

UNIVERSIDADE FEDERAL FLUMINENSE
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
CURSO DE ESTATÍSTICA

Marcela Martins da Serra Vilela Pinto

**Distribuição da riqueza imobiliária urbana quanto à
classe média no Brasil em 2011**

Niterói

2014

Marcela Martins da Serra Vilela Pinto

Distribuição da riqueza imobiliária urbana quanto à classe média no Brasil em 2011

Monografia apresentada ao Curso de Estatística da
UFF, como requisito para a obtenção do grau de BA-
CHAREL em Estatística.

Orientadora: Márcia Marques de Carvalho
Doutorado em Economia

Niterói
2014

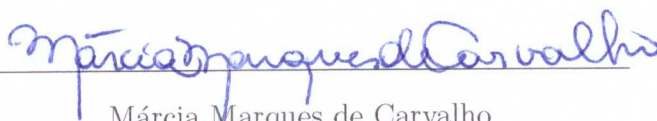
Marcela Martins da Serra Vilela Pinto

Distribuição da riqueza imobiliária urbana quanto à classe média no Brasil em 2011

Monografia apresentada ao Curso de Estatística da
UFF, como requisito para a obtenção do grau de BA-
CHAREL em Estatística.

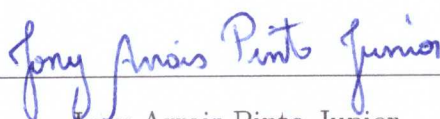
Aprovado em Janeiro de 2014

BANCA EXAMINADORA




Márcia Marques de Carvalho

Doutorado em Economia



Jony Arrais Pinto Junior

Mestrado em Estatística



Jéssica Kubrusly

Doutorado em Matemática

Dedico este trabalho, aos meus familiares e principalmente aos meus pais Andréa e Reynaldo, a quem devo toda esta caminhada, a eles o meu reconhecimento e carinho. Dedico também ao meu namorado Guilherme, pela sua ajuda e compreensão.

Resumo

Com o crescimento da classe média no Brasil, assuntos a respeito desta se tornaram cada vez mais importante. A compra da casa própria é também muito importante para o brasileiro. Nesse trabalho iremos analisar como se distribui a riqueza imobiliária quanto a classe média, usando o modelo de regressão logística e os microdados da PNAD (Pesquisa Nacional por Amostra de Domicílios) de 2011 do IBGE (Instituto Brasileiro de Geografia e Estatística).

Para definir a classe média foram utilizados dois critérios, um deles proposto pela SAE (Secretaria de Assuntos Estratégicos da Presidência da República) que usa a renda familiar per capita para determinar as classes, estabelecendo a renda da classe média entre R\$277,00 e R\$971,00, esse critério é utilizado pelos economistas. E o outro critério usa como principal meio de avaliação a ocupação do chefe, pois esta é capaz de vincular a educação à renda, este é o critério utilizado pelos sociólogos. A fim de entender como se comportam essas duas definições em relação a riqueza imobiliária, estas serão comparadas.

Palavras-chaves: Riqueza Imobiliária, Classe Média, Modelo de regressão logística

Agradecimentos

Agradeço primeiramente a minha orientadora Márcia Marques de Carvalho pelo o apoio, dedicação e principalmente pela paciência durante a realização desse projeto.

Agradeço aos meus professores e amigos durante essa longa caminhada.

Muito obrigada.

Sumário

Lista de Tabelas	6
1 Introdução	8
1.1 Objetivos	8
2 Riqueza Imobiliária e Classe Média	10
2.1 Riqueza Imobiliária	10
2.2 Classe Média	17
2.2.1 Critério utilizado pelos economistas	18
2.2.2 Critério utilizado pelos sociólogos	21
3 Metodologia: o modelo de regressão logística	26
3.1 Modelos Lineares Generalizados	27
3.2 Especificação do modelo de regressão logística	29
3.3 Estimação pelo Método da Máxima Verossimilhança	30
3.4 Interpretação	32
3.5 Inferência no Modelo	33
3.6 Qualidade do ajuste	35
3.7 Técnicas de diagnóstico	37

4	Resultados	40
4.1	Descrição das variáveis	40
4.2	Especificação do modelo	45
4.3	Modelos estimados	45
4.4	Qualidade do ajuste	50
4.5	Técnicas de Diagnóstico	51
5	Conclusão	53

Lista de Tabelas

2.1	Distribuição dos imóveis residenciais segundo condição de ocupação e situação censitária - Brasil 2011	11
2.2	Distribuição Percentual dos imóveis residências segundo condição de ocupação e situação censitária - Brasil 2011	11
2.3	Riqueza Imobiliária Urbana - Brasil 2011	12
2.4	Distribuição dos domicílios urbanos segundo características dos imóveis - Brasil 2011	12
2.5	Distribuição dos domicílios urbanos segundo características da pessoa de referência - Brasil 2011	15
2.6	Distribuição dos domicílios urbanos segundo tipo de família - Brasil 2011	16
2.7	Classes de Renda segundo o critério da Comissão da SAE no Brasil: 2001 a 2012	20
2.8	Classe média econômica	21
2.9	Distribuição do nível educacional segundo as classes econômicas	21
2.10	Evolução da Estratificação Social no Brasil: 1973 - 2006	23
2.11	Classe média sociológica	24
2.12	Distribuição da renda mensal per capita das classes sociológicas	24
2.13	Distribuição do nível educacional segundo as classes sociológicas	25

