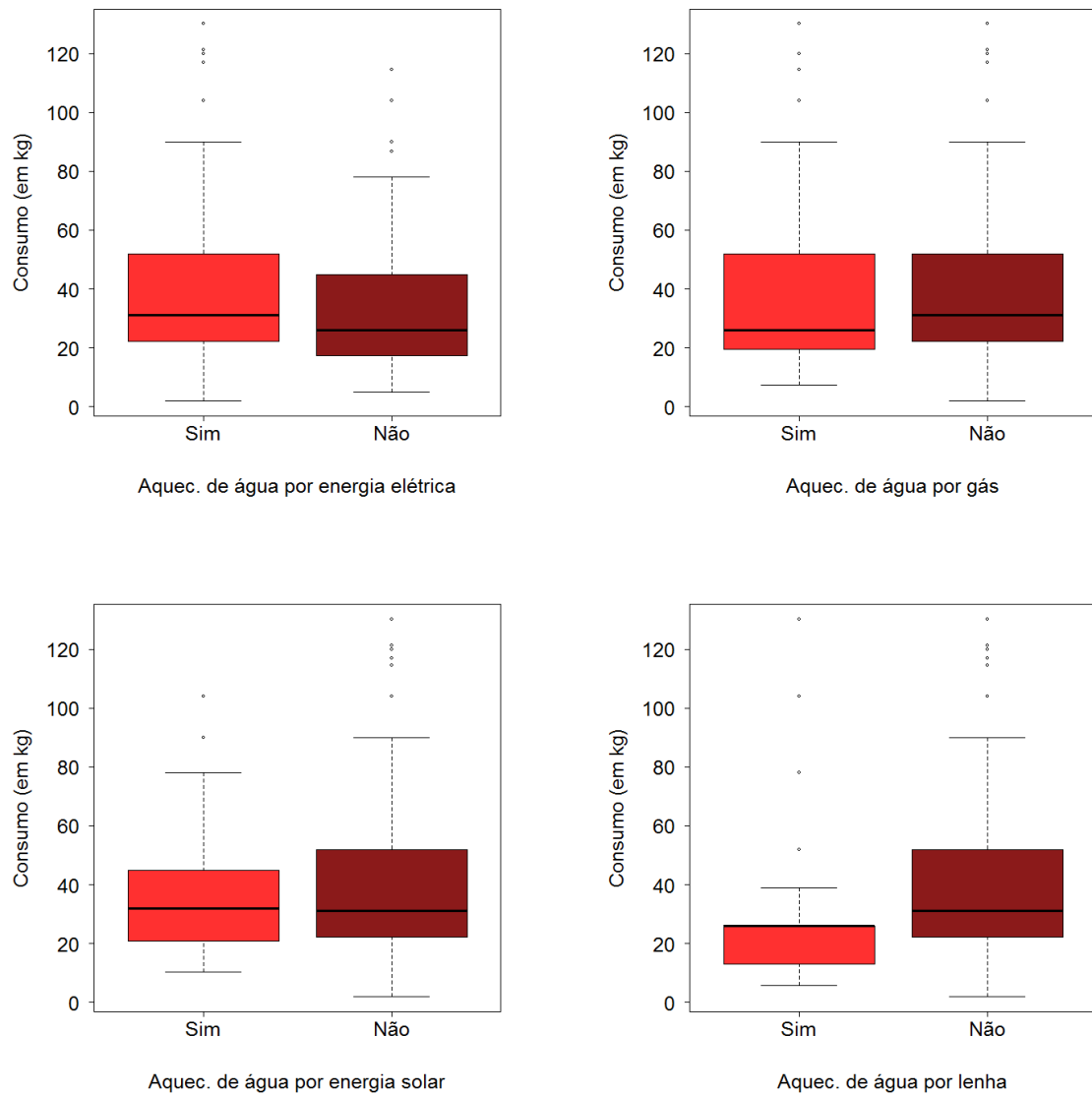


Figura A36 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de aquecimento de água por energia elétrica, gás, energia solar e lenha

Fonte: POF 2008-2009, IBGE. Elaboração própria

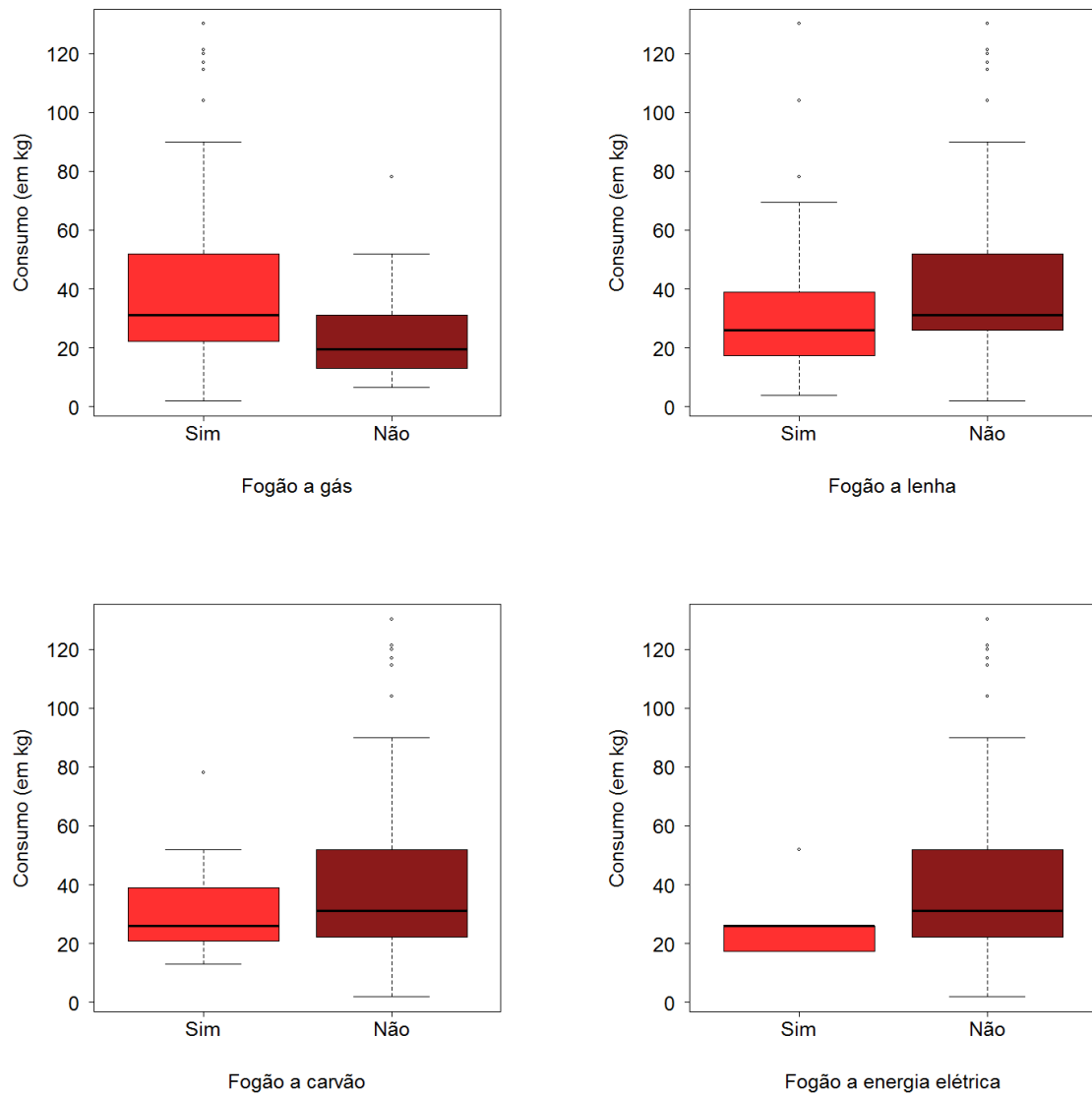


Figura A37 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de fogão a gás, a lenha, a carvão e a energia elétrica

