

MODELOS DE ESTIMAÇÃO DE RENDA FAMILIAR EM PONTES E LACERDA-MT

Oswaldo Martins de SOUZA¹

Eric Batista FERREIRA²

Agostinho Roberto de ABREU³

¹Professor assistente Mestre da Universidade do Estado de Mato Grosso (UNEMAT). E-mail: osvsou@gmail.com.

²Professor adjunto III, Universidade Federal de Alfenas (Unifal-MG). E-mail: eric.ferreira@unifal-mg.edu.br.

³Professor titular, Universidade Federal de Lavras (UFLA). E-mail: agostinho@dex.ufla.br.

Recebido em: 25/05/2014 - Aprovado em: 30/06/2014 - Disponibilizado em: 30/07/2014

RESUMO: O objetivo principal desse trabalho de mensuração não é estimar a renda das famílias da cidade de Pontes e Lacerda, mas desenvolver uma metodologia com variáveis objetivas que possam validar uma nova classificação socioeconômica de renda familiar. Esta classificação foi construída sob perspectiva prática na análise de classes sociais, inspirada nas contribuições do consumo mensal de água (m^3), energia elétrica (KWh) e área construída em (m^2). Esta nova metodologia traduz um esforço de aperfeiçoamento na técnica de estimar a renda familiar com precisão igual a métodos já existentes, tendo maior rapidez, menor custo e menos constrangimentos para os entrevistados.

Palavras-chave: Renda familiar. Modelo de regressão. Classe econômica. Amostragem.

Models for estimating the family income in Pontes e Lacerda-MT

ABSTRACT: The main goal of this work is not to estimate the income of families in the city of Pontes e Lacerda, but develop a methodology with objective variables that can validate a new socio-economic classification of household income. This classification was constructed under practical perspective in the analysis of social classes, inspired by the contributions of the monthly water consumption (m^3), electricity (kWh) and area in (m^2). This new methodology reflects an effort to improve the technique to estimate the family income with a precision of existing methods, and faster, lower cost and fewer constraints to respondents.

Keywords: Family income. Regression model. Economic class, Sampling.

INTRODUÇÃO

O Census Bureau é parte do Departamento de Comércio dos Estados Unidos da América. É a agência governamental encarregada pelo censo nos Estados Unidos define renda como o dinheiro de ordenados e salários, assim como o de pagamentos de juros e previdência social (UNITED STATES CENSUS BUREAU, 2014).

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). 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.

Evolução histórica da classificação social

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 de 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. E ponto final. 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. Em 1940, a revista "Fortune" resolveu fazer um levantamento com 5207 pessoas para que elas mesmas definissem a que classe pertencia. Cerca de 80 % dos americanos se autoclassificaram 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 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; SCALON, 2001).

A PNAD foi implantada no Brasil em 1967, como fruto do desenvolvimento dos sistemas de estatísticas domiciliares contínuas que vinham sendo testados e implantados na América do Norte desde os anos cinquenta e dos esforços do Inter American Statistical Institute (IASI) e da United States Agency for International Development (USAID) em levar

tal tipo de experiência aos países latino-americanos. Serviu como base conceitual para implantação das PNADs no Brasil, o chamado Plano Atlanta, desenvolvido pelo “Bureau of Census” norte americano. Os objetivos iniciais da PNAD se limitavam a criar um sistema de estatísticas domiciliares por amostragem que fosse barato e fácil de manter, permitindo, ao mesmo tempo, estudar o comportamento de algumas variáveis sócio-econômicas ligadas aos indivíduos, as famílias e aos domicílios, ao nível agregado. Demandas específicas da Secretaria de Planejamento da Presidência da República, bem como de diversos órgãos de desenvolvimento econômico regional, fizeram com que os estudos a posteriores planos de amostragem permitissem maiores desagregações, incorporando 7 regiões, que passaram a ser denominadas Regiões PNAD, além de estimativas para alguns Estados e Regiões Metropolitanas. Na década de setenta, a divulgação das PNADs era feita para o Brasil, para as 7 Regiões (embora não fossem consideradas as áreas rurais da Região VII) e para todas as Regiões Metropolitanas. Nos anos oitenta, as PNADs passaram a ser divulgadas para as 5 Macro-Regiões Fisiográficas (Norte, Nordeste, Sul, Sudeste e Centro-Oeste), para alguns Estados e para as Regiões Metropolitanas.

