

**PROPOSTA DE MODELOS PARA ESTIMAÇÃO DE
RENDA FAMILIAR**

OSVALDO MARTINS DE SOUZA

2007

OSVALDO MARTINS DE SOUZA

**PROPOSTA DE MODELOS PARA
ESTIMAÇÃO DE RENDA FAMILIAR**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

Orientador: Prof. Dr. Agostinho Roberto de Abreu

Lavras, Minas Gerais

2007

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Souza, Osvaldo Martins de
Proposta de modelos para estimação de renda familiar/ Osvaldo Martins de
Souza. -- Lavras : UFLA, 2007.
67 p. : il.

Orientador
Dr. Agostinho Roberto de Abreu.
Dissertação (Mestrado) - UFLA.
Bibliografia.

1. Renda familiar. 2. Modelo de regressão. 3. Classe econômica. 4. Amostragem. I. Universidade Federal de Lavras. II. Título.

CDD-519.542

OSVALDO MARTINS DE SOUZA

**PROPOSTA DE MODELOS PARA
ESTIMAÇÃO DE RENDA FAMILIAR**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

APROVADA em 16 de Julho de 2007

Prof. Dr. Fabyano Fonseca e Silva

UFV

Prof. Dr. Augusto Ramalho de Moraes

UFLA

Prof. Dr. Agostinho Roberto de Abreu
UFLA
(Orientador)

Lavras, Minas Gerais
2007

À Minha esposa **LUCILAINE**, e minha filha **PAULA**, pois, das coisas boas em minha vida, vocês são, certamente as melhores

OFEREÇO.

Aos meus pais, **IZAIAS** e **DORVALINA** e meus **IRMÃOS** (a),

DEDICO.

AGRADECIMENTOS

Ao Ser Superior, por estar sempre presente, me guiando e dando forças para vencer os obstáculos.

A minha família pelos constantes incentivos e pelo perdão nos muitos momentos de ausência.

À Universidade Federal de Lavras, pela oportunidade de aprimoramento acadêmico.

À Secretaria de Educação de Mato Grosso (Seduc), pelo incentivo, e todos os colegas e amigos do colégio 14 de Fevereiro.

Ao apoio, confiança, dedicação, críticas e ensinamentos durante o curso, além da amizade do orientador, Prof. Dr Agostinho R. de Abreu

Aos professores do Departamento de Ciências Exatas, pelos conhecimentos transmitidos, em especial ao Prof. Dr Daniel Furtado, que sempre colaborou para o esclarecimento de muitas de minhas dúvidas.

Aos meus amigos: Samir, da Secretaria de Planejamento de Pontes e Lacerda, MT e Carlos do IBGE de Cuiabá, MT que sempre se dispuseram de informações que auxiliaram nesse trabalho.

Aos colegas de turmas de 2004, 2005 e 2006 que sempre estiveram presentes nos momentos difíceis, em especial: Eric, Devanil e Verônica, e também aos funcionários do Departamento de Ciências Exatas.

Nesta página, eu não poderia deixar de mencionar, de maneira muito especial, ao doutorando Eric Batista Ferreira, um grande amigo, que provou sua amizade para comigo, atuando como um co-orientador sempre presente e que não mediu esforços para que esta dissertação viesse à luz de forma adequada . Obrigado pelo que você fez por mim.

RESUMO

SOUZA, Osvaldo Martins de. **Proposta de modelos para estimação de renda familiar**. Lavras: UFLA, 2007. 67 p. (Dissertação - Mestrado em Estatística e Experimentação Agropecuária) *

A estimação de renda familiar e classe econômica costuma ser constrangedora e/ou dispendiosa, uma vez que demanda entrevistas e aplicação de questionários. Este trabalho baseia-se na investigação de modelos de regressão baseados em covariáveis objetivas para estimar classe econômica e renda familiar na cidade de Pontes e Lacerda, MT. As variáveis que não dependem de entrevistas sugeridas são o consumo mensal de água (m^3), de energia elétrica (kWh) e a área construída m^2 . Um modelo híbrido, que leva em consideração o total de pontos conseguidos no Critério Brasil, mostrou-se como uma melhor performance e o modelo objetivo se mostrou tão eficiente quanto o Critério Brasil na estimação da classe econômica.

*Orientador Prof. Dr. Agostinho Roberto de Abreu - UFLA. Co-orientador: MSc. Eric Batista Ferreira.

ABSTRACT

SOUZA, Osvaldo Martins de. **Models for estimating familiar income.** Lavras: UFLA, 2007. 67 p. (Dissertation - Master in Statistics and Agricultural Experimentation) *

Estimating familiar income and economic class use to be a constraining and/or expensive task, since it demands interviews and application of questionnaires. This work is based on regression models constructed with objective covariates to esteem economic class and familiar income at the city of Pontes e Lacerda, MT. Suggested variables that do not depend on interviews are: the monthly consumption of water (m^3), of electricity (kWh) and the constructed area m^2 . An hybrid model, that takes in consideration the total of points obtained in the Critério Brasil, had the best performance and the objective model was so efficient as the Critério Brasil in estimating the economic class.

*Supervisor: Prof. Dr: Agostinho Roberto de Abreu - UFLA. Co-advisor: MSc. Eric Batista Ferreira.

SUMÁRIO

Resumo	i
Abstract	ii
Lista de Figuras	v
Lista de Tabelas	vii
1 Introdução	1
2 Referencial Teórico	3
2.1 Amostragem	3
2.1.1 Amostra Aleatória Simples (AAS)	7
2.1.2 Amostra Aleatória Estratificada (AAE)	10
2.1.3 Amostragem Aleatória por Conglomerado (AAC)	14
2.1.4 Amostragem Aleatória Sistemática (AS)	17
2.2 Regressão	19
2.2.1 Regressão linear simples	19
2.2.2 Regressão linear múltipla	22

2.2.3	Seleção de modelos	24
2.2.3.1	Stepwise	25
2.2.3.2	Forward	27
2.2.3.3	Backward	27
2.2.4	Qualidade de ajuste	28
2.3	Pesquisa de opinião pública	31
2.3.1	Renda, classe econômica e classe social	32
2.3.2	Classes econômicas no Brasil	36
2.3.3	Classe social e Status	37
2.3.4	Processos de compras	41
2.3.5	Poder de compra e renda	41
3	Metodologia	43
3.1	Pontes e Lacerda	43
3.1.1	História do município	44
3.1.2	Vias de acesso	45
3.2	A amostragem	46
4	Resultados e discussão	49
5	Conclusões	65
	Referências	66

LISTA DE FIGURAS

1	Esquema da retirada de uma amostra de n indivíduos de uma população de tamanho N	6
2	Esquema da retirada de um indivíduo, com probabilidade $1/N$, de uma população de tamanho N	8
3	Esquema da retirada de uma amostra proporcional ao tamanho dos estratos de uma população de tamanho N	11
4	Esquema da retirada de uma amostra aleatória por conglomerado de uma população de tamanho N	15
5	Esquema de uma população disposta em forma seqüencial e, em negrito, estão os elementos a serem amostrados.	18
6	Box-plot da renda, R\$, dos 23 bairros amostrados, ressaltando a amostra piloto (1 a 10).	51
7	Proporções das classes econômicas em Pontes e Lacerda, MT, destacando-se, na parte superior do retângulo hachurado, a estimação pontual, e a parte superior do retângulo em branco o limite superior do intervalo de confiança para proporção com 95% de confiança.	52
8	Histograma da renda (R\$) em Ponte de Lacerda, MT, destacando-se a moda, a média e a média.	53

9	Comparação entre as classes estimadas por meio do Critério Brasil (azul), modelo objetivo (verde), modelo híbrido (vermelho) e classes econômicas reais (preto).	64
10	Comparação entre os intervalos de confiança da raiz de erros quadráticos médios cometidos pelos modelos, Critério Brasil, objetivo e híbrido na estimação de classes econômicas.	64

LISTA DE TABELAS

1	Expressões para obtenção de intervalo de confiança e tamanho de amostra para população.	7
2	Representação esquemática de uma população estratificada, contendo os dados originais de cada estrato Y_h^* e seus respectivos totais (T_h), médias ($\mu_H = \bar{Y}_H$) e variâncias (σ_H^2 ou S_H^2).	11
3	Tabela de análise de variância para um modelo de regressão linear simples contendo as fontes de variação com os respectivos graus de liberdade (GL), somas de quadrados (SQ), e quadrados médios (QM) e a estatística do teste F para a hipótese de nulidade do coeficiente de regressão	30
4	Associação tabular entre renda e classe econômica usada pelo Critério Brasil.	36
5	Renda média familiar em R\$, variância da renda familiar e tamanho da amostra em dez bairros da cidade de Pontes e Lacerda, MT	49
6	Modelo Objetivo: comparação dos modelos 4.1 e 4.2, estimados por Stepwise e backward vs forward	55
7	Modelo Híbrido: comparação dos modelos 4.3 e 4.4, estimados por Stepwise e backward vs forward	56

8	Coeficiente de determinação ajustado (R_{aj}^2), erro quadrático médio na renda (EQM_{renda}) e p-valor do teste de significância dos coeficientes de determinação, múltiplos (R^2), nos 23 bairros estudados, para os modelos de regressão, baseados em covariáveis objetivas (Ob) e o modelo híbrido (Hb).	58
9	Presença (x) de efeitos no modelo objetivo final para os 23 bairros estudados.	60
10	Presença (x) de efeitos no modelo híbrido final para os 23 bairros estudados.	61
11	Ajuste do modelo baseado em covariáveis objetivas.	62
12	Ajuste do modelo híbrido, baseado em covariáveis objetivas e no total de pontos no Critério Brasil.	63

Capítulo 1: Introdução

1 INTRODUÇÃO

No mundo atual percebe-se que os indivíduos são diferentes e estas diferenças se baseiam em aspectos materiais, raça, sexo, cultura e outros. Os aspectos mais simples para se constatar que os homens são diferentes são: físicos ou sociais. Constatamos isso em nossa sociedade, pois nela existem indivíduos que vivem em absoluta miséria e outros que vivem em mansões rodeados de objetos luxuosos e com mesa farta todos os dias enquanto outros não têm o que comer durante o dia. Pelas próprias características, em cada sociedade existem essas desigualdades, elas assumem feições distintas, porque são constituídas de um conjunto de elementos econômicos, políticos e culturais próprios de cada sociedade.

O conhecimento da renda familiar se reveste de fator importante propiciando tomadas de decisões, investimentos, etc.

Este trabalho baseia-se, para tanto, nos resultados de uma pesquisa de variáveis objetiva: Área, água e Energia, que procuram validar uma nova classificação socioeconômica de renda familiar para o município de Pontes e Lacerda, MT, classificação esta construída de uma perspectiva prática na análise e divisão da classe econômica, inspirada nas contribuições do consumo mensal de energia elétrica (KWh), de água (m^3), e na área construída (m^2) de cada propriedade. Este modelo traduz um esforço de aperfeiçoamento da técnica de estimar a renda familiar com maior precisão

e menos constrangimento para os entrevistados.

Existem vários métodos de avaliação da renda familiar, entre esses mostra-se uma abordagem alternativa ao acompanhamento de indicadores (que se fundamentam nas características das famílias), buscando uma estimativa de renda, para definir a classe econômica.

2 REFERENCIAL TEÓRICO

2.1 Amostragem

É compreensível que o estudo de todos os elementos da população possibilita um preciso conhecimento das variáveis que estão sendo pesquisadas; todavia, nem sempre é possível obter as informações de todos os elementos da população. Limitações de tempo, custo e as vantagens do uso das técnicas estatísticas de inferência justificam o uso de planos amostrais. Torna-se claro que a representatividade da amostra dependerá do seu tamanho e de outras considerações de ordem metodológica. Isto é, o investigador procurará acercar-se de cuidados, visando à obtenção de uma amostra significativa, ou seja, que de fato represente “o melhor possível” toda a população (Ferreira, 2005) .

À medida que o tamanho da amostra aumenta, as distribuições do estimador da média \bar{x} e do total populacional T vão se aproximando da distribuição normal, de acordo com o Teorema de Limite Central (TLC), tanto para o caso da Amostragem Aleatória Simples com reposição (AASc), como para Amostragem Aleatória Simples sem reposição (AASs) (Bussab & Moretin, 2005)

A média amostral

$$\bar{x} = \frac{1}{n} \sum_{i \in s} Y_i = \hat{\mu},$$

sendo \bar{x} = média; n = tamanho da amostra; Y_i = valor da i -ésima observação; $\hat{\mu}$ média populacional,

é um estimador não viesado da média populacional μ dentro do plano AASc, e ainda

$$Var[\bar{x}] = \frac{\sigma^2}{n}$$

O uso de informações de uma amostra para inferir sobre a população faz parte da atividade diária da maioria das pessoas. Basta observar como uma cozinheira verifica se o prato que ela está preparando tem ou não quantidade adequada de sal. Ou, ainda, quando um comprador, após experimentar um pedaço de laranja numa banca de feira, decide se vai comprar ou não as laranjas. Essas são decisões baseadas em procedimentos amostrais (Bussab & Moretin, 2003) .

Um ponto deve ser salientado: amostras maiores dão resultados mais precisos do que amostras menores. Extraíndo uma amostra muito grande, pode-se confiar que o resultado da amostra esteja muito próximo da verdade sobre a população. A amostra da pesquisa populacional dos EUA de 60.000 residências, estima com bastante precisão a taxa nacional de desemprego. Naturalmente, só as amostras probabilísticas, isto é, aquelas escolhidas aleatoriamente, é que oferecem esta garantia. A utilização do planejamento amostral probabilístico e o cuidado de contornar as dificuldades práticas reduzem a tendenciosidade de uma amostra. O tamanho da amostra determina, então, quão próximo da verdade populacional, provavelmente, estará o resultado amostral (David, 2000) .

Os levantamentos por amostragem permitem obter informações a

respeito de valores populacionais desconhecidos, por meio da observação de apenas uma parte (amostra) do universo de estudo (população). Os elementos da população são as unidades de observação e de análise. A unidade amostral é a mesma unidade de observação e de análise, sendo a menor parte distinta da população (Ferreira, 2005) .

Dentre os vários processos existentes para a obtenção de amostras, a amostragem probabilística caracteriza-se por garantir, a priori, que todo elemento pertencente ao universo de estudo possua probabilidade, conhecida e diferente de zero, de pertencer à amostra sorteada. A identificação, direta ou indireta, dos elementos e o uso de sorteio fundamentam as propriedades matemáticas desse tipo de processo. Se por qualquer razão, alguns elementos da população não puderem pertencer à amostra sorteada, a amostragem é dita não- probabilística. Alguns tipos de amostragem não-probabilísticos podem ser empregados quando a população de estudo não é totalmente acessível, quando a amostragem é realizada a esmo, ou seja, sem sorteio, e quando a população é formada de material contínuo (líquido ou gás), em que o uso de sorteio não é possível (Ferreira, 2005).