3.1	Tabela de classificação	37
3.2	Possíveis valores para t_{Si} , t_{Di} e LD_i segundo as probabilidades ajustadas . .	39
4.1	Descrição das variáveis	41
4.2	Cruzamento das variáveis explicativas com a variável resposta	42
4.3	Estimativa dos parâmetros referentes ao modelo logístico utilizando as classes econômicas	47
4.4	Estimativa dos parâmetros referentes ao modelo logístico utilizando as classes sociológicas	49
4.5	Qualidade do ajuste	50
4.6	Tabela de classificação	50

1 Introdução

Os estudos sobre a distribuição pessoal ou familiar do estoque de riqueza imobiliária enfatizam seu papel relevante na determinação da desigualdade e na definição da estrutura de estratificação social (TAFNER & CARVALHO (2007)).

A respeito dos imóveis residenciais, o total destes no Brasil em 2011 era de mais de 61 milhões. Desses imóveis, 86,2% se encontram na zona urbana e 13,8% na zona rural. Esses imóveis são tipicamente de paredes de alvenaria (91,9%), com em média 1,89 cômodos servindo de dormitórios. Quanto à condição de ocupação, 70,1% são próprios já pagos, 4,7% são próprios ainda pagando e 17,3% são alugados, 7,4% são cedidos e 0,5% estão em outras condições, segundo a Pesquisa Nacional por Amostra de Domicílio (PNAD - 2011). O imóvel próprio é considerado uma riqueza imobiliária porque além de ser um ativo material, ele permite que o dinheiro que deveria ser gasto com aluguel seja utilizado com outras coisas, como por exemplo, educação e plano de saúde para os filhos.

A pergunta que esse trabalho pretende responder é: quem tem riqueza imobiliária urbana no Brasil? A hipótese que será testada é se quem tem riqueza imobiliária urbana no Brasil é a classe média. Esta hipótese será verificada utilizando duas definições de classe média: a nova classe média, utilizando o critério dos economistas baseado na renda e o critério dos sociólogos, utilizando as ocupações.

1.1 Objetivos

O objetivo geral desse trabalho é identificar como está distribuída a riqueza imobiliária urbana quanto a classe média no Brasil.

Objetivos específicos:

- Traçar o perfil das pessoas que possuem riqueza imobiliária urbana;
- Comparar a riqueza imobiliária através dos dois critérios distintos utilizados para definir a classe média;
- Examinar os determinantes da riqueza imobiliária no Brasil segundo, gênero, faixa etária, nível de escolaridade, raça e tipo de família) utilizando o modelo de regressão logística.

A base de dados utilizada tem abrangência nacional e é elaborada e coletada pelo IBGE. Além disso, ela possui informações sobre o domicílio, pessoas, trabalho e rendimentos, o que permite atingir os objetivos específicos.

A monografia está organizada em quatro capítulos, além desta introdução. O capítulo dois apresenta algumas informações sobre riqueza imobiliária e critérios para definir a classe média brasileira. O capítulo três apresenta a metodologia utilizada para o desenvolvimento do trabalho. O capítulo quatro expõe os resultados obtidos e também os modelos estimados. E por último, o capítulo cinco que relata as principais conclusões do trabalho.

2 Riqueza Imobiliária e Classe Média

Nesse capítulo, vamos traçar o perfil das pessoas que possuem riqueza imobiliária e também traçar o perfil da classe média. Além disso, vamos definir os dois critérios utilizados para definir a classe média. Será um capítulo descritivo, a partir da base de dados da PNAD de 2011.

2.1 Riqueza Imobiliária

A riqueza não pode ser definida de forma absoluta. Seu significado é distinto não apenas de um país para outro, mas também entre diferentes membros de uma mesma nação. A riqueza diz respeito a quantidade de bens que uma pessoa possui, pode ser em dinheiro, patrimônios móveis, imóveis e bens que se movem por si próprios, como o gado. Também é avaliada pela possibilidade do indivíduo ter acesso a serviços essenciais, como alimentação, educação, saúde, entre outros .

Nesse trabalho será considerada somente a riqueza imobiliária residencial no Brasil. De acordo com a Tabela 2.1, em 2011, 74,8% dos imóveis residenciais eram próprios, seguido dos alugados com 17,3%.

Com o auxílio da Tabela 2.2 é possível notar que, o percentual de imóveis de todos os tipos de condição de ocupação são maioria na zona urbana. O que mais chama a atenção é o percentual de imóveis alugados nessa área (97,8%).

Tabela 2.1: Distribuição dos imóveis residenciais segundo condição de ocupação e situação censitária - Brasil 2011

Situação Censitária	Próprios ¹	Alugados	Cedidos	Outros	Total
Urbano	39.147.587	10.352.686	3.100.765	200.177	52.801.215
Rural	6.695.307	232.484	1.465.834	97.199	8.490.824
Total	45.842.894	10.585.170	4.566.599	297.376	61.292.039
%	74,8	17,3	7,4	0,5	100,0

Fonte:IBGE/PNAD de 2011 Tabulação do autor

¹ Próprios = próprios pagos ou ainda pagando

Tabela 2.2: Distribuição Percentual dos imóveis residências segundo condição de ocupação e situação censitária - Brasil 2011

Situação Censitária	Próprios	Alugados	Cedidos	Outros	Total
Urbano	85,4	97,8	67,9	67,3	86,1
Rural	14,6	2,2	32,2	32,7	13,9
Total	100,0	100,0	100,0	100,0	100,0

Fonte:IBGE/PNAD de 2011 Tabulação do autor

Como os imóveis urbanos e rurais diferem muito em extensão, nesse estudo serão utilizados somente os imóveis urbanos.

