ANÁLISE CONDICIONADA DA DEMANDA DE ENERGIA ELÉTRICA UTILIZANDO MODELAGEM ROBUSTA

Luzia Maria Tarcitano de Araujo

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE PRODUÇÃO.

Aprovada por:

_____
Prof. Marcos Pereira Estellita Lins, D. Sc

_____
Profª. Angela Cristina Moreira da Silva, D.Sc

_____
Prof. Basílio de Bragança Pereira, Ph.D

_____
Prof. Anníbal Parracho Sant'anna, Ph.D

RIO DE JANEIRO, RJ – BRASIL
DEZEMBRO DE 2005

ARAUJO, LUZIA MARIA TARCITANO
Análise Condicionada da Demanda de Energia Elétrica Utilizando Modelagem Robusta [Rio de Janeiro] 2005
VI, 89 p. 29,7 cm (COPPE / UFRJ, M.Sc, Engenharia de Produção, 2005)
Dissertação – Universidade Federal do Rio de Janeiro, COPPE
1. Análise Condicionada da Demanda; 2. Robustez; 3. Consumo de Eletricidade
I. COPPE / UFRJ II. Título (série)

Meu pai Reynaldo (in memorian) e minha mãe Yolanda
Aos meus filhos Angélica e André Ricardo
A Arnaldo, com todo meu amor.

Agradecimentos

Agradeço especialmente à minha família pelo amor e apoio incondicional.

Agradeço aos meus orientadores Marcos Pereira Estellita Lins e Angela Cristina Moreira da Silva pela oportunidade e orientação.

Ao CNPq, pelo apoio financeiro para o desenvolvimento dessa dissertação.

Aos meus colegas de turma Vinícius de Melo Araújo Martins, Paulo Jorge Magalhães Teixeira e Edson Luiz de Carvalho Barbosa pelo constante incentivo e apoio na realização deste trabalho.

Aos professores e funcionários da COPPE/UFRJ, em especial a Andréia Lima, secretaria do Programa de Engenharia de Produção.

Ao Vinícius Brito Rocha pela colaboração no *software* S-Plus, para realização dos experimentos do trabalho.

Aos demais membros da banca, Professor Basílio de Bragança Pereira e Anníbal Parracho Sant'anna

E a DEUS por me encaminhar na trajetória da vida.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ANÁLISE CONDICIONADA DA DEMANDA DE ENERGIA ELÉTRICA UTILIZANDO MODELAGEM ROBUSTA

Luzia Maria Tarcitano de Araujo

Dezembro / 2005

Orientadores: Marcos Pereira Estellita Lins
Angela Cristina Moreira da Silva

Programa: Engenharia de Produção

A necessidade deste trabalho surgiu em 2001 frente à crise de energia sofrida neste país, sobretudo nas Regiões Nordeste, Centro-Oeste e Sudeste. O planejamento da demanda do consumo residencial se apresenta como uma opção para auxiliar na política nacional de energia a longo prazo. O método de Análise Condicionada da Demanda (Conditional Demand Analysis – CDA) será utilizado para estimar o consumo de energia por uso final a partir de estimadores robustos. A base de dados utilizada nesta análise faz parte do projeto financiado pelo CNPq no âmbito do edital do CT-Energ, desenvolvida de forma conjunta entre pesquisadores dos programas de Engenharia de Produção da COPPE/UFRJ e da UFPE, entitulada como “Pesquisa de Posse de Eletrodomésticos e Preferências de Consumo”. O plano amostral foi composto de 600 domicílios, representativos do município de Recife e o período da pesquisa compreendeu entre os meses de abril e outubro de 2003. Neste trabalho foram investigados um total de 35 equipamentos, com ênfase nos principais consumidores de energia.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

CONDITIONAL DEMAND ANALYSIS OF ELETRIC ENERGY DEMAND USING ROBUST MODELLING

Luzia Maria Tarcitano de Araujo

December/2005

Advisors: Marcos Pereira Estellita Lins
Angela Cristina Moreira da Silva

Department: Productive Engineering

The motivation for this work has arisen with the Brazilian energy crisis in 2001, which greatly impacted the Northeastern, Center and Southeastern regions of the country. The demand planning of the residential energy consumption constitutes an alternative for supporting the national energy policy in the long term. The methodology known as Conditional Demand Analysis (CDA) will be used to asses the end use consumption through robust estimates. The database resulted from a survey supported by CNPq/CTenerg and carried out by the Production Engineering Programs of UFRJ and UFPE, entitled “Pesquisa de Posse de Eletrodomésticos e Preferências de Consumo”. The sample comprises 600 households visited between April and October 2003. A total of 35 appliances was investigated, with an emphasis on the more energy intensive ones.

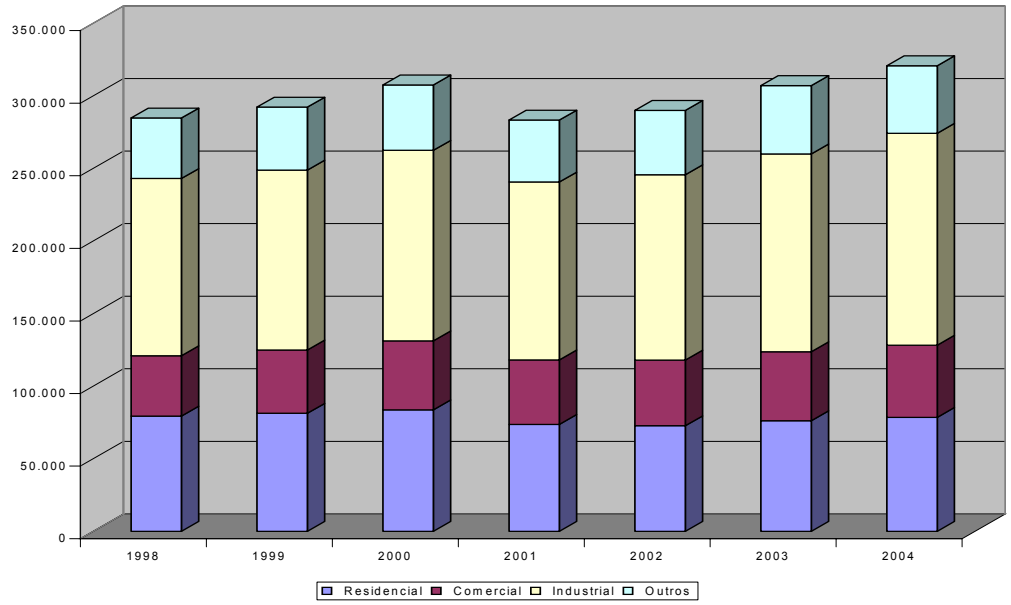
SUMÁRIO

1. INTRODUÇÃO

O Brasil vem desenvolvendo esforços para conservar energia, desde meados da década de 80, quando aconteceu o primeiro período de racionamento, em 1987, na Região Nordeste e parte do estado do Pará. A segunda fase de racionamento de energia vigorou de 01/06/2001 a 28/02/2002, penalizando agora além da Região Nordeste, o Sudeste e o Centro-Oeste do país. Como medida de contenção o Governo Federal criou a Câmara de Gestão da Crise de Energia Elétrica - GCE. Esta Câmara teve como objetivo propor e implementar medidas de natureza emergencial, decorrentes da situação hidrológica crítica para compatibilizar a demanda e a oferta de energia elétrica, sendo extinta em 30 de junho de 2002, substituída pela Câmara de Gestão do Setor Elétrico – CGSE, que passou a integrar o Conselho Nacional de Política Energética – CNPE. Mediante um processo de aprimoramento do novo modelo do Setor Elétrico Brasileiro, a GCE criou através da Resolução nº 18, de 22 de junho de 2001, o Comitê de Revitalização do Modelo do Setor Elétrico. Este comitê tinha como missão encaminhar propostas para corrigir as disfunções correntes e propor aperfeiçoamentos para o referido modelo, que passou a ser gerido pela CGSE, conforme o Decreto nº 4.505, de 11 de dezembro de 2002 (SILVA et al 2004).

Com base na evolução do consumo de energia elétrica (GWh), verifica-se que o mesmo já atingiu, em 2004, o valor de 320.701 GWh, patamar superior ao período pré-racionamento (307.449 GWh em 2000) (Figura 1.1). As taxas de crescimento de 2001 a 2004 foram 4,5%, 5,9% e 2,4%, respectivamente. No período 2003/2004 a região Norte foi a que apresentou maior taxa anual de crescimento, de 8,1%, seguidas pelas Regiões Nordeste com taxa de 5,5% e 4,1% para as regiões Sudeste/Centro-Oeste e de 3,2% para a Região Sul. Com relação aos setores, foi o industrial que apresentou a maior taxa de crescimento de consumo de energia no último período 2003/2004, 7,2%, seguido pelo setor comercial, com 4,5% e o residencial com 3,0%, já os demais setores tiveram uma queda de consumo de 1,1%. Veja a Tabela 1.1.

Figura 1.1 – Consumo de Energia Elétrica - (GWh) – BRASIL
1998-2004



Fonte: Eletrobrás

Tabela 1.1 – Consumo de Energia Elétrica por Setores e Regiões
1998-2004 - Brasil

Discriminação	1998	1999	2000	2001	2002	2003	2004	Var.(%) 2004/03
Por Setores								
Residencial	79.378	81.291	83.613	73.622	72.660	76.162	78.473	3,0
Comercial	41.579	43.588	47.510	44.434	45.251	47.531	49.691	4,5
Industrial	122.023	123.893	131.315	122.539	127.694	136.221	145.996	7,2
Outros	41.729	43.416	45.011	42.662	44.327	47.073	46.541	-1,1
Por Regiões								
Norte	14.336	14.877	16.033	23.048	17.016	26.934	29.104	8,1
Nordeste	46.103	47.334	49.617	37.463	47.334	42.438	44.758	5,5
Sudeste/Centro-Oeste	180.459	183.660	192.073	173.537	175.114	184.018	191.517	4,1
Sul	43.811	46.317	49.726	49.209	50.468	53.597	55.322	3,2
Brasil	**284.709**	**292.188**	**307.449**	**283.257**	**289.932**	**306.987**	**320.701**	**4,5**

Fonte: Eletrobrás

Com relação ao consumo de energia residencial, destacamos uma taxa de crescimento modesta, no período 2001 a 2004, se comparada com o consumo total, com taxas de 3,0%, 4,8% e –1,3%, respectivamente. Enquanto, o consumo total de energia do Brasil teve uma redução de 7,9%, no período 2000/2001, o setor residencial caiu 11,9% e até 2004 (78.473 GWh) este consumo não atingiu ao patamar de 2000 (83.613 GWh), período pré-racionamento.

O chamado “apagão” mudou a cultura do consumo de energia elétrica, hoje são comuns campanhas publicitárias de informação e estímulo à conservação de energia elétrica. Atualmente, observa-se um interesse crescente por equipamentos mais eficientes, do ponto de vista econômico, como as lâmpadas compactas fluorescentes e as geladeiras detentoras do selo Procel de economia de energia.

1.1. OBJETIVOS

O risco de um novo racionamento de energia elétrica, não está descartado e ocorrerá se houver grandes desequilíbrios oferta-demanda. Para que sejam tomadas as ações preventivas pertinentes, em caso de crescimento dos riscos de déficit de energia, é necessária avaliação permanente do comportamento do mercado tanto pela ótica da oferta, quanto da demanda.

Este estudo focaliza o problema energético pelo lado da demanda, onde o planejamento do mercado de energia por uso final no setor residencial e os programas de conservação de energia elétrica tem requerido informações sobre o consumo dos equipamentos nas condições em que são utilizados nos domicílios.

Nesta dissertação espera-se construir uma nova metodologia de estimação dos consumos de equipamentos de usos finais, através de métodos estatísticos. Essas estimativas são de suma relevância para auxiliar os estudos de planejamento de demanda e programas governamentais de economia de energia que visem o uso eficiente dos equipamentos.

Para isso será utilizado o método conhecido na literatura internacional como *Conditional Demand Analysis* – CDA (Análise Condicionada da Demanda), que através de dados de consumo, posse e hábitos de uso dos equipamentos elétricos, estabelecem estimação de consumo. A inovação desta dissertação será empregar e comparar os resultados obtidos pelos estimadores robustos e clássicos, obtidos pelo método dos Mínimos Quadrados Ordinários (OLS – *Ordinary Least Square*). Este estudo possibilitará analisar os fatores que determinam o comportamento dos consumidores residenciais, procurando quantificar os principais equipamentos no consumo de eletricidade residencial, na cidade do Recife, proporcionando implementar políticas de racionalização do consumo de energia elétrica, contribuindo, assim, para equacionar o problema energético ora enfrentado.

1.2. METODOLOGIA

Neste trabalho é proposta a aplicação da metodologia de Análise de Demanda Condicionada, onde é utilizada a regressão estatística para quebrar o consumo residencial em suas partes constituintes. Segundo SILVA (2000) o consumo de energia residencial é determinado pela composição do estoque de equipamentos e pelo uso final dos equipamentos, ambos podendo ser influenciados pela renda, número de habitantes, tamanho da família, etc (PARTI et al., 1980 e DUBIN et al., 1984).

No Brasil, a primeira aplicação desta técnica foi apresentada por LINS et al. (1996), utilizando dados de posse e consumo de energia de pesquisa realizada pelo PROCEL/ELETROBRÁS em 1988/1989.

O modelo pressupõe as relações lineares entre consumo médio de energia (CE) em cada domicílio amostrado i e a quantidade dos diversos equipamentos (j), representados por X variáveis aleatórias. Temos:

$$CE_i = \sum_{\mathrm{i}} \beta_j X_{ij} + \varepsilon_i \qquad (1.1)$$

Onde,

CE_i é o consumo médio de energia no domicílio i.(Variável dependente)

X_{ij} é a quantidade de equipamento X_j no domicílio i.(Variáveis independentes)

β_j é o coeficiente de regressão que estima o consumo do equipamento j.

ε_i é o termo aleatório.

O modelo aditivo assume a hipótese de que os coeficientes de regressão (β_j) são os mesmos para todos os domicílios, independente do número de moradores ou classe social. Segundo LINS 1996 esses modelos são preferíveis pela facilidade de sua interpretação física, que lhes dá a propriedade de simular o consumo de cada equipamento. Isto é, se tomarmos os coeficientes de regressão (β_j) do modelo ajustado e multiplicarmos pelo estoque dos equipamentos teremos o consumo residencial total. Desta forma, a participação de cada equipamento deve ser expressa como uma única parcela na equação do modelo de regressão, onde o intercepto representa o consumo dos equipamentos restantes, não incluídos na análise.

O diferencial desse método para isolar o consumo de uso final, usando um modelo de regressão linear, sem medição direta dos equipamentos, depende diretamente das diferenças dos padrões de posse dos equipamentos (LINS et al. 2003).

Existem algumas técnicas de estimação para o modelo linear de regressão. O método clássico de estimativa conhecido como dos Mínimos Quadrados Ordinários (OLS – Ordinary Least Square) é considerado o melhor estimador linear não viesado, mas precisam satisfazer alguns pressupostos básicos. Já os métodos Robustos apresentam-se como uma alternativa para estimação quando as hipóteses dos modelos clássicos de regressão são violadas. Como estamos utilizando dados de corte do tipo *cross-section* que apresentam heterocedasticidade nos erros, essa metodologia é indicada por suportar os *outliers* que distorcem os resultados das estimativas dos parâmetros.

1.3. DELIMITAÇÃO DO TRABALHO

A base de dados utilizada nesta análise faz parte do projeto financiado pelo CNPq no âmbito do edital do CT-Energ, desenvolvida de forma conjunta entre pesquisadores dos programas de Engenharia de Produção da COPPE/UFRJ e da UFPE, intitulada como "Pesquisa de Posse de Eletrodomésticos e Preferências de Consumo".

O trabalho contou com 600 domicílios entrevistados no município de Recife, em Pernambuco, no período compreendido entre os meses de abril e outubro de 2003. Neste levantamento foram investigados um total de 33 equipamentos, além das características de iluminação: lâmpadas fluorescentes e incandescentes, com ênfase nos seus principais equipamentos consumidores de energia, tais como refrigerador, freezer, chuveiro elétrico, condicionador de ar e televisão. A variável dependente do modelo é o consumo médio de energia elétrica, no período de 12 meses (março de 2002 e fevereiro de 2003), a qual se pretende explicar a partir da posse do conjunto dos equipamentos eletroeletrônicos. Essas variáveis explicativas são potenciais variáveis independentes do modelo, já que nem todas necessariamente fornecem informações relevantes para obter o consumo de energia domiciliar. Neste trabalho, utilizou-se dois métodos de seleção de variáveis, o método *Stepwise* e o *Teste t-student*, que visam determinar a significância dos parâmetros estimados. Note que o teste t tem sua maior relevância após o resultado da seleção *Stepwise*.

Com base nessa seleção desenvolveu-se o presente estudo de Análise Condicionada da Demanda comparando os resultados dos modelos de regressão linear clássico e o robusto. Uma inovação neste tipo de análise é o uso de metodologia robusta para estimação dos parâmetros e sua análise comparativa ao método clássico, de modo a obter-se o consumo total de energia de cada domicílio.

1.4. ESTRUTURA DO TRABALHO

Esta dissertação está estruturada em cinco capítulos, sendo o primeiro, introdutório, onde é apresentado o seu objetivo e metodologia. No segundo e terceiro capítulos são apresentados os modelos de Regressão Clássica pelo Método de Mínimos Quadrados Ordinários e os Métodos Robustos, respectivamente, que dão sustentação teórica para o trabalho. O quarto capítulo aborda a aplicação de Métodos Robustos a Análise Condicionada da Demanda. No quinto conclui-se o trabalho e são feitas recomendações para futuros trabalhos.

2. REGRESSÃO CLÁSSICA

Segundo WEISBERG (1947), a Análise de Regressão é comumente utilizada em estudos que envolvem variáveis mensuráveis. A regressão linear é usada em situações onde o relacionamento pode ser descrito geometricamente através de uma linha reta, ou generalizando, para mais de uma variável, por linhas retas de múltiplas dimensões. Este tipo de técnica é utilizado em vários campos de estudo, como ciências sociais, biológica, negócios, tecnologia e humanas. As razões para se ajustar um modelo de regressão linear são das mais variadas, assim como suas aplicações, mas as mais comuns são:

- Descrição do relacionamento entre variáveis;
- Previsão de valores;
- Estimação de parâmetros.

O objetivo dos modelos de regressão utilizados neste trabalho é o de estimar o consumo médio dos equipamentos eletroeletrônicos de uma região, a partir de uma função utilidade para cada equipamento, considerando o consumo naquele equipamento como uma função linear. Esta metodologia é conhecida como Análise Condicionada da Demanda, onde uma variável assume o significado de variável resposta, enquanto que as demais assumem o significado de variáveis preditoras[1]. O método de estimação para os parâmetros abordados neste capítulo é o da "soma de mínimos quadrados"; também conhecido como Mínimos Quadrados Ordinários (OLS – *Ordinary Least Square*).

2.1. REGRESSÃO LINEAR SIMPLES

No caso da análise de regressão linear simples a relação entre duas variáveis quantitativas ditas X e Y, observáveis em cada um dos "n" indivíduos sob estudo é explorada. Espera-se que o relacionamento entre estas duas variáveis possa ser descrito através de uma linha reta[2]. Os valores

[1] Também são conhecidas como variável dependente e independente, respectivamente.
[2] Quando a relação é não linear pode-se transformar as variáveis para obter este resultado.

observados são descritos através do par (x_i, y_i), onde i é a i-ésima observação.

Sejam duas variáveis X e Y, se uma relação existe entre elas, a mesma pode ser escrita por uma função matemática, supondo como hipótese nula - H_0 que X causa Y.

$$Y = f(X) \tag{2.1}$$

Através de uma amostra de dados coletados deseja-se estudar f, e a relação descrita entre X e Y. Para cada n unidades ou casos, observa-se x_i de X e y_i de Y, onde i=1,...,n .

$$y_i = f(x_i) + \in_i \tag{2.2}$$

O termo $\in_i$ é definido como erro de medida, relacionado ao fato de um modelo representar a simplificação da realidade. Por isso, variáveis omitidas e que juntamente com a variável independente se relacionam com a variável dependente podem ser incluídas no termo de erro. Se esses efeitos omitidos são pequenos, é razoável supor que o termo de erro é aleatório. Outra fonte de erro está associado ao processo de coleta das observações, provocado por algum tipo de coleta ineficiente.

Suponha agora que a forma desconhecida da função f possa ser aproximada por uma linha reta e que X é uma variável não-estocástica cujos valores são fixos. Então f(X) é estimada por β_0+β_1X para algum β_0 e β_1. Os parâmetros β_0 e β_1 são chamados de coeficientes de regressão.

$$f(x_i) = \beta_0 + \beta_1 x_i + \delta_i \tag{2.3}$$

Surge então o resíduo δ_i que é o valor da distância entre as observações e o valor da função f, refletindo o efeito da distância do modelo linear para a função f, ou seja, $\delta_i = f(x_i) - \beta_0 - \beta_1 x_i$. Para o modelo de regressão linear é interessante que δ_i seja o menor possível. Combinando 2.2 e 2.3 é possível definir $e_i = \in_i + \delta_i$ onde então obtêm-se o modelo de regressão linear simples, dado pela expressão a seguir:

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad i = 1, 2, ..., n \tag{2.4}$$

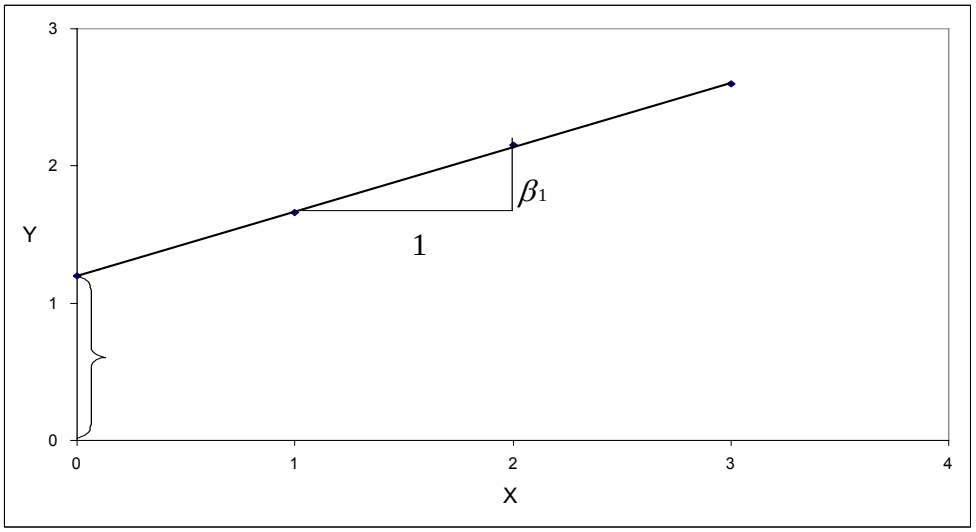
A partir de agora assume-se que os x_i's serão observados sem falha no processo de coleta de dados. O termo referente ao erro em (2.4) será equivalente ao resíduo de (2.3). Este procedimento é comumente adotado devido à aleatoriedade imputada pelo termo de erro em (2.2). Os erros estão associados à dificuldade em se obter o verdadeiro modelo de regressão, enquanto que os resíduos surgem do processo de estimação.