A amostragem probabilística (ou estatística) fundamenta-se na distribuição amostral e na seleção aleatória. As unidades amostrais possuem probabilidade conhecida e diferente de zero de serem sorteadas. Além disso, o processo de sorteio é sempre necessário, pois trata-se do método que garante cientificamente a aplicação das técnicas estatísticas de inferências. A figura 1 sugere que na população de N elementos são sorteados n para compor a amostra. Somente com base em amostragens probabilísticas é que se podem realizar inferências ou induções sobre a população a partir do conhecimento da amostra (Ferreira, 2005).

Levantamentos probabilísticos reúnem todas aquelas técnicas que usam mecanismos aleatórios de seleção dos elementos de uma amostra, atri-

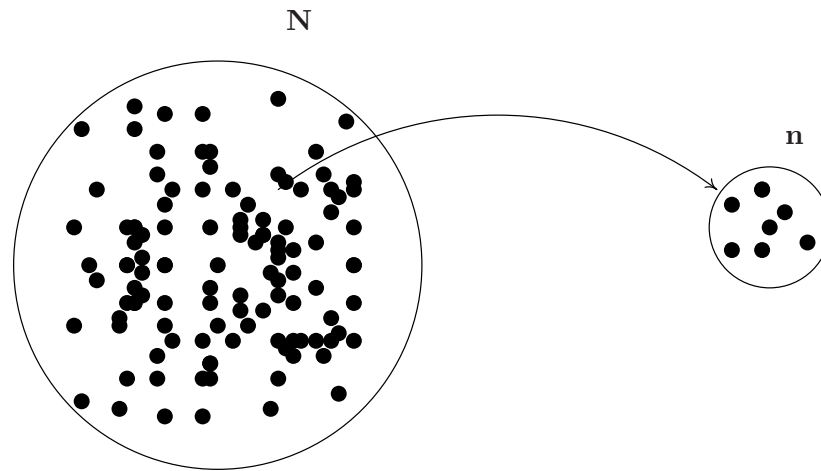


FIGURA 1: Esquema da retirada de uma amostra de n indivíduos de uma população de tamanho N .

buindo a cada um deles uma probabilidade, conhecida *a priori*, de pertencer à amostra (Bussab & Moretin, 2003).

Na amostragem probabilística, o tamanho da amostra é função:

- dos parâmetros a estimar;
- do nível de confiança desejável;
- do erro tolerável ou índice de precisão escolhidos; e
- do grau de dispersão da população;

podendo, ainda, depender do tamanho da população e de outros parâmetros específicos.

A amostragem probabilística permite, além da estimação pontual e do controle do erro, a estimação por intervalo (Tabela 1), ao qual se associa um grau de confiança desejável, ou seja, uma probabilidade de o verdadeiro

valor do parâmetro a estimar estar nele compreendido. É vantajosa a sua utilização, quando o objetivo é a obtenção de estimativas para a população (média, total, proporção e número de ocorrência), permitindo trabalhar com amostras de tamanho reduzido, em níveis de significância e confiabilidade adequados para assegurar a precisão desejada.

TABELA 1: Expressões para obtenção de intervalo de confiança e tamanho de amostra para população.

	População infinita	População finita
Intervalo de Confiança em função do		
Tamanho da Amostra	$Ic = \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$Ic' = \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$
Tamanho da amostra em função do		
intervalo de confiança	$n = \left(\frac{Z_{\frac{\alpha}{2}} \cdot \sigma}{Ic=d} \right)^2$	$n' = \frac{Z^2 \cdot \sigma^2 \cdot N}{d^2(N-1) + Z^2 \sigma^2}$

FONTE: Ferreira (2005)

N = Tamanho da população

$Z_{\frac{\alpha}{2}}$ = Valor de z para o nível de confiança $(1 - \alpha)$

n = Tamanho da amostra

n' = Tamanho da amostra corrigido para populações finita

Ic = Intervalo de confiança

Ic' = Intervalo de confiança corrigido para populações finita

Entre os tipos de amostragem probabilística os mais usados são: amostra aleatória simples, estratificada, sistemática e amostra aleatória por conglomerado.

2.1.1 Amostra Aleatória Simples (AAS)

Amostragem aleatória simples (AAS) é o método mais simples e mais importante para a seleção de uma amostra. Além de servir como um plano próprio, o seu procedimento é usado de modo repetido em procedimentos

de múltiplos estágios. Ele pode ser caracterizado através da definição operacional de uma lista com N unidades elementares, sorteiam-se com igual probabilidade n unidades, ou seja, se N for o tamanho da população, a probabilidade de cada elemento ser sorteado é $1/N$. (Bussab & Moretin, 2003).

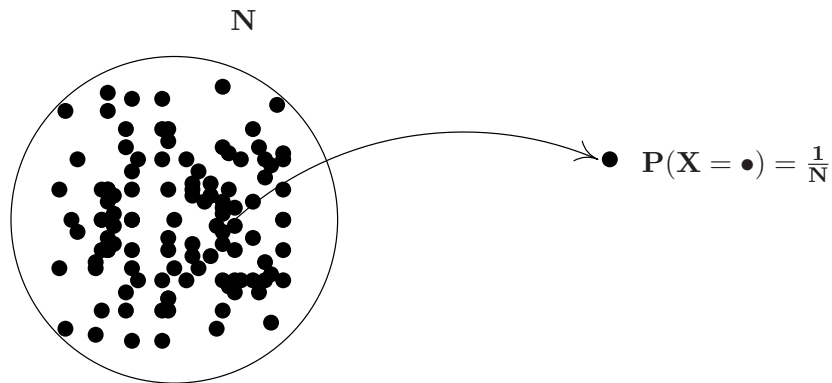


FIGURA 2: Esquema da retirada de um indivíduo, com probabilidade $1/N$, de uma população de tamanho N .

Uma amostra AAS não só dá a cada indivíduo igual chance de ser escolhido (evitando assim tendenciosidade na escolha) como também dá a toda amostra possível a mesma chance de ser extraída (David, 2000), conforme pode ser visto na Figura 2.

De acordo com Cochran (1977), o número possível de amostras de tamanho n que pode ser retiradas de uma população de tamanho N é dado por:

$$C_N^n = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Nesse tipo de amostragem, cada uma dessas combinações tem pro-

babilidade de $1/C_N^n$ de ser retirada. Na prática, cada elemento é amostrado por um processo aleatório que confere igual chance de ser sorteado a cada elemento da população. Sorteia-se um elemento e repete-se o processo para se selecionar o próximo elemento, dando sempre chances iguais para todos aqueles remanescentes na população. Repete-se tantas vezes esse procedimento até que a amostra de n elementos seja composta. A chance de um elemento ser sorteado num primeiro estágio é n/N ; a chance do segundo é $(n-1)/N-1$, e assim sucessivamente. Então, a chance de serem selecionados n elementos retiradas da população é:

$$\frac{n}{N} \cdot \frac{n-1}{N-1} \cdot \frac{n-2}{N-2} \cdots \frac{1}{N-n+1} = \frac{n!(N-n)!}{N!} = \frac{1}{C_N^n}$$

A partir do momento em que os elementos amostrados da população são removidos para as sucessivas retiradas subseqüentes, esse método é denominado de amostragem simples ao acaso sem reposição (Ferreira, 2005).

O processo de sorteio de uma amostra aleatória simples pode ser feito por meio de tabelas de números aleatórios, sorteio por funções de geradores de números aleatórios em programas de computadores, por uso de bolas enumerados em urnas ou papéis enumerado em algum tipo de recipiente. As tabelas de números aleatórios podem ser consideradas absoletas nos tempos atuais, devido à difusão dos computadores. O uso de papéis ou de bolas enumeradas em urnas ou saco não é operacionalmente satisfatório, principalmente se a população for muito grande. Inúmeros programas de computador e linguagens (compiladores) de programação possuem rotinas prontas para a geração de números aleatórios que são gerados em uma seqüência que só se repete após um ciclo superior à ordem dos trilhões de algarismo (Ferreira, 2005).

Determinação do Tamanho da Amostra

A determinação do tamanho da amostra n , contada por Bussab & Moretin (2003), é de tal forma que o estimador obtido tenha um erro máximo de estimação igual a ϵ , com determinado grau de confiança (probabilidade). De maneira mais específica, o problema consiste em determinar n , de modo que

$$P(|\bar{x} - \mu| \leq \epsilon) \simeq 1 - \alpha,$$

sendo em que ϵ equivale

$$\epsilon = Z_{\alpha} \frac{\sigma}{\sqrt{n}}.$$

Resolvendo em n , obtém-se

$$n = \left(\frac{Z_{\alpha} \sigma}{\epsilon} \right)^2.$$

Para a determinação do tamanho da amostra, é preciso fixar o erro máximo desejado (ϵ), com algum grau de confiança $1 - \alpha$ (traduzido pelo valor crítico z_{α}) e possuir algum conhecimento a priori da variabilidade da população (σ^2). Os dois primeiros são fixados pelo pesquisador e, quanto ao terceiro, a resposta exige mais trabalho. O uso de pesquisas passadas, “adivinhações” estatísticas, ou amostras piloto são os critérios mais usados (Bussab & Moretin, 2003)

2.1.2 Amostra Aleatória Estratificada (AAE)

Amostragem estratificada, (Figura 3), consiste na divisão de uma população em grupos (chamados estratos) segundo alguma(s) característica(s) conhecida(s) na população sob estudo, e de cada um desses estratos são selecionadas amostras em proporções convenientes (Tabela 2). A estratificação é usada principalmente para resolver questões como: melhoria da precisão

das estimativas; produção destas para toda população e subpopulações; de ordem administrativas, etc (Bussab & Moretin, 2005).

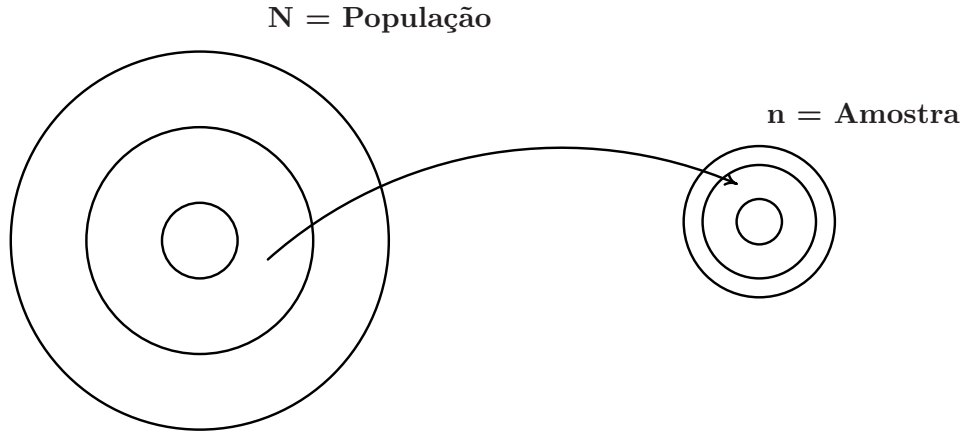


FIGURA 3: Esquema da retirada de uma amostra proporcional ao tamanho dos estratos de uma população de tamanho N .

TABELA 2: Representação esquemática de uma população estratificada, contendo os dados originais de cada estrato Y_h^* e seus respectivos totais (T_h), médias ($\mu_H = \bar{Y}_H$) e variâncias (σ_H^2 ou S_H^2).

Estrato	Dados	Total	Média	Variância
1	Y_1^*	τ_1	$\mu_1 = \bar{Y}_1$	σ_1^2 ou S_1^2
\vdots	\vdots	\vdots	\vdots	\vdots
h	Y_h^*	τ_h	$\mu_h = \bar{Y}_h$	σ_h^2 ou S_h^2
\vdots	\vdots	\vdots	\vdots	\vdots
H	Y_H^*	τ_H	$\mu_H = \bar{Y}_H$	σ_H^2 ou S_H^2

* sendo $Y_h' = (Y_{h1}, \dots, Y_{hN_h})$ é o vetor de dados no estrato, $h=1, \dots, H$.

FONTE: Bussab & Morettin (2003)

N_h : tamanho do estrato h ;

$$\tau_h \sum_{i=1}^{N_h} Y_{hi}: \text{total do estrado } h;$$

$$\mu_h = \bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} Y_{hi}: \text{ média do estrato } h;$$

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \mu_h)^2: \text{ variância do estrato } h;$$

$$\sigma_h^2 = \frac{1}{N_h} \sum_{i=1}^{N_h} (Y_{hi} - \mu_h)^2: \text{ variância do estrato } h;$$

$$N = \sum_{h=1}^H N_h: \text{ tamanho do universo};$$

$$W_h = \frac{N_h}{N}: \text{ peso (proporção) do estrato } h, \text{ com } \sum_{h=1}^H W_h = 1$$

$$\tau = \sum_{h=1}^H \tau_h = \sum_{h=1}^H \sum_{i=1}^{N_h} Y_{hi} = \sum_{h=1}^H N_h \mu_h: \text{ total populacional};$$

$$\mu = \bar{Y} = \frac{\tau}{N} = \frac{1}{N} \sum_{h=1}^H \sum_{i=1}^{N_h} Y_{hi} = \frac{1}{N} \sum_{h=1}^H N_h \mu_h = \sum_{h=1}^H W_h \mu_h: \text{ média populacional};$$

de modo que a média global é a média ponderada dos estratos (Busab & Moretin, 2003).

De acordo com Ferreira (2005), os valores paramétricos são definidos para o estrato h e para toda a população conforme observado na Tabela 2. Assim, a equação 2.1 refere-se à média populacional, e a equação 2.2 à variância populacional do estrato h

$$\mu_h = \frac{\sum_{i=1}^{N_h} X_{hi}}{N_h} \tag{2.1}$$

$$\sigma_h^2 = \frac{1}{N_h} \left[\sum_{i=1}^{N_h} X_{hi}^2 - \frac{\left(\sum_{i=1}^{N_h} X_{hi} \right)^2}{N_h} \right] \quad (2.2)$$

Os estimadores da média e da variância do estrato h são representados nas equações 2.3 e 2.4, respectivamente,

$$\bar{X}_h = \frac{\sum_{i=1}^{n_h} X_{hi}}{n_h} \quad (2.3)$$

$$S_h^2 = \frac{1}{n_h - 1} \left[\sum_{i=1}^{n_h} X_{hi}^2 - \frac{\left(\sum_{i=1}^{n_h} X_{hi} \right)^2}{n_h} \right] \quad (2.4)$$

Aumentando-se o tamanho da amostra, o erro padrão diminui. Se a população é muito heterogênea e as razões de custo limitam o aumento da amostra, torna-se impossível definir uma AASc da população toda com uma precisão razoável. Uma saída para esse problema é dividir a população em subpopulações internamente mais homogêneas, ou seja, grupos com variância σ^2 pequena que diminuirão o erro amostral global.