Para conhecer os atributos dos domicílios no Brasil, foram selecionadas as seguintes variáveis: tipo de imóvel, material das paredes, material da cobertura, número de dormitórios, existência de banheiro, proveniência da água, forma de escoadouro, destino do lixo e forma de iluminação.

Para todas as análises deste trabalho a partir daqui, serão considerados domicílios próprios aqueles que já estiverem quitados ou aqueles que ainda estão sendo pagos.

A Tabela 2.3 apresenta o total de pessoas de referência do domicílio (“chefes”) que possuem

riquezas imobiliária no Brasil em 2011: são aproximadamente 36 milhões de pessoas.

Tabela 2.3: Riqueza Imobiliária Urbana - Brasil 2011

Condição da ocupação	Domicílios	
	N	%
Próprios	36.764.410	73,7
Alugados/Cedidos/Outros	13.078.446	26,3
Total	49.842.856	100,0

Fonte: IBGE/PNAD de 2011 Tabulação do autor

Pela Tabela 2.4, é possível notar que no Brasil urbano, 87% dos domicílios são casas e 12,5% são apartamento. Cabe ressaltar que entre os domicílios próprios o percentual de casas é maior se comparados com o total.

Com relação ao material de construção, podemos destacar que a maior parte dos domicílios do Brasil são de parede de alvenaria e com telha como material de cobertura.

Quanto ao número de dormitórios, a maioria, tanto no total como nos próprios, possuem dois cômodos servindo de dormitório.

As variáveis existência de banheiro, forma de abastecimento da água, destino do lixo, forma de iluminação, estão distribuídas de forma bastante semelhante para total e para domicílios próprio.

Em suma, as características dos imóveis próprios são semelhantes ao total de imóveis no Brasil.

Tabela 2.4: Distribuição dos domicílios urbanos segundo características dos imóveis - Brasil 2011

Características do imóvel	Domicílios (%)	
	Total ²	Próprios ³
Tipo de Imóvel		
Casa	87,0	89,2

Apartamento	12,5	10,6
Cômodo	0,5	0,2
Total	100,0	100,0
Material das paredes		
Alvenaria	94,3	94,1
Madeira Aparelhada	4,9	5,1
Outros	0,8	0,8
Total	100,0	100,0
Material da cobertura		
Telha	70,5	71,6
Laje	27,5	26,4
Outros	2,0	2,0
Total	100,0	100,0
Número de dormitórios		
1	34,8	31,7
2	44,7	45,4
3 ou mais	20,5	22,9
Total		
Existência de banheiro sanitário		
Sim	99,0	99,0
Não	1,0	1,0
Total	100,0	100,0
Forma de abastecimento da água		
Rede geral	92,4	92,2
Poço ou nascente	5,2	5,4
Outros	2,4	2,4
Total	100,0	100,0
Forma de escoadouro		
Rede de esgoto	63,6	61,7
Fossa rudimentar	11,8	12,2

Fossa não ligada à rede	13,4	14,2
Fossa ligada à rede	8,6	9,2
Outros	2,6	2,7
Total	100,0	100,0
Destino do lixo		
Coletado diretamente	93,9	93,0
Coletado indiretamente	5,0	5,1
Queimado ou enterrado	1,2	1,4
Outros	0,5	0,5
Total	100,0	100,0
Forma de iluminação		
Elétrica (rede, gerador solar)	99,9	99,9
Outras	0,1	0,1
Total	100,0	100,0

Fonte: IBGE/PNAD de 2011 Tabulação do autor

² Total = próprios já pagos, próprios ainda pagando, alugados, cedidos, outros

³ Próprios já pagos e ainda pagando

Com relação ao perfil da pessoa de referência dos domicílios urbanos (Tabela 2.5), observamos uma maior incidência de chefes com até ensino fundamental incompleto e mais velhos como proprietários de domicílios.

Em suma, o perfil das pessoas que possuem riqueza imobiliária urbana é dado pelos chefes de domicílios de baixa instrução com idade entre 40 e 59 anos, que trabalham com carteira assinada.

Tabela 2.5: Distribuição dos domicílios urbanos segundo características da pessoa de referência - Brasil 2011

Características da pessoa de referência	Domicílios (%)	
	Total	Próprios
Nível de instrução		
Até Fundamental incompleto	43,4	47,4
Fundamental completo	15,2	14,3
Médio completo	28,5	23,8
Superior incompleto	3,3	2,9
Superior completo	11,8	11,7
Total	100,0	100,0
Renda domiciliar per capita		
Até $\frac{1}{2}$ salário mínimo	18,5	18,1
Mais de $\frac{1}{2}$ até 1 salário mínimo	26,7	27,0
Mais de 1 até 2 salários mínimos	27,5	27,4
Mais de 2 até 3 salários mínimos	9,7	9,4
Mais de 3 até 5 salários mínimos	6,5	6,4
Mais de 5 salários mínimos	5,8	5,7
Sem declaração	5,4	6,0
Total	100,0	100,0
Carteira assinada		
Sim	72,8	73,3
Não	27,2	26,7
Total	100,0	100,0
Idade		
Até 19 anos	0,7	0,4
De 20 a 29 anos	11,9	8,0
De 30 a 39 anos	21,6	19,3
De 40 a 49 anos	22,6	23,3