Fonte: POF 2008-2009, IBGE. Elaboração própria

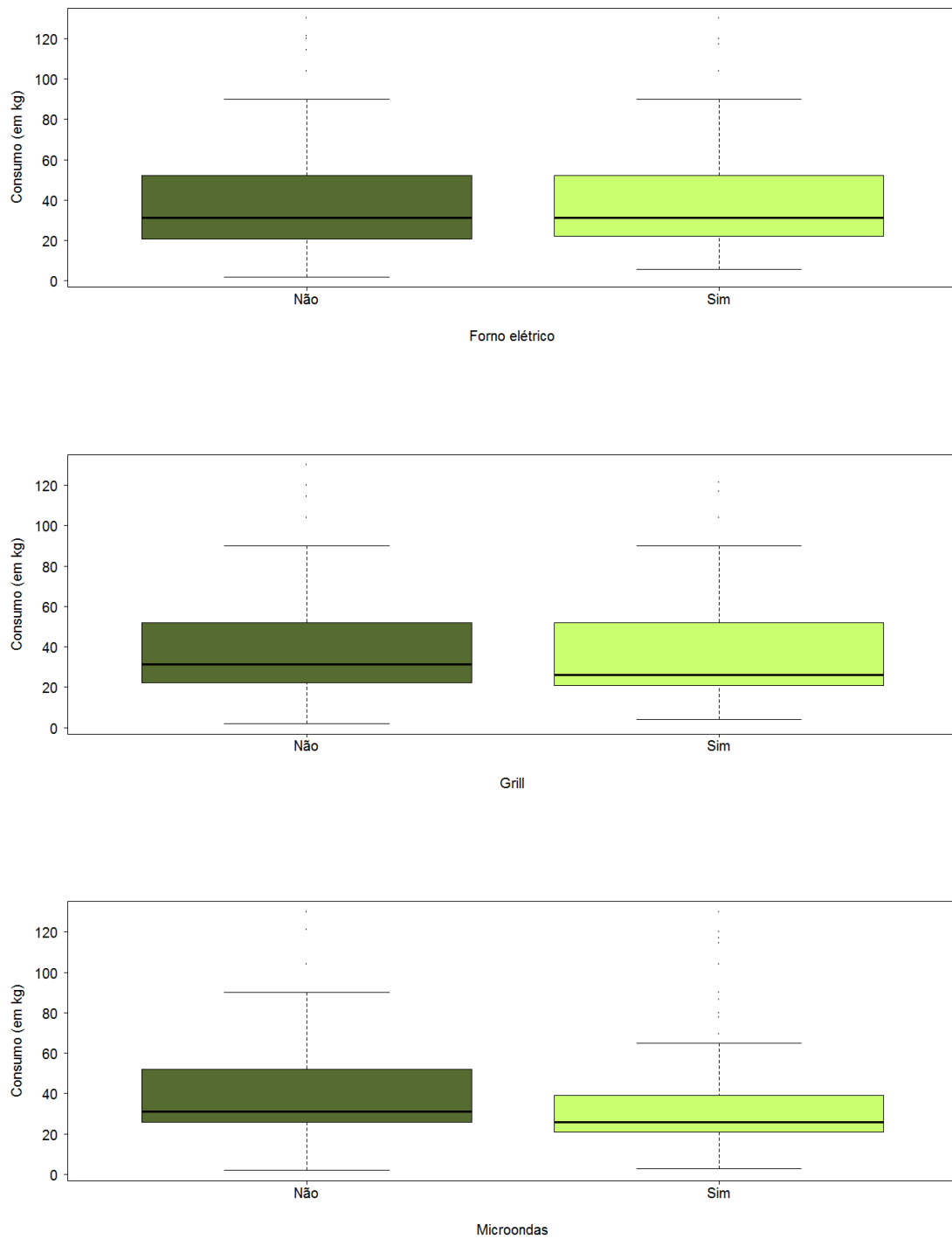


Figura A38 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de forno elétrico, *grill* e micro-ondas

Fonte: POF 2008-2009, IBGE. Elaboração própria

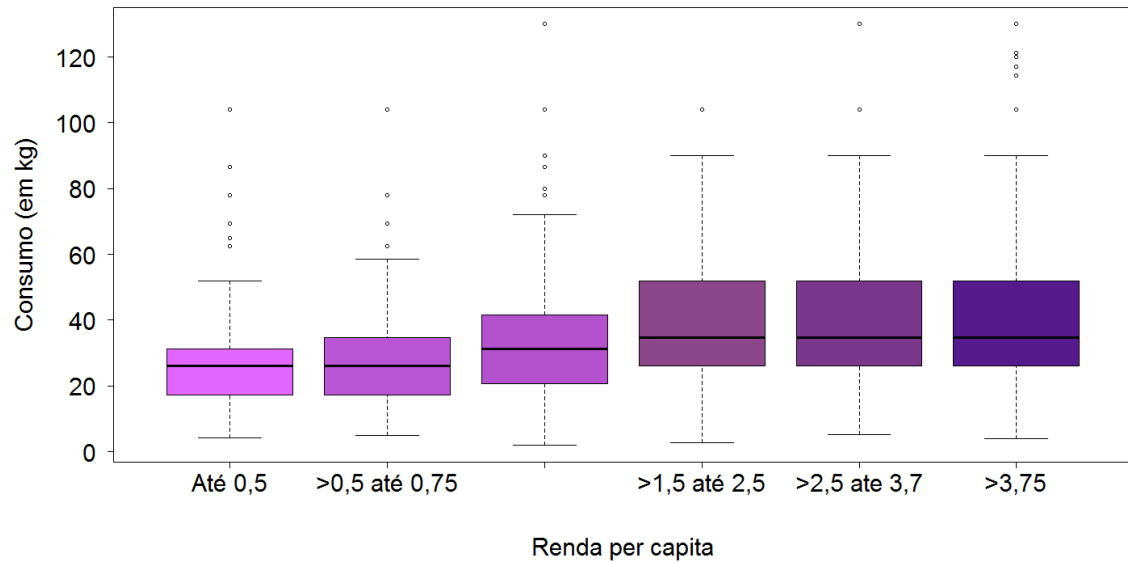


Figura A39 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de renda mensal per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

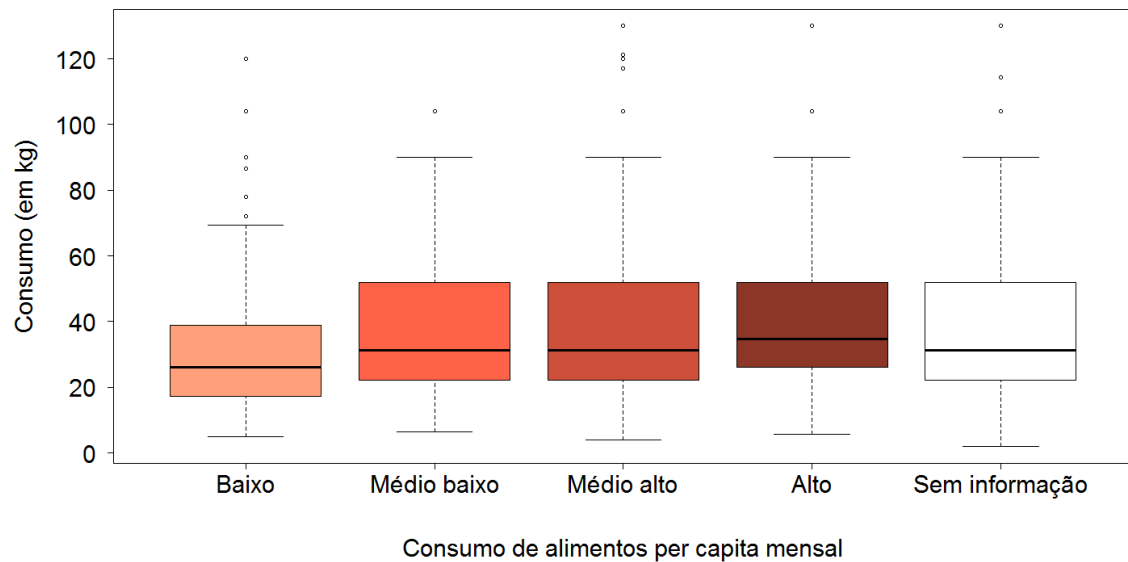


Figura A40 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de consumo de alimento per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

Região Sul

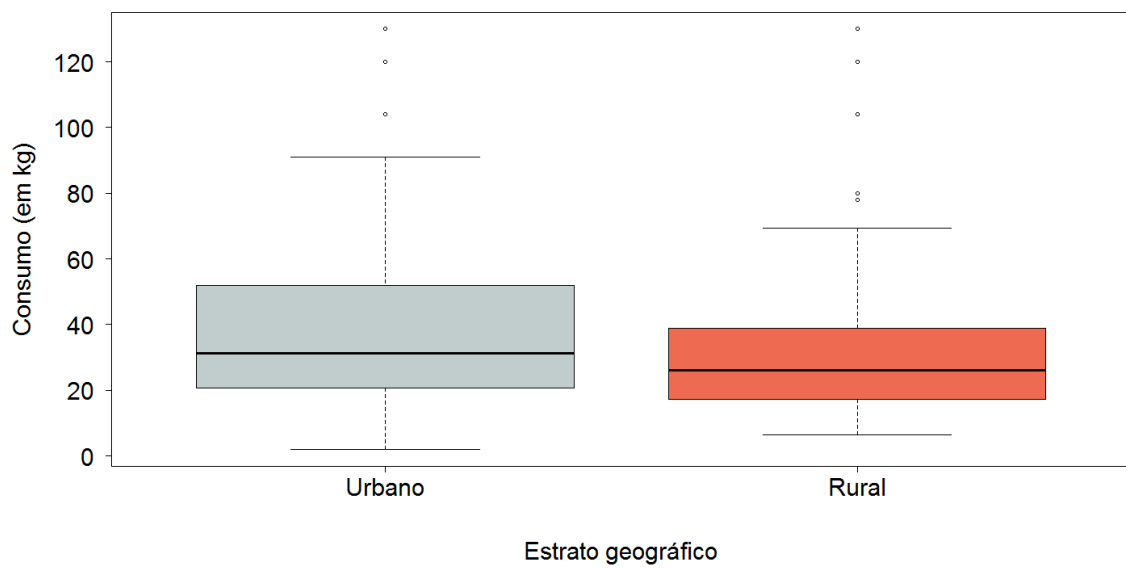


Figura A41 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo estrato geográfico