O parâmetro β_0 é denominado intercepto. O intercepto é o valor de Y quando X assume valor nulo. O β_1 informa as mudanças na média da distribuição de Y para cada unidade de medida de X, observe a figura 2.1. Todos os possíveis valores para esses parâmetros determinam todas as possíveis retas que descrevem a relação entre as variáveis.

Os valores dos parâmetros β_0 e β_1 e do termo aleatório e_i são desconhecidos na equação (2.4). Como e_i muda para cada observação y_i, fica difícil determiná-lo. Contudo, β_0 e β_1 permanecem fixos e embora não possam ser determinados sem examinarmos todos os possíveis valores de X e Y, podemos usar a informação obtida nas n observações para obtermos estimativas para β_0 e β_1,

Estes parâmetros são o objetivo principal da análise de regressão, visto que são geralmente desconhecidos, e precisam ser estimados através dos dados.

Figura 2.1 –Linha reta (Regressão linear simples)



Os dados reais quase nunca são descritos exatamente através de uma linha reta. A diferença entre os valores da variável resposta e dos valores obtidos através do modelo são chamados de erros estatísticos, isto é, ele é uma variável aleatória que quantifica a falha do modelo em se ajustar aos dados reais. Não deve ser confundido com erros provocados devido à escolha equivocada do modelo.

Segundo KMENTA (1988) o erro amostral é simplesmente a diferença entre o valor do estimador e o verdadeiro valor do parâmetro. Viés ou tendenciosidade é a diferença entre a média da distribuição amostral e o verdadeiro valor do parâmetro. Este valor é, para cada estimador determinado, um valor fixo que pode ou não ser igual a zero.

A interpretação para o modelo de regressão linear apresentado é que os valores observados de y_i podem ser determinados pelos valores observados de x_i através da equação (2.4), exceto por e_i que é um valor aleatório indeterminado. Existem três quantidades desconhecidas em (2.4), os parâmetros β_0 e β_1 e a variância dos resíduos σ^2. Os valores e_i's são quantidades não observáveis que são introduzidos no modelo para contemplar a falha dos valores observados a se ajustarem a reta de regressão. Somente os valores x_i e y_i são utilizados para estimar as quantidades desconhecidas citadas.

Ainda de acordo com KMENTA (1988) existem algumas propriedades desejáveis dos estimadores. Estas podem ser divididas em dois grupos, dependendo do tamanho da amostra. As propriedades de amostras finitas ou pequenas podem caracterizar estimativas calculadas a partir de qualquer número de observações. Por outro lado, as propriedades assintóticas ou de grandes amostras restringem-se a distribuições amostrais baseadas em amostras de tamanho infinito:

1. A primeira propriedade das pequenas amostras é a da **não-tendenciosidade**, onde estabelecemos que um estimador não-tendencioso é aquele cuja média é igual ao valor do parâmetro da população a ser estimada.

$$E\left(\hat{\theta}\right) = \theta \qquad (2.5)$$

Logo, θ é um estimador não-tendencioso.

2. Uma outra propriedade desejável é a **eficiência** do estimador, consideraremos um estimador eficiente se (e somente se) for não-tendencioso e ao mesmo tempo tiver variância mínima. Logo, um estimador $\hat{\theta}$ é um estimador eficiente de θ se as seguintes condições forem satisfeitas:
   - ✓ $\hat{\theta}$ for não-tendencioso;
   - ✓ $Var(\hat{\theta}) \leq Var(\tilde{\theta})$, onde $\tilde{\theta}$ é qualquer outro estimador não-tendencioso de θ.

Com relação às propriedades assintóticas dos estimadores, estas se referem à distribuição de um estimador quando o tamanho da amostra é grande e se aproxima do infinito. O Teorema do Limite Central afirma, em essência, que com o aumento do tamanho da amostra, a distribuição da média amostral se aproxima de uma distribuição normal. Então dizemos que a distribuição normal é uma distribuição assintótica da média. Geralmente, se a distribuição de um estimador tende a torna-se sempre mais semelhante na forma a alguma distribuição específica, com o aumento do tamanho da amostral, tal distribuição específica chama-se distribuição assintótica do estimador em questão. As três propriedades assintóticas de um estimador são:

1. A **não-tendenciosidade assintótica**, essa definição afirma que um estimador é assintóticamente não-tendencioso se ele se torna não-tendencioso quando o tamanho amostral se aproximar do infinito.
2. A segunda propriedade é com relação a **consistência** do estimador, a maneira de descobrirmos se um estimador é consistente é traçarmos o comportamento da tendenciosidade e da variância do estimador quando o tamanho da amostra se aproxima do infinito. Se o aumento no tamanho amostral for acompanhado de uma redução na tendenciosidade (se houver alguma) e na variância, e se isto continuar até que tanto a tendenciosidade como a variância se

aproximarem de zero quando $n \to \infty$, então o estimador em questão será consistente. Desde que a soma da tendenciosidade e da variância elevadas ao quadrado for igual ao erro médio quadrado, o desaparecimento da tendenciosidade e da variância quando $n \to \infty$ será equivalente ao desaparecimento do erro médio quadrado. Essa condição é, geralmente, condição suficiente, mas não necessária de consistência. Isto é, é possível achar estimadores cujo erro médio quadrado não se aproxima de zero quando $n \to \infty$ e contudo o estimador é consistente. Tal situação pode surgir quando a distribuição assintótica de um estimador é tal que sua média ou variância não existe. Isto complica o problema de se determinar se um estimador é ou não consistente. Felizmente não são freqüentes estimadores com médias assintóticas ou variâncias inexistentes. Se excluirmos expressamente tais estimadores da consideração e nos limitarmos a estimadores com médias e variâncias assintóticas finitas, então a condição abaixo representará uma condição necessária e suficiente de consistência. Se $\hat{\theta}$ for um estimador de θ e se $\lim_{n\to\infty} EMQ\left(\hat{\theta}\right) = 0$ então $\hat{\theta}$ é um estimador consistente de θ. Uma característica importante dos estimadores consistente é o fato de que qualquer função contínua de um estimador consistente é ela mesma um estimador consistente.

3. A última propriedade é a **eficiência assintótica** que se relaciona à dispersão da distribuição assintótica de um estimador. A eficiência assintótica só é definida para aqueles estimadores cuja a média e a variância assintótica existem, isto é, são iguais a alguns números finitos. Logo, um estimador $\hat{\theta}$ é um estimador assintoticamente eficiente de θ se todas as seguintes condições forem satisfeitas:
    - $\hat{\theta}$ tem uma distribuição assintótica com média finita e variância também finita;
    - $\hat{\theta}$ é consistente;
    - Nenhum outro estimador consistente de θ tem variância assintótica menor que $\hat{\theta}$.

Seja dada então a equação (2.4), as hipóteses assumidas no processo de estimação clássica, segundo KMENTA (1988), são:

Hipótese 1: Valor médio zero da distribuição e_i. Para cada valor de X, os erros distribuem-se em torno da média:

- $E(e_i) = 0;$ (2.6)

.

Hipótese 2: Homocedasticidade ou variância constante. Os erros têm a mesma variabilidade em todos os níveis da variável X:

- $Var(e_i) = \sigma^2;$ (2.7)

Hipótese 3: Ausência de autocorrelação entre as perturbações. Os erros são não correlacionados:

- $Cov(e_i, e_j) = 0, i \neq j;$ (2.8)

Hipótese 4: Normalidade: e_i tem distribuição normal.

- $e_i \overset{id}{\sim} Normal(0, \sigma^2), i = 1,..,n$ (2.9)

onde, id é uma distribuição identicamente distribuída.

Vale ressaltar que por se tratar de dados *cross section* a hipótese de autocorrelação (2.8) não é necessária para a construção do modelo.

2.1.1. ESTIMADORES DE MÍNIMOS QUADRADOS

O método discutido agora é conhecido como Mínimos Quadrados Ordinários (OLS – *Ordinary Least Square*), no qual os parâmetros, β_0 e β_1 são escolhidos de forma a minimizar a quantidade denominada soma dos resíduos ao quadrado. (RSS- *Residual Sum Square*).

É importante ressaltar a distinção entre estimativas de parâmetros e parâmetros. Analogamente esta diferença deve ser realizada também para variância σ^2 e sua estimativa. As estimativas são obtidas através dos

estimadores. Os estimadores são representados pelo mesmo caractere utilizado para representar o parâmetro, com a adição do símbolo "^". Uma extensão desta notação apresentada é utilizada para identificar os valores ajustados pelo modelo de regressão (2.10) e os resíduos obtidos (2.11).

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{2.10}$$

$$\hat{e}_i = y_i - \hat{y}_i \tag{2.11}$$

A diferença entre os métodos de estimação de Mínimos Quadrados Ordinários e o de Máxima Verossimilhança[3] é que para a estimação clássica de mínimos quadrados obtêm-se apenas expressões analíticas para os estimadores dos parâmetros, sem que ocorra suposição alguma a respeito da distribuição de probabilidade dos erros. Isto inviabiliza a obtenção de intervalos de confiança e a realização de testes de hipóteses. Já para a estimação utilizando o método de máxima verossimilhança, além das expressões analíticas para os estimadores, trabalha-se também com uma distribuição de probabilidade normal $(0,\sigma^2)$ para os erros, viabilizando as inferências estatísticas.

Segundo KMENTA (1988), os dois métodos de estimação, nos levam às mesmas estimativas para os parâmetros de regressão, pois os estimadores produzidos tanto para um, quanto para outro são iguais, $\hat{\beta}_{LS} = \hat{\beta}_M$.

Concluímos então, segundo KMENTA (1988), que os estimadores de OLS tem todas as propriedades assintóticas desejáveis, pois são os mesmos que os estimadores de máxima verossimilhança, e estes últimos são conhecidos como sendo assintoticamente não-tendenciosos, consistentes e assintoticamente eficientes. Portanto, os estimadores de Mínimos Quadrados Ordinários dos parâmetros de regressão do modelo linear têm todas as propriedades desejáveis, assintóticas e de amostras finitas. Em modelos de

[3] A base do Método de Estimação de Máxima Verossimilhança é o princípio de que diferentes populações geram amostras diferentes; qualquer amostra examinada tem mais probabilidade de provir de algumas populações do que de outras. O estimador de máxima verossimilhança maximiza a probabilidade de gerar as observações da amostra considerada.

regressão linear, os estimadores OLS são não viciados, normalmente distribuídos, e ainda possuem variância mínima possível entre qualquer outra classe de estimadores. Essas propriedades são aceitas como as melhores propriedades que uma classe de estimadores podem apresentar (MELNT). Segundo GREENE (1997), podem existir outros estimadores, também não viciados, mas esses são menos eficientes no sentido de suas variâncias excederem a variância dos estimadores de Mínimos Quadrados Ordinários.

2.1.2. CRITÉRIO PARA MINIMIZAÇÃO

Uma vez obtidos as estimativas de β_0 e β_1 o valor ajustado de y é dado pela equação (2.10), pode-se então obter os resíduos através de uma outra equação (2.11). No método de mínimos quadrados são escolhidos β_0 e β_1 de forma a tornar a soma dos resíduos ao quadrado RSS tão pequeno quanto for possível.

$$RSS = \sum \hat{e}_i^2 = \sum (y_i - \hat{y}_i) = \sum \left[y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right) \right]^2 \qquad (2.12)$$

Esta proposta de estimador não implica em nenhuma hipótese sobre os erros $e_{i,}$ as estimativas podem ser obtidas mesmo que o modelo adotado de regressão não seja correto para os dados em questão.

Obtêm-se os estimadores de β_0 e β_1 através da resolução das equações apresentadas a seguir, igualando-se as derivadas em relação a cada parâmetro a zero.

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2 \sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0 \qquad (2.13)$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = -2 \sum x_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0 \qquad (2.14)$$

Desta forma obtém-se:

$$\hat{\beta}_0 = \left(\frac{\sum y_i}{n} \right) - \hat{\beta}_1 \left(\frac{\sum x_i}{n} \right) \qquad (2.15)$$

$$\hat{\beta}_1 = \left(\frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} \right) \qquad (2.16)$$

2.1.3. ESTIMAÇÃO DE VARIÂNCIA

Segundo WEISBERG (1947), uma estimativa que não dependa propriamente do modelo ajustado é desejado. Em geral, esse tipo de estimativa é obtida somente em conjunto de dados onde os vários valores de y são representados por vários valores de x ou são provenientes de informação à priori. Não ocorrendo essas situações especiais, a estimativa usual de (σ^2) é dependente do modelo em função da soma dos resíduos ao quadrado RSS, equação (2.12). Como a variância (σ^2) é essencialmente a média ao quadrado das medidas dos e_i's é natural esperar que o estimador seja representado por $\hat{\sigma}^2$, obtido através da média de RSS. Assumindo que os erros (e_i's) são variáveis aleatórias não-correlacionadas com média zero e mesma variância, um estimador não viciado para a variância é obtido dividindo RSS pelos graus de liberdade[4] do experimento.

$$\hat{\sigma}^2 = \frac{RRS}{n-p} \qquad (2.17)$$

Ainda de acordo com WEISBERG (1947), caso seja escolhida uma especificação inadequada para o modelo de regressão, ou seja, uma formulação errada da equação de regressão e de proposições ou pressupostos referentes aos regressores ou ao termo de perturbação ocorrerá uma superestimativa da variância. Segundo KMENTA (1988), caso a hipótese violada seja variância constante, os estimadores de Mínimos Quadrados dos coeficientes de regressão não são os melhores estimadores lineares não-tendenciosos (MELNT). Isso significa que os estimadores de Mínimos Quadrados não têm a menor variância numa classe de estimadores não-tendenciosos e que, portanto eles não são eficientes. Isso significa que se estimarmos os parâmetros de regressão sob a falsa crença de que a

[4] Os graus de liberdades podem ser definidos como o número de observações menos o número de parâmetros utilizados pelo modelo. Para o modelo linear p=2 (β_0 e β_1)

perturbação seja homocedástica, nossas inferências sobre os coeficientes da população serão incorretas.

Por outro lado, caso as hipóteses que envolvem a perturbação estocástica e_i: normalidade, média zero, homocedasticidade e não autocorrelação sejam verificadas, é possível estabelecer uma relação entre o estimador da variância e a distribuição Qui-quadrado com *(n-2)* graus de liberdade, conforme equação a seguir:

$$(n-2)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-2) \qquad (2.18)$$

Sendo assim é possível realizar testes estatísticos para verificar as hipóteses e estabelecer um intervalo de confiança em torno da variância. Em particular este fato implica[5] que o estimador da variância (2.17) é não viciado:

$$E(\hat{\sigma}^2) = \sigma^2 \qquad (2.19)$$

2.1.4. ALGUMAS PROPRIEDADES PARA OS ESTIMADORES DO MODELO LINEAR DE REGRESSÃO.

A primeira propriedade a ser destacada é o fato da soma dos resíduos ser nula[6], devido à soma balanceada de valores positivos e negativos. Isto é conseqüência de modelos nos quais o ajuste é realizado incluindo o parâmetro β_0 (intercepto)[7]

$$\sum_{i=1}^{n} e_i = 0 \qquad (2.20)$$

Segundo WEISBERG (1947), como e_i's são variáveis aleatórias, então os estimadores de β_0 e β_1 também o são, isso porque dependem dos valores

[5] Não é necessário constatar normalidade para que o estimador seja não viciado.

[6] Em função disto verifica-se que o valor esperado para o erro é zero. $E(e) = 0$

[7] Quando o ajuste é feito passando pela origem o somatório em questão, normalmente é diferente de zero.

observados de y_i's e de e_i's. Caso o valor esperado do erro seja igual a zero, $E(e_i) = 0$; *i=1,..,n* e o modelo está correto, então é possível verificar que:

$$E(\hat{\beta}_0) = \beta_0 \tag{2.21}$$

$$E(\hat{\beta}_1) = \beta_1 \tag{2.22}$$

$$Var(\hat{\beta}_1) = \hat{\sigma}\frac{1}{SXX} \tag{2.23}$$

$$Var(\hat{\beta}_0) = \hat{\sigma}\left(\frac{1}{n} + \frac{\bar{x}^2}{SXX}\right) \tag{2.24}$$

$$\mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2\frac{\bar{x}}{SXX} \tag{2.25}$$

onde;

$$SXX = \sum(x_i - \bar{x})^2 = \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n} = \sum x_i^2 - n(\bar{x})^2 \tag{2.26}$$

SXX (Corrected sum of squares for the x_i*'s).*

Ou seja, os estimadores de mínimos quadrados para os parâmetros são não viciados. Para maiores explicações veja Weisberg, Sanford (1947), pg – 242.

2.1.5. ANÁLISE DE VARIÂNCIA

A análise de variância fornece uma maneira conveniente de comparar o ajuste de dois ou mais modelos para um mesmo conjunto de dados. Note que esta metodologia possui um sentido mais amplo quando aplicada à análise de regressão múltipla, de qualquer forma seus princípios podem ser ilustrados nesta parte do trabalho.

Para introduzir esta metodologia imagine inicialmente um modelo de regressão do tipo:

$$y_i = \beta_0 + e_i \tag{2.27}$$

Onde o ajuste é realizado sem considerar nenhuma variável independente, conseqüentemente o modelo que ajusta é uma linha paralela ao eixo X. O ajuste é dados por:

$$\hat{y}_i = \hat{\beta}_0 \tag{2.28}$$

Lembrando que o ajuste é realizado através da minimização dos quadrados, logo, a obtenção do estimativa de β_0 é obtida através da expressão $\sum\left(y_i - \hat{\beta}_0\right)^2$. Não é difícil verificar que $\hat{\beta}_0 = \bar{y}$, logo a expressão citada torna-se *SYY* (soma total de quadrados das diferenças em relação a média).

$$SYY = \sum\left(y_i - \bar{y}\right)^2 \tag{2.29}$$

Neste caso, devido à falta de variáveis independentes, *SYY* e *RSS* são iguais possuindo n-1 graus de liberdade, devido ao número de parâmetros do modelo. O modelo que inclui o parâmetro β_1 apresenta um *RSS* de acordo com a equação a seguir:

$$RSS = SYY - \frac{(SXY)^2}{SXX} \tag{2.30}$$

Porém este possui n-2 graus de liberdade, devido a apresentar dois parâmetros. A análise de variância realiza um procedimento com base na diferença entre as equações (2.29) e (2.30).

$$SSreg = SYY - RSS \tag{2.31}$$

Denominado de soma de quadrados da regressão, que terá seu significado explicado na seção seguinte. Da mesma forma os graus de liberdade devem ser subtraídos, ou seja, (n-1)–(n-2) = 1. Com base em todas essas informações é possível construir a tabela denominada de tabela de análise de variância.

Tabela 2.1 – Análise de Variância

	Graus Liberdade	Soma de Quadrados	Média	Estatística F
Reg em X	*1*	*SSreg*	*SSreg/1*	*Ssreg (n-2)/RSS*
Resíduo	*n-2*	*RSS*	*RSS/n-2*	
Total	*n-1*	*SYY*		

A última coluna da tabela 2.1 é referente à estatística F. Este valor é utilizado para a realização de um teste de hipótese, visando à verificação da significância da variável independente.

2.1.6. COEFICIENTE DE DETERMINAÇÃO R^2

Através da equação (2.31) é possível desenvolver o conceito de coeficiente de determinação, bastando apenas dividi-la pela estatística *SYY*. A equação (2.30), *SSreg*, é a variabilidade explicada pelo modelo de regressão em X, ou melhor, o quanto o modelo retrata a variável X para explicar a variável Y. O coeficiente de Determinação R^2 pode ser visto como uma padronização desta medida, pois para obtê-lo basta dividir a medida *SSreg* pela variação total em Y, *SYY*, ou como é chamada soma total de quadrados.

$$R^2 = \frac{SSreg}{SYY} = 1 - \frac{RSS}{SYY} \qquad (2.32)$$

Note que o R^2 pode ser calculado com as medidas obtidas da tabela de Análise de Variância. Variando entre zero a um, normalmente é visto na forma de percentual, é uma estatística muito popular e não sofre perda de generalidade para com a análise de regressão múltipla.

Uma observação importante a ser feita é o fato de que o coeficiente de determinação, para o caso de regressão univariado, é igual o coeficiente de correlação ao quadrado.

$$R^2 = \frac{SSreg}{SYY} = \frac{\left(SXY\right)^2}{\left(SXX\right)\left(SYY\right)} = r_{XY}^2 \tag{2.33}$$

2.1.7. INTERVALOS DE CONFIANÇA

Segundo WEISBERG (1947) quando os erros são independentes e normalmente distribuídos de acordo com (2.9) então os parâmetros estimados, os valores ajustados e as previsões também serão normais, isso porque todos estes valores são combinações lineares dos valores y_i's, que por sua vez também é uma combinação linear dos erros e_i's. Conseqüentemente, intervalos de confiança e testes de hipóteses podem ser baseados na distribuição *t-student.*

A partir de agora serão apresentadas às estatísticas dos desvios padrão e intervalos de confiança para cada termo resultado de combinações lineares dos e_i's.

$$dp(\hat{\beta}_0) = \hat{\sigma}\left(\frac{1}{n} + \frac{\overline{x}^2}{SXX}\right)^{\frac{1}{2}} \tag{2.34}$$

$$dp\left(\hat{\beta}_1\right) = \frac{\hat{\sigma}}{\sqrt{SXX}} \tag{2.35}$$

Com base no desvio padrão e na estatística *t-student*, estabelece-se os intervalos de confiança para o parâmetro β_0 e β_1.

$$\hat{\beta}_0 - t_{(\alpha,n-2)}dp\left(\hat{\beta}_0\right) \le \beta_0 \le \hat{\beta}_0 + t_{(\alpha,n-2)}dp\left(\hat{\beta}_0\right) \tag{2.36}$$

$$\hat{\beta}_1 - t_{(\alpha,n-2)}dp\left(\hat{\beta}_1\right) \le \beta_1 \le \hat{\beta}_1 + t_{(\alpha,n-2)}dp\left(\hat{\beta}_1\right) \tag{2.37}$$

Sendo (1-α) 100% o percentual de confiança do intervalo.