O resultado será mais eficaz quanto maior for a habilidade do pesquisador em produzir estratos homogêneos. O caso limite é aquele em que se consegue a homogeneidade máxima (variância nula) dentro de cada estrato. A simples estratificação por si só não produz necessariamente estimativas mais eficientes do que a AAS. A execução de um plano de amostragem estratificada (AE) exige os seguintes passos:

- i. divisão da população em sub-populações bem definidas (estratos);

- ii. de cada estrato retira-se uma amostra, usualmente independente;
- iii. em cada amostra, usam-se estimadores convenientes para os parâmetros do estrato;
- iv. monta-se para a população um estimador, combinando os estimadores de cada estrato, e determinam-se suas propriedades.

2.1.3 Amostragem Aleatória por Conglomerado (AAC)

De acordo com Bussab & Moretin (2003), quando os sistemas de referência não são adequados e o custo de atualizá-los é muito alto, ou ainda, quando a movimentação para identificar as unidades elementares no campo é cara e consome muito tempo, a tarefa amostral pode ser facilitada se forem selecionados grupos de unidades elementares, os chamados conglomerados. Por exemplo, uma amostra de eleitores pode ser obtida pelo sorteio de um número de domicílios; trabalhadores por uma amostra de empresas, ou estudantes por uma amostra de escolas ou classes. O que caracteriza bem o planejamento amostral de conglomerado é que a unidade amostral contém mais de um elemento populacional.

Quando os sistemas de referência não são adequados e o custo de atualizá-los é muito elevado, ou ainda, quando a movimentação para identificar as unidades elementares no campo é cara e consome muito tempo, a tarefa amostral pode ser facilitada se forem selecionados grupos de unidades elementares, os chamados conglomerados. O que caracteriza bem o planejamento amostral de conglomerados é que a unidade amostral contém mais de um elemento populacional.

Para determinar a renda média familiar de moradores de uma grande cidade, dificilmente se dispõe de uma lista de famílias, a unidade elementar

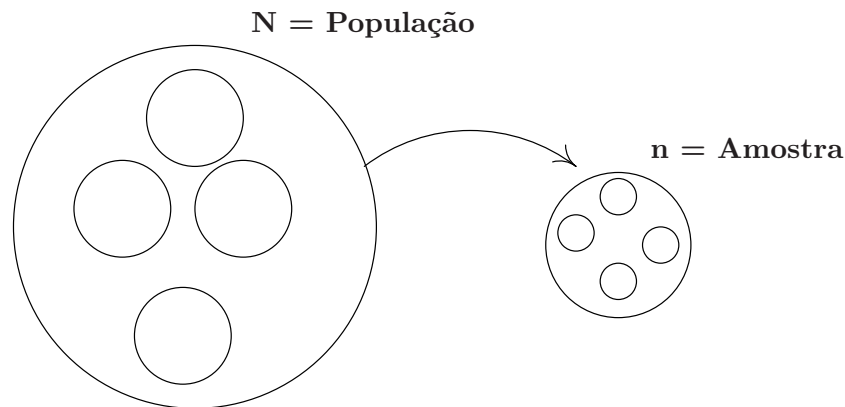


FIGURA 4: Esquema da retirada de uma amostra aleatória por conglomerado de uma população de tamanho N .

da população de interesse. Com alternativa, pode-se usar como sistema de referência a lista de Setores Censitários (SC) do IBGE. Um SC é uma área bem definida com cerca de 300 domicílios e são usados para fazer o recenseamento a cada 10 anos. Pode-se começar sorteando-se um certo número de SC, e a cada SC sorteado seriam sorteados quarteirões, dos quais sorteiavam-se os domicílios. Este é um plano amostral de conglomerados em três estágios (Bussab & Moretin, 2003).

Uma das inconveniências para o uso da amostragem de conglomerados prende-se ao fato de que as unidades, dentro de um mesmo conglomerado, tendem a ter valores parecidos em relação às variáveis que estão sendo pesquisadas, e isso torna estes planos menos eficientes (Bussab & Moretin, 2003).

Quando os elementos da população são reunidos em grupos que são sorteados para compor a amostra, o processo é denominado de amostragem por conglomerado (Ferreira, 2005). A razão de se usar um tipo de amostragem como esse é principalmente motivada por critério de ordem prática.

Dentre esses critérios destaca-se a ausência de uma listagem de todos os elementos populacionais (Ferreira, 2005).

Algumas populações não permitem, ou tornam extremamente difícil que se identifiquem seus elementos. Não obstante isso, pode ser relativamente fácil identificar alguns subgrupos da população. Em tais casos, uma amostra aleatória simples desses subgrupos (conglomerados) pode ser colhida, e uma contagem completa deve ser feita para o conglomerado sorteado. Agregados típicos são quarteirões, famílias, organizações, agências, edifícios, etc (Ferreira, 2005).

Em uma perspectiva sociológica, que valoriza o papel das relações e divisões sociais de classe, raça e gênero, a construção de "narrativas causais" deve incorporar o entendimento das interseções e interações causais, entre classe social e essas outras relações. As diversas dimensões da desigualdade social não podem ser reduzidas à desigualdade de classe, porém as relações de classe, ainda assim, jogam um papel decisivo na modelagem das demais formas de desigualdade (Wright, 1978). Erik Olin Wright propõe duas teses básicas para o estudo dos efeitos conjuntos de classe e raça na explicação de padrões sociais.

A primeira tese, chamada de mecanismos distintos, considera que classe e raça representam diferentes formas de divisão social e identificam tipos distintos de mecanismos causais, de modo que uma categoria não pode ser dissolvida na outra, como se não existissem efeitos independentes de ambas. A segunda tese, denominada de interação estrutural, considera que esses distintos mecanismos interagem no mundo social, pois a realidade não possui uma conformação meramente aditiva, de modo que o efeito da raça pode depender, em parte, da classe (Wright, 2002).

A formação dos grupos raciais, ao envolverem características comuns de comportamento imputado, pode vincular-se fortemente aos contextos e às

motivações de classe e de grupos de status. Raça representa, por si mesma, uma fonte potencialmente importante de unidade dentro do grupo e de divisão intra grupo, mas esse potencial pode requerer um conteúdo estrutural para ser ativado. Da mesma maneira, a conexão da raça com os fatores políticos e econômicos projeta o seu papel em relação às classes e seus conflitos, aos sistemas de Estado e à formação dos grupos de status (Rex, 1986 apud Santos, 2005, p. 16-17 e p. 35- 36). Os sistemas de crenças raciais influenciam os padrões de relações sociais, que vêm a constituir as relações raciais, mas esses sistemas de crenças, por sua vez, dependem de estruturas subjacentes, de amplitude limitada, que devem ser investigadas (Rex, 1983). As desigualdades de raça são, não apenas distintas, mas também diferentes em seus modos de operação social das desigualdades de classe, porém tenderiam a assumir uma expressão maior como desigualdades de classe. Nesse sentido, elas operariam, em grande parte, ainda que não apenas, por intermédio da colocação dos não-brancos em posições inferiores aos brancos na ordem da produção e da distribuição. As desigualdades de classe constituem estruturas fundamentais por meio das quais as desigualdades distintas de raça são articuladas (Westergaard, 1995 apud Santos, 2005, p. 144-147).

2.1.4 Amostragem Aleatória Sistemática (AS)

Trata se de uma variação de amostragem aleatória simples, conveniente quando a população está ordenada segundo algum critério, como fichas em um fichário, listas telefônicas, etc (Ferreira, 2005).

É um tipo de amostragem cujo plano é obtido por um critério pelo qual intervalos de mesmo tamanho entre unidades da amostra são tomados até compor uma amostra do tamanho n e toda a extensão da localização física da população alvo. Para implementar esse sorteio, os N elementos populacionais são tomados a cada $K = N/n$ elementos. O primeiro elemento

deve ser sorteado entre os K primeiros. Se, por exemplo, uma população de $N = 10.000$ elementos é considerada e deseja-se extrair uma amostra de tamanho $n = 500$, então k será de $10.000/500 = 20$. Assim, se o elemento 11 for o primeiro a ser sorteado entre os 20 primeiros, a amostra fica determinada da seguinte forma: 11, 31, 51, e assim por diante conforme ilustração da Figura 5. A seleção da primeira unidade define a amostra toda (Ferreira, 2005)



FIGURA 5: Esquema de uma população disposta em forma seqüencial e, em negrito, estão os elementos a serem amostrados.

Considere uma população com N elementos, na qual $N = Kn$ e K é um número inteiro. Considere também que a população está ordenada de 1 a N , formando o sistema de referências. Uma unidade é então selecionada aleatoriamente (segundo a AAS) entre as K primeiras unidades do sistema de referências. As unidades seguintes farão parte da amostra que serão obtidas a partir da primeira unidade selecionada em intervalos de comprimento K (Bussab & Moretin, 2003).

A vantagem principal da amostra sistemática (AS) é a facilidade de sua execução. Também, é bem menos sujeita a erros do entrevistador que os outros esquemas de amostragem. Por outro lado, quanto à sua precisão, existem situações em que ela é mais precisa que a AAS no entanto, na maioria dos casos, a sua eficiência é próxima da AAS, principalmente quando o sistema de referências está numa “ordem aleatória” (Bussab & Moretin, 2003)

2.2 Regressão

Os métodos que estabelecem relações lineares para os parâmetros de um modelo entre duas ou mais variáveis são denominados de métodos de regressão linear.

2.2.1 Regressão linear simples

Se for considerado o modelo funcional apresentado de forma geral na equação 2.5, então, o modelo de regressão é denominado de linear simples (Ferreira, 2005).

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.5)$$

Em que, X_i é denominado de variável independente e ε_i é o erro não observável associado a i -ésima observação. Para se estimar os parâmetros desse modelo (β_0 e β_1), que são o intercepto ou coeficiente linear e o coeficiente angular ou coeficiente de regressão, nenhuma pressuposição é feita acerca da distribuição da variável aleatória (ε_i). Todavia, se inferências forem realizadas, esses erros deverão assumir um modelo específico de probabilidade e devem ser homocedásticos (Ferreira, 2005).

O objetivo na regressão linear é determinar estimadores de β_0 e β_1 de tal forma que as distâncias médias entre a reta de regressão e os valores observados sejam minimizadas. Definindo $\varepsilon_i = Y_i - \beta_0 - \beta_1 X_i$, a partir do modelo 2.5, então, um critério interessante para determinar esses estimadores seria minimizar as somas de quadrados desses resíduos ao longo de todos os n pares (X_i, Y_i) . Essa soma de quadrados é dado por $\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$. Esse método é denominado de quadrados mínimos e é facilmente obtido derivando essa última equação em relação aos parâmetros β_0 e β_1 e igualando essas derivadas a zero (Ferreira, 2005).

O estimador é uma função qualquer dos elementos da amostra. Esse tipo de função real retorna um único valor para o parâmetro θ desconhecido a partir de uma amostra. Esse tipo de processo é conhecido por estimação pontual ou estimação por ponto de θ . A questão agora é conectar a informação do processo de estimação pontual com a distribuição amostral do estimador. O propósito dessa abordagem é permitir que se especifique um procedimento de estimação em que se precise um intervalo, o qual possua uma elevada probabilidade de conter o verdadeiro valor do parâmetro desconhecido. Esse tipo de procedimento é conhecido como estimação por intervalo (Ferreira, 2005).

Métodos de obter estimadores

Outros métodos podem ser igualmente apresentados, como, por exemplo, o método da máxima verossimilhança, o qual assume normalidade para a distribuição de ε_i , e o método das médias condicionais, que assume normalidade bivariada para a distribuição de (X_i, Y_i) . O método dos momentos centrados na média (μ_r) sendo que o índice r da expressão é a ordem do momento. Assim, para $r = 1$ tem-se o momento de primeira ordem, o qual é sempre igual a zero; para $r = 2$ o momento de ordem 2, que é a variância da população; para $r = 3$ o momento de assimetria de ordem 3; para $r = 4$ o momento de curtose de ordem 4; e assim por diante. O método dos quadrados mínimos é o menos restrito, muito embora seja necessário presumir normalidade para a distribuição dos resíduos, se houver interesse em realizar inferência (Ferreira, 2005).

Método dos quadrados mínimos

Além da normalidade citada anteriormente, outras pressuposições devem ser assumidas. A primeira delas é que a variável X é por hipótese

controlada e não está sujeita a variações aleatórias. Diz-se que X é uma variável fixa (ou sem erro ou determinística). Segundo, para dado valor x de X , os erros distribuem-se ao redor da média $\alpha + \beta x$ com média zero, isto é,

$$E(e_i | x) = 0 \quad (2.6)$$

Em terceiro lugar, deve-se supor que os erros tenham a mesma variabilidade em torno dos níveis de X , ou seja,

$$\text{var}(e_i | x) = \sigma_e^2 \quad (2.7)$$

E em quarto lugar, introduz-se a restrição de que os erros sejam não-correlacionados. Colhida uma amostra de n indivíduos, tem-se n pares de valores (x_i, y_i) , $i = 1, \dots, n$, que devem satisfazer ao modelo $y_i = E(Y | x_i) + e_i = \alpha + \beta x_i + e_i$ $i = 1, 2, \dots, n$, isto é,

$$y_i = \alpha + \beta x_i + e_i \quad i = 1, 2, \dots, n \quad (2.8)$$

Tem-se, então, n equações e $n + 2$ incógnitas $(\alpha, \beta, e_1, e_2, \dots, e_n)$. Precisa-se introduzir um critério que permita encontrar α e β . Vem-se adotar o critério que consiste em encontrar os valores de α e β que minimizam a soma dos erros, dados por

$$e_i = y_i - (\alpha + \beta x_i), i = 1, \dots, n. \quad (2.9)$$

Obtem-se, então, a quantidade de informações perdida pelo modelo ou a soma dos quadrados dos erros (ou desvios)

$$SQ(\alpha, \beta) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n \{y_i - (\alpha + \beta x_i)\}^2 \quad (2.10)$$

Para cada valor de α e β tem-se um resultado para essa soma de quadrados, e a solução de mínimos quadrados (MQ) é aquela que torna essa soma

mínima. Tem-se então, o problema de encontrar o mínimo de uma função de duas variáveis, α e β . Derivando em relação a α e β e igualando a zero, observa-se que as soluções $\hat{\alpha}$ e $\hat{\beta}$ devem satisfazer

$$\begin{aligned} n\hat{\alpha} + \hat{\beta} \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ \hat{\alpha} \sum_{i=1}^n x_i + \hat{\beta} \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i \end{aligned} \quad (2.11)$$

As quais produzem as soluções

$$\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta}\bar{x}, \\ \hat{\beta} &= \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \end{aligned} \quad (2.12)$$

Substituindo em $E(Y | x) = \mu(x) = \alpha + \beta x$, obtem-se o estimador para a média $\mu(x)$, dado por

$$\hat{\mu}(x_i) = \hat{\alpha} + \hat{\beta}x_i \quad i = 1, \dots, n, \quad (2.13)$$

que será indicar por

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i \quad (2.14)$$

ou, ainda, por

$$\hat{y}_i = \bar{y} - \hat{\beta}\bar{x} + \hat{\beta}x_i = \bar{y} + \hat{\beta}(x_i - \bar{x}). \quad (2.15)$$

2.2.2 Regressão linear múltipla

É o caso de regressão em que é sabido que K variáveis contribuem para a explicação de um fenômeno. A parte funcional da regressão é uma função linear (Guerra & Donaire, 1991):

$$Y = f(X_1, X_2, \dots, X_k) + \epsilon \quad (2.16)$$

em que: $f(X_1, X_2, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$.