De 50 a 59 anos	19,3	21,3
De 60 a 64 anos	7,4	8,4
65 anos ou mais	16,4	19,3
Total	100,0	100,0
Aposentado		
Sim	20,3	23,9
Não	79,7	76,1
Total	100,0	100,0

Fonte: IBGE/PNAD de 2011 Tabulação do autor

Com relação ao tipo de família com maior incidência entre os domicílios próprios (Tabela 2.6), pode-se destacar o de casais com todos os filhos maiores de 14 anos.

Tabela 2.6: Distribuição dos domicílios urbanos segundo tipo de família - Brasil 2011

Tipos de família	Domicílios (%)	
	Total	Próprios
Casal sem filhos	18,1	18,6
Casal com todos os filhos menores de 14 anos	20,4	18,3
Casal com todos os filhos de 14 anos ou mais	17,7	20,3
Casal com filhos menores de 14 anos e de 14 anos ou mais	8,2	8,5
Mãe com todos os filhos menores de 14 anos	3,3	2,5
Mãe com todos os filhos de 14 anos ou mais	10,5	11,3
Mãe com filhos menores de 14 anos e de 14 anos ou mais	2,0	1,8
Outros tipos de família	19,8	18,6
Total	100,0	100,0

Fonte: IBGE/PNAD de 2011 Tabulação do autor

2.2 Classe Média

É necessário caracterizar a classe média para que se possa relatar as mudanças na distribuição de rendas que ocorreram e irão ocorrer ao longo dos anos no Brasil e também para ajudar na criação de políticas públicas voltadas para essa classe. E para saber o real tamanho e as evoluções desta, é preciso identificar onde começa e termina a classe média brasileira.

Existem diversos critérios para delimitar essa classe. Abaixo algumas das definições:

- Quartis da distribuição de renda;
- Entorno da Mediana da renda;
- Polarização;
- Ponto Neutro e Linha de Pobreza;
- Consumo de calorias e padrão de despesas;
- Vulnerabilidade;
- Inserção no mercado de trabalho;
- Capital Cultural.

Nesse trabalho serão utilizados dois critérios envolvendo essa classe: um deles, proposto pela SAE (Secretaria de Assuntos Estratégicos, vinculada à Presidência da República), utilizado pelos economistas, leva em conta o grau de vulnerabilidade, usando a renda familiar per capita para determinar as classes; e o outro critério, utilizado pelos sociólogos que usa como principal meio de avaliação as ocupações.

Esses critérios foram escolhidos, pelo fato de serem os mais utilizados em trabalhos e pesquisas a respeito da classe média. O critério sociológico é mais antigo e mais tradicional, enquanto o da renda está expandindo cada vez mais. Além desses, existe um outro método bastante utilizado para pesquisas de mercado e principalmente pelo Instituto Brasileiro de

Opinião Pública e Estatística (IBOPE), que consiste em alocar a população nas classes de acordo com a posse de bens e acesso a serviços. Esse último, utiliza os dados da Pesquisa de Orçamentos Familiares (POF), do Instituto Brasileiro de Geografia e Estatística (IBGE) (UCHÔA & KERSTENETZKY(2012)).

2.2.1 Critério utilizado pelos economistas

O Brasil vem passando por diversas mudanças econômicas e sociais importantes e possivelmente a redução da extrema pobreza é a maior delas. Tal redução se deve principalmente, ao fato do desenvolvimento e implementação de um conjunto de programas sociais eficazes, com a inserção feminina no mercado de trabalho também contribuindo para essa mudança.

Em decorrência da diminuição da pobreza e do crescimento da renda dos mais pobres, houve um aumento da população na classe intermediária. Já que a renda das famílias da classe média e da classe alta não sofreram grandes mudanças, a classe média ficou dilatada, passando a representar mais da metade da população brasileira. O problema agora é responder a seguinte pergunta: Os programas sociais já implementados irão suprir a necessidade da nova classe média? Não é muito difícil responder tal questionamento, já que os interesses dessas duas classes são bem diferentes. Enquanto a classe mais pobre mira o presente e usa estratégias para sobreviver a tal situação, a classe média pensa no futuro, tentando encontrar maneiras para se manter nessas melhores condições (SAE, 2012).

Como é necessário que as políticas públicas sejam voltadas para a classe média e não mais para a classe baixa, é preciso um conhecimento mais profundo das percepções, valores, atitudes e principalmente dos interesses dessa classe. É possível identificar esses anseios através da análise de alguns elementos, como: modelo de consumo (gastos essenciais e supérfluos), bens e forma de inserção no mercado de trabalho.

Alguns dados importantes da SAE - 2012:

- Em nível mundial o Brasil é o quinto país mais populoso, abaixo apenas da China,

Índia, Estados Unidos e Indonésia. Assim, se a classe média brasileira fosse um país, ela seria o 12^o país mais populoso do mundo;

- A população brasileira em idade ativa que está ocupada somam 58%, enquanto na classe média essa proporção já chega a 61%;
- Durante a última década, a classe média cresceu bastante, passou de 38% em 2002 para 53% em 2012;
- De 2002 a 2012, 21% da população brasileira passou da classe baixa (pobres e vulneráveis) para à média, enquanto da classe média para alta apenas 6% da população;
- Em 2013, quase oito milhões de brasileiros da classe média, sonham com a casa própria.