O primeiro período da PNAD pode ser demarcado como o que vai do segundo

trimestre de 1967 até o primeiro trimestre de 1970. Este período, caracterizado pela apresentação trimestral dos resultados para as cinco primeiras regiões PNADs foi interrompido por ocasião da realização do Censo Demográfico de 1970.

O segundo período se inicia no quarto trimestre de 1971 a termina no quarto trimestre de 1973, mudando a periodicidade da pesquisa de trimestral para anual, a qual passava a ser realizada sempre no último trimestre de cada ano. É marcado pela aplicação de diferentes tipos de questionários.

O terceiro período corresponde as PNADs de 1976 a 1979. Estas passaram a se caracterizar por uma investigação anual, a partir da escolha de uma semana de referência para a pesquisa. A base da pesquisa passou a ser um questionário, que correspondia ao levantamento de dados relacionados aos domicílios, a família e ao indivíduo.

O quarto período 1981-1986 caracteriza-se pela imutabilidade do "Corpo Básico" e pela aplicação anual de suplementos especiais que investigaram temas como saúde, educação, previdência social, assistência ao menor e acesso aos serviços de saúde, suplementação alimentar, associativismo e anticoncepção. Outra mudança importante foi a redução da amostra realizada em 1986, que permitiu maior eficiência, na medida em que praticamente foi

mantido o coeficiente de variação para um tamanho de amostra significativamente menor.

Instrumento de segmentação da população segundo o seu poder de compra - e não de acordo com a renda familiar - o Critério de Classificação Econômica Brasil (CCEB ou Critério Brasil) 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, vídeo cassete e/ou DVD, geladeiras e freezers.

Aquele que tem em casa mais automóveis ou pagam os serviços de uma empregada somam mais pontos, pelo CCEB. Logo, têm mais chances de pertencer às classes de maior poder de compra.

De acordo com Matos (2007) o 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.

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 (ENGELS et al., 1995).

Levantamentos por amostragem

É compreensível que o estudo de todos os elementos da população possibilita 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. Procurou acercar-se de cuidados, visando à obtenção de uma amostra significativa, ou seja, que de fato represente “o melhor possível” toda à população.

O uso de informações de uma amostra para concluir sobre o todo 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)

Os levantamentos por amostragem (PNADs) 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. 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 (FERREIRA, 2005).

A amostragem probabilística permite, além da estimação pontual e do controle do erro, a estimação por intervalo, 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.

Para a determinação do tamanho da amostra, é preciso fixar o erro máximo desejado ε , com algum grau de confiança $1 - \alpha$, 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, 2005).

Embasamento estatístico

A estatística passou a ser usada diariamente para explicar resultados de pesquisa de forma simples e dinâmica. Tomou força no século XX, mas já era utilizada pelos povos antigos. Mas é nas organizações que ela demonstra toda a sua força. Gráficos e tabelas são apresentados na exposição de resultados de análise de dados. Censos demográficos ajudam o Governo a entender melhor sua população e a organizar seus gastos com saúde e assistência social. Com a velocidade da informação a estatística passou a ser uma ferramenta essencial na produção e atuação do conhecimento.

Regressão linear simples

Os métodos que estabelecem relações lineares para os parâmetros de um modelo entre duas ou mais variáveis são denominados de métodos de regressão linear. Se for considerado o modelo funcional apresentado de forma geral na equação $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, então, o modelo de regressão é denominado de linear simples (FERREIRA, 2005).

Seleção de modelos

Neste tópico discutem-se alguns procedimentos estatísticos específicos para a seleção de variáveis na regressão, baseados na obra de Draper e Smith (1998). Suponha que se deseje 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 x_1, \dots, x_r são todas as funções de um ou mais X , ou seja, é 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 considerado para tornar o modelo útil para fins preditivos. Poderia se desejar que o modelo contivesse o máximo de Z possível para manter os erros de viés pequenos, para que valores ajustados confiáveis pudessem ser determinados. Por outro lado:

- a) Para manter a variância dos preditores razoavelmente pequena (lembrando que a variância média de \bar{y}_i é $\frac{\rho\sigma_2}{n}$, em que ρ o número de parâmetros no modelo e n é o número de observações
- b) Devido aos custos envolvidos na obtenção da informação em um grande número de Z e depois para monitorá-los, deve-se preferir que a equação inclua o mínimo de Z possível.