Analogamente têm-se os desvios padrões para os valores ajustados e para os valores previstos, assim sendo possível obter os respectivos intervalos de confiança para esses valores.

$$dp.ajuste\left(\hat{y} \mid x\right) = \hat{\sigma}\left(\frac{1}{n} + \frac{\left(x-\overline{x}\right)^2}{SXX}\right)^{\frac{1}{2}} \tag{2.38}$$

$$dp.previ\left(\tilde{y} \mid \tilde{x}\right) = \hat{\sigma}\left(1 + \frac{1}{n} + \frac{\left(x-\overline{x}\right)^2}{SXX}\right)^{\frac{1}{2}} \tag{2.39}$$

$$\hat{y} - t_{(\alpha,n-2)} dp.ajuste\left(\hat{y} \mid x\right) < y < \hat{y} + t_{(\alpha,n-2)} dp.ajuste\left(\hat{y} \mid x\right) \tag{2.40}$$

$$\tilde{y} - t_{(\alpha,n-2)} dp.previ\left(\tilde{y} \mid x\right) < y < \tilde{y} + t_{(\alpha,n-2)} dp.previ\left(\tilde{y} \mid x\right) \tag{2.41}$$

O desvio padrão e o intervalo de confiança utilizado para realizar estimativas a respeito da previsão apresentam uma notação até então não mencionada, o $\tilde{y}$, que tem como objetivo destacar a diferença entre valores ajustados e valores previstos. Entendem-se como valores previstos para y, todo aquele obtido com valores de x não utilizados na estimação do modelo.

2.1.8. TESTES DE HIPÓTESE

Na seção anterior o critério de testes de hipótese foi bastante mencionado, isso porque existe uma forte ligação entre a obtenção de um intervalo de confiança e a realização de um teste de hipótese.

Isso se caracteriza mais quando o foco é a obtenção de parâmetros para o modelo de regressão. Os testes têm como finalidade verificar a

utilização de uma determinada variável proposta ao modelo, através da inclusão de seu parâmetro, para isso é verificada a seguinte hipótese.

$$\begin{aligned} &Hipótese\ Nula\ \ \beta_i = 0 \\ &Hipótese\ Alternativa\ \ \beta_i \neq 0 \end{aligned} \tag{2.42}$$

Note que até o momento abordou-se o tema de regressão linear simples, o que não torna este teste muito útil. Sua utilização é intuitiva em um modelo de regressão múltipla, onde existem "p" variáveis e nem todas são significativas.

O teste simplesmente compara a estatística estimada como seu valor proposto na hipótese acima, através da distribuição *t-student* com n-2 graus de liberdade e o nível de confiança desejado (1-α). Note que o *dp(estimador)* da equação (2.43) é o desvio padrão do parâmetro a ser testado, de acordo com as equações (2.33) e (2.34).

$$t = \frac{\hat{\beta} - \beta.hipótese}{dp\left(\hat{\beta}\right)} \tag{2.43}$$

Outro teste importante, para identificação das variáveis do modelo é o Teste F. Sua estatística está presente na tabela de análise de variância, citada anteriormente. A constituição de suas hipóteses, nula e alternativa segue a seguir:

$$\begin{aligned} &Hipótese\ Nula\ \ y_i = \beta_0 + e_i \\ &Hiipótese\ Alternativa\ \ \mathrm{y_i} = \beta_0 + \beta_1 \mathrm{x_i} + \mathrm{e_i} \end{aligned} \tag{2.44}$$

Sendo os e_i's distribuídos normalmente, com média zero e variância qualquer e de acordo com as hipóteses imputadas anteriormente, então a hipótese nula segue uma distribuição F com os graus de liberdade do numerador e denominador (1 e n-2; regressão simples).

2.1.9. ANÁLISE DOS RESÍDUOS

Os resíduos $\hat{e}_i$ possuem as informações sobre o erro e sobre a veracidade das hipóteses assumidas pelo modelo. Qualquer análise requer a verificação detalhada dos resíduos. Um gráfico muito importante nesta análise é *plot* resíduos versus valores ajustados e não com os valores observados porque os erros e os valores observados são usualmente correlacionados. Espera-se neste gráfico que os resíduos estejam distribuídos aleatoriamente em torno de zero. O comportamento dos resíduos em relação aos valores ajustados serve para examinar a suposição de variância constante (homocedasticidade). Geralmente, a falta de homogeneidade de variâncias tende a produzir um gráfico em forma de megafone. A curvatura deste gráfico pode dar indícios sobre a adequação do modelo ajustado aos dados se é correta ou não. Resíduos muito grandes podem ser indícios de *outliers.*

2.2. REGRESSÃO MÚLTIPLA

Segundo WEISBERG (1947) neste tipo de procedimento, ocorre a utilização de um conjunto de variáveis independentes entre si e supostamente[8] dependentes com a variável resposta. Da mesma forma realizada para regressão linear simples, coleta-se n observações para as variáveis denominadas independentes e para a denominada resposta. As independentes são chamadas comumente de $X_1,...,X_p$ onde p refere-se ao número de variáveis, a independente é chamada de Y. O conjunto de dados é formado por uma matriz de (p+1) x n.

A regressão múltipla, além das hipóteses as descritas na seção 2.1, requer mais uma hipótese, devido à inclusão de mais de uma variável independente. Esta hipótese acaba por justificar este nome. Ocorre a necessidade de que essas variáveis não apresentem multicolineariedade, ou seja, que não influenciem ou sejam influenciadas por nenhuma[9] outra. Caso isso ocorra é necessário que seja realizada algum tipo de transformação nas

[8] O termo "supostamente", faz referência ao fato de ocorrerem variáveis independentes que são utilizadas indevidamente no modelo, pois não tem grande significância para variável resposta.
[9] Exceto a variável Y, a qual deseja-se que seja altamente influenciada pelas variáveis X_i's.

variáveis, ou que estas sejam retiradas do modelo. Ocorre que em certos casos duas variáveis "diferentes" acabam contendo a mesma informação. A seguir a tabela 2.2 esquematiza como são dispostos observações para uma regressão múltipla.

Tabela 2.2 – Matriz de Dados para Regressão Múltipla.

	Y	X_1	X_2	...	X_p
1	y_1	x_{11}	x_{21}	...	x_{p1}
2	y_2	x_{21}	x_{22}	...	x_{p2}
3	y_3	x_{31}	x_{23}	...	x_{p3}
⋮	⋮	⋮	⋮	...	
n	y_n	x_{n1}	x_{n2}	...	X_{np}

Para o modelo de regressão múltipla, a equação que expressa a variável resposta é uma função linear de *p* variáveis independentes, onde os coeficientes desta equação são estimados a partir de dados observados, de acordo com a tabela acima. O modelo é dado de acordo com a expressão a seguir.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + e \qquad (2.45)$$

Onde, da mesma forma que visto na última seção, β é um parâmetro desconhecido, a ser estimado. E o valor "*e*", representa o erro aleatório, Y é a variável resposta ou dependente, e $X_1,...,X_p$. são as variáveis independentes.

Os dados e os parâmetros para o modelo de regressão múltipla são apresentados em notação matricial a seguir:

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \qquad (2.46)$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{2.47}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \tag{2.48}$$

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{2.49}$$

As estimativas de β são obtidas através do mesmo critério utilizado no modelo linear simples, através do Método de Mínimos Quadrados. Devido à notação matricial os estimadores apresentam-se de forma diferente.

$$RSS = \hat{e}^T \hat{e} = \left(Y - X\hat{\beta}\right)^T \left(Y - X\hat{\beta}\right) \tag{2.50}$$

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T Y \tag{2.51}$$

$$Var\left(\hat{\beta}\right) = \sigma^2 \left(X^T X\right)^{-1} \tag{2.52}$$

Para obter a variância estimada do estimador de β, basta substituir na equação anterior o valor de σ^2 por sua estimativa, dada pela seguinte equação:

$$\hat{\sigma}^2 = \frac{RSS}{n-(p+1)} \tag{2.53}$$

2.2.1.ANÁLISE DE VARIÂNCIA NO CASO DE REGRESSÃO MÚLTIPLA

Análise de variância no caso de regressão múltipla é uma técnica muito rica, utilizada para dividir a variabilidade e assim comparar modelos que possuem diferentes cenários. Um modelo completo, com todas as variáveis, é comparado a um modelo reduzido, sem nenhuma variável, somente com intercepto. Este procedimento é análogo ao apresentado anteriormente, porém como se trata de um modelo com múltiplas variáveis é necessário que este fato seja agregado a análise. Isto é feito, através dos graus de liberdades, que devem ser subtraídos pelo número de variáveis independentes utilizadas.

A tabela desta forma sofre algumas alterações que podem ser constatadas abaixo.

Tabela 2.3 – Análise de Variância com Múltiplas Variáveis

	Graus Liberdade	Soma de Quadrados	Média	Estatística F
Reg em X1,..,Xp	*p*	*SSreg*	*SSreg/p*	*SSreg(n-p-1)/RSS(p)*
Resíduo	*n-p-1*	*RSS*	*RSS/n-p-1*	
Total	*n-1*	*SYY*		

Conforme este procedimento a distribuição F deve conter (p, n-p-1) graus de liberdades. E a hipótese[10] estabelecida para o teste, neste caso é do tipo:

$$\begin{aligned} &Hipótese\ Nula \quad \beta_0 = \beta_1 = ... = \beta_p = 0 \\ &Hipótese\ Alternativa \quad \beta_0 = \beta_1 = ... = \beta_p \neq 0 \end{aligned} \tag{2.54}$$

Diferentemente do teste realizado com a estatística *t-student*, que foi apresentado na seção 2.1.8 denominada, teste de hipótese. A ANOVA realiza o teste de significância para todos os parâmetros do modelo simultaneamente.

[10] Esta hipótese é análoga à realizada para o caso simples.

3. REGRESSÃO COM MÉTODOS ROBUSTOS

No capítulo anterior tratou-se da abordagem clássica para o modelo linear de regressão, tanto para o caso simples quanto para o caso de múltiplas variáveis explicativas no modelo. Esta abordagem utilizou estimadores de mínimos quadrados ordinários para obter suas estimativas. Contudo, basta a presença de um *outlier*[11] e a violação dos pressupostos básicos para que esse estimador perca as suas propriedades de melhores estimadores lineares não-tendenciosos (MELNT).

O estimador de mínimos quadrados data de 1800, isso é um dos motivos de sua popularidade, pois o fato da ausência de computadores na época, exigia o uso de expressões analíticas ao invés de métodos iterativos. Apesar de não existir mais o problema computacional, a grande maioria dos *softwares* ainda utilizam esta mesma técnica, muitas vezes devido à tradição.

Segundo MENDES (1999), um trabalho de maior importância no desenvolvimento da teoria da robustez foi o trabalho de TUKEY (1960) e o *Statistical Research Group at Princeton*. Neste trabalho o papel da média aritmética foi revisto e criticado e novas propostas tais como a média podada e a média reponderada foram investigadas. Somente nas últimas décadas apareceram as primeiras formalizações da teoria da estimação robusta de forma sistemática. A primeira abordagem sistemática no desenvolvimento de procedimentos robustos, chamada de abordagem minimax, surgiu com HUBER (1964). Ainda segundo MENDES (1999), e no contexto da estimação de parâmetros de locação, HUBER define formalmente o que seria uma vizinhança de um modelo probabilístico, a qual deveria conter a distribuição assumida como verdadeira. Em 1973 HUBER estende suas idéias ao contexto da regressão.

Com a utilização dos Mínimos Quadrados Ordinários, constatou-se que grande parte das análises realizadas com dados observados, estes não atendiam as hipóteses básicas enumerados no capítulo anterior. Os efeitos disto podem ser vistos em Student 1927, Pearson 1931, Box 1953 e Tukey 1960. Podem ser citados como exemplo os *outliers* em relação à *Y*, também

conhecidos como *outliers* de regressão. Um único ponto tem capacidade de influenciar a reta de mínimos quadrados, de forma que esta fique completamente desajustada ao conjunto de dados (figura 3.2).

Figura 3.1 - Ajuste OLS Sem *Outliers*

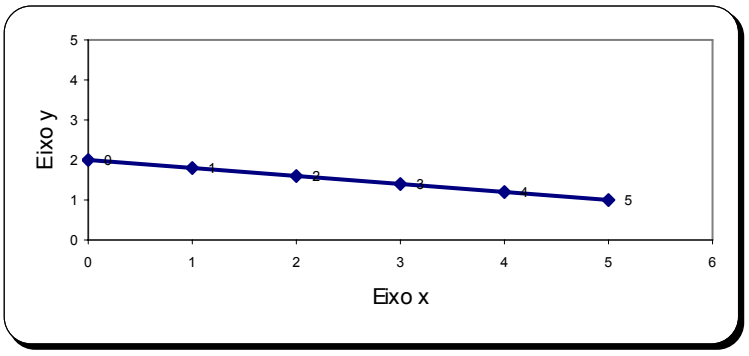
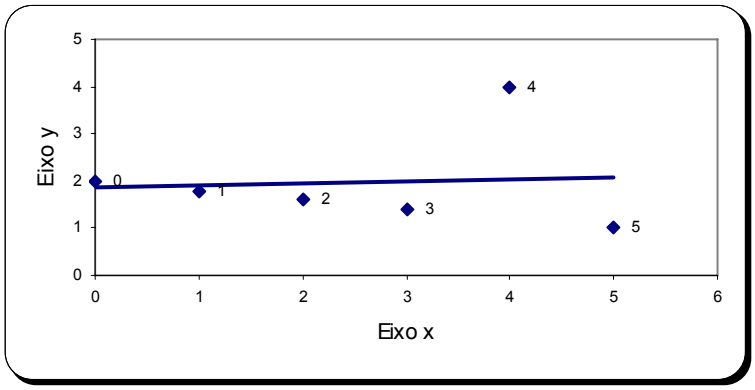


Figura 3.2 - Ajuste OLS Com *Outliers* em Y



De acordo com as figuras 3.1 e 3.2 é possível constatar o que foi dito a respeito de *outliers.* O caso em questão trata-se de um *outlier* em *Y*, graças a isto, os resíduos resultantes são grandes positiva ou negativamente. Note que a situação apresentada nestas figuras contempla apenas o problema[12] para a variável independente.

Contudo não há razão para considerar um *outlier* em *Y* como o pior dos casos, *outliers* em *X* podem provocar grandes distúrbios como pode ser visto nas figuras 3.3 e 3.4. É bem razoável que este tipo de problema ocorra com mais freqüência, do que o caso em Y. Isso porque se têm, na maioria das

[11] Entende-se por *outliers* observações discrepantes, com valores atípicos ao restante do conjunto de dados

[12] Em casos com *p* variáveis não é possível observar tão claramente os *outliers.*

vezes, *p* variáveis explicativas, o que proporciona maior probabilidade de erro.

Figura 3.3 - Ajuste OLS Sem *Outliers*

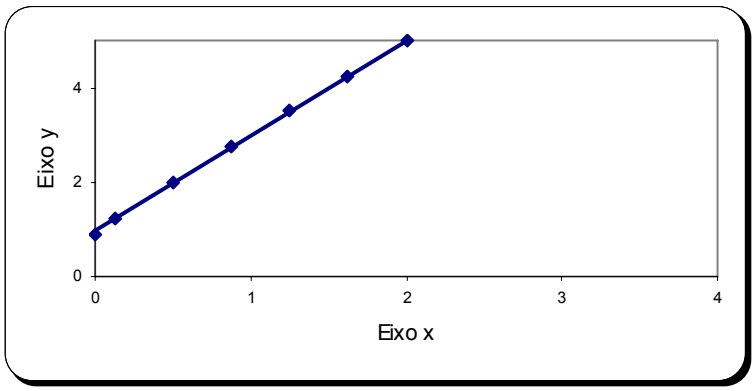
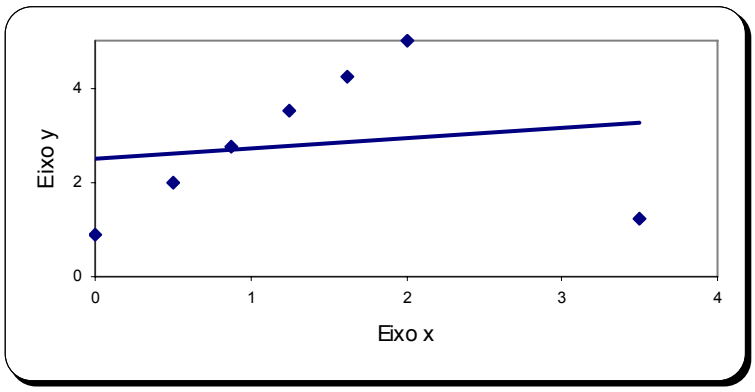


Figura 3.4 - Ajuste OLS Com *Outliers* em *X*



De acordo com a figura 3.4 um *outlier* em *X* pode provocar um completo "desajuste", bastando apenas um único valor discrepante. Este tipo de *outlier* é dito ponto de alavanca, nome dado em função da analogia com mecanismos de alavanca[13].

Segundo HAMPEL et al (1986) no caso de regressão múltipla um ponto de alavanca é definido pelo ponto $(x_{k1},...,x_{kp},y_k)$, *no qual* $(x_{k1},...,x_{kp})$ é atípico em relação ao vetor $(x_{i1},...,x_{ip})$ do conjunto de dados. Em determinadas situações onde se tem um pequeno número de *outliers* o método de mínimos quadrados ordinários funciona razoavelmente bem, bastando apenas retira-los ou corrigi-los. Caso contrário o método passa a ficar comprometido.

[13] Nem todo ponto de alavanca é ruim, é possível ter um grande valor de x e um grande valor de y, resultando em ponto próximo da reta ajustada, de acordo com a tendência dos dados.

Segundo ROUSSEEUW *e* LEROY (1987), uma outra abordagem para resolver o problema de estimação é a utilização de métodos robustos, que serão apresentados neste capítulo. Os métodos robustos são pouco sensíveis a *outliers*. Sendo assim ao analisar os resíduos, é possível detectar os pontos discrepantes com certa facilidade, o que não ocorre utilizando estimadores de Mínimos Quadrados Ordinários (OLS – *Ordinary Least Square*).

Os diagnósticos através de OLS, denominados clássicos, vêm a ter os mesmos objetivos de uma regressão robusta, porém em ordem inversa. Ao utilizar diagnósticos, tenta-se primeiro detectar e remover os *outliers* e então realizar um ajuste adequado através de OLS. Por outro lado, o método robusto ajusta o modelo aos dados, em sua grande maioria, e então detecta os *outliers*, através dos pontos que produzem grandes resíduos. É importante destacar que grandes resíduos podem ser provenientes de uma especificação inadequada do modelo.

Serão apresentados estimadores robustos, com a finalidade de utilizá-los no modelo de Análise Condicionada da Demanda. Este modelo foi proposto inicialmente tratando das estimativas para os parâmetros desconhecidos β, através de estimativas OLS. Porém, antes de apresentá-los é necessário introduzir os conceitos fundamentais de ponto de ruptura e eqüivariância. Estes são propriedades particulares dos estimadores robustos e serão decisores na escolha pelo que melhor atende as necessidades apresentadas por esta dissertação.

3.1. PONTO DE RUPTURA

Como já foi dito anteriormente, ao utilizar-se o método de OLS, uma única observação pode torná-lo ineficiente, como apresentado na figura 3.4.. Porém existem certos estimadores que promovem bons resultados, mesmo sob um percentual de observações discrepantes. Para formalizar este conceito introduz-se agora a definição de ponto de ruptura. Segundo ROUSSEEUW *e* LEROY (1987) esta abordagem foi realizada por Hodges (1967), para análise unidimensional e depois generalizada por Hampel

(1971). Porém esta última análise tem caráter assintótico. Donoho e Hurber (1983) introduziram este conceito de ponto de ruptura para amostras finitas. Seja uma amostra de *n* observações, representada pela equação 3.1.

$$Z = \left\{ \left(x_{11}, ..., x_{1p}, y_1 \right), ..., \left(x_{n1}, ..., x_{np}, y_n \right) \right\} \tag{3.1}$$

Onde T é um estimador[14] de regressão, aplicado à amostra Z. Implicando assim em um vetor de coeficientes, para um modelo linear de regressão, representado a seguir.

$$T\left(Z\right) = \hat{\theta} \tag{3.2}$$

Considere a partir de agora todas as possíveis amostras corrompidas, definidas por Z‘. Estas amostras são obtidas substituindo alguns *m* pontos dos dados originais por valores aleatórios, onde *m*<*n*. Denominamos *vício*(*m*;(T,Z)), a máxima contaminação causada por:

$$vício\left(m;\left(T,Z\right)\right) = \sup_{Z'} \left\| T\left(Z'\right) - T\left(Z\right) \right\| \tag{3.3}$$

Onde o supremo é obtido examinando-se todas as amostras corrompidas possíveis de Z‘.

Caso o *vício*(*m*,T;Z), seja infinito, significa que m *outliers* podem ter um grande efeito em T, que pode ser expresso através da definição de ruptura do estimador. Porém como a abordagem tem como foco amostras finitas o ponto de ruptura para um estimador T, para a amostra Z é definida como:

$$\varepsilon_n^*\left(T,Z\right) = \min\left\{ \frac{m}{n} : vício\left(m;T,Z\right) \right\} \ , \ vício\left(m;T,Z\right) \to \infty \tag{3.4}$$

Assim, o ponto de ruptura é a menor fração de contaminação que pode fazer com que o estimador T assuma valores arbitrários distantes dos valores

[14] Estimador sem critério ou método definido.

T(Z). Segundo MENDES (1999), isto quer dizer que m *outliers* têm o potencial de estragar por completo as estimativas produzidas pelo estimador T. O valor mais alto possível para o ponto de ruptura de estimadores de regressão com propriedades de equivariância é 0,5. Não faz sentido pensarmos em dados contendo fração maior que 50% de contaminação, já que neste caso não poderíamos distinguir quais os dados "ruins". É importante notar que a definição de ponto de ruptura não supõe uma determinada distribuição para os dados. Possuir ponto de ruptura de 50% é um critério global para a construção de estimadores robustos em relação a pontos de alavanca que são *outliers* de regressão.