A estimativa dessa equação de regressão será dada por:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k \quad (2.17)$$

em que as estimativas $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ dos coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ podem ser calculadas pelo método dos mínimos quadrados, partindo de hipóteses análogas àquelas adotadas para regressão linear simples.

O cálculo dos estimadores é feito a partir do seguinte sistema de equações:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + \dots + \beta_k X_{k1} + e_1 \\ y_2 &= \beta_0 + \beta_1 X_{12} + \beta_2 X_{22} + \dots + \beta_k X_{k2} + e_2 \\ &\vdots \\ y_n &= \beta_0 + \beta_1 X_{1n} + \beta_2 X_{2n} + \dots + \beta_k X_{kn} + e_3 \end{aligned}$$

Esse sistema de equações pode ser reescrito na forma matricial como:

$$Y = X\beta + \epsilon \quad (2.18)$$

cujo estimador de quadrados mínimos para o vetor β , é obtido por

$$\hat{\beta} = (X'X)^{-1}X'Y. \quad (2.19)$$

2.2.3 Seleção de modelos

Neste tópico discutem-se alguns procedimentos estatísticos específicos para a seleção de variáveis na regressão, baseados na obra de Draper & Smith (1998). Suponha que deseje-se estabelecer um modelo de regressão linear para uma dada variável resposta Y em termos das variáveis preditoras básicas X_1, \dots, X_k . Suponha ainda que Z_1, \dots, Z_r são todas as funções de um ou mais X s, ou seja, são o conjunto completo de variáveis de onde o modelo será escolhido já que esse conjunto contém qualquer função, tal como quadrados, interações, logaritmos, inversas, potências, que se possa desejar ou necessitar. Dois critérios de seleção de modelos, aparentemente contraditórios, são normalmente considerados:

1. Para tornar o modelo útil para fins preditivos poderia se desejar que o modelo contivesse o máximo de variáveis possíveis para manter os erros com viés pequenos, para que valores ajustados confiáveis pudessem ser determinados.
2. Por outro lado: (a) para manter a variância dos preditores razoavelmente pequena (lembrando que a variância média de \hat{Y}_i é $p\sigma^2/n$, em que p é o número de parâmetros no modelo e n é o número de observações), e (b) devido aos custos envolvidos na obtenção da informação em um grande número de variáveis e depois para monitorá-los, deve-se preferir que a equação inclua o mínimo possível de variáveis.

A implicação prática desses dois extremos é o que é comumente chamado de *seleção do melhor modelo de regressão*. Não há um único procedimento estatístico para este fim. Se se conhecesse a magnitude de σ^2 (a variação aleatória real das observações) para qualquer problema bem definido, a escolha do melhor modelo de regressão seria muito mais fácil.

Infelizmente, esta situação ocorre raramente, então uma grande dose de julgamento pessoal será parte necessária de qualquer dos métodos discutidos. Serão discutidos alguns procedimentos, cada um tendo suas justificativas. Para aumentar a confusão, eles não levam necessariamente a mesma conclusão quando aplicados em um mesmo problema, embora atinjam a mesma resposta para muitos problemas.

2.2.3.1 Stepwise

O procedimento de regressão Stepwise passo a passo começa pela escolha de uma equação contendo a melhor variável X (sozinha) e então tenta-se construir um modelo com várias adições de variáveis expressivas ou independentes, uma de cada vez, enquanto essas adições forem significativas. A ordem da adição é determinada pelo uso dos valores do teste F parcial para se selecionar qual variável deve entrar em seguida. O maior valor de F parcial é comparado com um valor de F a entrar (selecionado ou default). Após uma variável ser adicionada, a equação é examinada para ver se alguma variável deve ser deletada.

O procedimento básico é como segue: Primeiro seleciona-se o Z mais correlacionado com Y (suponha que seja Z_1) e encontra-se a equação de regressão linear de primeira ordem $\hat{Y} = f(Z_1)$. Checa-se se essa variável é significativa. Se não for, desiste-se e adota-se o modelo $Y = \bar{Y}$ como sendo o melhor; caso contrário, busca-se pela segunda variável preditora a entrar no modelo de regressão.

Examinam-se os valores de F parcial de todas as variáveis preditoras que não estão na regressão. A Z_j com o maior valor (por exemplo, Z_2) é agora selecionada e uma segunda equação de regressão $\hat{Y} = f(Z_1, Z_2)$ é ajustada. Checa-se a significância da regressão total, o aumento do R^2 é avaliado, e os valores de F parcial para *ambas as variáveis que estão*

agora na equação (e não apenas aquela que acabou de entrar) são examinados. O menor desses dois F's parciais é então comparado com um quantil apropriado de F, F-de-saída, e a variável preditora correspondente é retida na equação ou rejeitada, sendo o teste significativo ou não. Esse teste do "preditor" menos útil atualmente na equação é realizado em todo passo do procedimento Stepwise. Um preditor que pode ter sido o melhor candidato num estágio anterior pode, no estágio seguinte, ser supérfluo devido a sua relação com as outras variáveis, agora no modelo. Para checar isso, o valor do F parcial para cada variável na regressão a todo momento é calculado, e o menor desses F's parciais (que pode ser associado com a última entrada ou entradas anteriores) é então comparado com um quantil apropriado da distribuição F ou com um valor default de F. Isso promove um julgamento da contribuição da variável menos valiosa na regressão naquele momento, independente se ela foi a última a entrar no modelo ou se entrou em qualquer outro momento. Se a variável testada gerar uma contribuição não-significativa, ela é removida do modelo e a equação de regressão ajustada apropriada é então computada para todas as demais variáveis que permaneceram no modelo.

A melhor das variáveis que não está no modelo (isto é, aquela cujo valor de F parcial, dadas as preditoras já no modelo, é o maior) é então checada para ver se consegue passar pelo teste do F parcial de entrada. Se passar, é adicionada, e volta-se a testar todos os F's parciais das que estão no modelo. Se não passar, é feita mais uma tentativa de remoção. Eventualmente, quando nenhuma variável no modelo atual pode ser removida e a próxima melhor candidata não consegue garantir seu lugar na equação, o processo pára. À medida que cada variável é adicionada na regressão, o seu efeito no R^2 , o quadrado do coeficiente de correlação múltipla, é geralmente calculado e mostrado pelo programa.

Geralmente, é aconselhado que se escolha o mesmo nível de significância para os testes de entrada e saída. Se for escolhido um α menor para a saída do que para a entrada, um padrão recursivo pode ocorrer. O uso de um α maior para o teste de saída deixa o teste conservativo e pode implicar em variáveis cujas contribuições dificilmente serão retidas. Alguns pesquisadores acham que essa é uma característica desejável; isso é uma questão de preferência pessoal.

Draper & Smith (1998) dizem que esse é um dos melhores procedimentos de seleção de variáveis e recomendam seu uso, pela economia de recursos computacionais, o que evita trabalhar com mais X's do que o necessário; a medida que facilitam a equação a cada passo. Entretanto, dizem que o procedimento Stepwise pode ser usado exageradamente por estatísticos amadores. Como todo procedimento, o Stepwise também exige um julgamento criterioso e exame crítico do modelo por meio da averiguação de seus resíduos. “Existe uma tendência muito grande de colocar o software no piloto automático”, dizem Draper & Smith (1998).

2.2.3.2 Forward

Um procedimento mais simples, e menos eficiente, no qual apenas a última entrada no modelo é testada, é chamado *procedimento de seleção Forward*. Esse procedimento é ainda uma opção em muitas rotinas computacionais. O Forward garante que as variáveis que entram no modelo não sejam mais removidas, o que pode ser desejável em algumas aplicações específicas.

2.2.3.3 Backward

O método de eliminação Backward é também um procedimento econômico que *tenta* examinar apenas as melhores regressões contendo um certo

numero de variáveis. Os passos básicos nesse procedimento são:

1. A equação de regressão contendo todas as variáveis é computada.
2. O valor de F parcial é calculado para cada variável preditora *tratada como se tivesse sido a última a entrar no modelo de regressão*.
3. O menor valor de F parcial, digamos, F_L , é comparado com um nível de significância pré-selecionado ou default, digamos F_0 .
 - a. Se $F_L < F_0$, remove-se a variável Z_L , que gerou F_L , de consideração e recalcula-se a equação de regressão com as variáveis restantes; volta-se ao estágio (2).
 - b. If $F_L > F_0$, adota-se a equação de regressão como está.

De acordo com Draper & Smith (1998), esse é um bom procedimento, especialmente para aqueles estatísticos que gostam de ver todas as variáveis na equação para “não perder nada”. É muito mais econômico em termos computacionais e de pessoal do que o método de todas as regressões. Entretanto, se os dados de entrada gerarem uma matrix $X'X$ com ma condição, ou seja, aproximadamente singular, então a equação superparametrizada pode não fazer sentido devido aos erros de arredondamento. Com as rotinas modernas de inversão de matrizes, esse não é geralmente um problema sério. Pode-se reconhecer que, uma vez que uma variável seja eliminada nesse processo, ela não retorna para o modelo. Então, todos os modelos alternativos, usando-se as variáveis eliminadas, não são passíveis de avaliação.

2.2.4 Qualidade de ajuste

Um dos objetivos da regressão é desenvolver a equação que permitirá ao investigador fazer previsões dos valores da variável aleatória Y . Para isso,

é necessário ajustar a equação, ou seja, os valores dos parâmetros do modelo β_0 e β_1 e da variância residual σ^2 , consistentes com os dados disponíveis determinados. Um dos métodos que pode ser utilizado para isso é o dos quadrados mínimos (Ferreira, 2005).

Para verificar se um modelo é adequado, tem que investigar se as suposições feitas para o desenvolvimento do modelo estão satisfeitas. Para tanto, estuda-se o comportamento do modelo usando o conjunto de dados observados, notadamente as discrepâncias entre os valores observados e os valores ajustados pelo modelo, ou seja, faz-se uma análise de resíduos (Bussab & Moretin, 2003).

De acordo com Bussab & Moretin (2003), o i -ésimo resíduo é dado por

$$\hat{\epsilon}_i = y_i - \hat{y}_i \quad (i = 1, 2, \dots, n).$$

Existem várias técnicas formais para conduzir essa análise (Draper & Smith, 1998).

A análise de variância é uma forma de dividir a variância total em componentes devidos à regressão linear e ao resíduo. Essa partição é obtida para atender a uma série de objetivos. Um desses objetivos é verificar se a parte da variação total explicada pelo modelo é significativamente diferente de zero, nesse caso, pode se utilizar o teste F para a hipótese $H_0 : (\beta_1) = 0$. Assim, a estatística obtida por $F_c = \frac{QMReg}{QMRes}$ possui distribuição F (Ferreira, 2005), conforme a Tabela 3.

Outra forma de avaliar a qualidade do modelo ajustado é o coeficiente de determinação (R^2), cuja referência à quantidade da variação total de Y é explicada pelo modelo de regressão. Uma medida complementar é o que explica o coeficiente de não-determinação ($K^2 = 1 - R^2$) (Ferreira, 2005).

TABELA 3: Tabela de análise de variância para um modelo de regressão linear simples contendo as fontes de variação com os respectivos graus de liberdade (GL), somas de quadrados (SQ), e quadrados médios (QM) e a estatística do teste F para a hipótese de nulidade do coeficiente de regressão

Fonte de variação	(GL)	(SQ)	(QM)	F
Regressão	1	SQ Regressão	QM Regressão	$F_c = \frac{QM_{Reg}}{QM_{Res}}$
Resíduo	$V = n - 2$	SQ Resíduo	QM Resíduo	
Total	$n - 1$	SQ Total		

FONTE: Draper & Smith (1998)

De acordo com Schlich (1995), para se testar se o R^2 foi significativo, ou seja, se o modelo em questão é aceitável ou não, pode ser usada a estatística

$$F_c = \frac{(n - p) R^2}{p(1 - R^2)}, \quad (2.20)$$

que, sob H_0 : o modelo não se ajusta bem aos dados, tem distribuição F com $n - p$ e p graus de liberdade.

Outra medida da qualidade do ajuste de um modelo é o Erro Quadrático Médio (EQM). Essa grandeza expressa a média dos quadrados dos desvios entre as observações (y_i) e as estimativas das mesmas (\hat{y}_i), geradas pelo modelo em questão:

$$EQM = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}, \quad (2.21)$$

em que n representa o tamanho amostral.

Para facilitar as comparações, a raiz quadrada do EQM muitas vezes é utilizada por deixar tal quantidade na mesma unidade dos dados. Parti-

cularmente nesse estudo, utilizou-se da raiz do EQM na estimação da renda e na estimação da classe econômica.

A raiz do EQM-renda tem o objetivo de comparar o desempenho daqueles modelos que estimam renda, ou seja, os modelos de regressão múltipla objetivo e híbrido. Ele se expressa em unidades monetárias (reais):

$$\sqrt{EQM_{renda}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}. \quad (2.22)$$

Por sua vez, a raiz do EQM-classe, expressa em termos percentuais, mede o erro cometido, em média, em classe pelo modelo em questão. Essa quantidade tem por objetivo comparar os modelos do Critério Brasil, modelo objetivo e modelo híbrido na estimação da classe social das famílias,

$$\sqrt{EQM_{classe}} = \sqrt{\frac{\sum_{i=1}^k (fr_i - \hat{fr}_i)^2}{k} \left(\frac{k}{n}\right)}, \quad (2.23)$$

em que fr_i é a frequência relativa real da classe i ; \hat{fr}_i é a frequência relativa da classe i estimada pelo modelo em questão; k é o número de classes sociais existentes (7) e n é o tamanho amostral.