A implicação prática desses dois extremos é o que é comumente chamado de *seleção do melhor modelo* de regressão. Não ha um único procedimento estatístico para este fim. Se é conhecida 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 para muito problemas atinjam a mesma resposta.

Stepwise

O procedimento de regressão Stepwise começa pela escolha de uma equação contendo a melhor variável X (sozinha) e então se tenta construir um modelo com várias adições de X , um de cada vez, enquanto essas adições valerem a pena. A ordem da adição é determinada pelo uso dos valores do teste F parcial para selecionar qual variável deve entrar em seguida. O maior valor de F parcial é comparado com um valor de F a entrar (selecionado). 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 $\bar{y} = f(Z_1)$. 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 AZ_j com o maior valor, por exemplo, Z_2 é agora selecionada e uma segunda equação de regressão $Y = f(Z_1, Z_2)$ 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 parciais é então comparado com um quantil apropriado de F, F de saída, e a variável preditora correspondente é retida na equação ou rejeitada de acordo com o teste se for significativo ou não significativo.

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 critério do F parcial para cada variável na regressão à todo momento é calculado, e o menor desses F 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 estão 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 parciais das que estão no modelo. Se não passar, é feita mais uma tentativa de remoção. Eventualmente (a menos que os níveis de α para entrada e saída sejam mal escolhidos, a ponto de promover um efeito cíclico), quando nenhuma variável no modelo atual pode ser removida e a próxima melhor candidata não consegue garantir seu lugar na equação, o processo para a 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 e Smith (1998) dizem que esse é um dos melhores procedimentos de seleção de variáveis e recomendam seu uso. Dizem que ele faz um uso econômico dos recursos computacionais e que evita trabalhar com mais X do que necessário a medida que melhoram 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 e Smith (1998).

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 devem ser determinados. Um dos

métodos que pode ser utilizados, 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).

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

Teste do R^2

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, a estatística abaixo é conhecida

$$F(x) = \frac{(n-p)R^2}{p(1-R^2)}, \text{ que é distribuída como uma}$$

estatística de Fisher com $n - p$ e p graus de liberdade se o modelo não se ajusta bem aos dados (hipótese H_0).

Objetivos

Esse trabalho baseia-se, para tanto, nos resultados de uma investigação de variáveis objetiva que procura validar uma nova classificação socioeconômica de renda familiar para a cidade de Pontes e Lacerda-MT, inspirada nas contribuições do consumo mensal de energia elétrica (KWh), água (m^3) e área construída (m^2). Este modelo traduz um esforço de aperfeiçoamento da técnica de estimar a

renda familiar com maior precisão e menos constrangimento para os entrevistados.

METODOLOGIA

Em outubro de 2006, na cidade de Pontes e Lacerda-MT (Figura 1), uma pesquisa foi realizada em domicílios particulares (familiares) para investigar a renda familiar mensal (R\$) e o consumo médio de energia (KWh), o consumo mensal médio de água (m^3) e área construída (m^2) e o Critério Brasil.



Figura 1 – Foto de satélite do município de 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 $15^{\circ}13'34''$ sul e a uma longitude $59^{\circ}20'07''$ oeste, estando a uma altitude de 254 metros. Sua população estimada em 2004 era de 40 830 habitantes. Possui uma área de $13.169,5 \text{ km}^2$.

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.

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.

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.

Planejou-se um esquema amostragem capaz de cobrir toda a área urbana da cidade. Primeiramente, foi feita uma amostragem piloto, sorteando-se 10 bairros e amostrando-se, em média, 35 domicílios por bairro. Partiu-se da divisão da cidade em bairros, de acordo com a secretaria de planejamento do município. A Tabela 1 contém o número de domicílios amostrados por bairro, a média e a variância da renda. A amostra piloto foi feita de forma aleatória sistemática, dentro de cada bairro sorteado.

Tabela 1 - Amostra piloto de 10 bairros de Pontes e Lacerda MT.