Para o caso dos estimadores de Mínimos Quadrados Ordinários, tem-se que apenas um *outlier* é suficiente para fazer com que o estimador T fique distorcido. Seu ponto de ruptura é:

$$\varepsilon_n^*(T,Z) = \frac{1}{n} \qquad (3.5)$$

Que claramente tende a zero, quando o tamanho da amostra n cresce. Podendo dizer-se que estimador OLS possui um ponto de ruptura de 0%. O que reflete a extrema sensibilidade deste método para com *outliers.*

3.2. EQÜIVARIÂNCIA

O conceito de eqüivariância, do ponto de vista estatístico, tem como base definir estimadores que transformam corretamente suas estimativas de acordo com as mudanças realizadas nas observações. É importante destacar dois estimadores eqüivariantes, porém com baixo ponto de ruptura: Mínimos Quadrados Ordinários – OLS e Valores Absolutos dos Resíduos - L_1, equação (3.9).

Mediante as características dos estimadores abordados nesta dissertação serão apresentados três tipos de eqüivariâncias, são elas: *de Regressão, de Escala e de Afinidade ou Afim.*

Um estimador T é dito equivariante de regressão se satisfaz:

$$T\left(\left\{\left(x_i; y_i + x_i v\right); i = 1,...,n\right\}\right) = T\left(\left\{\left(x_i; y_i\right); i = 1,...,n\right\}\right) + v \,, \quad (3.6)$$

v é um vetor qualquer

O estimador T é dito equivariante de escala se satisfaz:

$$T\left(\left\{\left(x_i; cy_i\right); i = 1,...,n\right\}\right) = cT\left(\left\{\left(x_i; y_i\right); i = 1,...,n\right\}\right) \,, \quad (3.7)$$

c é uma constante

O estimador T é dito equivariante de afinidade:

$$T\left(\left\{\left(x_i A; y_i\right); i = 1,...,n\right\}\right) = A^{-1}T\left(\left\{\left(x_i; y_i\right); i = 1,...,n\right\}\right) \,, \quad (3.8)$$

A é uma matrix quadrada, não singular

Estas são propriedades matemáticas que garantem que as modificações realizadas nas observações serão transmitidas de alguma forma às estimativas produzidas pelos estimadores.

A seguir apresentamos uma análise de alguns estimadores robustos e de suas propriedades, fundamentadas em ROUSSEEUW *e* LEROY (1987).

3.3. ESTIMADORES ROBUSTOS

Para limitar a influência dos *outliers* na modelagem foram propostos vários tipos de estimadores robustos dos parâmetros do modelo. A partir de agora serão introduzidos os estimadores que foram utilizados até alcançar a proposta do MM-estimador, o qual foi utilizado nesta dissertação, visto que este estimador acumula os desenvolvimentos conceituais daqueles que os antecederam.

3.3.1. ESTIMADOR L_1

Primeiramente, para resolver o problema de sensibilidade dos estimadores clássicos, foi tentada a utilização de estimadores robustos por Edgeworth (1887), através de uma proposta de Boscovich. Acreditavam que a grande influência sofrida pelos *outliers*, no OLS, era causada pela utilização dos resíduos ao quadrado. Sendo assim propuseram um estimador que minimizasse os valores absolutos. Onde r_i é o valor do resíduo obtido entre a diferença do valor observado e o estimado.

$$\min_{\hat{\theta}} \sum_{i=1}^{n} |r_i| \tag{3.9}$$

Este método veio a ser conhecido como estimador de norma L_1, uma vez que os resíduos de regressão são elevados à primeira potência, o Método de Mínimos Quadrados tem norma L_2. Porém o estimador L_1 não apresenta diferença do L_2 em relação ao seu ponto de ruptura, que também é de 0%. Isso pode ser verificado nos gráficos a seguir.

Figura 3.5 – Estimador L_1 com *outliers* em Y

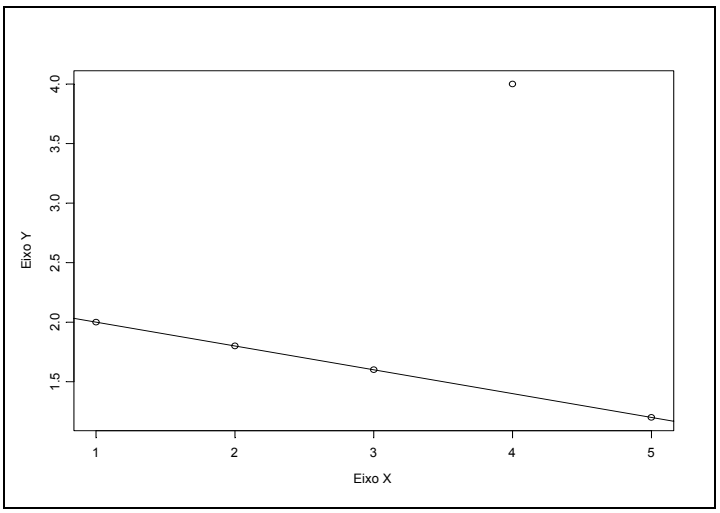
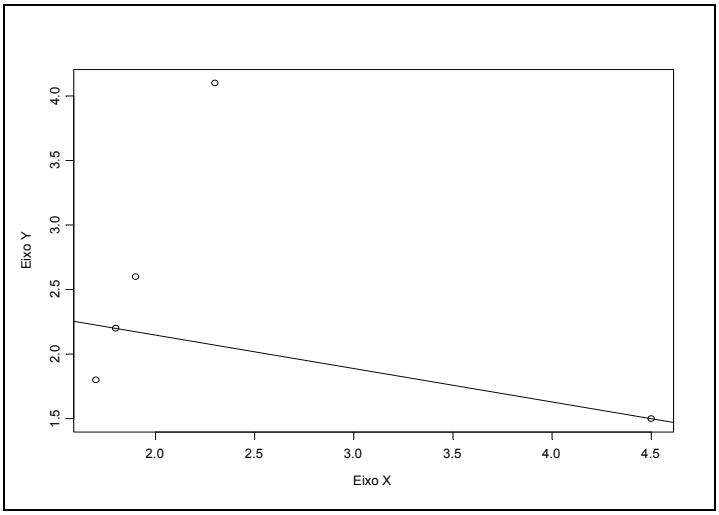


Figura 3.6 – Estimador L_1 com *outliers* em X



Como se pode perceber este estimador apresenta um bom resultado somente para *outliers* em Y (Figura 3.5), tornando-se preferível neste aspecto ao estimador do OLS. Porém note na figura 3.6 a existência de um ponto de alavanca, que resulta em um total desajuste da reta em relação aos dados, mostrando que basta apenas uma observação discrepante para inviabilizar a utilização deste método.

3.3.2. ESTIMADOR M – (Maximum likelihood)

O próximo passo na busca de um estimador adequado foi à utilização dos M-estimadores, propostos por HUBER (1973). Estes estimadores apresentam em sua idéia a substituição da função quadrática aplicada aos resíduos, por uma outra função denominada como função perda e definida pelo símbolo ρ. Esta função é simétrica, ou seja, $\rho(-t) = \rho(t)$ para todo t, apresentando um único mínimo no zero, note que ao decorrer deste capítulo será apresentado um exemplo da função de perda ρ.

Como feito anteriormente para o estimador de mínimos quadráticos, minimiza-se a expressão (3.10) para determinar os parâmetros do modelo de regressão. Para isso diferencia-se (3.10) e iguala-se o resultado a zero, obtendo a expressão (3.11).

$$Mínimo_{\hat{\theta}} \sum_{i=1}^{n} \rho(r_i) \tag{3.10}$$

Onde $\psi(r_i)$ é a derivada da função perda ρ em relação ao regressor θ_i.

$$\sum_{i=1}^{n} \psi(r_i) x_i = 0 \tag{3.11}$$

Note que tanto x_i quanto zero são vetores:

$$x_i = (x_{i1},...,x_{ip}) \tag{3.12}$$

$$0 = (0,..,0) \tag{3.13}$$

A solução da equação (3.11) é extremamente complexa devido a ser um sistema de *p* equações. Na solução deste problema são utilizadas soluções iterativas com base no método de mínimos quadrados re-ponderados, também conhecidos como algoritmo H, introduzido por HOLLAND & WELSCH (1977). Porém esta solução não apresentam as propriedades de eqüivariancia, em relação às modificações na variável resposta Y.

Para tratar este problema da eqüivariancia foi utilizada a medida de desvio padrão para padronizar os resíduos, sendo que a estimativa do desvio deve ser realizada simultaneamente com solução de (3.11) tomando a nova expressão descrita por (3.14). Para isso HUBER, motivado pelos argumentos minimax da variância assintótica, propôs a função (3.15).

$$\sum_{i=1}^{n} \psi\left(\frac{r_i}{\hat{\sigma}}\right) x_i = 0 \tag{3.14}$$

$$\psi(t) = \min(c, \max(t, -c)) \tag{3.15}$$

Os M-estimadores que utilizam a função citada acima são assintoticamente mais eficientes[15] do que o estimador do tipo L_1 apresentado por (3.9) e resolvem o problema de robustez em relação à variável resposta. Porém estes estimadores ainda apresentam um ponto de ruptura igual aos L_2, isso devido aos efeitos de *outliers* em relação à x.

3.3.3. ESTIMADOR GM (*Generalized Maximum Likelihood*)

Em função de toda esta vulnerabilidade que os pontos de alavanca impunha aos M-estimadores, introduziu-se os GM-estimadores, visando limitar estes *outliers* em x_i com base em uma função de peso *w*. Mallows (1975) propôs a mudança da equação (3.14) por:

$$\sum_{i=1}^{n} w(x_i)\psi\left(\frac{r_i}{\hat{\sigma}}\right)x_i = 0 \qquad (3.16)$$

Poucos anos depois Schweppe (veja Hill 1977) propôs:

$$\sum_{i=1}^{n} w(x_i)\psi\left(\frac{r_i}{w(x_i)\hat{\sigma}}\right)x_i = 0 \qquad (3.17)$$

Todos estes estimadores foram construídos na esperança de limitar a influência de uma única observação do tipo ponto de alavanca, este tipo de efeito é normalmente medido com a função denominada função de influência (Hampel 1974). Baseada no critério de escolha ótima para as funções ψ e w, feitas por (Hampel 1978, Krasker 1980, Krasker e Welsch 1982, Roncheti e Rousseeuw 1985 e Samarov 1985). Com tudo isso os GM estimadores, conhecidos atualmente como estimadores limite-influência, não se firmaram como estimadores adequados. Estes apresentam um ponto de ruptura que decresce com o aumento do número de variáveis independentes, coeficientes, que o modelo assume, quanto maior p, menor o ponto de ruptura.

[15] Desde que os resíduos sejam normalmente distribuídos.

3.3.4. ESTIMADOR LMS – (*Least Median Square*)

Vários outros estimadores foram propostos, baseados nos métodos de Wald (1940), Nair e Shrivastava (1942), Bartlett (1949) e Brown e Mood (1951), entre outros. Porém, para o modelo de regressão simples nenhum método apresentou um ponto de ruptura superior a 30%, sendo que alguns não foram definidos para uma quantidade parâmetros superior a dois (p>2).

Todos estes fatores sobre estimadores robustos convergiam para o único questionamento. Seria possível realizar uma regressão robusta com um estimador com ponto de ruptura de 50%? A resposta veio através de Siegel (1982) com o *repeated median estimator,* porém este estimador não é eqüivariante para transformações lineares nas variáveis independentes X's. Note que o valor máximo que o ponto de ruptura pode suportar é 50%, visto que após isso torna-se impossível distinguir o que seria a parte "boa" e a "ruim" de uma amostra.

Surge então o estimador LMS, para compreendê-lo é necessário voltar ao estimador L_2, ou seja, de Mínimos Quadrados Ordinários (OLS). Na verdade este estimador chama-se soma dos Mínimos Quadrados, porém aparentemente a palavra soma foi retirada de seu nome. Talvez devido a isto, alguns estudiosos na tentativa de construção de estimadores robustos vislumbravam apenas a substituição da função quadrática por uma outra. A proposta do estimador LMS, Rousseeuw (1984), propõe a troca do somatório da equação (3.10), mantendo a função ρ quadrática.

$$\underset{\hat{\theta}}{Min} \quad \underset{i}{Mediana} \; r_i^2 \tag{3.18}$$

Esta proposta foi essencialmente baseada na idéia apresentada por Hampel (1975). Este estimador é extremamente robusto em relação tanto aos *outliers* de regressão quanto aos pontos de alavanca, ou seja, tanto em Y quanto em X. Também é eqüivariante para transformações lineares nas variáveis explicativas, isso por que (3.18) utiliza somente os resíduos. Porém do ponto de vista de sua eficiência assintótica, este estimador é pobre, pois possui uma taxa lenta de convergência. Note que devido aos avanços computacionais isto não o torna um estimador ruim.

3.3.5. ESTIMADOR LTS – (*Least Trimed Square*)

Devido aos problemas apresentados pelo LMS, Rousseeuw (1983,1984), introduziu seu estimador LTS dados por:

$$\underset{\hat{\theta}}{Mínimo} \sum_{i=1}^{h} \left(r^2 \right)_{i:n} \tag{3.19}$$

Onde, $(r^2)_{1:n} \leq ... \leq (r^2)_{n:n}$, são os quadrados dos resíduos ordenados[16]. Esta minimização lembra a realizada pelo OLS. A diferença é que os maiores resíduos quadráticos não são utilizados no somatório. Este estimador apresenta uma taxa de convergência razoável e possui eficiência assintótica, assim como o LMS, este estimador é eqüivariante para transformações lineares sobre x_i. As propriedades robustas apresentam melhores resultados quando *h* é aproximadamente *n / 2*, neste caso seu ponto de ruptura é de 50%.

3.3.6. S - ESTIMADOR

Tanto LMS quanto LTS foram definidos com o propósito de minimizar uma medida robusta da dispersão dos resíduos. Generalizando este fato, Rousseeuw e Yohai (1984), introduziram um estimador denominado de *S-estimatores*, que corresponde a:

$$\underset{\hat{\theta}}{Min}\ S\left(\theta\right) \tag{3.20}$$

Onde $S(\theta)$ é um certo tipo de M-estimador robusto de escala dos resíduos $r_1(\theta),...,r_n(\theta)$. Este estimador tem a mesma eficiência assintótica que um M-estimador, com um ponto de ruptura de 50% e é afim eqüivariante com uma taxa de convergência de $n^{-1/2}$. Sua definição é feita pela minimização da dispersão dos resíduos:

[16] Primeiro são obtidos os quadrados e depois ocorre a ordenação.

$$Min \;\; s\left(r_1(\theta),...,r_n(\theta)\right) \tag{3.21}$$

com uma estimativa de escala,

$$\hat{\sigma} = s\left(r_1\left(\hat{\theta}\right),...,r_n\left(\hat{\theta}\right)\right) \tag{3.22}$$

o valor da dispersão, $s\left(r_1\left(\hat{\theta}\right),...,r_n\left(\hat{\theta}\right)\right)$, da equação (3.22) é encontrado através da resolução da seguinte equação:

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i}{s}\right) = E_{\Phi}\left[\rho\right] \tag{3.23}$$

onde, $E_{\Phi}[\rho]$ é o valor esperado para a função ρ, e Φ representa a distribuição normal padrão. Esta função é do mesmo tipo definido para um M-estimador, ou seja, uma função de perda, e deve satisfazer as seguintes propriedades:

a. ρ deve ser simétrica, contínua e diferenciável e $\rho(0)=0$;

b. $c>0 \perp \rho$ é uma função estritamente crescente em [0,c] e constante em [c,∞];

c. $\frac{E_{\Phi}\left[\rho\right]}{\rho(c)} = \frac{1}{2}$, o que garante o ponto de ruptura de 50%, quando c=1,547 este valor de ruptura varia, para valores diferentes de 1/2.

Caso ocorra mais de uma solução para 3.23 então $s(r_1,...,r_n)$ deve ser feito igual ao supremo das soluções:

$$s\left(r_1,...,r_n\right) = \sup\left\{s;\left(\frac{1}{n}\right)\sum_{i=1}^{n}\rho\left(\frac{r_i}{s}\right) = E_{\Phi}\left[\rho\right]\right\} \tag{3.24}$$

e caso não haja nenhuma solução então $s(r_1,...,r_n)=0$.

O estimador recebeu este nome por ser derivado de uma escala estatística, implicitamente (s dado na equação (3.19) é um M-estimador de escala).

A seguir, um exemplo, de uma função de perda muito utilizada na prática, conhecida como função de perda Tukey biponderada: MENDES (1999)

$$\rho(x)=\begin{cases}\dfrac{x^2}{2}-\dfrac{x^4}{2c^2}+\dfrac{x^6}{6c^4} & \text{se} \quad |x|\le c \\ \dfrac{c^2}{6} & \text{se} \quad |x|>c\end{cases} \tag{3.25}$$

onde a constante reguladora c >0 é também responsável pela eficiência.

Esta função ρ dá origem à função ψ de Tukey. (3.26)

$$\psi(x)=\begin{cases}\left[x\left(1-\left(x/c\right)^2\right)^2\right] & |x|\le c \\ 0 & |x|>c\end{cases} \tag{3.26}$$

Figura 3.7 – Função ρ de Tukey

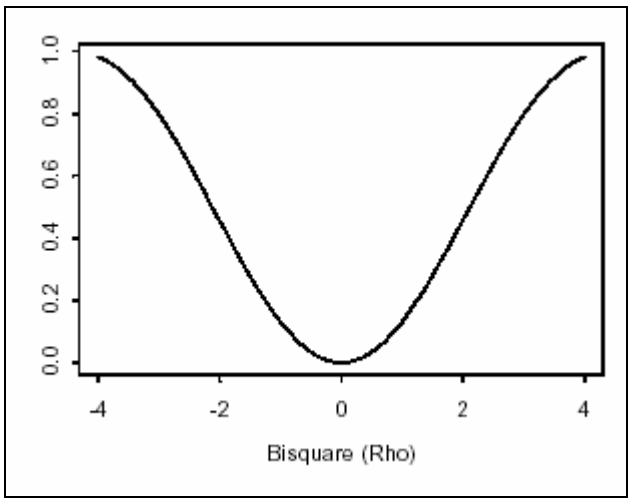
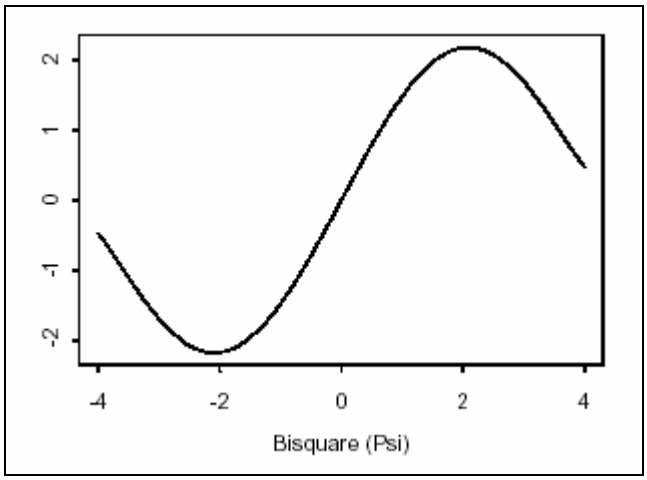


Figura 3.8 - Função ψ de Tukey



Para um valor de c=1,547 a função atende às propriedades a, b e c, mencionadas anteriormente e apresenta um ponto de ruptura de 50%. Este ponto de ruptura pode variar de acordo com o valor da constante c adotada. A seguir para a função ρ descrita acima, os valores da constante c e seus respectivos pontos de ruptura, eficiência assintótica e valor esperado de ρ, K=$E_{\phi}[\rho]$.

Tabela 3.1 – Pontos de Ruptura para S-Estimador (3.21)

Ponto de Ruptura	Eficiência Assintótica	Constante c	Valor Esperado K
50%	0,287	1,547	0,1995
45%	0,37	1,756	0,2312
40%	0,462	1,988	0,2634
35%	0,56	2,251	0,2957
30%	0,661	2,560	0,3278
25%	0,759	2,937	0,3593
20%	0,847	3,420	0,3899
15%	0,917	4,096	0,4194
10%	0,966	5,182	0,4475

3.3.7. MM – ESTIMADOR (Multiple - M Estimador)

Neste primeiro momento será apresentada a definição para o MM - estimador em sua estrutura independentemente de um modelo de regressão. Para que, posteriormente utilize-se a abordagem em estimativas deste modelo.

Segundo Rey, W. J. J. (1978), seja um espaço amostral qualquer, $\Omega \subset \Re^p$ e uma função de densidade de probabilidade f(x), $x \in \Omega$. Onde a distribuição de probabilidade é conhecida somente em um certo número de observações.

Sendo o vetor de parâmetros[17] $(\theta_1,...,\theta_n)$, que minimizam as $(M_1,...,M_g)$ funções, definidas da seguinte forma:

$$M_j = \min para\ \theta_j \text{ ; } j = 1,...,g \tag{3.27}$$

$$M_j = \int \rho_j(x;\theta_1,...,\theta_g)\, f(x)\, dx \tag{3.28}$$

Onde essa definição é uma generalização da teoria para os estimadores de máxima verossimilhança. Igualando (1) e (2) determina-se o EMV desejado.

No caso de desejar-se obter um estimador de escala, que esta em função de um estimador de posição ou locação. Como por exemplo, a variância $\hat{\sigma}$, que pode ser definida através da estrutura de um M-estimador com espaço amostral $\Omega \subset \Re$. Ou seja:

$$\int \left[(x-\mu)^2 - \sigma^2\right]^2 f(x)\, dx = \min \text{ para } \sigma^2 \tag{3.29}$$

Onde o parâmetro μ em (3.29) é a média e pode também ser estimada através da mesma estrutura anterior, de um M-estimador, da seguinte forma:

[17] O vetor de parâmetros nada tem haver com os coeficientes do modelo de regressão.

$$\int (x-\mu)^2 f(x)\,dx = \min \text{ para } \mu \tag{3.30}$$

Note que, as funções de perda $\rho(x)$, para as duas equações anteriores são dadas por:

$$\rho(x) = \left[(x-\mu)^2 - \sigma^2\right]^2 \tag{3.31}$$

$$\rho(x) = (x-\mu)^2 \tag{3.32}$$

De fato, ao se estimar μ, através de (3.30), o processo de estimação minimiza a variância, equação (3.33).

$$E\left[(X-\mu)^2\right] = V[X] = \int (x-\mu)^2 f(x)\,\mathrm{dx} \tag{3.33}$$

Para o procedimento em múltiplas etapas, são necessárias as duas hipóteses:

- Independência do espaço amostral Ω em relação aos parâmetros $\theta_1,...,\theta_2$.
- $$\begin{cases} \psi_j(x,\bullet) = \dfrac{\partial}{\partial\theta_j}\rho_j(x,\bullet) \\ \phi_{jk}(x,\bullet) = \dfrac{\partial}{\partial\theta_k}\psi_j(x,\bullet) \end{cases} \tag{3.34}$$

A minimização da equação (3.28), em função do parâmetro em questão, é resolvida através da solução de (3.35), o que justifica uma das hipóteses acima.

$$\int \psi_j(x,\theta_1,...,\theta_2)\, f(x)\, dx = 0 \tag{3.35}$$

Para o exemplo citado, em que se deseja obter estimativas para a média μ e a variância σ^2, as equações (3.29) e (3.30) passam a ser

representadas por uma estrutura semelhante à (3.35), onde ao utilizar uma função de distribuição de probabilidade empírica têm-se as equações:

$$\sum \omega_i \left(x_i - \mu \right) = 0 \tag{3.36}$$

$$\sum \omega_i \left[\left(x_i - \mu \right)^2 - \sigma^2 \right] = 0 \tag{3.37}$$

O parâmetro μ é obtido através de um M-estimador, enquanto o parâmetro σ^2 de um MM-estimador.