Como visto anteriormente, existem alguns critérios que podem ser utilizados para definir a classe média. O critério utilizado pela SAE é o grau de vulnerabilidade, que foi definido como sendo a probabilidade de retornar (ou permanecer) à condição de pobreza em algum momento dos próximos cinco anos. Tal probabilidade foi obtida através da observação dos movimentos de ascensão e queda de renda da população brasileira nos últimos anos (usando dados da PNAD). Para calcular esta probabilidade é necessário ter uma linha de pobreza estipulada, nesse caso será usada a linha de pobreza de R\$140,00 por mês, utilizada no contexto do programa Brasil Sem Miséria.

Para encontrar o ponto de corte de cada uma das classes, a população foi dividida em 100 partes de acordo com a renda domiciliar per capita. Ordenou-se os dados de maneira crescente pela renda, obtendo-se o que se chama de percentis da distribuição. Para cada uma das partes tem-se associada uma probabilidade de retorno (ou permanência) à condição de pobreza.

Após esse procedimento, aplicou-se o método da polarização, que consiste no seguinte problema estatístico: como dividir uma população em grupos, de forma que cada grupo seja o mais homogêneo possível dentro dele e o mais heterogêneo fora dele. Então, foi verificado os pontos de corte que dividem a população brasileira em 3 classes, baixa, média e alta. Através

de simulações os resultados foram os seguintes: a classe baixa termina no 34^o percentil, a classe média está entre o 34^o e 82^o percentil e a classe alta do 82^o em diante.

Com isso, a Comissão da SAE chegou nos seguintes valores para a classe média, renda domiciliar per capita entre R\$277,00 e R\$971,00 em 2011.

Tabela 2.7: Classes de Renda segundo o critério da Comissão da SAE no Brasil: 2001 a 2012

Classe Média	Limite inferior (R\$)	Limite superior (R\$)
2001	140	491
2002	154	538
2003	183	640
2004	194	679
2005	206	722
2006	214	749
2007	221	775
2008	234	819
2009	248	867
2010	261	914
2011	277	971
2012	291	1019

Fonte: Secretária de Estudos Estratégicos (SAE) - 2012 Tabulação do autor

Aplicando esse intervalo aos dados da PNAD de 2011 encontra-se que 52,2% dos domicílios são considerados classes média.

Estudos indicam que o acesso à classe média foi devido ao aumento do nível educacional das pessoas. A Tabela 2.9 apresenta o nível educacional dos chefes de domicílio segundo a classe média econômica.

Podemos observar pela Tabela 2.9 que conforme o nível educacional aumenta a porcentagem de indivíduos que pertencem a classe baixa vai diminuindo e de forma inversa, quanto mais o nível educacional aumenta a porcentagem de indivíduos que pertencem a classe alta

Tabela 2.8: Classe média econômica

Valor	N	%
Baixa	9.915.939	19,9
Média	52.801.215	52,2
Alta	13.924.524	27,9
Total	49.842.856	100,0

Fonte: IBGE/PNAD de 2011 Tabulação do autor

aumenta. Já a classe média, tem um percentual entre 50 e 60, exceto nos níveis superiores, com uma porcentagem bem mais baixa, entre 20% e 40%.

Tabela 2.9: Distribuição do nível educacional segundo as classes econômicas

Nível educacional	Classe econômica			Total
	Baixa	Média	Alta	
Sem instrução	43,5	49,3	7,1	100,0
Ensino fundamental incompleto	30,5	56,0	13,5	100,0
Ensino fundamental completo	20,8	59,8	19,4	100,0
Ensino médio incompleto	23,2	59,7	17,1	100,0
Ensino médio completo	12,0	56,7	31,3	100,0
Ensino superior incompleto	4,1	38,8	57,1	100,0
Ensino superior completo	1,7	20,2	78,1	100,0
Total	19,9	52,2	27,9	100,0

Fonte: IBGE/PNAD de 2011 Tabulação do autor

2.2.2 Critério utilizado pelos sociólogos

Existem pelo menos dois conceitos que definem as classes sociais pela ótica sociológica. Um deles, utiliza o conceito marxista, que se refere a um grupo estruturalmente bem delimitado, consciente de si e dotado de estilos de vida, padrões de comportamento e projetos de sociedade diferenciados em relação as outras classes. Numa vertente oposta, está o conceito weberiano,

que atém-se a características objetivamente mensuráveis, como a educação, a renda e a ocupação, entendidas como atributos individuais.

Até o começo do século XX, ter completado o ensino médio era sinônimo de boas oportunidades, porém, hoje é exigido mais. A demanda por educação vem crescendo, estimulada por retornos mais altos de renda, com isso a vantagem relativa de que usufruía a classe média alta está perdendo espaço. De fato, a ascensão da nova classe média está associada à queda de discrepância educacional e de renda, o que, tornou a educação um indicador menos preciso de posição social. Porém, a educação ainda é um fator muito importante, devido a associação com a ocupação exercida e as chances de mobilidade ocupacional num futuro. O acesso a escolaridade também corresponde a uma maior participação em redes sociais e outros tipos de organizações, e assim essas pessoas passam a ter um nível mais alto de informações. Por esses fatores, existe uma grande expectativa de que a classe média venha a contribuir para a democracia e economia do Brasil.