Fonte: POF 2008-2009, IBGE. Elaboração própria

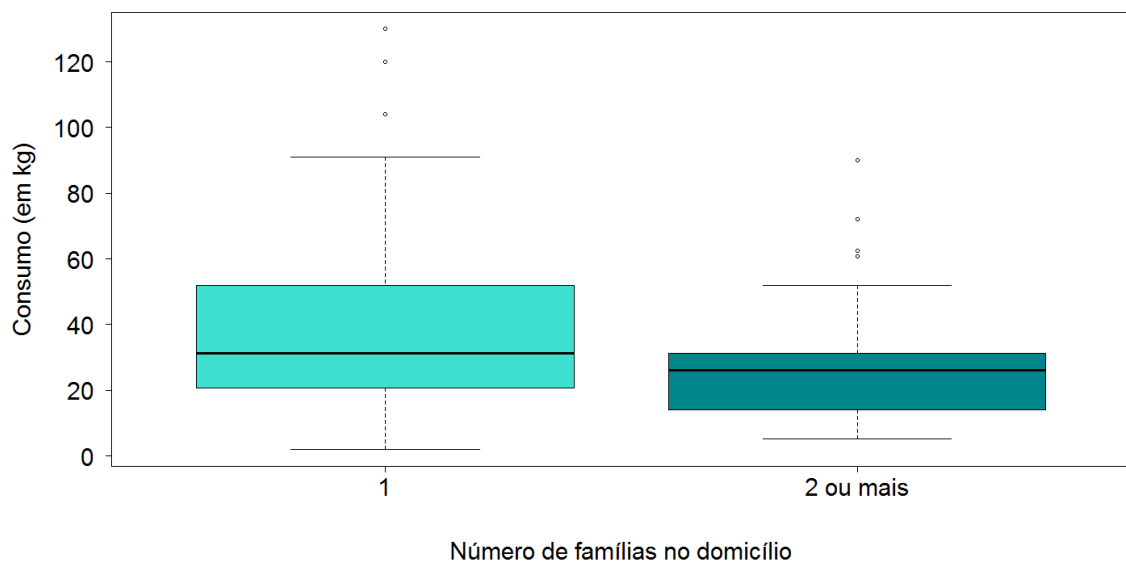


Figura A42 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo número de famílias por domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

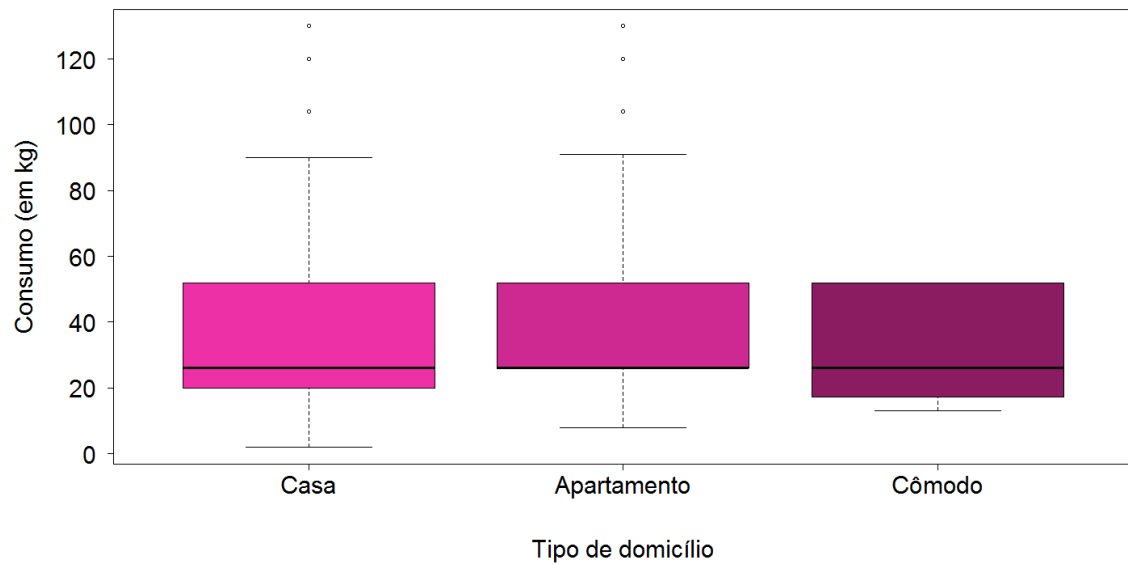


Figura A43 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo tipo de domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

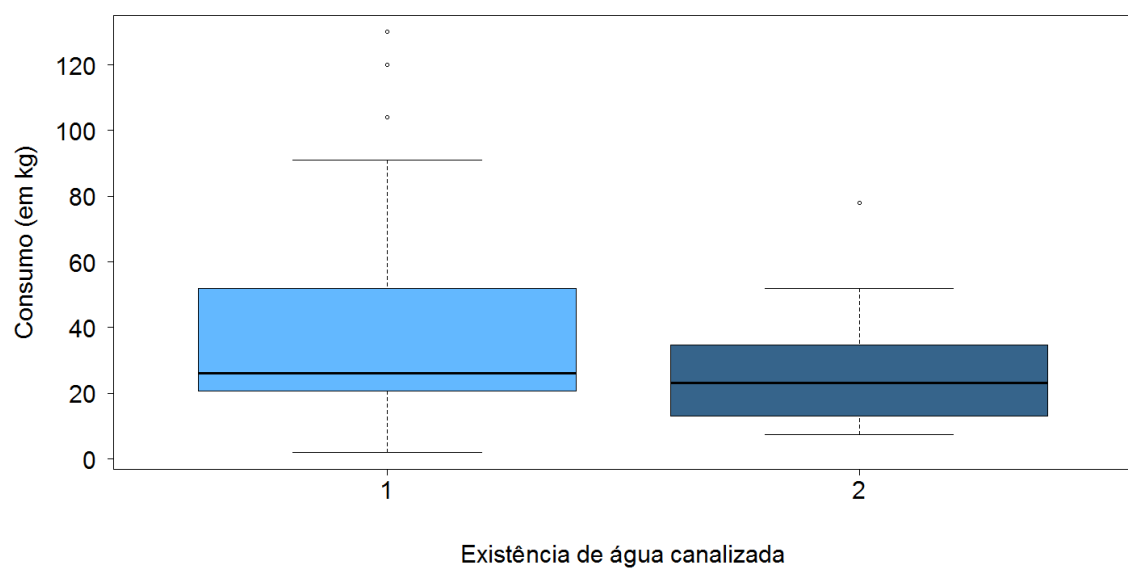


Figura A44 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de água canalizada no domicílio

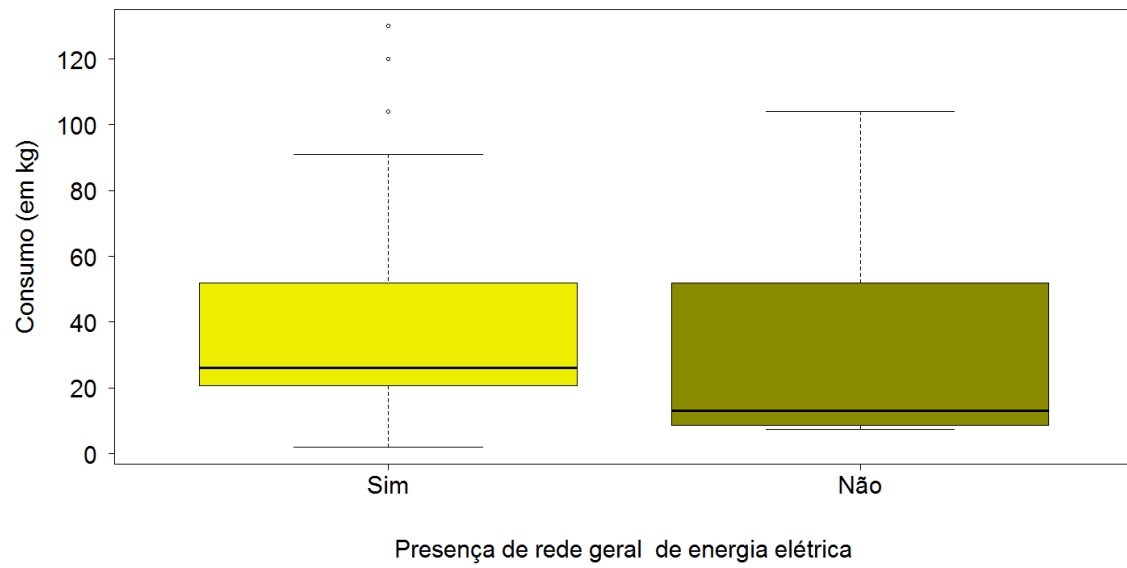


Figura A45 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de rede geral de energia elétrica no domicílio