2.3 Pesquisa de opinião pública

Uma pesquisa de opinião pública tem por finalidade investigar e retratar as atitudes e opiniões de indivíduos da população, de modo que esta seja conhecida. Também, permite o planejamento de projetos orientados à necessidade de cada cliente, tanto quantitativos quanto qualitativos, atendendo a prazos cada vez mais exíguos, com alto nível de precisão, e com menor custo possível.

A opinião pública tem sua raiz no social e, portanto, é socializadora, pois alimenta-se de suas relações. Para se levantar a opinião numa dada situação, é necessária uma avaliação histórica e atual dos fatores que envolvem a formação desta opinião (fase dialética), pois o conhecimento é o resultado de uma postura reflexiva que agrega valores subjetivos e concretos. No entanto, os métodos quantitativos representados pelos momentos funcionalistas da pesquisa são relevantes para dar credibilidade (consistência) a um trabalho reflexivo, uma vez que o mercado entende a realidade quantificável.

O conhecimento é apreendido somente se for apresentado por caminhos que lhe dêem segurança e condição de interação. Portanto, querer conhecer a opinião de uma comunidade sobre um assunto controverso é dar-lhe a oportunidade de quantificação de seus valores subjetivos.

Uma característica que vale ser referendada na pesquisa de opinião pública, como técnica, é que ela deve ser um instrumento de trabalho de Relações Públicas; não um simples instrumento de coleta de dados, mas de mudança. A importância de um modelo dialético para se pesquisar a opinião pública está em preparar pesquisadores conscientes de que este é um processo de transformação social (César et al., 2007).

2.3.1 Renda, classe econômica e classe social

O Census Bureau (2000) define renda como o dinheiro de ordenados e salários, assim como o de pagamentos de juros e previdência social.

As medidas oficiais de renda não incluem outros tipos de compensação como benefícios do empregador ou do governo. Tentar estimar os valores em dinheiro destes benefícios é difícil, e eles aumentam substancialmente a renda do consumidor. Mas deixar de incluí-los subestima mudanças

de renda em anos recentes.

O consumo total ou o potencial de mercado pode ser identificado ao analisar renda e sua alocação para a categoria de um produto por segmentos demográficos. É por isso que os profissionais de marketing dão tanta ênfase em saber quem tem poder de compra e como estes gastam seu dinheiro (Engel et al., 1995).

No mundo inteiro, a grande maioria dos consumidores é de renda baixa. No Canadá, Europa, Estados Unidos e outros países industrializados, a maioria dos clientes é de renda média, mas mesmo nestes países o número de consumidores no “mercado inferior” é substancial. Mesmo lojas que têm apelo para o mercado inferior precisam ser atraentes, ter estilo e tratar as pessoas com respeito. Standford Goodkin (apud Engel et al., 1995), explicou: “Ninguém quer ser lembrado de que não é rico. Lojas de descontos de sucesso ficaram famosas ao convencer seus clientes de que eles são inteligentes e especiais, não gentalha pobre”.

A divisão em classes sociais gera estruturas que impactam na vida das pessoas e na dinâmica das instituições. Apesar de seu objetivo ser apenas o de “didaticamente” segmentar as pessoas que vivem em uma sociedade em classes ou estratos que possuem características comuns, a existência de tais rótulos constantemente é encarada como constrangedora por parte dos cidadãos.

A Escala Chapin: O status social foi descrito por Francis Stuart Chapin, professor de sociologia norte-americano como a posição que o indivíduo ou a família ocupa com referência ao padrão médio predominante de posses culturais, renda efetiva, posses materiais e participação nas atividades da comunidade (Chapin, 1933). Num esforço para deduzir valores quantitativos úteis para o status social da forma como havia definido, Chapin construiu uma escala baseada nos equipamentos existentes na sala-de-estar

de residências urbanas (Fauze, 1997) .

Uma das escalas mais antigas de definição de status social das famílias foi baseada nos equipamentos existentes na sala de estar das residências no início dos 1930. Desenvolvido no intervalo de duas Grandes Guerras, pelo professor de sociologia norte-americano Francis Stuart Chapin, o método de estratificação social, muito criticado posteriormente - concedia mais pontos àqueles que tinham a melhor sala de visitas.

A presença de lareira, tapetes de bom tamanho, luz elétrica e piano, por exemplo, aumentava a pontuação total da família na análise geral. Excesso de pó ou mobília com manchas faziam com que a casa fosse preterida - artigos espalhados ou em desordem levavam a casa a perder dois pontos. Móveis consertados, menos dois pontos também, e uma sala "bizarra, chocante, desarmoniosa ou agressiva", pior ainda: quatro pontos eram tirados do total.

Descartado pelos estudiosos nos anos 40 - por retratar apenas o nível de conforto das famílias, e não escolaridade, por exemplo, ou o efeito da renda - esse modelo de estratificação social ainda foi testado nos EUA em 1942. Uma pequena amostra de 67 famílias negras em Minneapolis participou do teste, e chegou-se à conclusão de que era preciso fazer algumas alterações. Logo depois, caiu em desuso (Gomes, 2007).

A escala desenvolvida e testada por Chapin compreendia 21 ítems de avaliação, sendo 17 referentes aos equipamentos (ou ítems) encontrados na sala-de-jantar e quatro referentes à avaliação dos ítems. Para cada item de avaliação, havia diferentes pesos. A pontuação assim obtida refletiria o status social da família e seria resultante do efeito conjunto de uma somatória de variáveis intervenientes no processo de estratificação.

Nos anos seguintes, outros modelos apareceram. Em 1940, a revista

Fortune resolveu fazer um levantamento com 5207 pessoas para que elas mesmas definissem a que classe pertenciam. Cerca de 80 % dos americanos se autotransformaram de classe média. Como na pesquisa havia só três opções a escolher (classe alta, média ou baixa), os economistas criticaram a tentativa por considerá-la distorcida e tendenciosa. Foi então que se resolveu somar na lista de escolhas a opção "classe trabalhadora". Resultado: 51% achavam que pertenciam a esse grupo - a classe média se reduziu para 43%.

Nos anos 80, dois economistas concluíram que a pesquisa da Fortune tinha uma série de manipulações, de maneira que o pesquisador pudesse ter o resultado que bem entendesse.

Foi nos anos 60 que se chegou, nos EUA, a um modelo um pouco mais próximo do aceitável - porém estatísticos ainda o criticavam. Nele, variáveis como educação e renda passaram a ser as principais determinantes do status econômico do indivíduo na sociedade (Gomes, 2007).

Logo, este esquema de classes foi inicialmente proposto por Erickson, Goldthorpe e Portocarrero (1979) e por Goldthorpe, Llewellyn e Payne (1987) para o estudo da mobilidade social de acordo com a perspectiva da análise de classes. Nos círculos de pesquisa comparativa internacional ele é conhecido como EGP (iniciais dos formuladores: Erickson, Goldthorpe e Portocarrero). Segundo o sociólogo inglês John Goldthorpe, cujo empenho em definir os contornos teóricos da mobilidade de classes tem sido o mais consistente, o estudo da mobilidade social, usando o esquema de classes EGP e modelos estatísticos log-lineares, seria especialmente adequado para descrever empiricamente a formação de classes sociais e os padrões de fluidez social, ou, inversamente, os padrões de rigidez da estrutura de classes. (Ribeiro & Scalón, 2001).

2.3.2 Classes econômicas no Brasil

Os critérios que definem a que classe econômica o brasileiro pertence mudaram. Parte dos cidadãos poderá ganhar uma nova posição na pirâmide social, alguns serão considerados mais ricos, e outros, mais pobres.

É a primeira mudança de maior peso feita desde que esses princípios de classes econômicas foram instituídos, em 1996, por meio de uma pesquisa chamada CCEB (Critério de Classificação Econômica Brasil), desenvolvida pela Abep (Associação Brasileiro de Empresas de Pesquisa) e seguida como padrão pelo mercado. As alterações vigorarão neste ano 2007.

TABELA 4: Associação tabular entre renda e classe econômica usada pelo Critério Brasil.

Classe	Pontos	Limites das classes (R\$)	Renda Média Familiar (R\$)
A1	30 a 34	> 6.220,50	7.793,00
A2	25 a 29	[3.726,00;6.220,50)	4.648,00
B1	21 a 24	[2.236,50;3.726,00)	2.804,00
B2	17 a 20	[1.298,00;2.236,50)	1.669,00
C	11 a 16	[675,50;1.298,00)	927,00
D	06 a 10	[315,50;675,50)	424,00
E	0 a 05	[0,00;315,50)	207,00

FONTE: Associação Brasileira de Empresas de Pesquisa-ABEP (2003).

Instrumento de segmentação da população segundo o seu poder de compra - e não de acordo com a renda familiar - o CCEB separa os consumidores em classes econômicas (A, B, C, D, E). A divisão é feita por meio de pontuação obtida pela posse de determinados bens e pelo grau de instrução do chefe de família, em pesquisa feita no domicílio por técnicos. Os bens pontuados são: televisores, rádios, banheiros, automóveis, empregadas mensalistas, aspiradores de pó, máquinas de lavar, video cassete e/ou DVD, geladeiras e freezers.

Aquele que tem em casa mais automóveis ou paga os serviços de

uma empregada soma mais pontos, pelo CCEB. Logo, tem mais chances de pertencer às classes de maior poder de compra.

O CCEB, ou "Critério Brasil", precisa sofrer uma revisão. As transformações no perfil de compra do brasileiro nos últimos anos, com a perda de importância de determinados bens - e com outros ganhando espaços -, mostraram que era necessária uma revisão, ou um "refinamento" do modelo, como explica a Abep (Mattos, 2007).

2.3.3 Classe social e Status

As classes sociais mostram as desigualdades da sociedade capitalista. Cada tipo de organização social estabelece as desigualdades, de privilégios e de desvantagens entre os indivíduos. As desigualdades são vistas como coisas absolutamente normais, como algo sem relação com produção no convívio na sociedade. As divisões em classes se dá na forma em que o indivíduo está situado econômica sócio-politicamente em sua sociedade. Como no capitalismo, quem tinha condições para a dominação e a apropriação eram os ricos, quem trabalhava para estes eram os pobres. Pois bem, esses elementos eram os principais denominadores de desigualdade social. Os consumidores associam marcas de produtos e serviços a classes sociais específicas, mesmo que seja desconfortável pensar sobre status social e desigualdade, é necessário, porque tanto compra quanto consumo são diretamente afetados por tais divisões na sociedade (Engel et al., 1995).

Segundo Engel (1995), classes sociais e sistemas de status existem em todos os países do mundo. Na Europa, o conceito é tão importante para se compreender o comportamento do consumidor que a European Society for Opinion and Marketing Research (ESOMAR) elabora perguntas para que as pesquisas de classe econômica sejam comparáveis entre os países. Compreender o desenvolvimento das classes sociais é importante para entender

o consumo, pois os estilos de vida da classe média alta tendem a se filtrar e a se tornar geralmente aceitos pelo resto da sociedade.

Classes sociais são definidas como divisões relativamente permanentes e homogêneas numa sociedade, nas quais indivíduos ou famílias, partilhando valores semelhantes, estilos de vida, interesses e comportamento podem ser categorizados. Referem-se a um agrupamento de pessoas que são similares em seus comportamentos baseados em sua posição econômica no mercado, enquanto que **Grupos de Status** refletem as expectativas da comunidade por estilo de vida entre cada classe. De acordo com Max Weber e Karl Marx, pode ser considerado o pai da teoria de classe social, esclareceu a definição:

Com alguma super simplificação, pode-se dizer que as "classes" são estratificadas de acordo com suas relações na produção e aquisição de bens, enquanto "grupos de status" são estratificados de acordo com os princípios de seu consumo de bens representados por "estilos de vida" especiais (Engel et al., 1995).

O status de uma pessoa também pode ser influenciado por seu sucesso relativo ao de outros na mesma ocupação - pelo desempenho pessoal do indivíduo. Declarações tais como "Ela é a melhor advogada de defesa da cidade" ou Frank é o único programador em quem eu confio para fazer isto" são exemplos de avaliações de desempenho pessoal. Muito embora renda não seja um bom indicador de classe social geral, ela pode servir como um gabarito de desempenho pessoal dentro de uma ocupação.

Não cometa o erro de pensar em equiparar classe social à renda. Classe social não é determinada por renda, muito embora possa haver uma correlação devido ao relacionamento entre renda e outras variáveis que determinam a classe social. Um gari sênior, por exemplo, pode ganhar mais do que um professor assistente de história. Entretanto, seria atribuída ao

professor uma classe social mais alta (Engel et al., 1995).

Uma das variáveis políticas importantes de classe social é a consciência de classe. Consciência de classe refere-se ao grau em que as pessoas em uma classe social têm consciência de si mesma como um grupo distinto, com interesses políticos e econômicos partilhados. Os americanos geralmente declaram que não tem consciência de classe. Em algumas medidas, a classe social de uma pessoa é indicada pela consciência que essa pessoa têm de classe social. Indivíduo de classe social mais baixa pode reconhecer a realidade da classe social, mas pode não ser tão sensível a diferenças específicas. Assim, a propaganda de produtos para alvos de mercado de classe superior é geralmente rica em símbolos de classe social, mas os anúncios para alvos das classes sociais média e baixa podem não ser bem recebidos se usarem um apelo direto de classe.

Os pesquisadores do consumidor desenvolveram e tomaram emprestado uma variedade de métodos para medir e descrever classe social. Estes métodos foram desenvolvidos para relacionar variáveis dependentes, tais como uso de produtos, preferência de marca, atitudes, imagem da loja e preferência à variável independente de classe social. Com tal pesquisa, pode ser possível definir segmentos de mercado com base na classe social e entender padrões de consumo e compra desses segmentos.

Os procedimentos para segmentação de mercado incluem os seguintes passos:

1. Identificação de uso de classe social do produto.
2. Comparação de variáveis de classe social para segmentação com outras variáveis (renda, ciclo de vida, etc...).
3. Descrição de características de classe social identificadas no mercado-alvo.

4. Desenvolvimento de programa de marketing para maximizar a eficácia do mix de marketing com base em consistência em relação a atributos de classe social.