3.3.8.MM – ESTIMADOR NA ANÁLISE DE REGRESSÃO

Na seção anterior tratou-se de uma forma geral a questão das estimativas através de um MM - estimador. Para esta seção, a abordagem torna-se mais específica visando tratar a estimativa para o caso de um modelo linear de regressão, sendo assim o parâmetro θ_1 da seção anterior toma a forma do parâmetro β do modelo linear.

De acordo com Rey, W. J. J. (1978); não é possível obter um M-estimador robusto de regressão sem envolver, de alguma forma, um parâmetro de escala. Isso ocorre mesmo em um caso pontual, para uma estimativa de locação. Conseqüentemente, inclui-se um estimador de escala sempre que se soluciona um problema de regressão robusta. Marona (1976) foi quem definiu simultaneamente a estimação de escala e de locação, através de um MM-estimador.

Seja o vetor de resíduos $r_1,...,r_n$, produzidos em função da estimativa do parâmetro β. Podendo ser representado pela expressão a seguir:

$$r_i = y_i - \beta x_i \tag{3.38}$$

Através de certos critérios de minimização, produzem a estimativa de escala **s**. Essa estimativa é um parâmetro auxiliar para a estimativa final, desejada no modelo linear de regressão. Tendo então β como θ_1 e **s** como θ_2.

Mediante a equação (3.28) obtêm-se a equação (3.39).

$$\int \rho_2\left(x;\theta_1,\theta_2\right) f\left(x\right)\, dx = \min \text{ para } \theta_2 \tag{3.39}$$

Tendo em vista que θ_2 é definido como uma medida de escala para os resíduos:

$$\theta_2 = escala \text{ de } \left(r_1,...,r_n\right) \tag{3.40}$$

A função $\rho_2(.)$ deve ser escolhida de maneira que valha a propriedade da invariância de escala, ou seja, sendo (3.40) valha (3.41):

$$\left|\lambda\right|\theta_2 = escala \text{ de } \left(\lambda r_1,...,\lambda r_n\right),\ \ \lambda \in \mathbb{R} \tag{3.41}$$

O mesmo princípio de invariância é aplicado voltando ao MM - estimador para parâmetro θ_1, onde:

$$\int \rho_1\left(x;\theta_1,\theta_2\right) f\left(x\right)\, dx = \min \text{ para } \theta_1 \tag{3.42}$$

Sendo que neste momento $\rho_1(.)$ é determinada de tal modo a:

$$\theta_1 = estimativa \text{ de } regressão \text{ sobre } \left(x_1,...,x_n\right) \text{ e} \tag{3.43}$$

$$\theta_1 = estimativa \text{ de } regressão \text{ sobre } \left(\lambda x_1,...,\lambda x_n\right),\ \lambda \in \mathbb{R}_+ \tag{3.44}$$

E conseqüentemente, valha a propriedade da invariância de regressão.

As duas condições nas quais as funções $\rho_1(.)$ e $\rho_2(.)$ estão determinadas podem ser apresentadas de uma forma mais natural conforme a seguir:

$$\rho_1\left(x,\theta_1,\theta_2\right)=\rho_1\left[\frac{\left(y-\theta_1 x\right)}{\theta_2}\right]=\rho_1\left(\frac{r}{s}\right) \tag{3.45}$$

A forma na qual de fato os resultados para esta tese de mestrado foram obtidos e como o método tem sua aplicação pratica estão claramente descritas a seguir, na próxima subseção.

A descrição apresentada até aqui para o MM-estimador apenas formaliza seu conceito.

3.3.8.1. MM – ESTIMADOR – UMA ABORDAGEM APLICADA AO PROBLEMA DE REGRESSÃO ROBUSTA

O MM-Estimador de regressão robusta trabalha em três etapas, ou estágios, são elas: estimação do parâmetro β, minimização da estimativa de escala s(β) e estimação do parâmetro β[1]. O MM estimador será utilizado na obtenção das estimativas para o modelo de Análise Condicionada da Demanda.

De acordo com manual de estatística do software S-PLUS 2000, *s*eu algoritmo realiza o cálculo para o primeiro e o segundo estágio conjuntamente, visando obter os parâmetros β's iniciais e ao mesmo tempo minimizando a estimativa de escala s, de acordo como descrito para os S-estimadores, na seção 3.3.6. Na obtenção das estimativas dos parâmetros β's visa-se realizar a minimização da M-estimativa de escala[18] *s* em (3.46).

$$\frac{1}{n-p}\sum_{i=1}^{n}\rho\left(\frac{y_i-x_i^T\beta}{\hat{s}(\beta)}\right)=E_{\Phi}(\rho) \tag{3.46}$$

[18] M-Estimador de escala é também conhecido como S-Estimador.

A terceira e última etapa constitui-se na minimização da equação (3.47), para a obtenção da estimativa-M de regressão β^1.

$$\min \sum_{i=1}^{n} \rho\left(\frac{y_i - x_i^T \beta^1}{s}\right) \tag{3.47}$$

Em termos computacionais, o cálculo é realizado através da resolução da equação (3.48), ao invés de minimizar (3.47), onde a função ψ é a derivada da função[19] ρ, definida em (3.25).

$$\sum_{i=1}^{n} \psi\left(\frac{r_i\left(\beta^1\right)}{s}\right) x_i = 0 \tag{3.48}$$

Para que as estimativas dos β^1,encontradas através da equação (3.48), sejam utilizadas, é necessário que esta seja mais eficiente que as estimativas obtidas pelo S estimador (3.46), os estimadores encontrados no primeiro e segundo estágio, ou seja, os β's iniciais. Uma forma de escolher os estimadores é realizar a comparação a seguir:

$$S\left(\beta^1\right) \leq S\left(\beta\right) \tag{3.49}$$

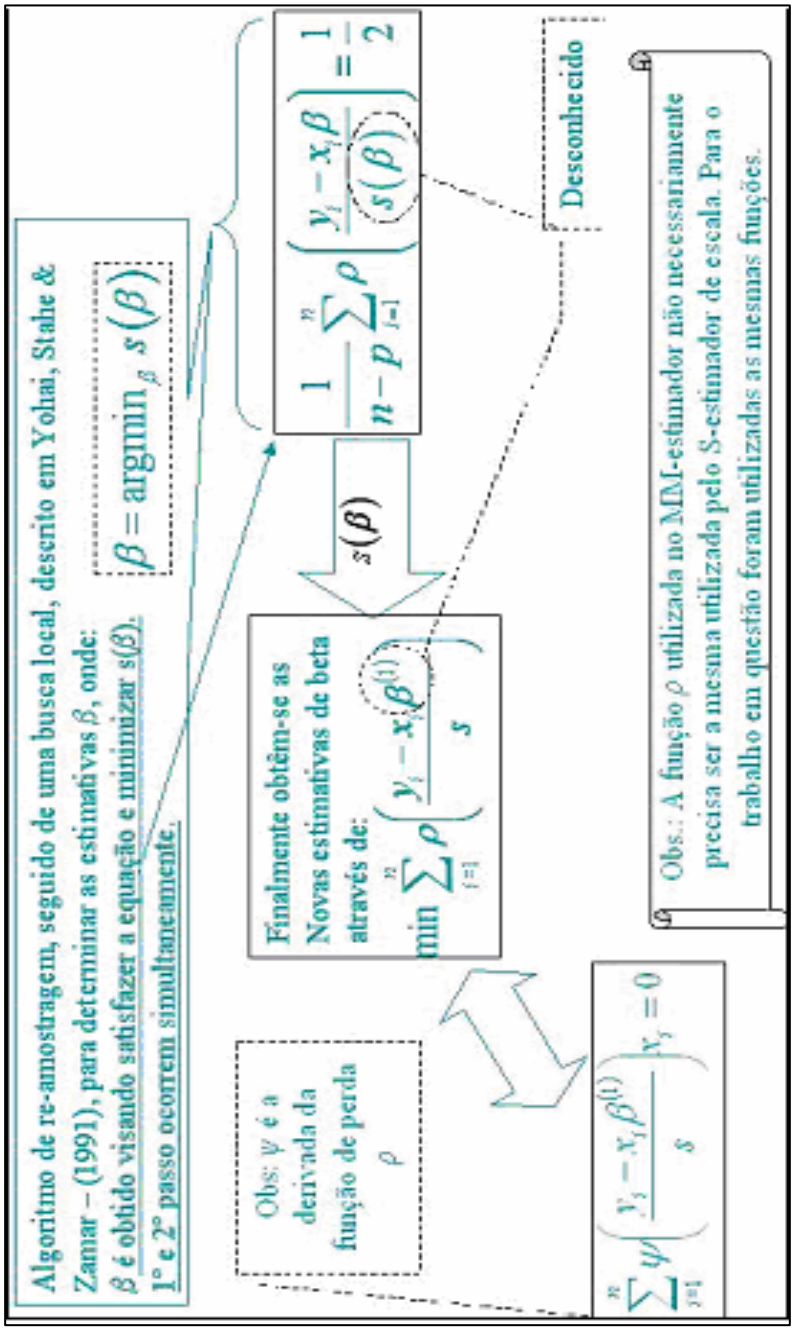
onde a função S é dada por,

$$S\left(\theta\right) = \sum_{i=1}^{n} \rho\left(\frac{r_i\left(\theta\right)}{s}\right) \tag{3.50}$$

O que é interessante observar é que ao aplicar β^1 na função acima, este é o valor que irá minimizá-la, isto de acordo com a equação (3.47). Caso β produza um resultado que seja menor, aplicado em (3.49), existe uma incoerência para com o terceiro estágio. Por isso opta-se pela utilização das estimativas produzidas entre as duas primeiras etapas.

[19] A função ρ não necessariamente precisa ser a (3.25), porém não é qualquer função. Quando ρ é uma função monótona o ponto de ruptura do estimador torna-se zero, é o caso da função quadrática, que dá origem ao estimador OLS – mínimos quadrados ordinários.

Gráfico 3.1 – Esquema do MM – estimador



4. ANÁLISE CONDICIONADA DA DEMANDA - ESTUDO DE CASO DO RECIFE

Será introduzido neste capítulo a aplicação do conceito de Análise Condicionada da Demanda, primeiro utilizado por PARTI et al. (1980), para obter estimativas dos consumos individuais de cada equipamento eletroeletrônico, a partir do consumo total domiciliar. Utilizando, como base os dados do levantamento da Pesquisa de Posse de Eletrodomésticos e Preferência de Consumo[20], realizada no município de Recife.

A referida base de dados partiram de uma amostra aleatória, extraída do cadastro de consumidores residenciais da CELPE – Companhia de Eletricidade de Pernambuco. O questionário apresentado no apêndice I foi elaborado com base em estudos anteriores de "Posse e Hábitos de Consumo de Energia" (SILVA et al, 2002 e 2004a), onde foi possível incluir os equipamentos mais relevantes, ou seja, aqueles que apresentaram percentual do consumo mais significativos. Foram investigados um total de 33 equipamentos, além das características de iluminação: lâmpadas fluorescentes e incandescentes.

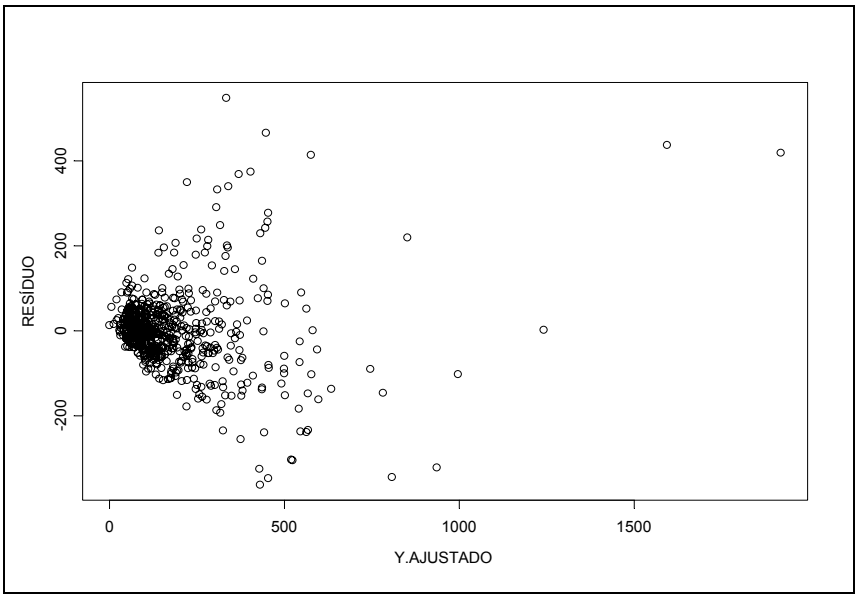
A concessionária de energia também cedeu ao Projeto o consumo mensal dos domicílios para um período de 12 meses, datando de março de 2002 a fevereiro de 2003. Nesta dissertação trabalhamos com o consumo médio. Sendo essa a variável dependente do modelo, ou seja, aquela que desejamos estimar com base no número de equipamentos existentes no domicílio, essas variáveis explicativas são potenciais variáveis independentes do modelo.

O tipo de dados coletados é conhecido como de corte do tipo *cross-section*, que apresentam como vantagem a ausência de autocorrelação entre as perturbações, ou seja, os erros são não correlacionados. Segundo GUJARATI (2000) esse tipo de dados tem como principal problema à heterogeneidade dos erros, isso ocorre porque, segundo SILVA et al (2004b),

[20] O plano amostral e o questionário são partes integrantes do projeto desenvolvido no âmbito do Projeto "Modelagem de Apoio à Decisão na Seleção de Instrumentos de Racionalização de Energia no Setor Residencial" . Processo : 5514162001-7.

um domicílio com maior número de equipamentos de um determinado tipo apresentam uma maior variância do consumo, do que domicílios com apenas um equipamento (ou nenhum) deste tipo. Do mesmo modo que famílias com renda mais alta tendem a apresentar maior variância do consumo do que famílias de baixa renda, como podemos verificar no gráfico de valores ajustados versus resíduos, figura 4.1. Segundo KMENTA (1990) quando existe heterocedasticidade os estimadores Mínimos Quadrados Ordinários (OLS) ainda são não tendenciosos e consistentes, mas não tem variância mínima, e portanto não são eficientes e nem assintoticamente eficientes.

Figura 4.1 - Valores Ajustados Vs. Resíduos



Quando os pressupostos básicos são violados, uma alternativa para superar esse problema é utilizar o método de regressão linear robusto, pois estes estimadores, em sua grande maioria, apresentam ponto de ruptura de 50%, que suporta a influência dos *outliers*, para estimar os coeficientes de regressão.

4.1. SELEÇÃO DE VARIÁVEIS.

Mesmo tendo uma proposta robusta para a questão de estimação dos parâmetros, é necessário reduzir o número de variáveis utilizadas no modelo, pois nem todas conseguem apresentar melhora significativa e quanto maior for o número de variáveis menor é o número de graus de liberdade apresentado no modelo.

Para resolver este problema utiliza-se a metodologia conhecida como *stepwise*, indicada para os problemas de modelos de regressão. Segundo Chaterjee *et al.* (1977) o Método *Stepwise* é utilizado no caso de existir uma grande quantia *j* de variáveis explicativas e mostra-se computacionalmente eficaz, por não considerar todas as combinações possíveis de variáveis.

Esta metodologia tem como fundamento selecionar as variáveis independentes que melhor expliquem a variável resposta, através do teste da estatística F e da inclusão e/ou exclusão de variáveis. Existem três tipos diferentes de abordagens *stepwise*, são elas: *Forward Selection, Backward Elimination e Stepwise Regression.* MONTGOMERY, D., C. *Et al.* (1982)

A metodologia *Forward Selection* inicia-se com apenas uma variável e se sucede com a inclusão dos regressores de maior correlação com Y, após sua inclusão é calculado a estatística F, conforme mostrado no capítulo 2. Caso a estatística recém calculada seja maior que a estatística pré-estabelecida[21] F_{IN} a variável permanece no modelo. Em um segundo momento o regressor que apresentar maior correlação parcial com os valores ajustados de Y, dado a primeira variável selecionada, é utilizado para a verificação do teste F. Esta correlação parcial nada mais é do que a correlação simples entre os resíduos provenientes dos modelos a seguir.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 \tag{4.1}$$

$$\hat{x}_j = \hat{\alpha}_{0j} + \hat{\alpha}_{1j} x_1, \quad j = 2,3,...,K \tag{4.2}$$

[21] Esta variável também conhecida como $F_{ENTRADA.}$

Note que este segundo modelo (4.2) nada mais é do que a regressão entre a primeira[22] variável selecionada versus as outras possíveis variáveis. Selecionando a próxima variável com maior correlação, mediante esta metodologia obtêm-se a estatística F através de*:*

$$F = \frac{RSS\left(x_2 \mid x_1\right)}{\frac{RSS\left(x_1, x_2\right)}{n-p}}, \quad \text{p=número de parâmetros} \qquad (4.3)$$

Outra vez, caso a estatística obtida seja maior que F_{IN}, esta é agregada ao modelo. Generalizando, em cada passo realizado o regressor deve ter a maior correlação parcial com a variável Y (consumo médio de energia), equivalentemente a maior estatística F parcial, dados outros regressores já presentes no modelo. Assim a variável em questão é inserida, caso sua estatística parcial F seja maior que o valor pré-determinado F_{IN}. O procedimento é dado como encerrado quando a estatística parcial F, para um passo em particular não exceda F_{IN}.

Backward Elimination: o procedimento é inverso, inicia-se com todas as *p* variáveis e estas são retiradas passo a passo. A estatística parcial F é calculada, conforme o *Forward Selection*, para cada regressor como se esta fosse a última variável a entrar no modelo. O processo de decisão é feito comparando a menor destas estatísticas com o valor pré-determinado, agora denominado F_{OUT}. Caso a estatística seja menor que F_{OUT} o regressor é retirado do modelo. Passa-se a ter um novo modelo, porém agora trabalhando com *p-1* variáveis regressoras e outra vez um modelo é ajustado, a estatística F-parcial é calculada e o processo é repetido. O critério de parada para este modelo é dado quando a menor das estatísticas de F-parcial não é menor do que F_{OUT} pré-determinado.

O método *Stepwise Regression* a última metodologia da técnica *stepwise*, apresenta um comportamento que mescla os dois processos anteriores, também é conhecida como *Stepwise Efroymson* (1960). Este método é a modificação do *Stepwise Forward Selection, onde* uma variável regressora inserida em um passo anterior, pode-se tornar redundante devido a sua relação com as demais variáveis já presentes no modelo, sendo assim,

[22] Para facilitar assume-se que a primeira variável selecionada foi o x_1 e a segunda o x_2.

torna-se propensa a ser removida. Para este método são necessários dois valores[23] pré-determinados, F_{IN} e F_{OUT}. Devido a sua utilização através do *software* **S-PLUS**, esses valores já são pré-estabelecidos em rotina interna.

Em função de sua eficácia, o método utilizado por este trabalho foi o *Stepwise Regression.* A estatística decisiva para incluir, ou não, uma variável regressora é a F. Esta estatística é obtida por métodos via minimização da soma de quadrados, como visto no capítulo 2. Os estimadores deste tipo apresentam ponto de ruptura zero, ou seja, suscetíveis a *outliers*, isso podendo levar a inclusão, ou exclusão de variáveis de importância para o estudo.

A solução utilizada foi inicialmente realizar estimativas para os parâmetros, através de um estimador robusto, neste caso utilizamos o estimador *Least Trimed Square* - LTS – visto no Capítulo 3. Este tipo de estimador trabalha de forma que os resíduos são minimizados, porém somente os h primeiros são utilizados, visando garantir um ponto de ruptura de 50%, utilizamos $h \cong \frac{n}{2}$ na fórmula abaixo (4.4) .

$$\underset{\hat{\theta}}{Mínimo} \sum_{i=1}^{h} \left(r^2 \right)_{i:n} \qquad (4.4)$$

Em função do procedimento citado acima foi possível identificar os *outliers*, pois este algoritmo atribui pesos zero ou um as observações, os quais foram utilizados para realizar um tratamento prévio a base de dados e então utilizar a metodologia *Stepwise Regression.* Foram detectados 65 *outliers*, ou seja, pesos zero, que foram removidos da amostra, restando 535 observações a serem utilizadas. A seleção de variáveis via *Stepwise*[24] foi realizada, implicando em uma redução de 12 variáveis. Segundo o critério adotado, 23 variáveis apresentaram relação significativa com o consumo médio de energia e estão relacionadas na tabela 4.1..

[23] Certas análises assumem valores iguais para F_{IN} e F_{OUT}. Isto não é uma regra, comumente usa-se $F_{IN} > F_{OUT.}$ Tornando a inclusão de uma variável mais difícil que a exclusão.

[24] O método *Stepwise* Regression ou *Stepwise* Efroymson será tratada apenas como *Stepwise.*

Tabela 4.1 – Resultado do método *Stepwise* para 535 domicílios

Variáveis	Coeficientes	Erro Padrão	Estatística-t	Pr(>|t|)
Intercepto	23,34	8,81	2,65	0,0083
Lampâdas Incandescentes	4,29	0,52	8,22	0,0000
Lampâdas Fluorescentes	3,77	0,59	6,42	0,0000
Geladeira 1 porta	17,32	6,70	2,59	0,0100
Geladeira 2 portas	49,12	7,96	6,17	0,0000
Freezer	39,34	5,53	7,11	0,0000
ChuveiroXpessoas	6,70	1,73	3,87	0,0001
Ar	22,85	4,42	5,17	0,0000
TV	17,66	3,52	5,01	0,0000
Vídeo	12,43	4,14	3,01	0,0028
Microcomputador	15,61	5,81	2,69	0,0075
Game	34,75	7,20	4,83	0,0000
Secadora	305,46	26,14	11,69	0,0000
Microondas	-10,51	6,46	-1,63	0,1044
Liquidificador	-16,29	7,34	-2,22	0,0269
Cafeteira	-11,20	7,49	-1,49	0,1356
Exaustor	-15,92	8,26	-1,93	0,0546
Ventilador	3,97	2,00	1,99	0,0472
Bomba	19,05	6,14	3,10	0,0020
Gelo	15,21	6,69	2,27	0,0234
DVD	24,34	8,53	2,85	0,0045
Fax	29,81	14,41	2,07	0,0391
Gril	15,50	5,97	2,60	0,0097
Churrasqueira	63,71	34,66	1,84	0,0667

Em uma segunda etapa, inicia-se o processo de estimação dos parâmetros utilizando o modelo de regressão linear robusto. Existem atualmente vários estimadores robustos que podem ser utilizados em um modelo de regressão linear. Porém o problema que ainda dificulta a sua utilização é o cálculo da maioria deles. Certos *softwares* estatísticos trazem um módulo sobre Estimadores Robustos. O S-Plus é um deles e será a ferramenta computacional utilizada nessa dissertação.