Fonte: POF 2008-2009, IBGE. Elaboração própria

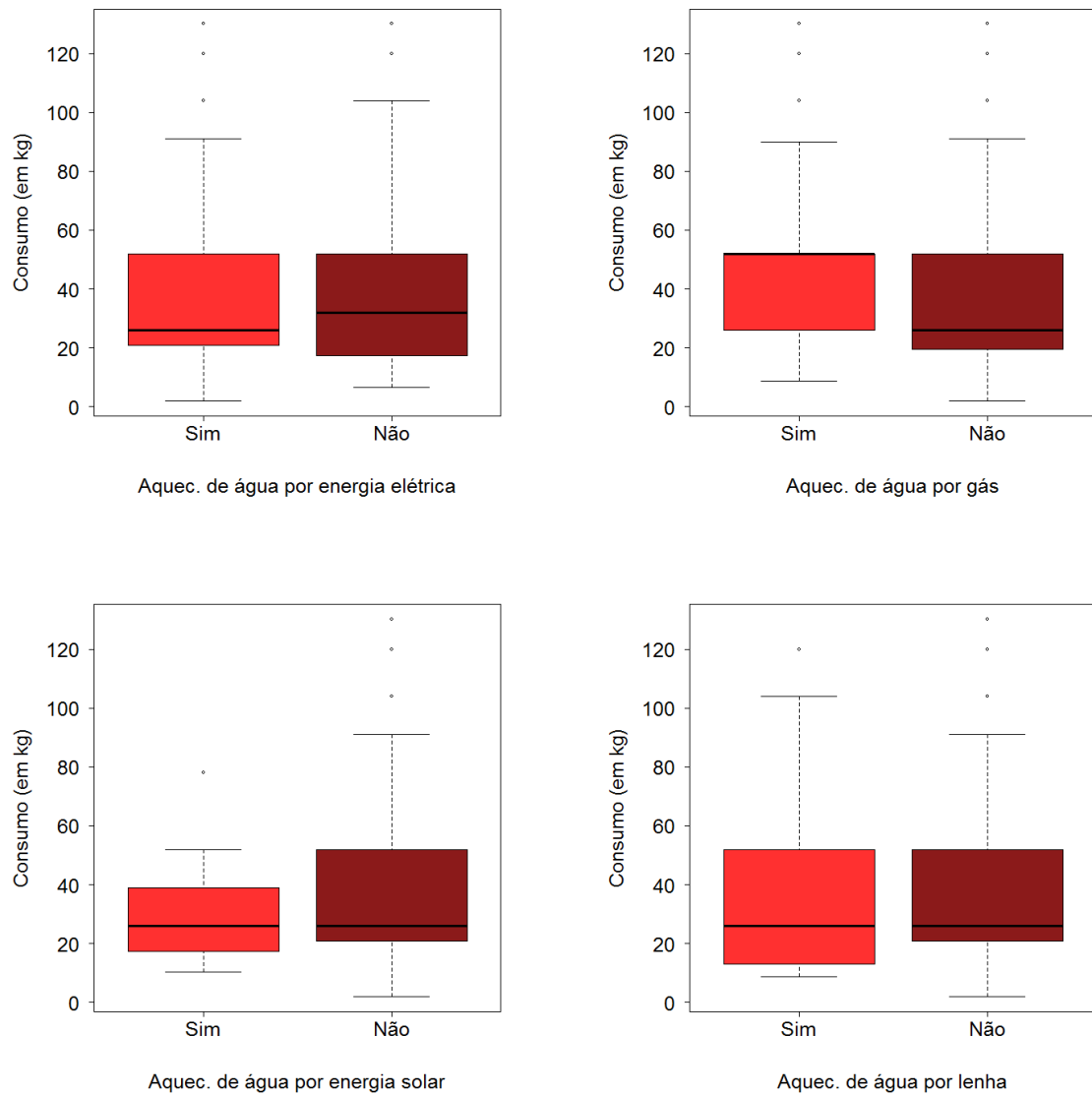


Figura A46 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de aquecimento de água por energia elétrica, gás, energia solar e lenha

Fonte: POF 2008-2009, IBGE. Elaboração própria

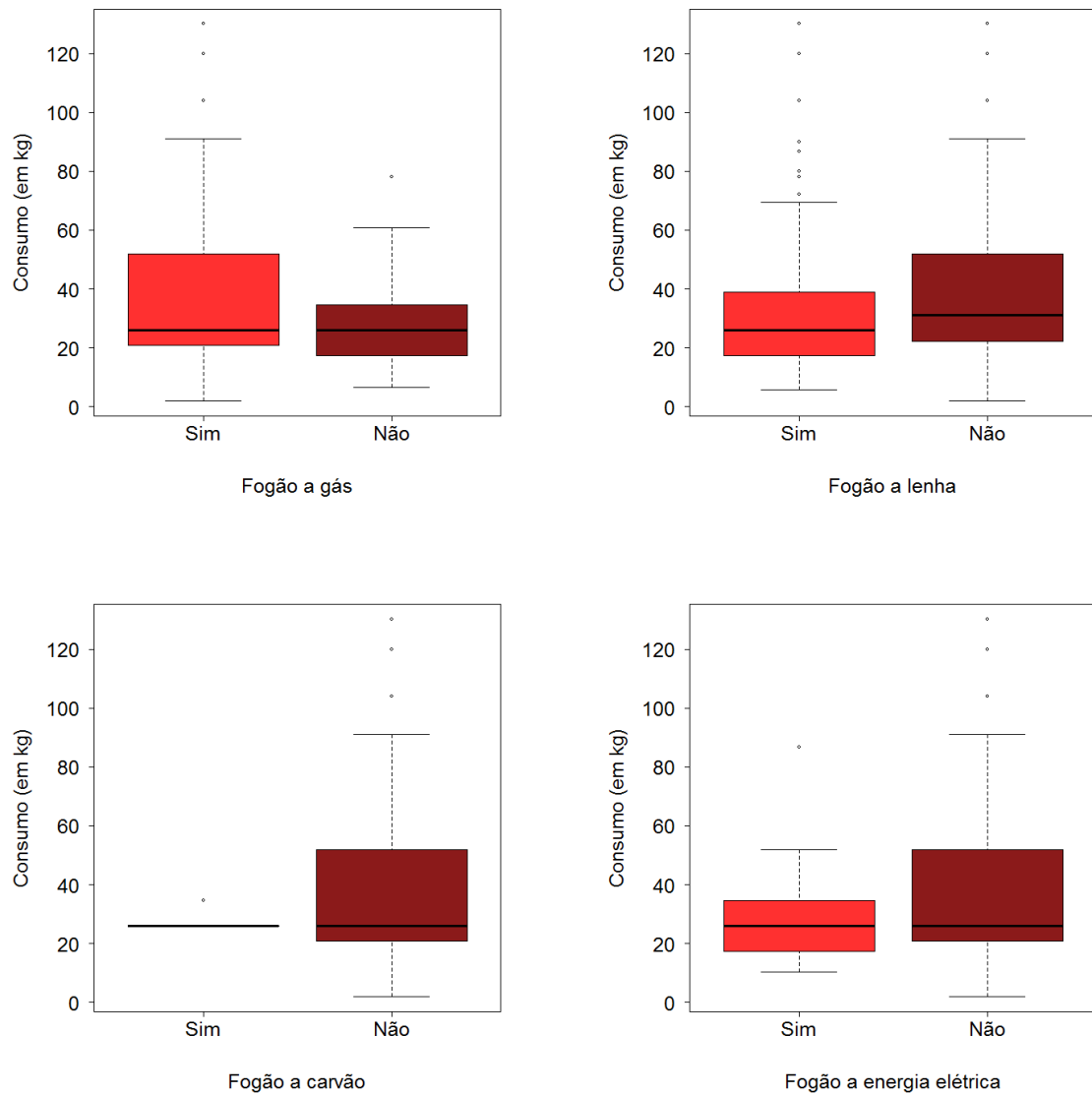


Figura A47 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de fogão a gás, a lenha, a carvão e a energia elétrica