A análise de segmentos de mercado por perfis socioeconômicos ajuda no desenvolvimento de um programa de marketing abrangente para combinar as preferências e comportamentos do mercado-alvo.

A tomada de decisão do consumidor é influenciada pela classe social da pessoa, especialmente na determinação de necessidades e critérios de avaliação. O tipo, a qualidade e o estilo de roupas que uma pessoa usa estão intimamente ligados à classe social dessa pessoa. Os critérios usados pelos consumidores para mobiliar a casa estão estreitamente relacionados à classe social. A classe social afeta o lazer de várias formas. O tipo de lazer preferido baseia-se em atividades que ocorrem principalmente com as pessoas nos mesmos níveis de status ou nos níveis adjacentes próximos, a proporção de renda familiar gasta em lazer pode não variar muito entre as classes sociais, mas o tipo de recreação varia enormemente. Pólo e Squash é classe alta; bridge, tênis e raquete é um jogo de classe média à alta; bingo e boxe é de classe baixa (Engel et al., 1995).

A quantidade e tipo de atividade empreendida e o processamento de informação por um indivíduo variam por classe social. Infelizmente, as classes sociais mais baixas geralmente têm fontes de informação limitadas e podem estar em desvantagem para filtrar informações incorretas e fraude numa sociedade urbana complexa.

Os padrões de linguagem de indivíduos estão intimamente correlacionado com sua classe social. As revistas e os jornais têm mais apelo para classes altas do que baixas, e os meios de comunicação podem influenciar a maneira como as classes sociais percebem a realidade social.

2.3.4 Processos de compras

O status social influencia onde e como as pessoas pensam que devem fazer compras. Pessoas de status mais baixos preferem lugares próximos, cara a cara, onde obtêm um serviço simpático e crédito fácil - geralmente no próprio bairro onde moram. O consumidor médio alto sente-se mais confiante de sua capacidade de compra. Ele se aventurará em novos lugares para comprar e percorrerá a loja inteira para encontrar o que quer.

A loja de descontos tradicionalmente apela às classes médias porque elas são cuidadosas e orientadas para a economia em suas compras. Os consumidores têm uma imagem de qual classe social uma loja atrai e têm uma compreensão do que deve ser, fazer compras numa loja que apela para a sua própria classe social. As pessoas de classes superiores querem uma atmosfera agradável de loja, apresentando displays excitantes e serviços excelentes. As classes inferiores enfatizam a aquisição de artigos para a casa ou roupas como a parte divertida de fazer compras.

Os pesquisadores de mercado medem a classe social como uma variável independente para determinar sua associação com variáveis dependentes de interesse para o marketing (Engel et al., 1995).

2.3.5 Poder de compra e renda

O poder aquisitivo do consumidor (possibilidade de adquirir mercadorias e serviços), é determinado não apenas por sua renda, mas também pelos preços. O poder de compra do consumidor poderia ser dobrado tanto em virtude da duplicação de sua renda como de uma redução, pela metade, de todos os preços das mercadorias que viesse, a adquirir.

Dadas as preferências e as restrições orçamentárias, pode-se então determinar como os consumidores escolhem quanto comprar de cada mer-

cadoria. Suponha que eles façam essa escolha de maneira racional; com isso diz-se que eles decidem a quantidade de cada bem visando a maximizar o grau de satisfação que poderão obter, considerando o orçamento limitado do que dispõem (Pendick & Rubinfeld, 2002). .

Se aumenta a renda de um consumidor, este normalmente desejará gastar mais e demandará maior quantidade de bens (Troster & Mochon, 1994).

3 METODOLOGIA

3.1 Pontes e Lacerda

Pontes e Lacerda é um município brasileiro do estado de Mato Grosso que se localiza à margem direita do Rio Guaporé a uma latitude 15°13'34" sul e a uma longitude 59°20'07" oeste, estando a uma altitude de 254 metros. Sua população estimada em 2004 era de 40 830 habitantes. Possui uma área de 13.169,5 km^2 .

A sede do município distancia-se 430 km da capital Cuiabá e 1025 km de Porto Velho, capital de Rondônia, sendo cortada pela Br-174 que une as duas capitais.

O relevo é composto pelo Planalto dos Parecís com depressão do Guaporé, Planícies e Pantanaís do Médio e Alto Guaporé. O solo predominante é o podizólico eutrófico, acompanhado por latossolo, com areias quartizosas e hidromórficos.

O município é banhado pelo Rio Guaporé, que guarda o ponto de estrangulamento oeste do maior divisor de águas na América Latina. A média da temperatura máxima diária é de 38° C e a média da mínima é de 14°, e com temperatura média de 24° C. A altitude em relação ao nível do mar é de 254 metros. (Anônimo, 2002).

O censo do IBGE de 2000 divulgou um mapa do município de Pontes

e Lacerda dividido em setores (e não bairros) e seus respectivos números de domicílios.

3.1.1 História do município

Os primeiros habitantes da região foram os índios do povo nambikwára. Os paulistas devassadores da região, nos primeiros tempos, denominaram esses índios de cabichis ou cavichis.

Apesar da dizimação da tribo por preias dos paulistas e problemas da aculturação, ainda hoje uma parte do povo nambikwára mantém vida organizada no município, em área indígena denominada Sararé. Nesta mesma área, em 1987, habitavam 41 índios.

Em 1906 a região passou a ser objeto de trabalho da Comissão de Linhas Telegráficas Estratégicas do Mato Grosso ao Amazonas, (Comissão Rondon). Pontes e Lacerda era estação telegráfica na linha variante entre Cáceres (cidade então denominada São Luís de Cáceres) e Mato Grosso (Vila Bela da Santíssima Trindade).

A denominação Pontes e Lacerda é recente, mas tem origem nos nomes dos astrônomos e cartógrafos Antonio Pires da Silva Pontes, que era mineiro e Francisco José de Lacerda e Almeida, que era paulista. Ambos eram diplomados pela Universidade de Coimbra - Portugal. O binômio Pontes e Lacerda honram os dois astrônomos.

A Lei Estadual nº 4.167, de 29 de dezembro de 1979, de autoria do deputado Ubiratan Spinelli e sancionada pelo governador Frederico Campos, criou o município de Pontes e Lacerda, com território desmembrado do município de Vila Bela da Santíssima Trindade. A instalação oficial deu-se somente em janeiro de 1981, ocasião em que tomou posse o Sr. Gercino Rodrigues de Souza, na ocasião, primeiro prefeito municipal nomeado. Dois

anos depois entregou o cargo ao Sr. Dionir de Freitas, primeiro prefeito eleito.

3.1.2 Vias de acesso

Rodovia

Br-364, situada no Km 227. Esta rodovia passa dentro do perímetro urbano do município na região norte, numa extensão de 80 Km e segue em direção a Porto Velho, capital do estado de Rondônia.

Br-473. Ramificada da Br-174, corta o município na região sul numa extensão de 150 km possibilitando acesso às principais áreas de produção pecuária, inclusive com acesso até a Bolívia.

Br-174. Rodovia de acesso ao município de Vila Bela da Santíssima Trindade numa extensão de 75 Km.

Hidrovias

Hidrovia Paraguai - Paraná. Faz a integração com o Continente Americano. Possui 3.442 Km de extensão de Cáceres a Buenos Aires/Argentina. Distância de Pontes e Lacerda a Cáceres 220 Km.

Hidrovia Madeira Amazonas. Utiliza-se a Br-364 para levar as mercadorias a Porto Velho (RO) onde são embarcadas e despachadas através do Rio Madeira até Itacoatiara (AM).

Aerovia

Aeroporto Municipal. Com capacidade de pouso de aeronaves de grande porte e operação de linhas aéreas comerciais.

3.2 A amostragem

Em outubro de 2006, na cidade de Pontes e Lacerda, uma pesquisa foi realizada em domicílios particulares (familiares), para investigar o consumo médio de energia, de água, área construída, renda familiar (R\$) e as variáveis do Critério Brasil (número de televisores, rádios, banheiros, automóveis, empregadas mensalistas, aspiradores de pó, máquinas de lavar, video cassete e/ou DVD, geladeiras e freezers; e grau de instrução do chefe de família).

Coletou-se, preliminarmente, uma *amostra piloto*, sorteando-se dez bairros, de acordo com a Tabela 5. Dentro de cada bairro sorteado, foi feita uma amostragem aleatória sistemática dos domicílios. Por sua vez, em cada um desses domicílios foi feita uma entrevista com o chefe de família.

Logo, pode-se determinar alguns parâmetros sobre a renda da população, como: média, mediana, moda e variância. Com as estimativas desses parâmetros, calculou-se a variância combinada (S_p^2) e definiu-se o tamanho da amostra com um erro de 5% da média (renda) e 95% de confiança.

A partir das informações descritivas da amostra piloto, a amostragem foi expandida para mais treze bairros da cidade; os bairros foram classificados em três grupos, de acordo com a localização.

Informações do estudo piloto sugerem que a renda varia em função da distância deles ao centro da cidade. Por isso, os bairros foram informalmente divididos em três grupos de acordo com a Figura 6:

- 1) bairros afastados do centro: menor classe econômica
- 2) bairros centralizados: maior classe econômica
- 3) bairros intermediários: classes econômicas intermediárias

A amostra piloto determina que o plano amostral deve ser composto

por localização de bairros associados à distância do centro da cidade (estratos).

Como base para compor o banco de dados, adotou-se esquema de amostragem aleatória estratificada informal, constituindo três grupos diferentes por localização dos bairros, centro (2, 3, 5, 8, 13, 18), periférico (4, 7, 9, 12, 17, 23) e intermediário (1, 6, 10, 11, 14, 15, 16, 19, 20, 21, 22), e dentro de cada grupo (conglomerado), sortearam-se bairros dos quais foram amostrados 750 domicílios. Embora nunca se tenha pretendido atingir o nível de sofisticação dessas investigações, conduzidas em nível nacional por órgãos especializados na execução de censos, procurou-se seguir uma metodologia similar que viesse facilitar uma posterior expansão do experimento para outras cidades do Estado de Mato grosso-MT.

O delineamento da amostragem conduz a uma amostra probabilística, cujo tamanho é função:

- do(s) parâmetro(s) a estimar;
- do nível de confiança desejável;
- do erro tolerável ou índice de precisão escolhidos; e
- do grau de dispersão da população.

Podendo, ainda, depender do tamanho da população e de outros parâmetros específicos.

A amostragem foi realizada em dois estágios:

- 1) Amostra piloto em 10 bairros da cidade de Pontes e Lacerda-MT
- 2) Complementação da amostra em mais 13 bairros da cidade, completando um total de 23 bairros amostrados

A técnica de amostragem utilizada consistiu na divisão da popula-

ção da cidade em bairros (chamados estratos) seguindo alguma(s) característica(s) conhecida(s) na população (estudo piloto), variância dos bairros e, de cada um desses estratos, foram selecionadas amostras em proporções convenientes.

No plano amostral utilizou-se de amostragem sistemática em cada estrato para a extração das amostras probabilística de domicílios, estratificada e conglomerada em dois estágios, para cada bairro de abrangência da pesquisa. Foi feita a seleção das unidades primárias de amostragem (bairros) e posteriormente das unidades secundárias de amostragem. As unidades primárias de amostragem da pesquisa são as estratificação da cidade de Pontes e Lacerda por bairros, as unidades secundárias de amostragem foi a seleção dos domicílios através de amostragem sistemática, com probabilidade proporcional ao total de domicílios ocupados obtida pelo Censo Demográfico de 2006 realizado nessa cidade pela Prefeitura Municipal. Após a seleção dos bairros, e com base na listagem atualizada de domicílios nestes estratos, faz-se, então, a seleção dos mesmos através de amostragem sistemática simples. A seleção dos domicílios da amostra é feita a partir do sorteio de um domicílio e, a partir deste, manter-se intervalos equidistantes de seleção de domicílios até a composição da amostra

Técnica de coleta

Nos domicílios selecionados pela pesquisa, o chefe da família foi submetido a um questionário através de entrevista.

Tamanho da amostra

O tamanho da amostra, fornecido pela amostra piloto, foi determinado pela variabilidade da renda, com um nível de confiança de 95% e um erro de 5% da média, de acordo com o Censo Demográfico de 2006 realizado nessa cidade, pela Prefeitura Municipal.

4 RESULTADOS E DISCUSSÃO

De acordo com estudo piloto, Tabela 5, pesquisou-se a renda de 367 chefes de família de Pontes e Lacerda-MT, determinando-se a estimativa da renda média por classe econômica. Calculou-se a variância combinada S_p^2 , para o dimensionamento do tamanho da amostra, que totalizou-se em 1088 domicílios. Portanto, voltou-se a campo para a complementação da amostra, ou seja, para serem observados mais 735 domicílios.

TABELA 5: Renda média familiar em R\$, variância da renda familiar e tamanho da amostra em dez bairros da cidade de Pontes e Lacerda, MT

Bairro	Média	Variância	Tamanho da amostra
1	1.658,33	1.519.842	37
2	2.451,00	5.459.881	33
3	2.708,89	3.165.210	36
4	647,70	172.785	44
5	2.168,00	586.011	40
6	1.568,70	2.745.078	40
7	912,30	632.334	39
8	2.068,60	2.886.067	40
9	1.089,70	1.397.463	29
10	949,60	464.824	29

FONTE: Levantamento amostral, março de 2007.

A variância combinada (S_p^2) foi estimada por:

$$\begin{aligned}
S_p^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \dots + (n_{10} - 1)s_{10}^2}{n_1 + n_2 + \dots + n_{10} - 10} \\
&= \frac{\sum_{i=1}^{10} (n_i - 1)s_i^2}{\sum_{i=1}^{10} n_i - 10} \\
&= 1.863.163 \text{ reais}^2
\end{aligned}$$

A renda média familiar, que pertence à classe média B2, foi estimada em R\$ 1622,30 reais pelo Critério Brasil. O tamanho da amostra para se cometer 5% de erro na estimação da média foi estimado por

$$\begin{aligned}
n &= \left(\frac{Z_{\frac{\alpha}{2}} S_p}{d} \right)^2 \\
&= 1.088 \text{ domicílios.}
\end{aligned}$$

sendo,

$Z_{\frac{\alpha}{2}}$ = Valor crítico da distribuição normal

S_p^2 = Desvio-padrão combinado

d = diferença a ser determinada, obtida como 5% da média geral (81, 08 reais)

Pode-se observar, nos primeiros dez bairros da Figura 6, amostra piloto, que a similaridade de posicionamento da mediana dos box-plot sugere a existência de três grupos que se diferenciam quanto à renda. Por exemplo, os bairros 4, 7 e 9 parecem determinar um conjunto de bairros com menor renda mediana; os bairros 2, 3, 5 e 8, de maior renda; e 1, 6 e 10, bairros de renda intermediária.