Em função das propostas apresentadas no capítulo 3, suas vantagens e desvantagens, juntamente com o fator custo computacional, optou-se pelo Estimador-MM, no decorrer deste trabalho será comparado ao método OLS, para que seja então verificado sua precisão, através do desvio padrão de suas estimativas de consumo de equipamentos.

4.1.1. *OUTLIERS* NA SELEÇÃO DE VARIÁVEIS

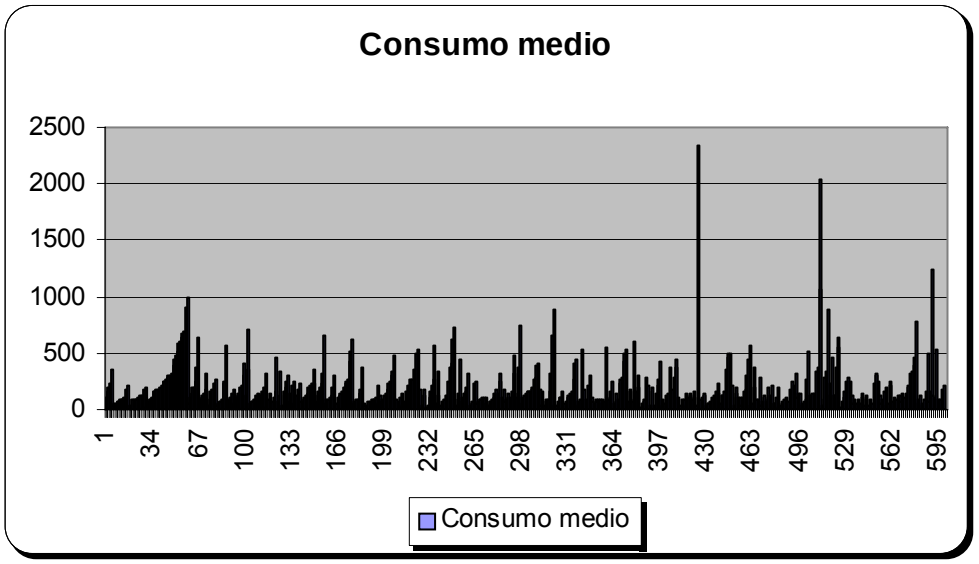
Os *outliers* são retirados por basicamente dois motivos. Quando a média do consumo de energia é demasiadamente alta, ou quando é demasiadamente baixa. Para qualquer um dos casos o número de equipamentos apresentados pelo domicílio pesquisado, deve ser condizente com o consumo médio, caso contrário pode estar ocorrendo as seguintes situações:

- Um consumo médio muito baixo e uma grande quantidade de equipamentos são um indício de fraude do domicílio pesquisado para com a companhia de energia elétrica ou caracterizar uma família com mais de um domicílio.
- Um alto consumo médio por parte do domicílio pesquisado e uma pequena quantidade de equipamentos são um indício de erro na cobrança da conta.

Como já mencionado anteriormente, ocorreram 65 observações caracterizadas como *outliers*, que receberam pesos nulos, para não influenciarem na análise. O método utilizado na detecção foi o estimador LTS – capítulo 4, através de uma regressão robusta. Outra forma para detecção de *outliers* pode ser realizada através de análises gráficas. Isto, porém, não é um método muito eficiente, mesmo assim será utilizado visando corroborar a análise feita com LTS. Os resultados a seguir para esta subseção foram obtidos considerando toda a base de dados, ou seja, as 600 observações.

A figura 4.2 apresenta um gráfico de consumo médio de energia versus o número da observação pesquisada, o que mostra picos de consumo médio da ordem de 2000 kWh. Um exemplo é caso de um domicílio pesquisado que apresentou consumo médio de 2031 kWh, ao analisa-lo individualmente, verificou-se que o mesmo apresentava uma alta quantidade de lâmpadas, 107, uma geladeira de duas portas com capacidade de 410 litros e 7 aparelhos de TV, entre outros equipamentos. Mesmo que a grande quantidade de aparelhos apresentados validem o consumo, é importante destacar, que este domicílio influi com uma grande variabilidade para a estimativa do consumo médio por equipamento.

Figura 4.2 – Consumo Médio de Energia em kWh



Note que, é possível estabelecer um critério com base no 1º e 3º quartil, de acordo com a tabela 4.2.

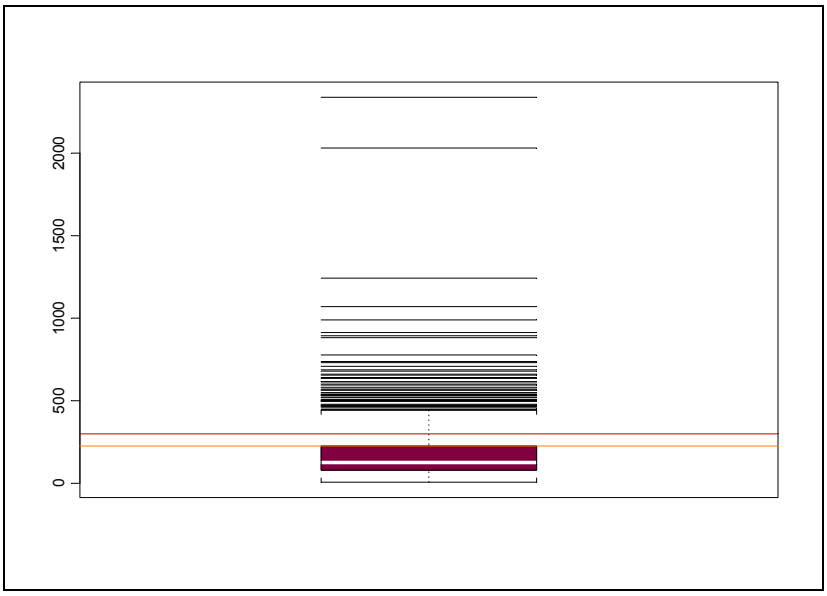
Tabela 4.2 – Estatísticas do Consumo Médio de Energia

600 Observações

Min.	1st	Median	Mean	3rd	Max.
7	79,75	126,5	189,1	226	2338

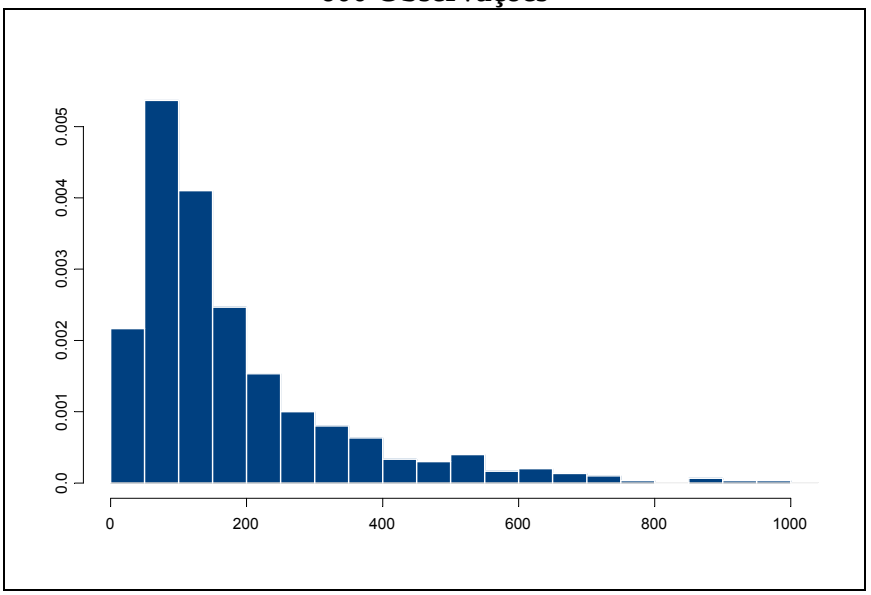
A figura 4.3 apresenta um gráfico do tipo Box Plot, visto que, este apresenta uma forma de analisar a distribuição da variável - consumo médio de energia - graficamente através dos seus quartis. O problema deste tipo de análise é que não leva em consideração o número de equipamentos do domicílio pesquisado. O que pode levar a conclusões erradas. De acordo com DeGROOT (1986), pode-se classificar esta situação como erro tipo 1, ou seja, descartar uma observação não sendo esta um *outlier.* Ou erro tipo 2, aceitar uma observação por estar entre o primeiro e terceiro quartil, quando esta de fato é um *outlier*. Um exemplo que se enquadra na situação do tipo 2, é o domicílio que apresenta um consumo médio de 107 kWh, onde o individuo está dentro dos limites estabelecidos mas de fato é um *outlier*. – Um possível caso de fraude contra a companhia ou residência de verâneio. Estes domicílios foram identificados através do estimador robusto LTS e retirados da análise final.

Figura 4.3 – Box Plot Consumo Médio de Energia em kWh



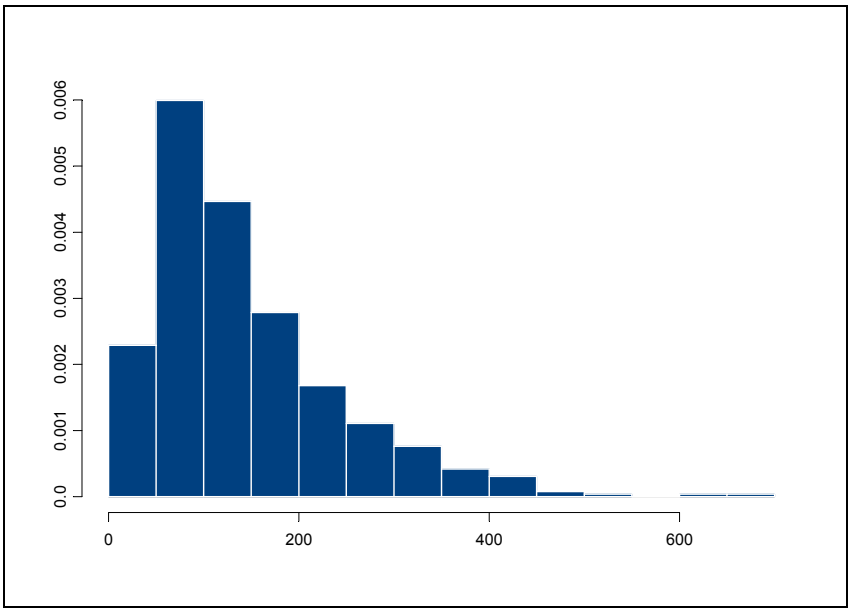
Na seção 4.1, foi mencionado a utilização do estimador LTS como ferramenta na detecção de *outliers*, e a partir disso passou-se a trabalhar com apenas 535 observações ao invés das 600.

Figura 4.4 – Histograma do Consumo Médio de Energia
600 Observações



A figura 4.4 apresenta a distribuição do consumo médio de energia para as 600 observações. É clara a assimetria positiva do consumo, apresentando a mediana de 126 kWh e a média de 189 kWh por mês.

Figura 4.5 – Histograma do Consumo Médio de Energia
535 Observações



Através da retirada das observações identificadas como *outliers*, a distribuição do consumo continua assimétrica, porém a variabilidade diminui sensivelmente. Isso é visto claramente no histograma da figura 4.5 e na tabela 4.3, a seguir, através de seus valores de máximo e de mínimo e da redução do número de classes do histograma.

Tabela 4.3 – Estatísticas do Consumo Médio de Energia
535 Observações

Min.	1st	Median	Mean	3rd	Max.
7	75	116	144,6	189,5	656

4.2. ESTIMATIVAS

Conforme o procedimento apresentado no capítulo 3, subseção 3.3.8.1, referente ao MM Estimador, ao realizar o procedimento para a escolha entre os β's iniciais estimados entre os estágios um e dois, conhecido como estimador S, e os β^1's resultantes do terceiro estágio, o teste indicou que

para a primeira estimativa, utilizando as 535 observações os β's utilizados foram o do terceiro estágio, ou seja, os β^{1}'s. Já para a última estimativa do modelo CDA, os estimadores finais apresentam vício em relação aos obtidos no processo intermediário. Sendo assim as estimativas utilizadas foram as obtidas via o S-estimador.

Tabela 4.4 – Primeira estimativa via MM-estimador com 535 observações

Variáveis	Coeficientes	Erro Padrao	Estatística t	Pr(>|t|)
Intercepto	23,63	10,07	2,35	0,0193
Lâmpadas Incandescentes	4,28	0,59	7,29	0,0000
Lâmpadas Fluorescentes	3,43	0,70	4,89	0,0000
Geladeira 1 porta	17,63	7,75	2,28	0,0233
Geladeira 2 portas	51,23	9,05	5,66	0,0000
Freezer	41,52	7,27	5,71	0,0000
ChuveiroXpessoas	7,20	2,21	3,25	0,0012
Ar	23,46	5,22	4,49	0,0000
TV	16,68	4,01	4,16	0,0000
Video	12,23	4,62	2,64	0,0084
Microcomputador	15,69	6,61	2,38	0,0179
Game	32,29	8,16	3,95	0,0001
Secadora	302,29	30,94	9,77	0,0000
Microondas	-10,33	7,37	-1,40	0,1615
Liquidificador	-16,24	8,12	-2,00	0,0462
Cafeteira	-7,89	8,46	-0,93	0,3514
Exaustor	-15,49	9,74	-1,59	0,1123
Ventilador	4,39	2,28	1,92	0,0549
Bomba	17,27	6,92	2,50	0,0129
Gelo	16,04	7,63	2,10	0,0361
DVD	25,75	9,77	2,64	0,0087
Fax	27,87	17,03	1,64	0,1024
Gril	15,50	5,97	2,60	0,0238
Churrasqueira	63,71	34,66	1,84	0,0951

É importante destacar alguns aspectos desta primeira estimativa com o estimador robusto (tabela 4.4). A primeira delas é que os desvios padrão das estimativas utilizando os estimadores MM são menores do que as primeiras estimativas obtidas via *Stepwise*, ou seja, utilizando os estimadores de Mínimos Quadrados Ordinários, sendo de 45,5 para o primeiro e 47,38 para os estimadores clássicos, ambos com 511 graus de liberdades. O fato mais importante a se destacar são os coeficientes com valores negativos, que tanto para um método como outro foram produzidos. De acordo com a estatística *t-student*, o teste realizado para identificar a significância dos

coeficientes, mostra claramente que para um nível de significância de 5%, essas variáveis, que apresentaram seus coeficientes com valores negativos devem ser retiradas do modelo.

A estatística *t-student* corroborando com a retirada das variáveis de coeficiente negativo mostra que o modelo proposto está correto, visto que, segundo o modelo de Análise Condicionada Demanda de energia não podem ocorrer estimativas negativas, pois estas são referentes ao consumo médio por equipamentos, o que seria uma incoerência apresentar consumos negativos. Apresentamos a seguir a estimativa final para o modelo, lembrando que foram utilizados os β's iniciais, isto é, as estimativas do S Estimador (tabela 4.5)

Tabela 4.5 – Última estimativa via MM-estimador com 535 observações

Variáveis	Coeficientes	Erro Padrao	Estatística t	Pr(>|t|)
Intercepto	11,18	8,02	1,39	0,1637
Lâmpadas Incandescentes	2,59	0,60	4,33	0,0000
Lâmpadas Fluorescentes	0,58	0,71	0,82	0,0411
Geladeira 1 porta	33,75	8,17	4,13	0,0000
Geladeira 2 portas	81,73	9,94	8,22	0,0000
Freezer	59,71	6,95	8,59	0,0000
ChuveiroXPessoas	16,22	2,22	7,30	0,0000
Tv	17,95	4,25	4,23	0,0000
Ar	36,23	5,03	7,20	0,0000
Vídeo	5,80	4,85	1,20	0,0223
Game	18,17	9,34	1,94	0,0524
Secadoura	229,57	28,19	8,15	0,0000
Bomba	6,81	7,18	0,95	0,0429
Gelo	43,73	8,11	5,39	0,0000

Ocorreu uma redução de 35 variáveis iniciais, para 13 variáveis finais, que estão descritas a seguir, na tabela 4.6

Tabela 4.6 – Descrição das Variáveis do Modelo

Variáveis	Descrição
Consumo Médio – Variável dependente	Consumo médio de eletricidade por domicílio (KWh/mês)
Lâmpadas Incandescentes	Número de lâmpadas incandescentes no domicílio
Lâmpadas Fluorescentes	Número de lâmpadas fluorescentes no domicílio
Geladeira 1 porta	Número de refrigeradores 1 porta
Geladeira 2 portas	Número de refrigeradores 2 portas
Freezer	Número de freezers
ChuveiroXPessoas	Chuveiro elétrico X número de moradores que utilizam chuveiro elétrico
TV	Número de televisores
Ar	Número de ar condicionado
Vídeo	Número de videocassete
Game	Número de videogame
Secadora	Número de secadora de roupa
Bomba	Número de bombas d'agua
Gelo	Número de gelo água

4.3. COMPARAÇÃO MM E OLS

Com o intuito de assegurar a coerência dos resultados do Estimador MM, foi adotado um procedimento para realizar a devida comparação entre os dois métodos em questão, Robusto e Clássico. Como dito anteriormente as 65 observações mais discrepantes deste conjunto de dados foram retiradas com intuito de utilizar a metodologia *Stepwise* (seleção de variáveis). Vamos utilizar agora os dois conjuntos de dados para comparar os processos de estimação citados.

Tabela 4.7 – Comparação MM Estimadores X Mínimos Quadrados Ordinários

Variáveis	Resultados com 535 observações		Resultados com 600 observações	
	MM estimadores	OLS	MM estimadores	OLS
Intercepto	11,18	12,77	13,32	2,15
Lâmpadas Incandescentes	2,59	4,00	2,63	9,65
Lâmpadas Fluorescentes	0,58	3,99	0,78	5,82
Geladeira 1 porta	33,75	18,13	31,65	3,51
Geladeira 2 portas	81,73	50,85	79,05	41,43
Freezer	59,71	41,63	59,66	53,80
ChuveiroXPessoas	16,22	8,82	15,17	5,86
TV	17,95	27,32	17,48	20,36
AR	36,23	19,19	37,26	51,78
Vídeo	5,80	12,75	8,17	5,83
Game	18,17	36,76	20,38	29,82
Secadora	229,57	280,95	229,72	230,57
Bomba	6,81	18,17	5,51	18,53
Gelo	43,73	20,24	40,12	15,83

A tabela 4.7 apresenta praticamente os mesmos resultados para os valores dos coeficientes (β) para o MM – Estimador, com as duas bases, corroborando a robustez do estimador, que tem ponto de ruptura de 50%. O mesmo não se pode verificar para as estimativas que foram obtidas via OLS, onde encontramos estimativas muito diferentes para os dois conjuntos de dados, com e sem *outliers*.

Quando utilizamos as 600 observações além de que todos os desvios padrões via estimador robusto serem menores que os apresentados utilizando os OLS. Outras medidas também favorecem a metodologia robusta e estão expostas na tabela 4.8

Tabela 4.8 – Medidas de Variação com 600 observações

	Erro Padrão	Proporção de Explicação do modelo
Ajuste OLS	103,3	0,74
Ajuste Robusto	52,19	0,61
Graus de Liberdades	586	586

A estimativa de R^2, apresentada no capítulo 2, para o OLS, são análogas à proporção de explicação do modelo, para o caso robusto. Segundo a tabela 4.8, pode-se concluir que OLS possui um maior poder de explicação que MM, mas como já mencionados, este valor está superestimado. De

acordo com SILVA (2000) foi diagnósticado que em estudos CDA (LAFRANCE (1994) e EPRI (1989)), os valores de R^2, por se tratar de dados *cross section,* normalmente variam entre 0,55 e 0,70. O resultado obtido pelo estimador robusto está dentro dessa faixa de aceitação. Agora tomando como prioridade a redução do desvio padrão[25] é notório que o modelo robusto apresente uma medida de menos que a metade da obtida para o OLS.

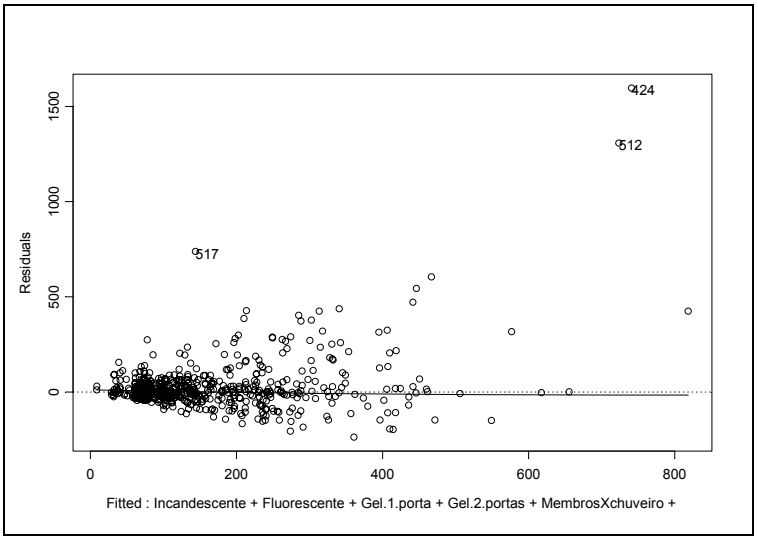
Quando comparamos os resultados das estimativas com 535 observações, obtemos uma melhora significativa nos estimadores clássicos, onde o desvio padrão do estimador de Mínimos Quadrados Ordinários aproxima-se do valor encontrado pelo estimador robusto, sendo de 43,3 e 49,1, respectivamente. Com relação ao R^2 o estimador clássico mantém o melhor índice de explicação do modelo, conforme mostrado na tabela 4.9.