Bairro	Média	Variância	Domicílios
1	1.658,33	1.519,84	37
2	2.451,00	5.459,88	33
3	2.708,89	3.165,21	36

4	647,70	172,78	44
5	2.168,00	586,01	40
6	1.568,70	2.745,07	40
7	912,30	632,33	39
8	2.068,60	2.886,06	40
9	1.089,70	1.397,46	29
10	949,60	464,82	29

Com a amostragem piloto, alguns parâmetros da variável renda da população puderam ser estimados, tais como: média, mediana, moda e variância da renda. Considerando as variâncias da renda iguais ao longo dos bairros, calculou-se a variância amostral combinada (S_p^2), dada por

$$S_p^2 = \frac{S_1^2(n_1 - 1) + S_2^2(n_2 - 1) + \dots + S_{10}^2(n_{10} - 1)}{\sum_{i=1}^{10} n_i - 10}.$$

Com a estimativa da média, a variância combinada e considerando-se um coeficiente de confiança de 95%, definiu-se o tamanho da nova amostra, tal que se cometesse um erro de apenas 5% da média:

$$n = \frac{z_{\alpha/2}^2 S_p^2}{e^2},$$

em que n é o tamanho da nova amostra, $z_{\alpha/2}^2$ é o quantil $\alpha/2$ da distribuição normal padrão, para $\alpha = 0,05$; S_p^2 é a variância amostral combinada; e e é o erro que se pretende cometer em relação à média, $e = 5\% \bar{x}$.

Informações do estudo piloto sugerem que a renda varia em função da distância deles ao centro da cidade. Por isso, os n novos bairros

foram arbitrariamente divididos em três grupos (estratos):

- bairros afastados do centro (menor classe econômica);
- bairros próximos do centro (maior classe econômica);
- bairros intermediários (classe econômica intermediária).

Optou-se, posteriormente, por uma amostra em que os bairros da cidade pudessem participar com pesos proporcionais ao tamanho de suas populações no censo municipal de 2006 realizado nessa cidade, pela Prefeitura Municipal. Como base para compor o banco de dados, adotou-se esquema de amostragem aleatória estratificada informal, constituindo três grupos diferentes por localização dos bairros e dentro de cada grupo (conglomerado), fez-se o sorteio para compor a amostra com 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, onde o tamanho da amostra é função:

- 1) do(s) parâmetro(s) a estimar;
- 2) do nível de confiança desejável;
- 3) do erro tolerável ou índice de precisão escolhidos; e
- 4) 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 população 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 estratos 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 estratificações 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 obtido pelo Censo Demo gráfico 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

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

Tamanho da amostra

O cálculo do tamanho da amostra é uma das partes do projeto de pesquisa na qual são descritos os itens que foram levados em consideração para que o estudo tivesse o tamanho da amostra utilizado.

O tamanho da amostra foi determinado pela variabilidade de domicílios, fornecido pela

amostra piloto com um nível de confiança de 95% e um erro de 5% da média.

Todas as análises e gráficos foram feitas no software estatístico R (R CORE TEAM, 2014).

RESULTADOS

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

De acordo com estudo piloto, pesquisou-se a renda de 367 chefes de família de Pontes e Lacerda-MT, determinando-se a estimativa da renda média por classe econômica, a partir dos dados explicitados. Calculou-se a variância combinada S_p^2 para o dimensionamento do

tamanho da amostra, que totalizou 1088 domicílios. Portanto, voltou-se a campo para a complementação da amostra, ou seja, para serem observados, mais 735 domicílios.

De acordo com a interpretação gráfica da Figura 2 da amostra piloto de dez bairros da cidade de Pontes e Lacerda-MT, observam-se três grupos que se diferenciam entre si, por renda. Ainda verificou-se que os bairros mais centralizados têm rendimentos médios maiores que os bairros mais afastados do centro da cidade, o que, de um modo geral, é verificado em diversas cidades. Uma ideia para contornar o problema da estratificação por renda é a estratificação informal dos bairros quanto a distância dos bairros em relação ao centro da cidade.

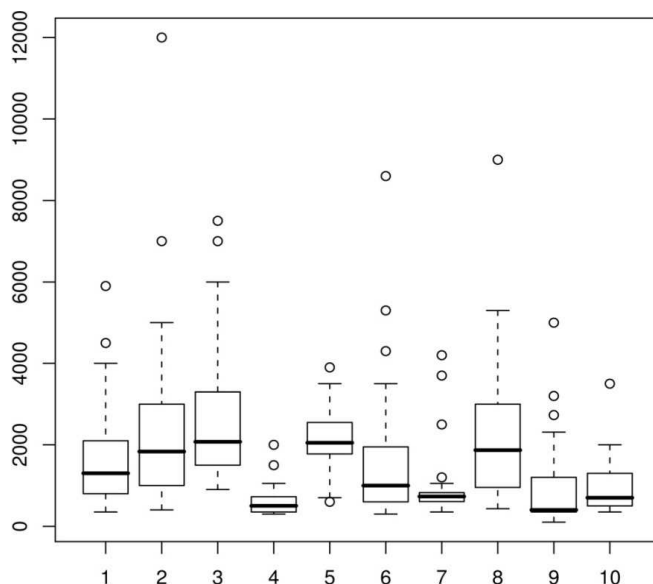


Figura 2 - Amostra piloto que define 3 grupos diferentes de renda de acordo com a localização referente ao centro da cidade

A Figura 3 mostra as proporções estimadas da população em Pontes e Lacerda-MT, por

classe econômica, e seus respectivos limites superiores dos intervalos de confiança 95%.