Além disso, a maior proximidade entre o segundo quartil e a mediana de cada box-plot, do que entre a mediana e o terceiro quartil, sugere que a distribuição da renda sempre se apresenta de forma assimétrica à direita. Vale ressaltar que o segundo e terceiro quartis de um box-plot, no desenho, são representados pelo início e final do retângulo que determinam a caixinha.

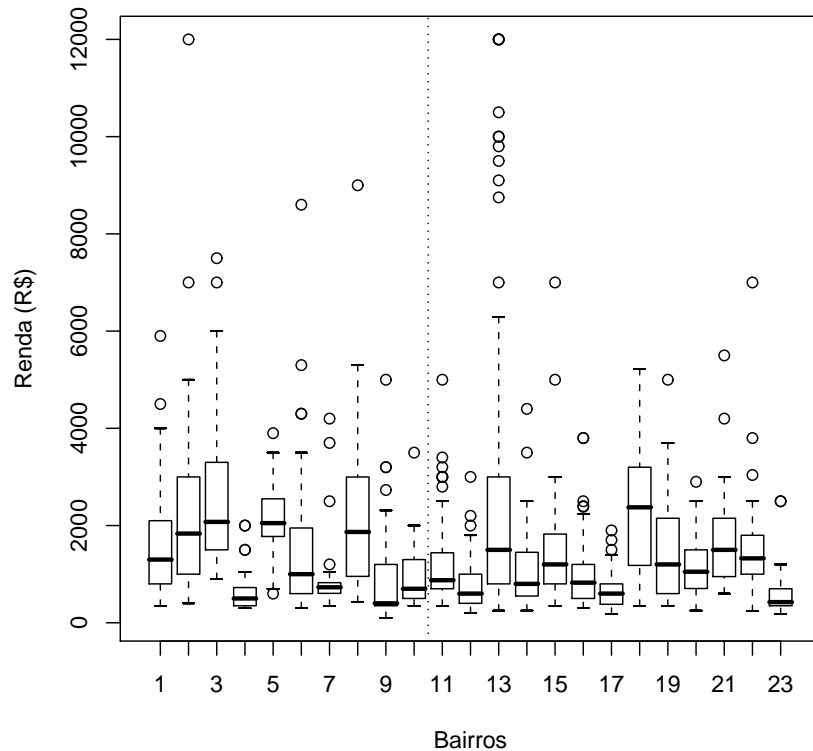


FIGURA 6: Box-plot da renda, R\$, dos 23 bairros amostrados, ressaltando a amostra piloto (1 a 10).

erificou-se, ainda, que os bairros mais centralizados têm rendimentos médios maiores que os mais afastados do centro da cidade, o que, de um modo geral, é verificado em diversas cidades. Uma idéia para contornar o problema da estratificação por renda, é a estratificação informal dos bairros, quanto a distância destes em relação ao centro da cidade.

De acordo com a Figura 6, os 10 primeiros bairros, que foram selecionados para esse trabalho, são referentes a uma amostra piloto. Observa-se que existe uma distribuição assimétrica à direita dentro de cada bairro com maior aglomeração de renda no primeiro e segundo quartil.

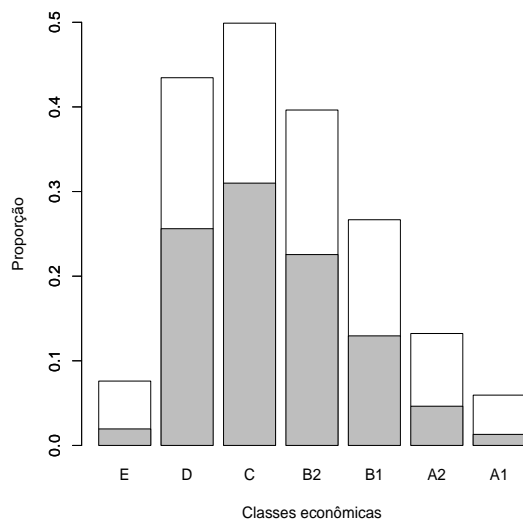


FIGURA 7: Proporções das classes econômicas em Pontes e Lacerda, MT, destacando-se, na parte superior do retângulo hachurado, a estimativa pontual, e a parte superior do retângulo em branco o limite superior do intervalo de confiança para proporção com 95% de confiança.

A Figura 7 mostra as proporções estimadas da população em Pontes e Lacerda-MT, por classe econômica, e seus respectivos limites superiores dos intervalos de confiança (95%).

As classes A1, A2 e E, de acordo com a Figura 7, apresentam intervalos sobrepostos, apresentando as menores proporções, ao passo que as demais classes referenciadas possuem o maior percentual.

A Figura 8 é um histograma da renda (R\$) em Pontes e Lacerda, que destaca a assimetria da distribuição dessa variável e a não conhecida da média, mediana e moda.

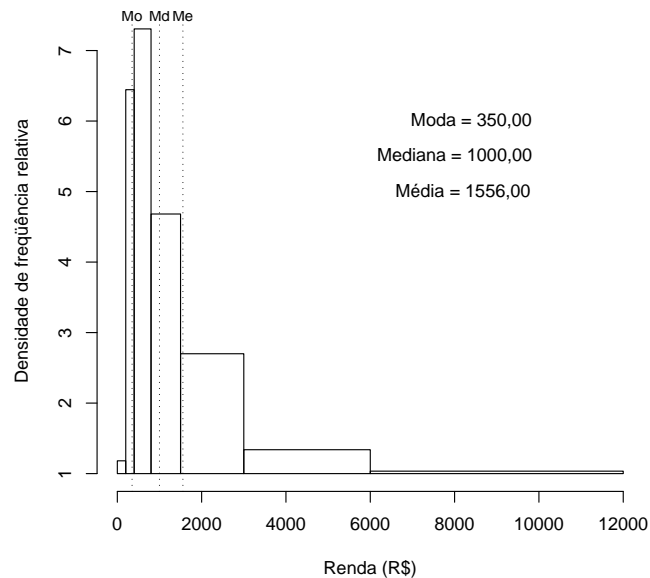


FIGURA 8: Histograma da renda (R\$) em Ponte de Lacerda, MT, destacando-se a moda, a mediana e a média.

Observou-se ainda que a média não consegue dar uma idéia da renda da população de Pontes e Lacerda, MT, logo, não é uma boa medida da realidade da população. De acordo com a Figura 8, percebe-se que a média expressa um valor muito acima da renda modal. O mesmo acontece em relação à mediana, porém em menor intensidade. Uma vez verificada a assimetria, assume-se que a melhor medida de posição é a moda.

A Tabela 6 e 7 mostra os modelos objetivo e híbrido para toda a cidade de Pontes e Lacerda-MT.

Os procedimentos Stepwise, Backward e Forward aplicados por meio do software (R Development Core Team (2007), foram usados para a seleção dos melhores modelos. No caso do modelo objetivo, os procedimentos

Stepwise e Backward levaram ao mesmo modelo (4.1) e o Forward levou a um modelo menos parcimonioso (4.2).

O modelo objetivo a seguir descrito foi obtido por Stepwise e Backward:

$$\begin{aligned}
y_i = & -718,9 + 55,62x_{i1} - 0,8146x_{i1}^2 + 2,8x_{i2} + 0,006523x_{i2}^2 + 9,324x_{i3} - \\
& - 0,01432x_{i3}^2 + 0,1122x_{i1}x_{i3} + 639,9b_{i2} + 773,2b_{i3} + 656,4b_{i5} - \\
& - 613,9b_{i10} - 532,5b_{i12} + 778,9b_{i18},
\end{aligned} \tag{4.1}$$

em que y_i é a renda estimada da i -ésima família; x_{i1} representa o i -ésimo consumo de água; x_{i2} representa o i -ésimo consumo de energia; x_{i3} representa a área construída da i -ésima residência; b_{i2} , b_{i3} , b_{i5} , b_{i10} , b_{i12} , b_{i18} são variáveis dicotômicas que indicam a presença (1) ou ausência (0) da i -ésima residência nos bairros 2, 3, 5, 10, 12 e 18, respectivamente.

O modelo objetivo a seguir descrito foi obtido por Forward:

$$\begin{aligned}
y_i = & -789,7 + 64,46x_{i1} - 0,8372x_{i1}^2 + 2,72x_{i2} + 0,006523x_{i2}^2 + 8,68x_{i3} - \\
& - 0,01689x_{i3}^2 + 0,119x_{i1}x_{i3} - 0,0523x_{i1}x_{i2} + 0,007824x_{i2}x_{i3} + \\
& + 633,4b_{i2} + 853,4b_{i3} + 695,8b_{i5} - 590,6b_{i10} + 410,6b_{i13} + \\
& + 864,3b_{i18},
\end{aligned} \tag{4.2}$$

em que y_i é a renda estimada da i -ésima família; x_{i1} representa o i -ésimo consumo de água; x_{i2} representa o i -ésimo consumo de energia; x_{i3} representa a área construída da i -ésima residência; b_{i2} , b_{i3} , b_{i5} , b_{i10} , b_{i13} , b_{i18} são variáveis dicotômicas que indicam a presença (1) ou ausência (0) da i -ésima residência nos bairros 2, 3, 5, 10, 13 e 18, respectivamente.

Para se decidir qual modelo adotar, foi feito um teste F ($p < 0,05$) para compará-los. Em ambos os modelos (objetivo e híbrido), aqueles selecionados por Stepwise foram considerados os melhores e mais parcimoniosos. Também, por isso, apenas o procedimento Stepwise foi usado na seleção de modelos dentro de cada bairro.

TABELA 6: Modelo Objetivo: comparação dos modelos 4.1 e 4.2, estimados por Stepwise e backward vs forward

GLResíduo	SQResíduos	GL	SQ	F	p-valor
1072	1179113695				
1070	1175358220	2	3755476	1,7094	0,1815

FONTE: Modelos estimados, com dados da pesquisa, março de 2007.

Analogamente para o modelo híbrido, os procedimentos de Stepwise e Backward levaram ao mesmo modelo:

$$\begin{aligned}
 y_i = & -1315 + 11,8x_{i1} + 0,005975x_{i2}^2 + 4,802x_{i3} - 0,007903x_{i3}^2 + \\
 & + 0,007194x_{i2}x_{i3} + 629,4b_{i2} + 610,9b_{i3} + 487,5b_{i5} + 506,6b_{i6} + \\
 & + 536,8b_{i13} + 432,7b_{i15} + 734,3b_{i18} + 441,1b_{i19} + 457,3b_{i20} + \\
 & + 123,5x_{i4}, \tag{4.3}
 \end{aligned}$$

em que y_i é a renda estimada da i -ésima família; x_{i1} representa o i -ésimo consumo de água; x_{i2} representa o i -ésimo consumo de energia; x_{i3} representa a área construída da i -ésima residência; b_{i2} , b_{i3} , b_{i5} , b_{i6} , b_{i13} , b_{i15} , b_{i18} , b_{i19} , b_{i20} são variáveis dicotômicas que indicam a presença (1) ou ausência (0) da i -ésima residência nos bairros 2, 3, 5, 6, 13, 15, 18, 19 e 20, respectivamente; e x_{i4} os pontos no Critério Brasil da i -ésima família;

Entretanto, novamente o procedimento Forward levou a um modelo menos parcimonioso:

$$\begin{aligned}
y_i = & -1299 + 27,15x_{i1} - 0,3878x_{i1}^2 - 1,132x_{i2} + 0,00855x_{i2}^2 + 4,733x_{i3} - \\
& - 0,00977x_{i3}^2 + 0,05571x_{i1}x_{i3} - 0,01507x_{i1}x_{i2} + 0,006293x_{i2}x_{i3} + \\
& + 614,5b_{i2} + 589,4b_{i3} + 473,8b_{i13} + 652,2b_{i18} + 124x_{i4}, \quad (4.4)
\end{aligned}$$

em que y_i é a renda estimada da i -ésima família; x_{i1} representa o i -ésimo consumo de água; x_{i2} representa o i -ésimo consumo de energia; x_{i3} representa a área construída da i -ésima residência; b_{i2} , b_{i3} , b_{i13} , b_{i18} , são variáveis dicotômicas que indicam a presença (1) ou ausência (0) da i -ésima residência nos bairros 2, 3, 13 e 18, respectivamente; e x_{i4} os pontos no Critério Brasil da i -ésima família;

TABELA 7: Modelo Híbrido: comparação dos modelos 4.3 e 4.4, estimados por Stepwise e backward vs forward

GLResíduo	SQResíduos	GL	SQ	F	p-valor
1073	983470142				
1069	979626204	4	3843938	1,0487	0,3809

FONTE: Modelos estimados, com dados da pesquisa, março de 2007.

Portanto, de acordo com as Tabelas 6 e 7, os modelos 4.1 e 4.3 foram eleitos os mais parcimoniosos para estimar a renda de domicílios em Pontes e Lacerda. Eles foram selecionados os modelos objetivo e híbrido, respectivamente.

De acordo com o tópico 2.2.4, foi feito o teste de significância para o R^2 ($p < 5\%$). O resultado desse teste é mostrado na coluna 5 da Tabela 8, para os 23 bairros da cidade de Pontes e Lacerda, onde foram analisados os p-valores para cada bairro, destacando-se em vermelho os R^2 que não foram significativos (quinta coluna da Tabela 8).

Nota-se que ambos os modelos, objetivo e híbrido, se ajustaram com R^2 significativos para a grande maioria dos bairros de Pontes e Lacerda-MT,

o que pode ser considerado um bom resultado, já que os valores preditos pelo modelo objetivo não ficam distantes dos valores esperados.

De acordo com a coluna 3 da Tabela 8, observa-se que o R^2 , ajustado nos 23 bairros de Pontes e Lacerda, para os modelos de regressão, baseados em covariáveis objetivas: modelo objetivo (Ob) e variáveis objetivas e subjetivas: modelo híbrido (Hb), mostra um índice de desvio equivalente, erram muito ou pouco dentro de um mesmo bairro, o que pode ser considerado um bom resultado, visto que os valores preditos pelo modelo objetivo não ficam distantes dos valores esperados, sabido que o modelo híbrido tem efeito do questionário Critério Brasil.

A Tabela 8 traz a raiz quadrada do erro quadrático médio (EQM), da renda de Pontes e Lacerda-MT, para os modelos objetivo e híbrido, baseado em um teste de χ^2 . O teste mostra que não se pode afirmar que o EQM dos três modelos se difere estatisticamente, o que nos leva a concluir que não fazer entrevista gera um erro estatístico igual a fazer, com 95% de confiança

De acordo com a Tabela 9 e 10, o modelo objetivo e o modelo híbrido não são representados pelas mesmas variáveis. No modelo objetivo, as variáveis que mais se destacaram foram: e , e^2 , a , ag explicada na maioria dos bairros, enquanto que, no modelo híbrido, as mais significativas foram: e , e^2 , ea , cb , isso mostra que o efeito de água e área construída foram substituídos pelo efeito Critério Brasil.