Tabela 4.9 – Medidas de Variação com 535 observações

	Erro Padrão	Proporção de Explicação do modelo
Ajuste OLS	49,1	0,81
Ajuste Robusto	43,3	0,65
Graus de Liberdades	521	521

A seguir serão introduzidos os diagnósticos gráficos que auxiliaram na escolha dos modelos.

Figura 4.6– Valores Estimados Robustos Vs. Resíduos



[25] Residual Scale

Veja como na figura 4.6 os *outliers*, como por exemplo, à observação 424, não detêm influência sobre os valores ajustados. O mesmo pode ser visto na figura 4.7. É importante ressaltar que estas estimativas robustas não apresentam heterocedasticidade, vide figura 4.6., o mesmo não pode ser verificado pelas estimativas via OLS, constatado segundo a figura 4.8. onde claramente a dispersão propaga-se com o aumento do consumo. Um fator a ser destacado para as estimativas por OLS é sua influência mediante *outliers*, figura 4.9.

Figura 4.7 – Valores Ajustados Vs. Valores Observados

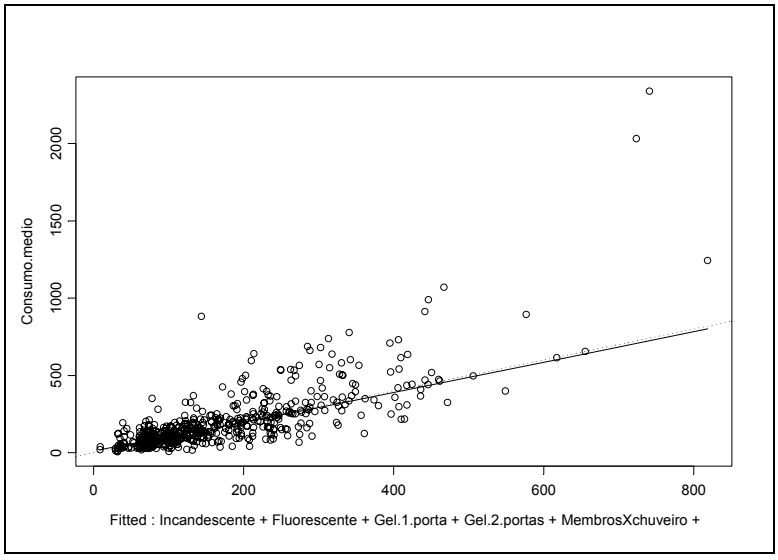


Figura 4.8- Resíduos V.s.Valores Ajustados

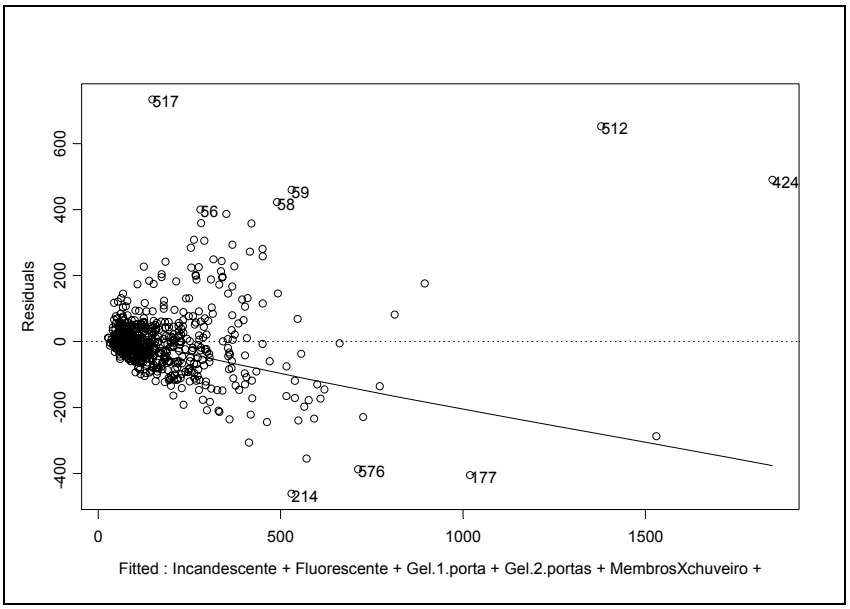
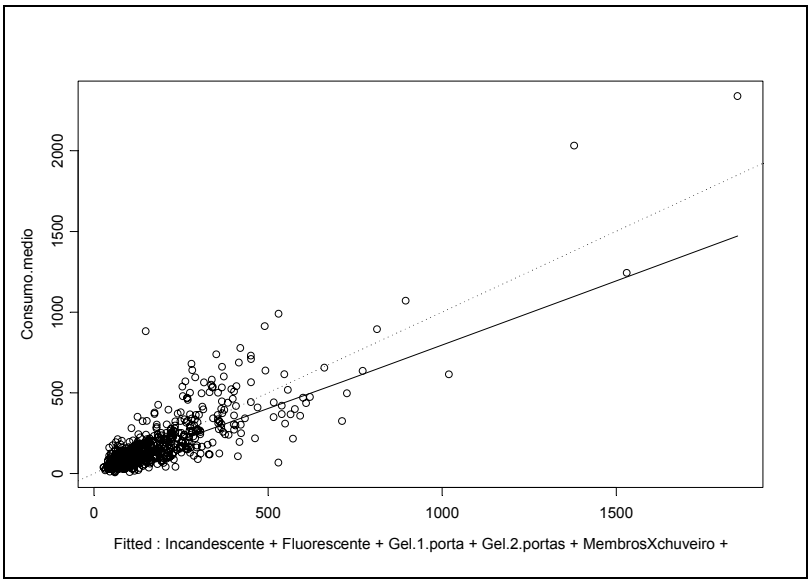


Figura 4.9 – Valores Ajustados V.s. Observados



5. CONCLUSÃO

O objetivo deste trabalho foi o de incorporar a metodologia de Análise Condicionada da Demanda um novo procedimento metodológico para obter os estimadores de consumo dos equipamentos eletrodomésticos. Para efeito de validação foi feita a comparação entre os resultados obtidos com o modelo de regressão linear clássico (OLS) e o modelo robusto (MM – Estimadores).

Os dados coletados são do tipo *cross-section*, que apresentam como vantagem à ausência de autocorrelação entre as perturbações. Porém, esse tipo de dados tem como principal problema apresentar comportamento heterocedástico, ou seja, variâncias do termo do erro do modelo que não são constantes para todas as observações. Segundo KMENTA (1988) na presença de heterocedasticidade o estimador de Mínimos Quadrados Ordinários - OLS dos parâmetros de regressão lineares permanecem não-tendencioso (isto é, em média é igual ao valor do parâmetro da população a ser estimado) e consistente (ou seja, converge em probabilidade para o parâmetro verdadeiro à medida que o número de observações $n \to \infty$). Contudo, os estimadores de Mínimos Quadrados Ordinários dos coeficientes de regressão não são melhores estimadores lineares não-tendenciosos (MELNT), quando não vale o pressuposto de homocedasticidade. Isso significa que os estimadores OLS não têm a menor variância numa classe de estimadores não-tendenciosos e que, portanto não são eficientes ou assintoticamente eficientes. Logo se procedermos com nossa análise de regressão, aceitando que o termo do erro do modelo apresentem variâncias constantes, quando não as tem, nossas inferências sobre os coeficientes de regressão serão incorretas.

Sendo assim, uma alternativa para o estudo da regressão do modelo linear de regressão, proposto nesse trabalho, foram a utilização dos métodos robustos, que são resistentes a influência dos *outliers*. Existem alguns estimadores robustos que substituem a soma dos resíduos ao quadrados por outro critério que minimize com menos influência os *outliers*, como os estimadores L_1,os M-estimadores, o estimador LMS, o estimador LTS, os S-estimadores e os MM-estimadores. O estimador robusto utilizado nessa dissertação foi o MM-Estimador, por ser indicado para dados heterocedásticos,

contando com as três propriedades de eqüivariância e ter ponto de ruptura de 50%.

A utilização de Modelos Robustos a Análise Condicionada da Demanda desagregando o consumo domicíliar de energia entre os equipamentos eletroeletrônicos se mostrou como uma alternativa satisfatória aos métodos tradicionais de estimação, como podemos constatar com os resultados obtidos na comparação dos desvios padrão obtidos pelos dois métodos e do valor do R^2, o percentual da variância explicada pelo modelo. É possível verificar, na tabela 5.1, que os erros padrões do Modelo Robustos para os parâmetros estimados, são significativamente menores. Mesmo tendo apresentado uma proporção de explicação menor que a obtida com a estimativa clássica. Note que o valor obtido de 74,2%, aparentemente melhor para a variação é comprometido, visto que o estimador de MQO é tendencioso, pois os dados apresentam heterocedasticidade e possuem uma grande quantidade de *outliers* em X e em Y.

Tabela 5.1 – Medidas de Variação com 600 observações

	Erro Padrão	Proporção de Explicação do modelo
Ajuste OLS	103,3	0,74
Ajuste Robusto	52,19	0,61
Graus de Liberdades	586	586

Utilizamos o modelo aditivo devido à facilidade de sua interpretação física, que possibilita simular o consumo final de cada equipamento. Desta forma basta utilizarmos os valores dos coeficientes de regressão de cada equipamento encontrado no modelo e multiplicarmos pelo total de equipamentos que obteremos o consumo residencial total.

O procedimento utilizado para seleção de variáveis baseou-se no método *Stepwise,* sendo posteriormente excluídas as variáveis cujo p-valor foi superior a 5,0%. Os resultados do modelo robusto dos parâmetros de regressão β_j ,para o consumo dos equipamentos mensais estão apresentados na tabela 5.2, onde o intercepto representa os equipamentos restantes não incluídos na análise.

Este resultado colabora para escolhermos o estimador Robusto (MM Estimador), que apresentam valores bastante parecidos dos β_j , tanto para a base com os dados contaminados (600 observações, sem a retirada dos *outlier*) como para a base com os dados limpos (535 observações).

Tabela 5.2 – MM estimador com 535 e 600 observações

Variáveis	β_j com 535 observações	Estatística t	β_j com 600 observações	Estatística t
Intercepto	11,18	1,39	13,32	1,72
Lâmpadas Incandescentes	2,59	4,33	2,63	4,50
Lâmpadas Fluorescentes	0,58	0,82	0,78	1,16
Geladeira 1 porta	33,75	4,13	31,65	4,06
Geladeira 2 portas	81,73	8,22	79,05	8,34
Freezer	59,71	8,59	59,66	8,98
ChuveiroXPessoas	16,22	7,30	15,17	7,28
TV	36,23	7,20	17,48	7,69
Ar	17,95	4,23	37,26	4,30
Vídeo	5,80	1,20	8,17	1,74
Game	18,17	1,94	20,38	2,32
Secadoura	229,57	8,15	229,72	8,62
Bomba	6,81	0,95	5,51	0,80
Gelo	43,73	5,39	40,12	5,21

Um outro resultado é quanto ao conjunto de equipamentos altamente representativos do consumo de energia domiciliar, são eles: lâmpadas incandescente e fluorescente, geladeira de 1 porta e 2 portas, freezer, chuveiro elétrico, televisão, ar condicionado e secadora de roupa. Destacamos que a variável chuveiro elétrico foi a única que incorporou na sua função de utilidade o número de pessoas que utilizam efetivamente o chuveiro no domicilio, pois como verificado em LINS et al. (1996), essa variável apresenta uma significância maior quando regredida com o número de usuários do equipamento. De fato, o banho é a única atividade que não é, geralmente, compartilhada simultaneamente por diversos moradores. Há também, um conjunto de equipamentos não significativos: como microcomputador e ferro elétrico, apesar deste último ser considerado, tradicionalmente, como um importante consumidor de energia. Isto deve ter ocorrido, provavelmente devido a hábitos culturais na região em análise. A interpretação dos coeficientes no modelo final se refere ao consumo de cada equipamento.

Temos assim que as geladeiras de uma porta consomem em média 33 kWh/mês, as geladeiras de duas portas consomem 82 kWh/mês e assim sucessivamente.

Com base na experiência com o desenvolvimento da metodologia de Análise Condicionada da Demanda, aplicada aos dados amostrais do Brasil vistos em Lins et al (2003), observamos que utilizar o consumo por equipamento regredido com o número de habitantes não apresenta uma melhora na estimativa dos consumos dos equipamentos, com exceção do chuveiro elétrico. Quando no estudo foi incorporada a variável renda dos moradores, o modelo de regressão resultou numa elevada variância não explicada, inviabilizando a sua incorporação no estudo.

Finalmente, é importante destacar alguns pontos que devem ser estudados em um trabalho futuro utilizando a modelagem de Análise Condicionada da Demanda com estimativas Robustas. Embora as equações descrevam bem os consumos residenciais, em questão, há alguns fatores a serem considerados quando de sua utilização. Utilizamos o modelo aditivo para que simplificações fossem introduzidas e o estudo fosse viável. Como, por exemplo, não foi incorporada na análise a renda dos moradores, tamanho dos domicílios, dados ambientais e etc., o que em um estudo futuro possibilite um refinamento do modelo agregar essas variáveis ao estoque de equipamentos.

6. REFERÊNCIAS BIBLIOGRÁFICAS

CHATTERJEE, S., PRICE, B. “Regression Analysis by Example”, *John Wiley & Sons,* 1977.

DeGROOT M. H. “Probability and Statistics – Second Edition”, Addison-Wesley Publishing Company, 1986.

DUBIN, J. & MCFADDEN, D. “An Econometric Analysis of Residential Eletric Appliance Holdings and Consumption”, *Econometrica* , v. 52, n. 2, pp. 345-362,1984.

GUJARATI, Damodar N. “Econometria Básica”- Pearson Education do Brasil – Terceira Edição - 1996.

GREENE, W.H. “Econometric Analysis.Third Edition. Prentice Hall. New Jersey. 1075p, 1997.

HAMPEL, F.R. RONCHETTI, E.M. ROUSSEEUW, P.J. STAHEL, W.A. R “ Robust Statistics – The Approach Based on Influence Functions” – 1986.

KMENTA, J., “Elementos de Econometria - Teoria Econométrica Básica”, *Editora Atlas S.A*., v. 2, 1990.

MONTGOMERY, D., C. PECK, VINING G.G. “ Introduction to Linear Regression Analysis” – Third Edition -1982.

LINS, M.P.E. ; SILVA, A.C.M. . Conditional Demand Analysis for Estimating Regional Variation in Appliance specific electricity Consumption for Brazilian Household Sector. In: Congresso Latino-Iberoamericano de Investigatión Operativa, 1996, Rio de Janeiro. VIII CLAIO . The First Forum on Energy (EULAFER), 1996.

LINS,.M. P. E., SILVA, A.C.M., PINGUELLI, L. R. "Regional Variations in Energy Consumption of Appliances: Conditional Demand Analysis Applied to Brazilian". *Annals of Operations Research*, v.117, pp.235-246, 2003.

MENDES, B.V.M. "Regressão Robusta: Conceitos, Aplicações e Aspectos Computacionais", 6ª Escola de Modelos de Regressão, 1999.

PARTI, M. and PARTI, C. "The total and appliance specific conditional demand for electricity in the household sector", *Bell Journal of Economics*, v.11, n.1, pp.309-321,1980.

REY, W. J. J., "Robust Statistical Methods", Springer-Verlag, 1978.

ROUSSEEUW, P.J. e LEROY, A.M. "Robust Regression and Outlier Detection" Wiley, New York, 1987.

SILVA, A.C.M, Tese de Doutorado : Análise Condicionada da Demanda de Energia no Setor Residencial Brasileiro – 2000.

SILVA, A. C. M. ; ALMEIDA, A. T. ; GODOY, M.V. . Modelagem de Apoio a Decisão na Seleção de Instrumentos de Racionalização Energética no Setor Residencial. In: Simpósio Brasileiro de Pesquisa Operacional, 2004, São João del-Rei. Anais do XXXVI SBPO, 2004.

SILVA, A. C. M. ; ROSA,L.P. . Análise Condicionada da Demanda com Correção de Heterocedasticidade. In: Simpósio Brasileiro de Pesquisa Operacional, 2004, São João del-Rei. XXXVI - SBPO, 2004.

SILVA, A. C. M. ; LINS, M. P. E. ; ALMEIDA, A. T. . O uso de modelos CDA para apoio à decisão na seleção de instrumentos de racionalização energética. In: X Congresso Brasileiro de Enegia, 2004, Rio de Janeiro. Anais do X CBE, 2004. v. 1. p. 129-138.

SILVA, A. C. M. ; ROSA,L.P. ; LINS, M. P. E. . Análise Condicionada da Demanda de Energia Elétrica para a Mesorreigão Metropolitana de Fortaleza e seu Efeito Sazonal. In: XI CLAIO Congreso Latino Iberoamericano de

Investigacion de Operaciones, 2002, Concepción. Anais XI CLAIO, 2002. p. 1-10.

S-PLUS. “S-Plus 1997: guide to statistics”, volume 1 e 2. 1997. Disponível em :http://www.mathsoft.com.

WEISBERG, S, “ Applied Linear Regression” (1947)

7. APÊNDICE

PESQUISA DE POSSE DE ELETRODOMÉSTICOS E PREFERÊNCIA DE CONSUMO

1. IDENTIFICAÇÃO: **CODUNC** ☐☐☐☐☐☐☐☐

CÓDIGO DE IDENTIFICAÇÃO: ☐ ☐☐ ☐☐☐ TEMPO RESIDÊNCIA:

☐☐a☐☐m

NOME DO CONSUMIDOR:_____
ENDEREÇO:_____

BAIRRO:_____ CEP: ☐☐☐☐☐ - ☐☐☐

01.1 - NOME DO ENTREVISTADOR: _____

01.2 - NOME DO ENTREVISTADO:_____

01.3 - TELEFONE DO DOMICÍLIO: ☐☐☐☐☐☐☐☐

01.4 - DATA DA ENTREVISTA: ☐☐/☐☐/☐☐ 01.5 - HORA DE **INÍCIO**: ☐☐:

01.6 - LISTE AS PESSOAS QUE MORAM NESTE DOMICÍLIO, ESPECIFICANDO GRAU DE PARENTESCO OU RELAÇÃO COM O(A) CHEFE DA FAMÍLIA, IDADE, SEXO, NÍVEL DE INSTRUÇÃO E PERÍODO HABITUAL DE PERMANÊNCIA NO DOMICÍLIO:

MORADORES DO DOMICÍLIO	CONDIÇÃO NO DOMICÍLIO (1)	IDADE	SEXO		NÍVEL DE INSTRUÇÃO (2)	PERÍODO HABITUAL DE PERMANÊNCIA NO DOMICÍLIO (3)				NÚMERO DE DIAS DE PERMANÊNCIA NO DOMICÍLIO
			F	M		M	T	N	MA	
1)										
2)										
3)										
4)										
5)										
6)										
7)										
8)										
9)										
10)										
11)										

CHAMADA:(1)

(**1**) PESSOA RESPONSÁVEL
(**2**) CÔNJUGE/COMPANHEIRO(A)
(**3**) FILHO OU ENTEADO
(**4**) OUTRO PARENTE
(**5**) AGREGADO
(**6**) PENSIONISTA
(**7**) EMPREGADO DOMÉSTICO
(**8**) PARENTE DE EMPREGADO
(**9**)HÓSPEDE

CHAMADA: (2)

(0) ANALFABETO
(1) 1ª A 4ª INCOMPLETO
(2) 1ª A 4ª COMPLETO
(3) 5ª A 8ª INCOMPLETO
(4) 5ª A 8ª COMPLETO
(5) MÉDIO INCOMPLETO
(6) MÉDIO COMPLETO
(7) SUPERIOR INCOMPLETO
(8) SUPERIOR COMPLETO

CHAMADA: (3)

(**M**) MANHÃ
(**T**) TARDE
(**N**) NOITE
(**MA**) MADRUGADA

2. CARACTERIZAÇÃO DO DOMICÍLIO:

0 2.1 - TIPO DE DOMICÍLIO:

1. □ CASA 2. □ APARTAMENTO 3. □ OUTRO

0 2.2 - PAREDES

1. □ ALVENARIA 2. □ MADEIRA 3. □ TAIPA NÃO REVESTIDA
4. □ MATERIAL APROVEIT. 5. □ PALHA 6. □ OUTRO

0 2.3 - PISO

1. □ MADEIRA 2. □ CERÂMICA 3. □ CIMENTO
4. □ MATERIAL APROVEIT. 5. □ TIJOLO 6. □ TERRA
7. □ OUTRO

→0 2.4 - COBERTURA (não deve ser respondido quando o tipo de domicílio for apartamento)

1. □ LAJE DE CONCRETO 2. □ TELHA DE BARRO 3. □ TELHA DE AMIANTO
4. □ ZINCO 5. □ MADEIRA 6. □ PALHA
7. □ MATERIAL APROVEIT. 8. □ OUTRO

→0 2.5 - ÁREA CONSTRUÍDA (m^2):

1. □ ATÉ 50 3. □ DE 76 a 100 5. □ DE 151 a 200
2. □ DE 51 a 75 4. □ DE 101 a 150 6. □ ACIMA DE 200

→2.6 - CONDIÇÃO DE OCUPAÇÃO DO DOMICÍLIO:

1. □ PRÓPRIO JÁ PAGO 2. □ PRÓPRIO AINDA PAGANDO
3. □ ALUGADO 4. □ CEDIDO POR EMPREGADOR
5. □ CEDIDO POR PARTICULAR 6. □ OUTRA, ESPECIFIQUE:_____

SE O DOMICÍLIO FOR <u>PRÓPRIO AINDA PAGANDO</u> OU <u>ALUGADO</u>, RESPONDA AO ITEM 2.7, SE NÃO, VÁ PARA O ÍTEM 2.8

→2.7 - QUAL FOI O VALOR MENSAL DO ALUGUEL OU DA PRESTAÇÃO PAGA, OU QUE DEVERIA TER SIDO PAGA, NO MÊS DE REFERÊNCIA DA PESQUISA (EM REAIS, EXCLUINDO O CONDOMÍNIO E O IPTU) ? _____ .

≠2.8 - O DOMICÍLIO POSSUI SISTEMA DE ABASTECIMENTO DE ÁGUA :

COM CANALIZAÇÃO INTERNA **SEM CANALIZAÇÃO INTERNA**

1. REDE GERAL

2. POÇO OU NASCENTE 6.

3. CARRO PIPA

7. CARRO PIPA

4. OUTRA FORMA

8. OUTRA FORMA

3. ILUMINAÇÃO

≠0 3.1. CARACTERÍSTICAS E HÁBITOS DE USO

TIPO DE CÔMODO	LÂMPADAS		TIPO DE USO DAS LÂMPADAS (**EVENTUAL OU HABITUAL**)	
	Total	Tipo (1)	EVENTUAL (X)	HABITUAL (número de horas estimada de uso)
Sala de estar, jantar E TV				
Quarto 1				
Quarto - 2				
Quarto- 3				
Quarto - 4				
Banheiro -1				
Banheiro -2				
Banheiro -3				
Corredores				
Copa/Cozinha				
Área de Serviço				
Garagem				
Área Externa				

NOTA: (1) Na sala e na copa/cozinha deve ser verificada a potência na própria lâmpada, nos demais cômodos essa medida pode ser feita por declaração.