Fonte: POF 2008-2009, IBGE. Elaboração própria

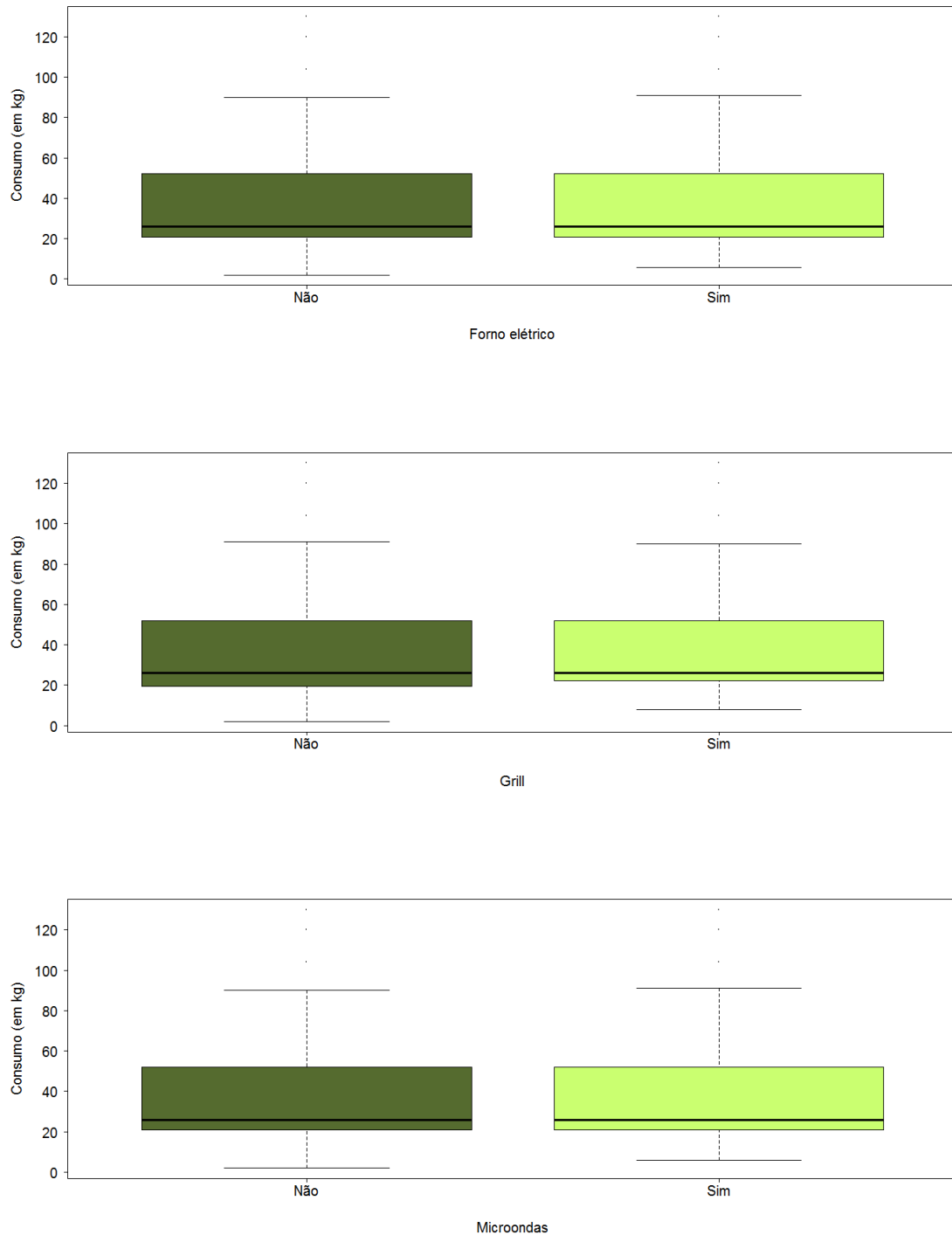


Figura A48 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo a existência de forno elétrico, *grill* e micro-ondas

Fonte: POF 2008-2009, IBGE. Elaboração própria

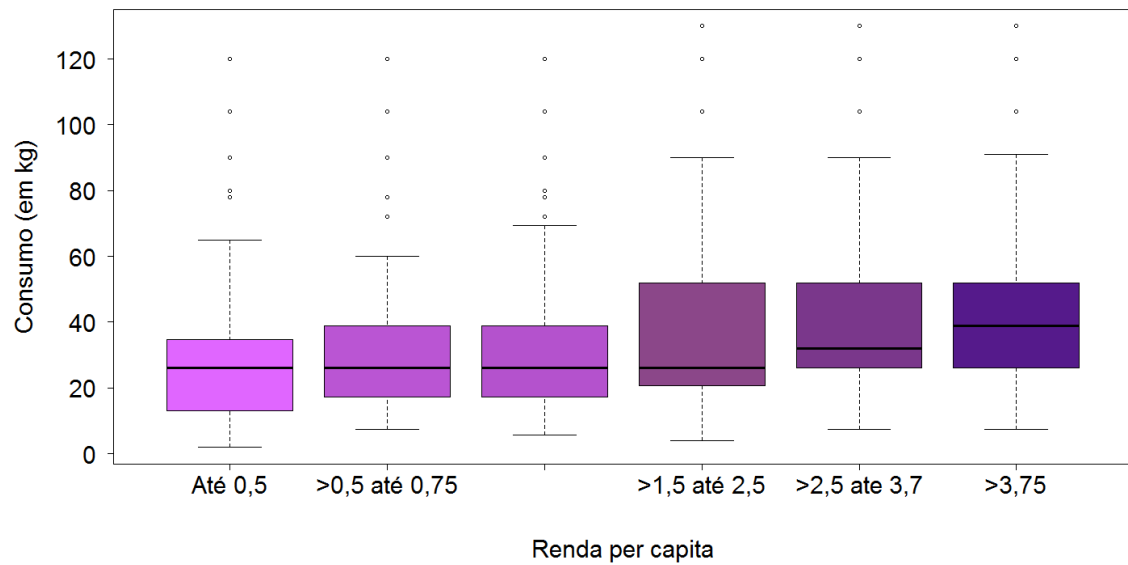


Figura A49 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de renda mensal per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

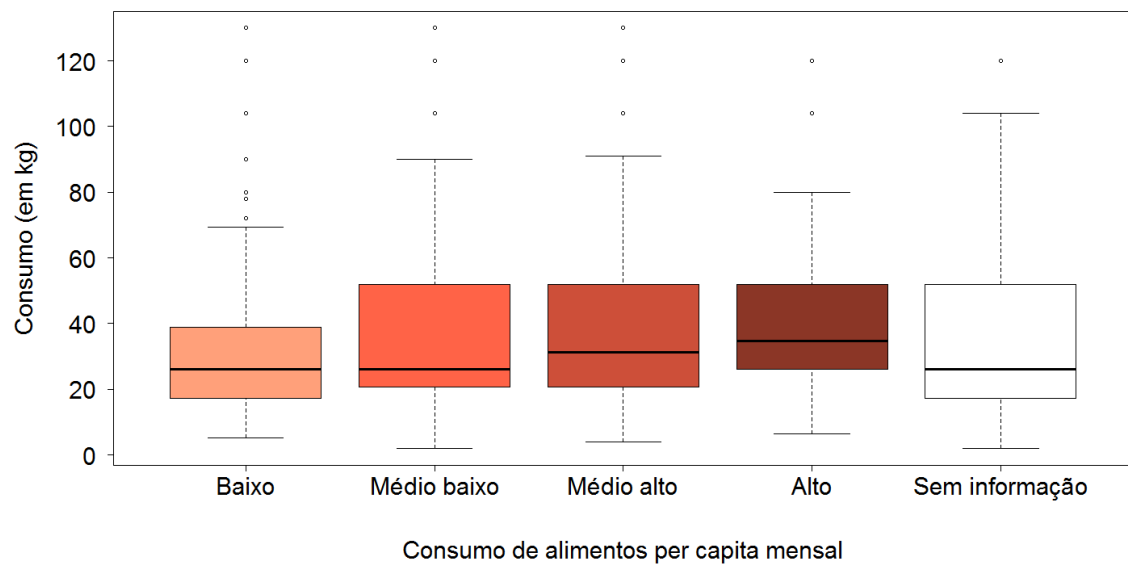


Figura A50 – *Boxplots* do consumo de gás anual domiciliar per capita (em kg), segundo as categorias de consumo de alimento per capita

Fonte: POF 2008-2009, IBGE. Elaboração própria

Anexo 1 – Rotina em R para a extração da base de dados

```

library(bit64)
library(data.table)
library(descr)
library(xtable)
library(reshape2)
library(plyr)
library(dplyr)
library(tidyr)
library(pastecs)
library(robustbase)
library(ggplot2)
library(grid)
library(gridExtra)
library(Rcpp)
library(png)

options(OutDec= ",")
options(stringsAsFactors=FALSE)
options(scipen=999)

#####
#####

####POF_P1_Q2 E POF_P2_Q14

## Criando o dicionário a partir das três primeiras colunas da planilha
dic_pof_p1_q2 <- read.csv(file = 'DOMICILIO_POF1_QUADRO2.csv', header=F, sep=";")
dic_pof_p1_q2 <- dic_pof_p1_q2[complete.cases(dic_pof_p1_q2),]
colnames(dic_pof_p1_q2) <- c('variavel', 'tamanho', 'inicio')

dic_pof_p2_q14 <- read.csv(file = 'INVENTARIO_POF2_QUADRO14.csv', header=F, sep=";")
dic_pof_p2_q14 <- dic_pof_p2_q14[complete.cases(dic_pof_p2_q14),]
colnames(dic_pof_p2_q14) <- c('variavel', 'tamanho', 'inicio')

## Parâmetro com o final de cada campo
end_dom = dic_pof_p1_q2$inicio + dic_pof_p1_q2$tamanho - 1
end_inv = dic_pof_p2_q14$inicio + dic_pof_p2_q14$tamanho - 1

#### NESTE PONTO O BANCO DE DADOS BAIXADOS DA POF (ARQUIVO TXT - FORAM UTILIZADOS
####OS NOMES ORIGINAIS) É LIDO E SALVO EM .CSV

## Converte o microdado para um arquivo csv
fwf2csv(fwffile='Dados/T_DOMICILIO_S.txt', csvfile='dom_p1_q2.csv',
        names=dic_pof_p1_q2$variavel, begin=dic_pof_p1_q2$inicio, end=end_dom)

fwf2csv(fwffile='Dados/T_INVENTARIO_S.txt', csvfile='inv_p2_q14.csv',
        names=dic_pof_p2_q14$variavel, begin=dic_pof_p2_q14$inicio, end=end_inv)

# Deletando objetos auxiliares