Observando-se as Tabelas 11 e 12, é fácil verificar alguns bairros com efeitos significativos, isto é, com valores acima da média. Logo, verifica-se que moradores residentes nesses bairros possuem uma classe econômica acima da moda.

Observa-se que o modelo híbrido é mais rigoroso, na definição dos bairros com médias superiores à moda.

TABELA 8: Coeficiente de determinação ajustado (R_{aj}^2), erro quadrático médio na renda (EQM_{renda}) e p-valor do teste de significância dos coeficientes de determinação, múltiplos (R^2), nos 23 bairros estudados, para os modelos de regressão, baseados em covariáveis objetivas (Ob) e o modelo híbrido (Hb).

Bairro	Modelo	R_{aj}^2	$\sqrt{EQM_{renda}}$	<i>p-valor</i>
1	Ob	0,5554	773,2161	0,021423
	Hb	0,5554	773,2161	0,021423
2	Ob	0,3832	1703,2260	0,050463
	Hb	0,5062	1549,9870	0,032933
3	Ob	0,7576	799,5534	0,000681
	Hb	0,8136	735,4409	0,003744
4	Ob	0,6199	231,8183	0,001136
	Hb	0,6852	208,0019	0,000309
5	Ob	0,6646	408,7070	0,001706
	Hb	0,8198	299,6223	0,000174
6	Ob	0,8047	668,6115	0,000230
	Hb	0,8324	619,2347	0,000135
7	Ob	0,9356	179,9099	0,000001
	Hb	0,9593	140,6249	0,000000
8	Ob	0,7846	724,3264	0,001541
	Hb	0,8105	679,4584	0,001133
9	Ob	0,9581	210,4092	0,000001
	Hb	0,9583	205,1091	0,000000
10	Ob	0,7163	343,8315	0,011432
	Hb	0,8542	236,8403	0,000494
11	Ob	0,2478	693,4695	0,014344
	Hb	0,4340	598,2080	0,001184
12	Ob	0,4167	445,3432	0,042184
	Hb	0,7427	283,1737	0,000667
13	Ob	0,5518	1814,5800	0,000414
	Hb	0,6160	1673,3050	0,000084
14	Ob	0,1848	773,4800	0,153865
	Hb	0,5401	564,0989	0,006901

Bairro	Modelo	R_{aj}^2	$\sqrt{EQM_{renda}}$	p -valor
15	Ob	0,5504	712,6766	0,002463
	Hb	0,5893	694,8945	0,004779
16	Ob	0,1600	709,1946	0,137638
	Hb	0,2464	665,1251	0,061043
17	Ob	0,6493	209,4731	0,000779
	Hb	0,6796	197,7666	0,000337
18	Ob	0,7706	589,2603	0,000103
	Hb	0,8106	541,5353	0,000104
19	Ob	0,3695	780,8816	0,029599
	Hb	0,3985	772,6590	0,029342
20	Ob	0,6404	374,9098	0,003688
	Hb	0,7355	316,7667	0,000739
21	Ob	0,2759	822,0142	0,081749
	Hb	0,2759	822,0142	0,081749
22	Ob	0,4631	708,5775	0,006929
	Hb	0,6536	561,6297	0,000439
23	Ob	0,5515	293,1831	0,003429
	Hb	0,5728	286,1396	0,002684

FONTE: Dados da pesquisa, março de 2007.

De acordo com comparações feitas na figura 9, entre as classes estimadas por meio do Critério Brasil (azul), modelo objetivo (verde), modelo híbrido (vermelho) e classes econômicas reais (preto), a Figura 9 mostra formas conceituais de quatro curvas. É possível ajustar um modelo baseado em covariáveis objetivas para estimação de renda familiar (em reais), na cidade de Pontes e Lacerda-MT, tal modelo em média tem um ajuste considerável, de acordo com a Tabela 8.

De acordo com a figura 10, observa-se que os modelos não se diferenciam entre si, são estatisticamente iguais, podendo então utilizar o modelo objetivo para estimar a renda familiar de Pontes e Lacerda-MT, com uma precisão equivalente ao modelo híbrido.

TABELA 9: Presença (x) de efeitos no modelo objetivo final para os 23 bairros estudados.

Bairro	e	e^2	a	a^2	ag	ag^2	ea	eag	aga
1	x						x		
2		x	x	x					
3		x	x		x		x		x
4	x	x	x		x	x	x		x
5	x				x	x	x		x
6		x		x			x	x	x
7	x	x			x	x	x	x	x
8					x	x	x		
9		x	x			x	x	x	x
10		x		x					
11		x	x	x	x			x	
12						x		x	
13	x		x		x			x	
14	x					x		x	
15		x	x	x	x		x		
16	x	x							
17	x		x	x	x	x		x	
18	x	x	x		x		x		x
19	x	x	x		x				
20	x				x		x		x
21	x					x			
22	x	x	x		x		x	x	x
23		x	x	x	x			x	x

FONTE: Dados da pesquisa, março de 2007.

TABELA 10: Presença (x) de efeitos no modelo híbrido final para os 23 bairros estudados.

Bairro	e	e^2	a	a^2	ag	ag^2	ea	eag	aga	cb
1	x						x			
2							x			x
3		x								x
4	x	x	x		x	x	x		x	x
5	x					x	x		x	x
6		x			x		x		x	x
7	x	x			x	x	x	x	x	x
8		x					x			x
9	x	x	x			x	x	x		x
10			x			x			x	x
11		x		x	x		x	x		x
12	x	x		x			x			x
13	x			x	x				x	x
14		x	x			x		x		x
15		x					x			x
16	x	x								x
17	x		x	x	x	x		x		x
18	x	x					x		x	x
19		x					x			x
20	x						x	x	x	x
21	x					x				
22	x	x	x		x		x	x	x	x
23	x	x		x	x				x	x

FONTE: Dados da pesquisa, março de 2007.

TABELA 11: Ajuste do modelo baseado em covariáveis objetivas.

Efeitos	Coefficiente	p-valor
Intercepto	-1.315,0000	0,000000 ***
ag	11,8000	0,009137 **
e^2	0,0060	0,000011 ***
a	4,8020	0,000027 ***
a2	-0,0079	0,026093 *
ea	0,0072	0,055007 .
ba2	629,4000	0,006416 **
ba3	610,9000	0,009640 **
ba4	218,1000	0,335728
ba5	487,5000	0,028036 *
ba6	506,6000	0,023433 *
ba7	9,2760	0,967773
ba8	327,9000	0,160551
ba9	421,5000	0,091351 .
ba10	-129,4000	0,592219
ba11	-28,4800	0,880687
ba12	-85,2400	0,703073
ba13	536,8000	0,003673 **
ba14	105,6000	0,633746
ba15	432,7000	0,038512 *
ba16	334,8000	0,118424
ba17	313,4000	0,150969
ba18	734,3000	0,000602 ***
ba19	441,1000	0,047084 *
ba20	457,3000	0,041648 *
ba21	424,2000	0,051398 .
ba22	268,9000	0,215683
ba23	118,3000	0,598146
totalcb	123,5000	0,000000 ***

FONTE: Dados da pesquisa, março de 2007.

TABELA 12: Ajuste do modelo híbrido, baseado em covariáveis objetivas e no total de pontos no Critério Brasil.

Efeitos	Coefficiente	p-valor
Intercepto	-1.299,0000	0,000000 ***
ag	27,1500	0,041725 *
ag^2	-3.8,7800	0,094437 .
e	-1,1320	0,271698
e^2	0,0086	0,000435 ***
a	4,7330	0,000288 ***
a^2	-0,0098	0,014156 *
aga	0,0557	0,276156
eag	-0,0151	0,679751
ea	0,0063	0,171897
ba2	614,5000	0,007984 **
ba3	589,4000	0,014090 *
ba4	167,4000	0,465457
ba5	428,1000	0,056881 .
ba6	439,7000	0,052237 .
ba7	-57,4900	0,805965
ba8	249,4000	0,292206
ba9	334,6000	0,186394
ba10	-173,7000	0,476371
ba11	-69,4600	0,717396
ba12	-145,4000	0,520479
ba13	473,8000	0,011841 *
ba14	56,6400	0,800010
ba15	385,1000	0,068381 .
ba16	281,3000	0,193391
ba17	239,6000	0,279178
ba18	652,2000	0,003055 **
ba19	368,1000	0,102481
ba20	408,7000	0,071903 .
ba21	361,7000	0,100841
ba22	214,1000	0,331449
ba23	475,7000	0,834449
totalcb	124,0000	0,000000 ***

FONTE: Dados da pesquisa, março de 2007.

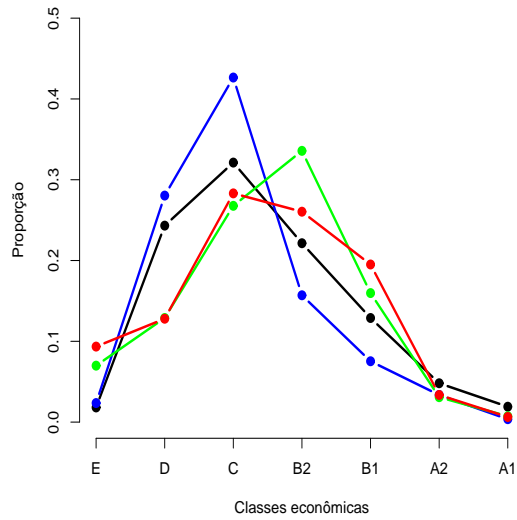


FIGURA 9: Comparação entre as classes estimadas por meio do Critério Brasil (azul), modelo objetivo (verde), modelo híbrido (vermelho) e classes econômicas reais (preto).

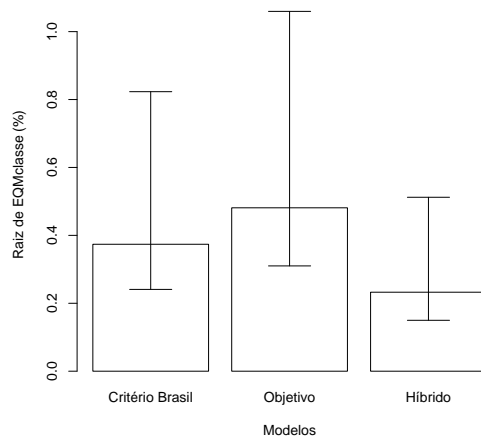


FIGURA 10: Comparação entre os intervalos de confiança da raiz de erros quadráticos médios cometidos pelos modelos, Critério Brasil, objetivo e híbrido na estimação de classes econômicas.

5 CONCLUSÕES

A dispensa da entrevista em uma pesquisa para a estimação da classe econômica pode representar uma enorme economia de recursos (tempo e dinheiro). Os modelos objetivo, híbrido e Critério Brasil proporcionaram erros quadráticos médios semelhantes, ou seja, não há evidências para assumi-los como diferentes. Portanto, usando o modelo objetivo, mesmo sem haver entrevistas, pode-se cometer o mesmo erro que o Critério Brasil normalmente comete em Pontes e Lacerda, embora o mesmo tenha apresentado comportamento razoável, pode ser melhorado com a inserção de outras covariáveis objetivas importantes.

Por sua vez, se entrevistas forem feitas e o Critério Brasil for aplicado, parece ser vantajosa a estimação das classes econômicas via modelo híbrido e, aconselha-se também este modelo se o foco de interesse for a renda.

REFERÊNCIAS

- ANÔNIMO. **Fórum regional de desenvolvimento sustentável.**
- BUSSAB, W. O.; HELENO, B. *Elementos de Amostragem.* Editora ABE, São Paulo, S.P., 2005. 269p.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica. 5. ed.** São Paulo:.
- CESAR, R. C. E.; et al. A pesquisa de opinião como método de investigação em relações públicas: Caso aterro do lago igapó. Disponível em: <<http://www.portal-rp.com.br/bibliotecavirtual/opiniaopublica/0181.htm>>. Acesso: 06 de Fevereiro de 2007.
- DAVID, M. *A Estatística Básica e Sua Prática.* Editora LTC, Rio de Janeiro, R.J., 2000. 482p.
- DRAPER, N. R.; SMITH, H. *Applied regression analysis. 3rd Ed.* John Wiley, New York, 1998. 706 p.
- ENGEL, J. F.; BLACKWELL, R. D.; MINIARD, P. W. **Comportamento do consumidor. 8ª Edição.** LTC Editora, Rio de Janeiro, 1995. 641 p.
- FAUZE, N. M. **Análise crítica dos métodos de estratificação social utilizados em marketing e pesquisas de marketing.** *Anais do 2º SEMEAD*, 1997. Disponível em: <<http://fauze.com.br/artigo03.htm>>. Acesso: 26 de outubro de 2006.
- FERREIRA, D. F. *Estatística Básica.* Editora UFLA, Lavras, M.G., 2005. 664p.
- GOMES, R. **Excesso de pó em móveis com mancha na sala já foram critério de classificação.** *Folha de São Paulo*, 07 de janeiro de 2007.
- GUERRA, M. J.; DONAIRE, D. *Estatística Indutiva. 5ª ed.* Editora LCTE, São Paulo, 1991. 311 p.

MATTOS, A. **Institutos vão mudar definição de classe social.** *Folha de São Paulo*, 07 de janeiro de 2007.

PINDICK, R. S.; RUBINFELD, D. L. *Microeconomia*. 5^a ed.. Editora Prentice Hall, São Paulo, 2002. 711 p.

RIBEIRO, C. A. C.; SCALON, M. C. Mobilidade de classe no brasil em perspectiva comparada. *Dados*, v.44(n.1), 2001. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0011-52582001000100004&lng=ptr&nrm=iso>. Acesso em: 26 de outubro de 2006. 36p.

SANTOS, J. A. F. Efeitos da classe na desigualdade racial no brasil. *Dados*, v.48(n.1), 2005. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0011-52582001000100004&lng=pt&nrm=iso>. Acesso em: 26 de outubro de 2006. 44p.

SCHLICH, P. *Preference Mapping: relating consumer preferences to sensory or instrumental measurements*. In: *Bioflavour 95*. Institut National de la Recherche Agronomique, Dijon, France, 1995. 135-150.

TROSTER, R. L.; MOCHON, F. *Introdução à Economia*. 2^a ed. Editora Makron Books, São Paulo, 1994. 391 p.