Outra forma de definir a classe média é pela renda. Existem duas vertentes a respeito da renda, a primeira incluiria as famílias que auferem a renda média da sociedade. Porém, o problema é que variações na distribuição de renda podem modificar a concepção de classe média ao longo do tempo. E a segunda, seria o critério usando o valor absoluto que evitaria esse problema, mas excluiria também pessoas que, mesmo sendo reconhecidas como de classe média, ganham abaixo de determinada faixa de renda.

Quanto à ocupação, esse é um indicador que permite entender melhor determinadas tendências de longa duração. Mesmo que os rendimentos variem ao longo da vida, os indivíduos tentam manter, na média, o mesmo padrão de vida.

Como a posição ocupacional é, em boa parte, função de aprendizagem orientada para o desempenho de funções específicas, pode-se dizer que é a ocupação que vincula a educação à renda. Por esse fato, o critério escolhido para definir a classe média sociológica será baseado na ocupação da pessoa de referência. (SOUZA(2012))

A estratificação social na Tabela 2.10 mostra a evolução de sete grandes estratos ocupa-

cionais ao longo dos últimos 40 anos. (SOUZA(2012))

Tabela 2.10: Evolução da Estratificação Social no Brasil: 1973 - 2006

Estratos Ocupacionais	1973 ⁴ (%)	1996 (%)	2006 (%)
Baixo-inferior (Trabalhadores rurais)	32,1	20,8	18,3
Baixo-superior (Serviço doméstico e vendedores ambulantes)	16,1	13,3	12,6
Médio-inferior (Trabalhadores manuais em indústrias tradicionais e serviços e médios proprietários rurais)	23,7	26,5	27,2
Médio-ascendente (Microempresários, trabalhadores manuais em indústrias modernas)	-	9,2	9,5
Médio-médio (Técnicos, artistas e supervisores do trabalho manual)	18,4	13,3	12,9
Médio-superior (Outros profissionais e funções não manuais técnicas e administrativas)	6,3	9,9	12,1
Alto (Grandes empresários, profissionais liberais e executivos)	3,5	6,9	7,3

Fonte: IBGE, Pesquisa Nacional por Amostra de Domicílios - Tabulação de Souza(2010)

⁴ Extraído de José Pastore e Nelson do Vale Sulva, Mobilidade social no Brasil, São Paulo: Makron Books, 2000, p.61

É conveniente agregar as categorias ocupacionais em número menor de grupos, para que correspondam mais de perto ao que se percebe como classes sociais. Portanto, as classes ficariam divididas da seguinte forma: baixa (baixo superior e baixo inferior), média inferior, média ascendente, média média, média superior e alta.

Utilizando os dados da PNAD 2011 e somente os domicílios urbanos, teremos:

Tabela 2.11: Classe média sociológica

Valor	N	%
Baixa	17.134.368	42,4
Média-inferior	12.928.626	32,0
Média-ascendente	2.293.077	5,7
Média-média	2.756.653	6,8
Média-superior	3.083.029	7,6
Alto	2.170.331	5,4
Total	40.366.083	100,0

Fonte: IBGE/PNAD de 2011 Tabulação do autor

Para se ter uma ideia da renda mensal domiciliar per capita em cada classe e para comparar com o critério anterior, segue a Tabela 2.12 com a distribuição da renda em cada categoria.

Tabela 2.12: Distribuição da renda mensal per capita das classes sociológicas

Classes	Média	Desvio-Padrão	Mediana	Máximo
Baixa	603,3	611,2	466,0	25.000
Média-inferior	692,5	1.110,7	500,0	100.000
Média-ascendente	1.118,3	2.890,6	766,0	125.000
Média-média	1.371,9	1.655,2	891,0	49.600
Média-superior	2.663,8	3.016,9	1.725,0	40.000
Alta	2.440,8	3.198,9	1.500,0	70.000
Total	969,8	1.700,0	566,0	125.000

Fonte: IBGE/PNAD de 2011 Tabulação do autor

Na Tabela 2.13 também foi feita a comparação para as classes sociológicas com relação ao nível educacional.

Podemos observar pela Tabela 2.13 que conforme o nível educacional aumenta a porcentagem de indivíduos que pertencem a classe baixa vai diminuindo e de forma inversa, quanto mais o nível educacional aumenta a porcentagem de indivíduos que pertencem a classe alta aumenta. Já a classe média, tem um percentual entre 40 e 60, exceto nos níveis superiores, com uma porcentagem bem mais alta, entre 70% e 80%.

Tabela 2.13: Distribuição do nível educacional segundo as classes sociológicas

Nível educacional	Classe econômica			Total
	Baixa	Média	Alta	
Sem instrução	58,1	40,8	1,1	100,0
Ensino fundamental incompleto	56,5	41,3	2,2	100,0
Ensino fundamental completo	51,2	45,1	3,7	100,0
Ensino médio incompleto	46,4	49,2	4,4	100,0
Ensino médio completo	34,2	59,1	6,7	100,0
Ensino superior incompleto	12,1	74,3	13,6	100,0
Ensino superior completo	5,8	79,3	14,9	100,0
Total	42,4	52,2	5,4	100,0

Fonte: IBGE/PNAD de 2011 Tabulação do autor

Comparando as Tabelas 2.9 e ??, podemos notar que existe uma diferença entre elas que chama a atenção, que é o percentual de pessoas da classe média nos níveis superiores. Enquanto na Tabela 2.9 a porcentagem de pessoas nos níveis superiores é menor se comparado aos demais níveis, na Tabela ?? acontece o contrário, ou seja, o percentual de pessoas de classe média com nível superior é maior que nos demais níveis de instrução.