```

```
rm(dic_pof_p1_q2,end_dom)
rm(dic_pof_p2_q14,end_inv)
```

```
#####
```

```
## Efetua a leitura do conjunto de dados com o fread do data.table
pof_p1_q2 <- fread(input='dom_p1_q2.csv', sep='auto', sep2='auto', integer64='double')
pof_p2_q14<- fread(input='inv_p2_q14.csv', sep='auto', sep2='auto', integer64='double')
```

```
#view(pof_p1_q2)
#view(pof_p2_q14)
```

```
#pof_p1_q2$CODIGODAUF[pof_p1_q2$CODIGODAUF == 60] <- NA
#pof_p1_q2$CODIGODAUF[pof_p1_q2$CODIGODAUF == 99] <- NA
```

```
#Individualizar domicílios
```

```
pof_p2_q14$DOM_IND <- pof_p2_q14[,CODIGODAUF*1000000 + NUMEROSEQUENCIAL*1000
                                + DVDSEQUENCIAL*100 + NUMERODODOMICILIO]
```

```
pof_p1_q2$DOM_IND <- pof_p1_q2[,CODIGODAUF*1000000 + NUMEROSEQUENCIAL*1000
                                + DVDSEQUENCIAL*100 + NUMERODODOMICILIO]
```

```
#### CONTAR O NÚMERO DE DOMICÍLIOS AMOSTRADOS NO BANCO DE DADOS (EM UMA DETERMINADA
#### COLUNA)
```

```
length(unique(pof_p1_q2$DOM_IND))
length(unique(pof_p2_q14$DOM_IND))
```

```
pof_p1_q2[, CODIGODAUF := factor(CODIGODAUF)]
```

```
unique(unlist(pof_p1_q2$CODIGODAUF, use.names = FALSE))
```

```
levels(pof_p1_q2$CODIGODAUF) = c("RO","AC","AM","RR","PA", "AP","TO","MA","PI","CE",
                                "RN","PB","PE", "AL","SE","BA","MG","ES","RJ","SP",
                                "PR","SC", "RS", "MS","MT","GO","DF")
```

```
pof_p2_q14$ITEM <- pof_p2_q14[, NUMEROQUADRO*100000 + CODIGODOITEM]
```

```
### Ver documento "Cadastro de Produtos POF 2008-2009" para os produtos
```

```
pof_p2_q14$ITEM[pof_p2_q14$ITEM == 1400101] <- "FOGAO"
pof_p2_q14$ITEM[pof_p2_q14$ITEM == 1400801] <- "GRILL"
pof_p2_q14$ITEM[pof_p2_q14$ITEM == 1401001] <- "FORNO_ELETRICO"
pof_p2_q14$ITEM[pof_p2_q14$ITEM == 1402601] <- "MICROONDAS"
#### criar um data.frame com as colunas de interesse pof_p1_q2
pof_p1_q2_red <- subset(pof_p1_q2, select =
                        c("CODIGODAUF", "ESTRATOGEOGRAFICO",
                          "FATORDEEXPANSAO2(AJUSTADOP/ESTIMATIVAS)",
                          "QUANTIDADEDEMoradores", "QUANTIDADEDEFAMILIAS",
                          "TIPODEDOMICILIO", "COMODOSSERVINDODEDORMITORIO",
                          "EXISTENCIADEAGUACANALIZADA", "PROVENIENCIAAAGUA",
                          "ESCOADOUROSANITARIO", "RENDATOTALMENSALDODOMICILIO",
```

```

      "REDEGERALDEENERGIAELETRICA",
      "AQUECIMENTODEAGUAPORENERGIAELETRICA",
      "AQUECIMENTODEAGUAPORGAS",
      "AQUECIMENTODEAGUAPORENERGIASOLAR",
      "AQUECIMENTODEAGUAPORLENHA/CARVAO",
      "AQUECIMENTODEAGUAPOROUTRAFORTE",
      "FOGAOAGAS", "FOGAOALENHA", "FOGAOACARVAO",
      "FOGAOENERGIAELETRICA", "FOGAOCOMOUTRAFORTE", "DOM_IND"))

setnames (pof_p1_q2_red, "FATORDEEXPANSAO2(AJUSTADOP/ESTIMATIVAS)", "Peso_para_estimativa")

#### criar um data.frame com as colunas de interesse pof_p2_q14

pof_p2_q14_intermediario <- subset(pof_p2_q14, select =
      c("DOM_IND", "ITEM", "QUANTIDADEDOITEM"))

#### Selecionar apenas as linhas que possuem itens de interesse

pof_p2_q14_intermediario <- subset(pof_p2_q14_intermediario, ITEM == "FOGAO"|
      ITEM == "FORNO_ELETRICO"|
      ITEM == "GRILL"|
      ITEM == "MICROONDAS")

#### Reshape o data.frame

pof_p2_q14_red <- dcast(pof_p2_q14_intermediario, DOM_IND ~ ITEM)

####Excluir data.frames desnecessários

rm(pof_p1_q2)
rm(pof_p2_q14)

gc()

#####

####POF_P3

## Criando o dicionário a partir das três primeiras colunas da planilha

dic_pof_p3 <- read.csv(file = 'CADERNETA_POF3.csv', header=F, sep=";")
dic_pof_p3 <- dic_pof_p3[complete.cases(dic_pof_p3),]
colnames(dic_pof_p3) <- c('variavel', 'tamanho', 'inicio')

## Parâmetro com o final de cada campo
end_cad = dic_pof_p3$inicio + dic_pof_p3$tamanho - 1

## Converte o microdado para um arquivo csv

fwf2csv(fwffile='Dados/T_CADERNETA_DESPESA_S.txt', csvfile='inv_p3.csv',
      names=dic_pof_p3$variavel, begin=dic_pof_p3$inicio, end=end_cad)

```

```

# Deletando objetos auxiliares

rm(dic_pof_p3,end_cad)

#####

## Efetua a leitura do conjunto de dados com o fread do data.table

pof_p3<- fread(input='inv_p3.csv', sep='auto', sep2='auto', integer64='double')

#Individualizar domicílios

pof_p3$DOM_IND <- pof_p3[,CODIGO_DA_UF*1000000 + NUMERO_SEQUENCIAL*1000
                        + DV_DO_SEQUENCIAL*100 + NUMERO_DO_DOMICILIO]

#### Identificar alimentos.

pof_p3$alimento_individualizado <- pof_p3[,NUMERO_DO_GRUPO_DE_DESPESA*100000+CODIGO_DO_ITEM]

pof_p3$QUANTIDADE_FINAL_EM_KG_ANUAL<-pof_p3[,QUANTIDADE_FINAL_EM_KG*FATOR_DE_ANUALIZACAO]

####Codigo individual dos produtos é formado pelas colunas B e C do arquivo
####"Cadastro de Produtos POF 2008-2009"

#### Apesar de o código do alimento ser composto por 7 dígitos (grupo
#### de despesa - 2 dígitos + código do item) - (ver linha 192), neste ponto reduzo para 5
#### dígitos para promover maior agregação.

pof_p3$grupo_alimento_individualizado<- floor(pof_p3$alimento_individualizado/100)

#### Identificar grupos de alimentos mais consumidos

consumo_por_alimento <- aggregate(QUANTIDADE_FINAL_EM_KG_ANUAL~
                                grupo_alimento_individualizado,data=pof_p3,sum)

#### Ordenar de forma decrescente

consumo_por_alimento <- consumo_por_alimento[order
(consumo_por_alimento$QUANTIDADE_FINAL_EM_KG_ANUAL,decreasing=TRUE),]

gc()
#### são contabilizados os grupos de alimentos mais consumidos entre
#### os 20 grupos que exigem cocção

#### Subset das observações dos grupos de alimento de interesse

pof_p3_grupos <- pof_p3[pof_p3$grupo_alimento_individualizado == 63001|
                        pof_p3$grupo_alimento_individualizado == 79001|
                        pof_p3$grupo_alimento_individualizado == 79002|
                        pof_p3$grupo_alimento_individualizado == 63034|
                        pof_p3$grupo_alimento_individualizado == 65014|
                        pof_p3$grupo_alimento_individualizado == 78001|
                        pof_p3$grupo_alimento_individualizado == 78002|
                        pof_p3$grupo_alimento_individualizado == 63021|
                        pof_p3$grupo_alimento_individualizado == 64001|