CHAMADA (1) :TIPO DE LÂMPADA

(1) 25 W- Incandescente
(2) 40 W - Incandescente
(3) 60 W - Incandescente
(4) 100 W - Incandescente
(5) 150 W - Incandescente
(6) 20 W- Fluorescente Tubular
(7) 40 W- Fluorescente Tubular
(8) Fluorescente Compacta
(9) Fluorescente Circular
(10) Dicróica
(11) PL
(12) OUTRO

→3.2 – QUAL A TONALIDADE DE ILUMINAÇÃO DE SUA PREFERÊNCIA?

1. □ AMARELA 2. □ BRANCA. 3. □ NÃO TEM PREFERÊNCIA 4. □ NÃO SABE.

→3.3 – VOCÊ TROCARIA A LÂMPADA INCANDESCENTE (QUE TEM MAIOR CONSUMO, PORÉM DE MENOR CUSTO DE AQUISIÇÃO) PELA FLUORESCENTE COMPACTA (QUE TEM MENOR CONSUMO, PORÉM DE MAIOR CUSTO DE AQUISIÇÃO)?

1. □ SIM 2. □. NÃO 3. □ NÃO SABE

EXPLICAÇÃO DAS CONSEQÜÊNCIAS: CONSIDERE QUE A LÂMPADA FLUORESCENTE COMPACTA CUSTA R$ 16,00 E A LÂMPADA INCANDESCENTE R$ 1,00. NESTE CASO, VOCE TERIA O RETORNO DA DIFERENÇA DE PREÇO EM 6 MESES E AINDA TERIA MAIS 6 MESES ADICIONAIS DE GANHO NO CONSUMO

4. REFRIGERADOR

→0 4.1. CARACTERÍSTICAS

Nº DE REFE-RÊNCIA DO APARE-LHO	TIPO DE APARELHO (1)				UTILIZ. (2)	TERMOSTATO			ESTIMATIVA DE IDADE DO APARELHO (em anos)	PROBLEMAS OCORRIDOS NOS ÚLTIMOS 12 MESES (3)	MEDIDAS DO APARELHO (cm)		
	MARCA	Nº DE PORTAS	FROST FREE – S/N	CAPACID litros		MÍN	MÉD	MÁX			ALTURA	LARGURA	PROFUND
1													
2													
3													

CHAMADA (1): VEJA NO **CARTÃO 3**.

CHAMADA (2)
(**1**) USO PERMANENTE
(**2**) **)** DESLIGADO
(**3**) USO PARTE DO DIA
(**4**) SÓ LIGADO EVENTUALMENTE

CHAMADA (3):
(**1**) MOTOR COM DEFEITO OU RUÍDO EXCESSIVO ?
(**2**) PORTA COM DIFICULDADE PARA FECHAR ?
(**3**) CONGELADOR FAZENDO GELO DEMAIS OU DE MENOS?
(**4**) OUTROS PROBLEMA

OBS.: ESTA QUESTÃO ADMITE RESPOSTAS MÚLTIPLAS

NOTA: A ÚLTIMA COLUNA REFERENTE AS MEDIDAS DO APARELHO SÓ DEVE SER PREENCHIDA CASO NÃO TENHA SIDO POSSÍVEL IDENTIFICAR O "TIPO DE APARELHO".

→4.2 - DO(S) REFRIGERADOR(ES) ACIMA CITADO(S), ADQUIRIU ALGUM NOVO (NA LOJA) NOS ÚLTIMOS DEZ ANOS?

1. □ NÃO 2. □ SIM QUANTOS? ☐

CASO "NÃO TENHA ADQUIRIDO" VÁ PARA O ITEM 5

→4.3. NESSA AQUISIÇÃO LEVOU EM CONSIDERAÇÃO O CONSUMO DO APARELHO MOSTRADO NA ETIQUETA AO ADQUIRÍ-LO?

1. □ SIM 3. □ NÃO SABIA DA EXISTÊNCIA DE ETIQUETA

2. □ NÃO ME INTERESSEI 4. □ NÃO TINHA ETIQUETA

OBS.: ESTA QUESTÃO ADMITE RESPOSTAS MÚLTIPLA

→4.4. NO CASO DE UMA FUTURA TROCA DE GELADEIRA VOCÊ COMPRARIA UMA COM SELO PROCEL?

1. □ SIM 2. □ NÃO PORQUE? _____.

5. FREEZER

→0 5.1. CARACTERÍSTICAS

Nº-REFERÊNCIA DO APARELHO	TIPO DE APARELHO (1)				UTILIZ. (2)	ESTIMATIVA DE IDADE DO APARELHO (em anos)	MEDIDAS DO APARELHO (cm)		
	MARCA	VERT/ HORIZ	FROST FREE - S/N	CAPACID litros			ALTURA	LARGURA	PROFUND
1									
2									
3									

CHAMADA (1): VEJA NO **CARTÃO 4**

CHAMADA (2) :

(1) USO PERMANENTE. **(3)** USO PARTE DO DIA.

(2) DESLIGADO. **(4)** SÓ LIGADO EVENTUALMENTE.

NOTA: A ÚLTIMA COLUNA REFERENTE AS MEDIDAS DO APARELHO SÓ DEVE SER PREENCHIDA CASO NÃO TENHA SIDO POSSÍVEL IDENTIFICAR O "TIPO DE APARELHO".

6. AQUECIMENTO DE ÁGUA PARA BANHO

→6.1 - COMO ESQUENTA A ÁGUA ?

AQUECIMENTO ELÉTRICO

AQUECIMENTO A GÁS

1. ☐ CHUVEIRO ELÉTRICO 4. ☐ GLP (GÁS DE BOTIJÃO)
2. ☐ BOILER
5. ☐ BOILER
3. ☐ AQUECIMENTO CENTRAL 6. ☐ AQUECIMENTO CENTRAL

7. ☐ AQUECEDOR SOLAR
8. ☐ NÃO ESQUENTA (BANHO FRIO)
9. ☐ OUTROS FORMAS DE AQUECIMENTO QUAL:_____

OBS.: ESTÁ QUESTÃO ADMITE RESPOSTAS MÚLTIPLAS

CASO UTILIZE **CHUVEIRO ELÉTRICO** PREENCHA OS ITENS 6.2 E 6.3

≠0 6.2 - CARACTERÍSTICAS

Nº REFERÊNCIA DO APARELHO	TIPO DE APARELHO (1)		NÚMERO DE PESSOAS QUE USAM	POSIÇÃO EM QUE SE ENCONTRA A CHAVE DO APARELHO NO VERÃO			DURANTE OS MESES DE INVERNO A CHAVE FICA NA POSIÇÃO		
	MARCA	POTÊNCIA (WATTS)		VERÃO (morno)	INVERNO (quente)	DESLIGADA (frio)	VERÃO (morno)	INVERNO (quente)	DESLIGADA (frio)
1									
2									
3									

CHAMADA: (1) VEJA NO **CARTÃO 5**

NOTA: PERGUNTAR OU IDENTIFICAR O MODELO DE CHUVEIRO ELÉTRICO DE ACORDO COM O CARTÃO 5.

≠6.3 - SE A TARIFA FOSSE O DOBRO DAS 17:30 ÀS 20:30 VOCÊ ACHA QUE A SUA FAMÍLIA EVITARIA TOMAR BANHO DE CHUVEIRO ELÉTRICO (QUENTE) NESTE HORÁRIO ?

1. ☐ SIM 2. ☐ NÃO 3. ☐ NÃO SEI

7. CONDICIONADOR DE AR

≠0 7.1 - CARACTERÍSTICAS

Nº REFERÊN-CIA DO APARELHO	TIPO DE APARELHO		ESTIMATIVA DA IDADE DO APARELHO	ESTE CÔMODO RECEBE SOL?			A LIMPEZA DO FILTRO É FEITA NO INÍCIO DE UM PERÍODO DE USO PROLONGADO DO APARELHO ?		MEDIDAS DO APARELHO	
	MARCA (1)	BTU	(em anos)	MANHÃ	TARDE	NÃO	SIM	NÃO	ALTURA	LARGURA
1										
2										
3										

CHAMADA (1): CARTÃO 6

NOTA: CASO NÃO SEJA POSSÍVEL IDENTIFICAR NO APARELHO SUA MARCA E BTU PREENCHA A LACUNA REFERENTE AS MEDIDAS DO APARELHO.

+7.2 - HÁBITOS DE USO DE ACORDO COM O CLIMA

Nº REFERÊNCIA DO APARELHO	USA O APARELHO NO CLIMA.....?(**S**IM OU **N**ÃO)		GRAU DE UTIL. (1)
1	Quente	( )	
	Ameno	( )	
	Frio	( )	
2	Quente	( )	
	Ameno	( )	
	Frio	( )	

CHAMADA (1) :

G (GRANDE)- UTILIZADO MAIS DE 4 VEZES POR SEMANA. **P** (PEQUENA) - MENOS DE UMA VEZ POR MÊS.

M (MÉDIA) - DE 1 A 3 VEZES POR SEMANA. **N** (NENHUMA) - NÃO UTILIZA

R (REGULAR) - DE 1 A 3 VEZES POR MÊS .

NOTA: DEVE SER EXCLUÍDO O PERÍODO EM QUE O CONDICIONADOR DE AR É UTILIZADO APENAS NA VENTILAÇÃO.

8. TELEVISÃO

+0 8.1 - CARACTERÍSTICAS E HÁBITOS DE USO

Nº REFERÊNCIA DO APARELHO	TIPO DE APARELHO		ESTIMATIVA DE IDADE (anos)	GRAU DE UTILIZAÇÃO (2)	MEDIDA NA DIAGONAL (cm)
	MARCA (1)	TAMANHO (POLEGADAS)			
1					
2					
3					
4					
5					

CHAMADA (1): VEJA NO **CARTÃO 7**

CHAMADA (2) :

G (GRANDE)- UTILIZADO MAIS DE 4 VEZES POR SEMANA. **P** (PEQUENA) - MENOS DE UMA VEZ POR MÊS.

M (MÉDIA) - DE 1 A 3 VEZES POR SEMANA. **N** (NENHUMA) - NÃO UTILIZA.

R (REGULAR) - DE 1 A 3 VEZES POR MÊS .

NOTA: PERGUNTE AO ENTREVISTADO A MARCA E O TAMANHO DO TELEVISOR, CASO ELE NÃO SAIBA OU HAJA DÚVIDA DA RESPOSTA PREENCHA A COLUNA CORRESPONDENTE À MEDIDA NA DIAGONAL

9. OUTROS ELETRODOMÉSTICOS

+9.1. CARACTERÍSTICAS E HÁBITOS DE USO

APARELHO	QUANTIDADE	GRAU DE UTILIZAÇÃO (1)		
		APARELHO 1	APARELHO 2	APARELHO 3
1. APARELHO DE SOM				
2. RÁDIO ELÉTRICO				
3. VIDEOCASSETE				
4. MICROCOMPUTADOR				
5. IMPRESSORA				
6. VIDEOGAME				
7. FERRO				
8. LAVA ROUPA				
9. LAVA LOUÇA				
10. SECADORA DE ROUPA				
11. FORNO DE MICROONDAS				
12. FORNO ELÉTRICO				
13. LIQUIDIFICADOR				
14. BATEDEIRA				
15 CAFETEIRA ELÉTRICA				
16. PANELA ELÉTRICA				
17. EXAUSTOR				
18.VENTILADOR/CIRCULADO				
19. AQUECEDOR DE				
20. ENCERADEIRA				
21. ASPIRADOR DE PÓ				
22. BOMBA D'AGUA				
23. GELO ÁGUA				
24. DVD				
25. PURIFICADOR				
26. SECADOR DE CABELO				
27. FAX				
28. SANDUICHEIRA/GRIL				

CHAMADA (1):

G (GRANDE)- UTILIZADO MAIS DE 4 VEZES POR SEMANA.

M (MÉDIA) - DE 1 A 3 VEZES POR SEMANA.

R (REGULAR) - DE 1 A 3 VEZES POR MÊS .

P (PEQUENA) - MENOS DE UMA VEZ POR MÊS.

N (NENHUMA) - NÃO UTILIZA.

NOTA: (1) Se no domicílio houver outro(s) equipamento(s) com o uso pelo menos "regular" - 1 a 3 vezes por mês - ele deve ser incluido na lista.

+ 9.2 - QUAIS DOS SEGUINTES EQUIPAMENTOS FORAM COMPRADOS PARA ESTE DOMICÍLIO NOS ÚLTIMOS DOIS ANOS ? INDIQUE O MÊS E O ANO DE AQUISIÇÃO?

GELADEIRA		FREEZER		LAVA ROUPA		LAVA LOUÇA		TELEVISÃO		AR CONDI CIONADO		COMPU TADOR		VIDEO CASSETE		APARELHO DE SOM		FORNO DE MICROONDAS		OUTRO, QUAL?_____	
MÊS	ANO	MÊS	ANO	MÊS	ANO	MÊS	ANO	MÊS	ANO	MÊS	ANO	MÊS	ANO	MÊS	ANO	MÊS	ANO	MÊS	ANO	MÊS	ANO

+ 9.3 - PRETENDE COMPRAR ALGUM ELETRODOMÉSTICO NOS PRÓXIMOS SEIS MESES, PARA ESTE DOMICÍLIO?

1. □ NÃO 2. □ SIM 3. QUAL(IS) (IDENTIFIQUE NO CARTÃO 8) :

+ 9.4 - CASO PRETENDA COMPRAR UM EQUIPAMENTO QUE JÁ POSSUA, QUAL DESTINO SERÁ DADO AO EQUIPAMENTO ANTIGO?

1. □ NÃO SABE 2. □ JOGAR FORA 3. □ DAR PARA ALGUÉM 4. □ FICAR COM OS DOIS 5. □VENDER

10. INFORMAÇÕES SOBRE ENERGIA ELÉTRICA

+10.1 - NESSA RESIDÊNCIA SÃO ADOTADAS MEDIDAS PARA ECONOMIZAR ENERGIA ELÉTRICA?

1. □ SIM
2. □ NÃO Vá para a pergunta 10.4
3. □ Não sabe Vá para a pergunta 10.4

+10.2 – COM RELAÇÃO A MEDIDAS DE EFICIÊNCIA, QUAIS VOCÊS ADOTAM?

(**LOCALIZAR** NO **CARTÃO 1** MARCANDO UM "X" NOS ESPAÇOS CORRESPONDENTES, ACEITANDO RESPOSTAS MÚLTIPLAS.)

1	2	3	4	5	6	7	8	9	10	11

+10.3 - QUAIS DESTAS MEDIDAS SÃO ADOTADAS? (MOSTRAR O CARTÃO 1 - MARCANDO UM "X" NOS ESPAÇOS CORRESPONDENTES, ACEITANDO RESPOSTAS MÚLTIPLAS.)

1	2	3	4	5	6	7	8	9	10	11

+10.4 – DENTRE OS ASSUNTOS RELACIONADOS NESTE CARTÃO, QUAIS SÃO OS ITENS, SOBRE OS QUAIS INTERESSA RECEBER INFORMAÇÕES ? (MOSTRAR CARTÃO 2 - ACEITANDO RESPOSTAS MÚLTIPLAS)

6. □ Outros. Quais ?_____

+10.5 - QUAL APARELHO ABAIXO O(A) SR(A) UTILIZA NO HORÁRIO DE 17:30 ÀS 20:30?

APARELHOS	1 - NÃO TEM O APARELHO	2- TEM, MAS NÃO USA	3 - USA
CHUVEIRO ELÉTRICO			
MICROONDAS			
AR CONDICIONADO			
FERRO ELÉTRICO			
FORNO ELÉTRICO DE PAREDE			
MÁQUINA DE LAVAR ROUPA			
MÁQUINA DE LAVAR LOUÇA			
FREEZER			
GELADEIRA			

NOTA: PARA A QUESTÃO 10.6, CONSIDERE APENAS AS LINHAS (APARELHOS) CUJAS RESPOSTAS À QUESTÃO ANTERIOR FORAM "USA".

+ 10.6. SE FOR OFERECIDO UM DESCONTO DE 10% NO VALOR DA SUA CONTA DE ENERGIA ELÉTRICA, DESDE QUE OS APARELHOS INDICADOS NÃO SEJAM LIGADOS, DURANTE TRÊS HORAS, NO HORÁRIO DE 17:30 ÀS 20:30. O(A) SR(A) CONCORDA COM A INSTALAÇÃO POR CONTA DA CONCESSIONÁRIA, DE UM DISPOSITIVO PARA IMPEDIR SEU FUNCIONAMENTO ?

APARELHOS	1 – SIM	2 - NÃO	3- SE NÃO ATÉ QUE % ACEITA	4 – NÃO SABE
CHUVEIRO ELÉTRICO				
MICROONDAS				
AR CONDICIONADO				
FERRO ELÉTRICO				

FORNO ELÉTRICO DE PAREDE				
MÁQUINA DE LAVAR ROUPA				
MÁQUINA DE LAVAR LOUÇA				
FREEZER				
GELADEIRA				

+10.7 - CASO A CONCESSIONÁRIA OFEREÇA ENERGIA ELÉTRICA MAIS BARATA PARA REDUZIR O SEU CONSUMO, O(A) Sr.(a) ESTARIA DISPOSTO(A) A REDUZÍ-LO ?

1. □ SIM 3. □ DEPENDE DO DESCONTO

2. □ NÃO 4. □. NÃO SABE

CASO, A RESPOSTA DA QUESTÃO 10.7 SEJA 1 OU 3, PREENCHER A QUESTÃO 10.8

+10.8 -:QUAL O DESCONTO QUE O SR (SRA) ACHARIA RAZOÁVEL PARA REDUZIR O SEU CONSUMO DE ENERGIA NOS SEGUINTES PERCENTUAIS?

REDUÇÃO DO CONSUMO DE ENERGIA EM	MENOR DESCONTO QUE VOCÊ ACHARIA RAZOÁVEL? (EM %)	NÃO ACEITARIA REDUZIR O CONSUMO NESTE NÍVEL (MARQUE UM X)
1 – 20%	1-	1-
2 – 10%	2-	2-
3 – 25%	3-	3-

O VALOR DE 20% CORRESPONDE AO ESFORÇO REALIZADO DURANTE O ÚLTIMO RACIONAMENTO

+10.9 - CONSIDERE QUE VAI HAVER UM AUMENTO DE TARIFA. O(A) SR.(A) ACEITARIA REDUZIR O CONSUMO EM 20%, PARA MANTER A TARIFA ATUAL?

*1. □ SIM 3. □ DEPENDE DO AUMENTO DA TARIFA

2. □ NÃO 4. □ NÃO SABE

CASO A RESPOSTA DA QUESTÃO 10.9 SEJA 2 ou 3, RESPONDA À PERGUNTA 10.10:

+10.10 - SE A TARIFA AUMENTASSE EM 20%, O(A) Sr.(a) ESTARIA DISPOSTO(A) A REDUZIR O CONSUMO DE ENERGIA EM 20%?

*1. □ SIM *3. □ SÓ SE O AUMENTO DA TARIFA FOSSE MAIOR, QUAL SERIA O VALOR LIMITE? ☐

2. □ NÃO 4. □ NÃO SABE

CASO A RESPOSTA DA QUESTÃO 10.9 SEJA 1, OU A RESPOSTA DA QUESTÃO 10.10 SEJA 1 OU 3, RESPONDA À PERGUNTA 10.11:

*+10.11- SE O(A) SR.(A) ACEITASSE REDUZIR O CONSUMO, SEJA EM FUNÇÃO DE UM DESCONTO DE SEU INTERESSE OU PARA EVITAR UM AUMENTO INDESEJÁVEL DE TARIFA, EM QUE SERVIÇO ATUARIA PRIORITARIAMENTE

SERVIÇO	ESCOLHA DOIS EM QUE VOCÊ ATUARIA PRIORITARAMENTE? (INDICAR 1 PARA O PRIMEIRO E 2 PARA O SEGUNDO NA ORDEM)	INDIQUE EM QUAL VOCÊ SÓ ATUARIA EM ÚLTIMO CASO
ILUMINAÇÃO		
TELEVISÃO		
GELADEIRA		
FERRO ELÉTRICO		
CHUVEIRO ELÉTRICO		
MÁQUINA DE LAVAR ROUPA		
FREEZER		
MICROONDAS		
AR CONDICIONADO		
MÁQUINA DE LAVAR LOUÇA		

10.12- Como você atuaria no serviço 1 para reduzir o consumo?

_____

_____

10.13- Como você atuaria no serviço 2 para reduzir o consumo?

_____

_____

11. DADOS SÓCIO-ECONÔMICOS

11.1 - ITENS DE CONFORTO FAMILIAR

ITENS	NÃO TEM	QUANTIDADE					
		1	2	3	4	5	6 e +
RÁDIO							
BANHEIRO							
AUTOMÓVEL							
EMPREGADA MENSALISTA							

11.2. RENDA DOMICILIAR (Piso nacional de salários)

1. □ <1 4. □ 3 a 4 7. □ 7 a 10 10. □ 20 a 30 13. □ NÃO SABE
2. □ 1 a 2 5. □ 4 a 5 8. □ 10 a 15 11. □ 30 a 40
3. □ 2 a 3 6. □ 5 a 7 9. □ 15 a 20 12. □ > 40

+ 11.3. NESTE DOMICÍLIO É FEITO ALGUM TIPO DE TRABALHO PARA SER COMERCIALIZADO ? (OLHAR NO **CARTÃO 9)** 1. □ NÃO 2. □ SIM 3. QUAL(IS):_____

+ 11.4. QUAIS SÃO OS EQUIPAMENTOS ELÉTRICOS UTILIZADOS NESTE(S) TRABALHO(S) (IDENTIFIQUE OS EQUIPAMENTOS NO **CARTÃO 8)?** ____ , ____ , ____ , ____

011.5 - REGIÃO DO DOMICÍLIO: □ LUXO □ CLASSE MÉDIA □ POBRE

011.6 - PRÓXIMO A FAVELA: □ SIM □ NÃO □ NA FAVELA

011.7 - HORA DE **TÉRMINO** DA ENTREVISTA: □□: □□