3 Metodologia: o modelo de regressão logística

Um dos objetivos específicos deste trabalho é examinar os determinantes da riqueza imobiliária no Brasil. Em outras palavras, estamos procurando as variáveis explicativas (\mathbf{x}) que estão relacionadas com a variável Y , construída da seguinte forma:

$$Y_i = \begin{cases} 1, & \text{se o indivíduo } i \text{ possui riqueza imobiliária,} \\ 0, & \text{se o indivíduo } i \text{ não possui riqueza imobiliária.} \end{cases}$$

Como Y é uma variável categórica binária, existem três modelos que relacionam Y com \mathbf{x} , são eles:

- modelo de probabilidade linear (MPL);
- modelo de regressão logística ou logit binário;
- modelo probit binário.

O modelo de probabilidade linear consiste em estimar uma reta aos dados, através do método de mínimos quadrados ajustados. Esse modelo tem uma falha pois pode prever valores de Y menores que 0 e maiores que 1, que não é bom.

Já os modelos de regressão logística e probit não possuem esse problema, por isso são preferidos para estimar variáveis do tipo binária (ou seja, assumem somente valores 0 ou 1). Os valores ajustados dos modelos de regressão logística e probit binário estão sempre entre 0 e 1.

O modelo de regressão logística (ou logit binário) tem uma vantagem adicional: a interpretação dos parâmetros através da razão de chances, por isso, o modelo adotado neste

trabalho para estimar os determinantes da riqueza imobiliária será o modelo de regressão logística, que será detalhado nas seções seguintes.

A metodologia discutida aqui é baseada em DAVISON(2003), HOSMER & LEMESHOW(1989), ANNETTE(1945) e PAULA(2004).

O modelo de regressão logística é um caso particular de modelos lineares generalizados. Portanto, vamos iniciar esse capítulo metodológico com os MLG's.

3.1 Modelos Lineares Generalizados

Os modelos lineares generalizados (MLGs) foram propostos por Nelder e Wedderburn (1972). Esta classe de modelos é definida por uma distribuição de probabilidade para a variável resposta, um conjunto de variáveis explicativas e uma função de ligação.

Definições:

- Componente aleatória: Composta de uma variável aleatória \mathbf{Y} com n valores independentes (Y_1, \dots, Y_n) , um vetor de média $\boldsymbol{\mu} = E(\mathbf{Y})$ e uma distribuição pertencente à família exponencial.

Pertencer à família exponencial significa ter sua função de probabilidade, $P(Y=y, \theta)$ no caso discreto ou função densidade de probabilidade, $f(y, \theta)$ no caso contínuo, escrita da seguinte forma:

$$P(Y=y, \theta) = e^{[a(y_i) \cdot b(\theta) + c(\theta) + d(y_i)]}$$

$$f(y, \theta) = e^{[a(y_i) \cdot b(\theta) + c(\theta) + d(y_i)]}$$

$$\forall i = 1, 2, \dots, n$$

- Componente sistemática: composta por variáveis explicativas $x_{i1}, x_{i2}, \dots, x_{ip}$ que geram um preditor linear η_i , dado por:

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$$

onde:

$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^T$ é um vetor de parâmetros desconhecidos a serem estimados.

- Função de ligação: função monótona e diferenciável, denotada por $g(\mu_i)$, que relaciona a média da variável resposta $\mu_i = E(Y_i)$ ao preditor linear η_i . Ou seja, $g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, $\forall i=1,2,\dots,n$

Casos especiais de MLG's:

- Modelos de regressão linear normal;
- Modelo de Análise de Variância;
- Modelo de Análise de Covariância;
- Modelo de regressão logística ou logit binário;
- Modelo Log-linear de Poisson;
- Modelos para dados categóricos.

Observações importante:

- Uma característica dos MLGs é a hipótese de independência (ou não correlação) das observações.
- Esse método é bastante utilizado nas áreas da saúde e social.
- O modelo de regressão logística, que será utilizado neste trabalho pertence à classe dos MLGs (Modelo Lineares Generalizados).

3.2 Especificação do modelo de regressão logística

Tal modelo é uma técnica estatística que tem como objetivo produzir a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável binária, ou seja, que assume valores 0 ou 1, a partir de variáveis explicativas. Este tipo de modelo difere do modelo de regressão linear, principalmente pelo fato da variável resposta ser categórica e pelo fato da relação entre $E[Y]$ e \mathbf{x} não ser linear.