```

```

pof_p3$grupo_alimento_individualizado == 65010|
pof_p3$grupo_alimento_individualizado == 78033|
pof_p3$grupo_alimento_individualizado == 67057|
pof_p3$grupo_alimento_individualizado == 71017|
pof_p3$grupo_alimento_individualizado == 63007|
pof_p3$grupo_alimento_individualizado == 65006|
pof_p3$grupo_alimento_individualizado == 65034|
pof_p3$grupo_alimento_individualizado == 71016|
pof_p3$grupo_alimento_individualizado == 64006|
pof_p3$grupo_alimento_individualizado == 82025|
pof_p3$grupo_alimento_individualizado == 79036,]

#### Conferir os grupos de alimentos selecionados

unique(unlist(pof_p3_grupos$grupo_alimento_individualizado, use.names = FALSE))

#### Agregar consumo dos alimentos selecionados por domicílio. Não preciso saber mais que tipo
#### de alimento é, apenas a quantidade consumida (em kg)

pof_p3_grupos <- aggregate(QUANTIDADE_FINAL_EM_KG_ANUAL~DOM_IND,
                           pof_p3_grupos,sum)

#### Apagar bancos desnecessários
rm(pof_p3)
rm(consumo_por_alimento)
gc()

#####
#####

####POF_P2_Q7

## Criando o dicionário a partir das três primeiras colunas da planilha

dic_pof_p2_q7 <- read.csv(file = 'DESPESA_90DIAS.csv', header=F, sep=";")
dic_pof_p2_q7 <- dic_pof_p2_q7[complete.cases(dic_pof_p2_q7),]
colnames(dic_pof_p2_q7) <- c('variavel', 'tamanho', 'inicio')

## Parâmetro com o final de cada campo
end_DESPESAS = dic_pof_p2_q7$inicio + dic_pof_p2_q7$tamanho - 1

## Converte o microdado para um arquivo csv

fwf2csv(fwffile='Dados/T_DESPESA_90DIAS_S.txt', csvfile='inv_p2_q7.csv',
        names=dic_pof_p2_q7$variavel, begin=dic_pof_p2_q7$inicio, end=end_DESPESAS)

# Deletando objetos auxiliares

rm(dic_pof_p2_q7,end_DESPESAS)

#####

## Efetua a leitura do conjunto de dados com o fread do data.table

pof_p2_q7<- fread(input='inv_p2_Q7.csv', sep='auto', sep2='auto', integer64='double')
```



```
#### Subset linhas que se referem ao quadro de interesse

pof_p2_q7_red <- subset(pof_p2_q7, NUMERO_DO_QUADRO== 7)

#### Subset linhas que se referem aos quadro de interesse

pof_p2_q7_red <- subset(pof_p2_q7_red, CODIGO_DO_ITEM == 101|
                        CODIGO_DO_ITEM == 102|
                        CODIGO_DO_ITEM == 103|
                        CODIGO_DO_ITEM == 104)

#### ver os resultados da coluna CODIGO_DO_ITEM
pof_p2_q7_red[, unique(unlist(CODIGO_DO_ITEM))]]

####Criar coluna com domicílios individualizados

pof_p2_q7_red$DOM_IND <- pof_p2_q7_red[,CODIGO_DA_UF*1000000 + NUMERO_SEQUENCIAL*1000
                                     + DV_DO_SEQUENCIAL*100 + NUMERO_DO_DOMICILIO]

#### Quantidade de glp (kg) adquirido por domicílio em um trimestre

gas_por_domicilio <- aggregate(QUANTIDADE_FINAL~DOM_IND,
                               pof_p2_q7_red,sum)

#### Anualizar quantidade de glp (kg) adquirido por domicílio

gas_por_domicilio$QUANTIDADE_GAS_FINAL_ANUAL <- (gas_por_domicilio$QUANTIDADE_FINAL_*4)
gas_por_domicilio$QUANTIDADE_FINAL_ <- NULL

#####
#####
#Criação da base de dados única

pre_banco_final <- merge(pof_p1_q2_red,pof_p2_q14_red,by="DOM_IND",all=TRUE)

pre_banco_final2 <- merge(pre_banco_final,pof_p3_grupos, by="DOM_IND",all=TRUE)

banco_final <- merge (pre_banco_final2, gas_por_domicilio,by="DOM_IND",all=TRUE)

rm(gas_por_domicilio); rm(pof_p1_q2_red); rm(pof_p2_q14_red);rm(pof_p2_q7);rm(pof_p2_q7_red);
rm(pof_p3_grupos);rm(pre_banco_final);rm(pre_banco_final2); rm(pof_p2_q14_intermediario)
gc()

#### Renda per capta

banco_final$RENDAPERCAPTA_MENSAL <- (banco_final$RENDATOTALMENSALDODOMICILIO/
                                     banco_final$QUANTIDADEDEMODADORES)

#### ESTRATO GEOGRÁFICO (1 PARA URBANO, 2 PARA RURAL)

banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="RO" & banco_final$ESTRATOGEOGRAFICO <=
6] <- 1
```

[illegible]

```

banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="BA" & banco_final$ESTRATOGEOGRAFICO <=
21] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="BA" & banco_final$ESTRATOGEOGRAFICO >=
22] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="MG" & banco_final$ESTRATOGEOGRAFICO <=
27] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="MG" & banco_final$ESTRATOGEOGRAFICO >=
28] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="ES" & banco_final$ESTRATOGEOGRAFICO <=
9] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="ES" & banco_final$ESTRATOGEOGRAFICO >=
10] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="RJ" & banco_final$ESTRATOGEOGRAFICO <=
30] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="RJ" & banco_final$ESTRATOGEOGRAFICO >=
31] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="SP" & banco_final$ESTRATOGEOGRAFICO <=
30] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="SP" & banco_final$ESTRATOGEOGRAFICO >=
31] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="PR" & banco_final$ESTRATOGEOGRAFICO <=
18] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="PR" & banco_final$ESTRATOGEOGRAFICO >=
19] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="SC" & banco_final$ESTRATOGEOGRAFICO <=
13] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="SC" & banco_final$ESTRATOGEOGRAFICO >=
14] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="RS" & banco_final$ESTRATOGEOGRAFICO <=
18] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="RS" & banco_final$ESTRATOGEOGRAFICO >=
19] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="MS" & banco_final$ESTRATOGEOGRAFICO <=
8] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="MS" & banco_final$ESTRATOGEOGRAFICO >=
9] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="MT" & banco_final$ESTRATOGEOGRAFICO <=
10] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="MT" & banco_final$ESTRATOGEOGRAFICO >=
11] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="GO" & banco_final$ESTRATOGEOGRAFICO <=
17] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="GO" & banco_final$ESTRATOGEOGRAFICO >=
18] <- 2
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="DF" & banco_final$ESTRATOGEOGRAFICO <=
7] <- 1
banco_final$ESTRATOGEOGRAFICO [banco_final$CODIGODAUF=="DF" & banco_final$ESTRATOGEOGRAFICO >=
8] <- 2

```

CRIAR REGIOES

```

banco_final$regiao[banco_final$CODIGODAUF=="RO"|banco_final$CODIGODAUF=="AC"|
banco_final$CODIGODAUF=="AM"|banco_final$CODIGODAUF=="RR"|
banco_final$CODIGODAUF=="PA"|banco_final$CODIGODAUF=="AP"|
banco_final$CODIGODAUF=="TO"] <- "Norte"

```

```

banco_final$regiao[banco_final$CODIGODAUF=="MA"|banco_final$CODIGODAUF=="PI"|
banco_final$CODIGODAUF=="CE"|banco_final$CODIGODAUF=="RN"|

```

```

    banco_final$CODIGODAUF=="PB"|banco_final$CODIGODAUF=="PE"|
    banco_final$CODIGODAUF=="AL"|banco_final$CODIGODAUF=="SE"|
    banco_final$CODIGODAUF=="BA"] <- "Nordeste"

banco_final$regiao[banco_final$CODIGODAUF=="MG"|banco_final$CODIGODAUF=="ES"|
    banco_final$CODIGODAUF=="RJ"|
    banco_final$CODIGODAUF=="SP"] <- "Sudeste"

banco_final$regiao[banco_final$CODIGODAUF=="PR"|banco_final$CODIGODAUF=="SC"|
    banco_final$CODIGODAUF=="RS"] <- "Sul"

banco_final$regiao[banco_final$CODIGODAUF=="MT"|banco_final$CODIGODAUF=="MS"|
    banco_final$CODIGODAUF=="GO"|
    banco_final$CODIGODAUF=="DF"] <- "Centro-Oeste"

banco_final[, unique(unlist(regiao))]]

####(remoção das linhas com NA na variável resposta)

banco_final<-banco_final[-which(is.na(banco_final$QUANTIDADE_GAS_FINAL_ANUAL)), ]

names(banco_final)[17]<-"AQUECIMENTODEAGUAPORLENHA"

attach(banco_final)
banco_final$consumo_per_capita<-QUANTIDADE_GAS_FINAL_ANUAL/QUANTIDADEDEMODADORES
banco_final$alimento_per_capita<-QUANTIDADE_FINAL_EM_KG_ANUAL/QUANTIDADEDEMODADORES
banco_final$renda_per_capita_anual<-RENDA_PER_CAPTA_MENSAL*12

####(remoção das linhas com consumo per capita superior a 130kg)

banco_final<-banco_final[-which(consumo_per_capita>130), ]

detach(banco_final)

#Criar variável com as categorias para aquecimento de água.
attach(banco_final)
banco_final$tipo_aquecimento<-AQUECIMENTODEAGUAPORENERGIAELETRICA[]*10000 +
    AQUECIMENTODEAGUAPORGAS[]*1000 +
    AQUECIMENTODEAGUAPORENERGIASOLAR[]*100 +
    AQUECIMENTODEAGUAPORLENHA[]*10 +
    AQUECIMENTODEAGUAPOROUTRAFORTE

detach(banco_final)
attach(banco_final)

banco_final$categorias_aquecimento[tipo_aquecimento[]==12222|tipo_aquecimento[]==12112|
    tipo_aquecimento[]==12122|tipo_aquecimento[]==12212|
    tipo_aquecimento[]==12221]<-"Energia eletrica (e outras
fontes)"

banco_final$categorias_aquecimento[tipo_aquecimento[]==21122|tipo_aquecimento[]==21212|
    tipo_aquecimento[]==21221|tipo_aquecimento[]==21222]<-
"Gas (e outras fontes)"