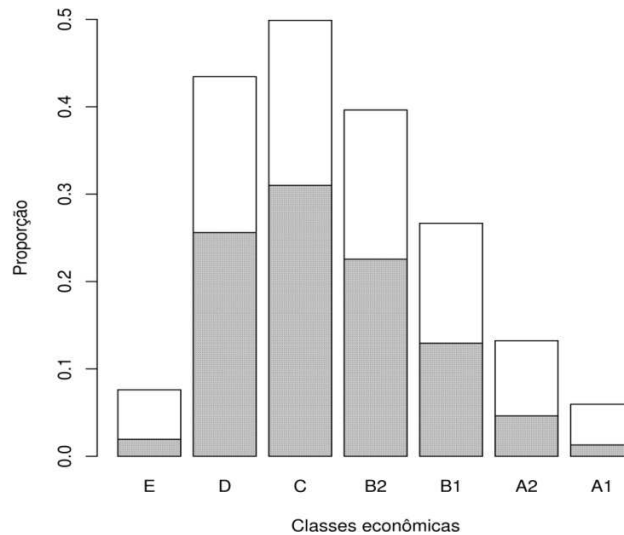


Figura 3 - Proporções das classes econômicas em Pontes e Lacerda

De acordo com a Figura 3, as classes A1, A2 e E, apresentam intervalos sobrepostos, logo podem ser consideradas estatisticamente iguais. Além disso, elas apresentam as menores proporções, enquanto as demais classes referenciadas possuem o maior percentual.

De acordo com o gráfico 4 observou-se ainda que a média não consegue dar uma ideia da renda per capita da população de Pontes e Lacerda-MT, logo, não é uma boa medida da realidade da população. De acordo com a Figura 4, percebe-se que a média expressa um

valor muito acima da renda modal. O mesmo acontece em relação a mediana, porém em menor intensidade. A melhor medida de posição nesse caso é a moda, isso fica claro no histograma da renda em Pontes e Lacerda, que destaca a assimetria da distribuição dessa variável e a não coincidência da média, mediana e moda.

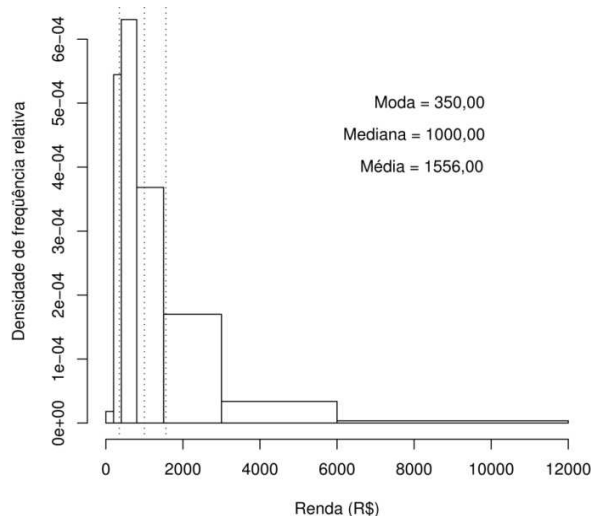


Figura 4 - Histograma da renda

Seleção de modelos

O modelo objetivo a seguir descrito foi obtido por Stepwise

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

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} e b_{i18} são variáveis dicotômicas que indicam a presença (1) ou ausência (0) da i -ésima residência nos bairros 2, 3, 5, 10, 12, 18, respectivamente.

Os procedimentos Stepwise, Backward e Forward foram usados para a seleção dos melhores modelos objetivo e híbrido. Stepwise e Backward levaram ao mesmo modelo, enquanto que Forward levou a um modelo diferente. Por isso, foi feito um teste F a 5% de

significância 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.