```

```

banco_final$categorias_aquecimento[tipo_aquecimento[]==11122|tipo_aquecimento[]==11212|
                                tipo_aquecimento[]==11222]<-"Energia eletrica e gas (e
outras fontes)"

banco_final$categorias_aquecimento[tipo_aquecimento[]==22122|tipo_aquecimento[]==22212|
                                tipo_aquecimento[]==22221]<-"Outras fontes (exceto
energia eletrica e gas)"

banco_final$categorias_aquecimento[tipo_aquecimento[]==22222]<-"Nenhum"

detach(banco_final)

```

#Criar variável com as categorias para fogão.

```

attach(banco_final)
banco_final$tipo_fogao<-FOGAOAGAS[]*10000 +
                        FOGAOLENHA[]*1000 +
                        FOGAOACARVAO[]*100 +
                        FOGAOENERGIAELETRICA[]*10 +
                        FOGAOCOMOUTRAFORTE

detach(banco_final)
attach(banco_final)
banco_final$categorias_fogao[tipo_fogao[]==12222|tipo_fogao[]==12121|
                             tipo_fogao[]==12122|tipo_fogao[]==12212|
                             tipo_fogao[]==12221]<-"Gas e outras fontes (exceto lenha)"

banco_final$categorias_fogao[tipo_fogao[]==21122|
                             tipo_fogao[]==21222]<-"Lenha e outras fontes (exceto gas))"

banco_final$categorias_fogao[tipo_fogao[]==11122|tipo_fogao[]==11212|
                             tipo_fogao[]==11221|tipo_fogao[]==11222]<-"Gas, lenha e outras
fontes"

banco_final$categorias_fogao[tipo_fogao[]==22122|tipo_fogao[]==22212|
                             tipo_fogao[]==22221]<-"Outras fontes (exceto gas e lenha)"

banco_final$categorias_fogao[tipo_fogao[]==22222]<-"Nenhum"

detach(banco_final)
attach(banco_final)

```

#####

#Cria quantidade_de_familias_categorizada

```

banco_final$quantidade_de_familias_cat<-QUANTIDADEDEFAMILIAS
banco_final$quantidade_de_familias_cat[QUANTIDADEDEFAMILIAS[]==3]<-2
banco_final$quantidade_de_familias_cat[QUANTIDADEDEFAMILIAS[]==4]<-2
banco_final$quantidade_de_familias_cat[QUANTIDADEDEFAMILIAS[]==5]<-2
banco_final$quantidade_de_familias_cat[QUANTIDADEDEFAMILIAS[]==6]<-2
detach(banco_final)
attach(banco_final)

```

#Cria comodoss_cat

```

banco_final$comodos_cat<-COMODOSSERVINDODEDORMITORIO

```

```

banco_final$comodos_cat[COMODOSSERVINDODEDORMITORIO[]==4]<-3
banco_final$comodos_cat[COMODOSSERVINDODEDORMITORIO[]==5]<-3
banco_final$comodos_cat[COMODOSSERVINDODEDORMITORIO[]==6]<-3
banco_final$comodos_cat[COMODOSSERVINDODEDORMITORIO[]==7]<-3
banco_final$comodos_cat[COMODOSSERVINDODEDORMITORIO[]==8]<-3
detach(banco_final)
attach(banco_final)

```

```
#####
```

```

#Cria dormitórios_categoria
banco_final$dormitorios_categorias<-COMODOSSERVINDODEDORMITORIO
banco_final$dormitorios_categorias[COMODOSSERVINDODEDORMITORIO[]==1]<-"1"
banco_final$dormitorios_categorias[COMODOSSERVINDODEDORMITORIO[]==2]<-"2 ou 3"
banco_final$dormitorios_categorias[COMODOSSERVINDODEDORMITORIO[]==3]<-"2 ou 3"
banco_final$dormitorios_categorias[COMODOSSERVINDODEDORMITORIO[]==4]<-"4 ou mais"
banco_final$dormitorios_categorias[COMODOSSERVINDODEDORMITORIO[]==5]<-"4 ou mais"
banco_final$dormitorios_categorias[COMODOSSERVINDODEDORMITORIO[]==6]<-"4 ou mais"
banco_final$dormitorios_categorias[COMODOSSERVINDODEDORMITORIO[]==7]<-"4 ou mais"
banco_final$dormitorios_categorias[COMODOSSERVINDODEDORMITORIO[]==8]<-"4 ou mais"

```

```

detach(banco_final)
attach(banco_final)

```

```
#####
```

```
#Cria categorias de renda
```

```

banco_final$renda_cat[RENDAPERCAPTA_MENSAL[]<=207.5]<-"_Ate 0,5"
banco_final$renda_cat[RENDAPERCAPTA_MENSAL[]> 207.5 &
                      RENDAPERCAPTA_MENSAL[]<=311.25]<-"> 0,5 ate 0,75"
banco_final$renda_cat[RENDAPERCAPTA_MENSAL[]> 311.25 &
                      RENDAPERCAPTA_MENSAL[]<=622.50]<-"> 0,75 ate 1,5"
banco_final$renda_cat[RENDAPERCAPTA_MENSAL[]> 622.50 &
                      RENDAPERCAPTA_MENSAL[]<=1037.50]<-"> 1,5 ate 2,5"
banco_final$renda_cat[RENDAPERCAPTA_MENSAL[]> 1037.50 &
                      RENDAPERCAPTA_MENSAL[]<=1556.25]<-"> 2,5 ate 3,75"
banco_final$renda_cat[RENDAPERCAPTA_MENSAL[]> 1556.25]<-"> 3,75"
detach(banco_final)
attach(banco_final)

```

```
#####
```

```
#Cria categorias de alimento
```

```

#Usamos os quantis ponderados para categorizar a aquisição de alimentos
#wtd.quantile(alimento_per_capita, weights=Peso_para_estimativa, probs=c(0,.25,.50,.75,1))

```

```

banco_final$alimento_cat<-alimento_per_capita
banco_final$alimento_cat[is.na(alimento_per_capita)]<-"Sem informação"
banco_final$alimento_cat[alimento_per_capita[]<=50]<-"1-Baixo"
banco_final$alimento_cat[alimento_per_capita[]> 50 &
                      alimento_per_capita[]<=100]<-"2-Médio baixo"
banco_final$alimento_cat[alimento_per_capita[]> 100 &

```

```
        alimento_per_capita[] <= 200] <- "3-Médio alto"
banco_final$alimento_cat[alimento_per_capita[] > 200] <- "4-Alto"

detach(banco_final)
attach(banco_final)

#table(alimento_cat)

#Ajusta as variáveis de GRILL, MICROONDAS, FORNO_ELETRICO

banco_final$FORNO_ELETRICO[FORNO_ELETRICO[] == 2] <- -1
banco_final$MICROONDAS[MICROONDAS[] == 2] <- -1
banco_final$GRILL[GRILL[] == 2] <- -1

detach(banco_final)
attach(banco_final)
```

Anexo 2 – Número de setores selecionados e domicílios esperados, selecionados e entrevistados, por Unidades da Federação – POF 2008-2009

Unidades da Federação	Número de setores selecionados	Número de domicílios na amostra		
		Esperados	Selecionados	Entrevistados
Brasil	4 696	59 548	68 373	55 970
Rondônia	73	952	1 090	907
Acre	66	848	975	863
Amazonas	105	1 356	1 531	1 344
Roraima	55	700	868	644
Pará	156	2 048	2 375	1 894
Amapá	44	568	704	689
Tocantins	102	1 308	1 489	1 270
Maranhão	209	2 656	3 072	2 562
Piauí	153	1 956	2 202	2 056
Ceará	143	1 876	2 178	1 861
Rio Grande do Norte	113	1 428	1 592	1 342
Paraíba	128	1 620	1 846	1 628
Pernambuco	193	2 440	2 823	2 367
Alagoas	246	3 032	3 345	2 712
Sergipe	141	1 716	1 956	1 654
Bahia	245	3 164	3 600	3 050
Minas Gerais	439	5 488	6 333	5 028
Espírito Santo	330	4 036	4 543	3 489
Rio de Janeiro	171	2 156	2 509	1 938
São Paulo	294	3 780	4 290	3 623
Paraná	231	2 904	3 272	2 477
Santa Catarina	182	2 304	2 602	2 029
Rio Grande do Sul	189	2 412	2 703	2 210
Mato Grosso do Sul	166	2 116	2 670	2 247
Mato Grosso	208	2 680	3 126	2 423
Goiás	197	2 532	2 976	2 686
Distrito Federal	117	1 472	1 703	977

Fonte: IBGE, Diretoria de Pesquisas, Coordenação de Trabalho e Rendimento, Pesquisa de Orçamentos Familiares 2008-2009