De acordo com a Tabela 2 o teste R^2 , com 5% de significância, observado na coluna 5, para os 23 bairros da cidade de Pontes e Lacerda, defini-se os resultados desse teste onde foram analisados os valores-p para cada bairro, destacando-se em vermelho os R^2 que não foram significativos.

Tabela 2 - Coeficiente de determinação ajustado R_{aj}^2 , erro quadrático médio na renda e valor-p do teste de significância dos coeficientes de determinação

Bairro	Modelo	R_{aj}^2	$\sqrt{QMR_{renda}}$	Valor-p
1	Ob	0,5554	773,2161	0,021423
	Hb	0,5554	773,2126	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	0,95811	210,4092	0,00000
	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
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	665,125	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

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, significando que os valores preditos pelo modelo objetivo, não ficam distantes dos valores esperados.

De acordo com a coluna 3 da Tabela 5, observa-se que o R^2 ajustado nos 23 bairros de Pontes e Lacerda, para os modelos de regressão baseados em covariáveis objetivas (Ob) e o modelo híbrido (Hb), mostra um índice de desvio equivalente, ou erram muito ou pouco dentro de um mesmo bairro, o que pode ser considerado um bom resultado, significando

que os valores preditos pelo modelo objetivo, não fica distantes do valores esperados, sabido que o modelo híbrido tem efeito do questionário Critério Brasil.

De acordo com comparações feitas na Figura 5, entre as classes estimadas por meio do Critério Brasil (azul), modelo objetivo (verde), modelo híbrido (vermelho) e classes econômicas reais (preto) a Figura 5 mostra formas conceituais de quatro curvas. É possível ajustar um modelo baseados 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 2.

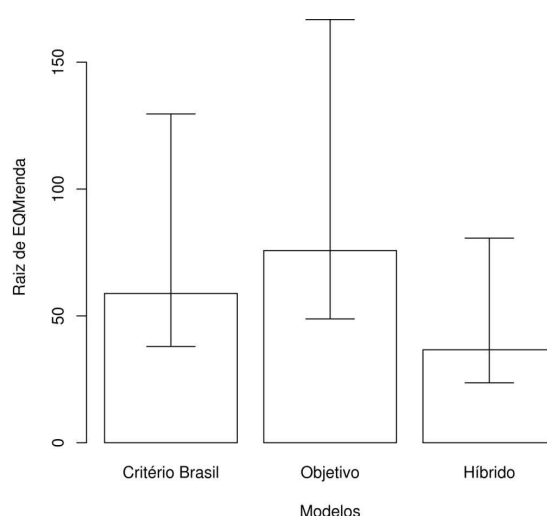


Figura 5 - Intervalo de confiança.

De acordo com a Figura 5, 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.

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 BIBLIOGRÁFICAS

- BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. 5. ed., São Paulo: Editora SARAIVA, 2003. 525 p.
- DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3rd Ed., 1998, 706 p., New York : John Wiley.
- ENGEL, J. F.; BLACKWELL, R. D.; MINIARD, P. W. **Comportamento do consumidor**. Rio de Janeiro: LTC, 1995. 641p
- FAUZE, N. M. Análise crítica dos métodos de estratificação social utilizados em marketing e pesquisas de marketing. **Anais do 2º SEMEAD**, Disponível em: <<http://fauze.com.br/artigo03.htm>>. Acesso: 26 de outubro de 2006, 1997.
- FERREIRA, D. F. **Estatística Básica**, Lavras : Editora UFLA. 2005. 664 p.
- 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.
- MATTOS, A. Institutos vão mudar definição de classe social **Em Pauta** Jornal folha de São Paulo, SP, 07 de jan de 2007
- R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2014.
- RIBEIRO, C. A. C.; SCALON, M. C. **Mobilidade de Classe no Brasil em Perspectiva Comparada**, Dados, v.44, n.1, Disponível em: <http://www.scielo.br/scielo.php?script=sci_text&pid=S0011-

52582001000100004\&lng=ptr\&nrm=iso>.

Acesso em: 26 de outubro de 2006. 36p. 2001.

SÃO PAULO Seminário em administração

Fórum regional de desenvolvimento

sustentável São Paulo, 1997 Disponível em:

<<http://www.ponteselacerda.mt.gov.br/cidade/localizacao.htm>>.

Acessado em: 06 jan 2007.

UNITED STATES CENSUS BUREAU.

Disponível em:

<http://pt.wikipedia.org/wiki/United_States_Census_Bureau>. Acesso em: 30 mai 2014.