A variável dependente é da seguinte forma:

$$Y_i = \begin{cases} 1, & \text{se ocorrer sucesso,} \\ 0, & \text{se ocorrer fracasso.} \end{cases}$$

Ou seja, $Y_i \sim \text{Bernoulli}(\pi(\mathbf{x}))$, com função de probabilidade dada por:

$$P(Y = y_i) = \pi^{y_i} \cdot (1 - \pi)^{1-y_i}, \quad \text{onde } \pi(\mathbf{x}) = P(Y_i = 1) = P(\text{sucesso}) \quad (3.1)$$

A relação entre $\pi(\mathbf{x})$ e a variável explicativa, é dada através da função de ligação para o modelo logístico:

$$g(x_i) = \log \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p \quad (3.2)$$

ou equivalentemente, temos a função logística:

$$\pi(\mathbf{x}) = P(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p)}{1 + \exp(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p)} \quad (3.3)$$

Em que $\mathbf{x} = (1, x_1, x_2, \dots, x_p)^T$ contém os valores observados de variáveis explicativas.

3.3 Estimação pelo Método da Máxima Verossimilhança

O método de máxima verossimilhança nos fornece valores para os parâmetros desconhecidos que maximizam a probabilidade de se obter determinado conjunto de dados.

Como $Y_i \sim \text{Bernoulli}(\pi(\mathbf{x}))$, sua função de probabilidade está representada pela equação 3.1. Assim, sua função de verossimilhança é dada por:

$$L(\beta|\mathbf{y}) = \prod_{i=1}^n P(Y = y_i) = \pi^{y_1} \cdot (1 - \pi)^{1-y_1} \cdot \dots \cdot \pi^{y_n} \cdot (1 - \pi)^{1-y_n} \quad (3.4)$$

Aplicando \ln na Equação 3.4, temos o que chamamos de função log-verossimilhança:

$$\ln L(\beta|\mathbf{y}) = \sum_{i=1}^n [y_i \cdot \ln(\pi(\mathbf{x}_i)) + (1 - y_i) \cdot \ln(1 - \pi(\mathbf{x}_i))] \quad (3.5)$$

Ou seja, $\ln L(\beta|\mathbf{y}) = y_1 \cdot \ln(\pi(\mathbf{x}_1)) + (1 - y_1) \cdot \ln(1 - \pi(\mathbf{x}_1)) + \dots + y_n \cdot \ln(\pi(\mathbf{x}_n)) + (1 - y_n) \cdot \ln(1 - \pi(\mathbf{x}_n))$

Para encontrar os estimadores de Máxima Verossimilhança, é necessário diferenciar a Equação 3.5 com relação à $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$ e igualar as expressões resultantes a zero. Ou seja,

$$\frac{\partial}{\partial \beta_0} \ln P(\beta|\mathbf{y}) = 0 \quad (3.6)$$

$$\frac{\partial}{\partial \beta_j} \ln P(\beta|\mathbf{y}) = 0, \quad \text{para } j=1, \dots, p \quad (3.7)$$

Das Equações 3.6 e 3.7, obtêm-se as seguintes expressões:

$$\begin{cases} \sum_{i=1}^n [y_i - \pi(x)] = 0 & \text{para } j=0 \\ \sum_{i=1}^n x_j \cdot [y_i - \pi(x)] = 0, & \text{para } j=1, \dots, p-1 \end{cases}$$

Esse é o sistema de equações de verossimilhança, porém se trata de um sistema não linear e como não é possível resolvê-los de forma analítica, devemos utilizar métodos iterativos para encontrar uma aproximação para a solução e estimar o β .

Um dos métodos que existem para solucionar esse “problema” é o processo iterativo de Newton Raphson. Com ele obtemos as estimativas de máxima verossimilhança para β expandindo a função escore U_β em torno de um valor inicial $\beta^{(0)}$, tal que

$$U_\beta \cong U_\beta^{(0)} + U_{\beta}^{\prime(0)}(\beta - \beta^{(0)}), \quad (3.8)$$

onde U_{β}^{\prime} se refere a primeira derivada de U_β com respeito a β^T , sendo $U_{\beta}^{\prime(0)}$ e $U_\beta^{(0)}$, respectivamente, essas quantidades avaliadas em $\beta^{(0)}$. Repetindo o procedimento descrito acima, chegamos no processo iterativo

$$\beta^{m+1} = \beta^m + \{(-U_{\beta}^{\prime})^{-1}\}^{(m)} U_{\beta}^{(m)}, m = 0, 1, \dots \quad (3.9)$$

Como a matriz $-U_{\beta}^{\prime}$ pode não ser positiva definida, a aplicação do método de escore de Fisher substituindo a matriz $-U_{\beta}^{\prime}$ pelo correspondente valor esperado $\mathbf{K}_{\beta\beta}$ pode ser mais conveniente. Isso resulta no seguinte processo iterativo:

$$\beta^{m+1} = \beta^m + \{\mathbf{K}_{\beta\beta}^{-1}\}^{(m)} U_{\beta}^{(m)}, m = 0, 1, \dots \quad (3.10)$$

Mexendo na equação 3.10, obtemos um processo iterativo de mínimos quadrados ponderados

$$\beta^{m+1} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)}, m = 0, 1, \dots \quad (3.11)$$

onde, $\mathbf{z} = \boldsymbol{\eta} + \mathbf{W}^{-1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu})$. \mathbf{z} desempenha o papel de uma variável dependente modificada, enquanto \mathbf{W} é uma matriz de pesos que se modifica a cada passo do processo iterativo. O processo descrito em 3.11 converge, em geral, com um número finito de passos, independente dos valores iniciais utilizados. É usual iniciarmos 3.11 com $\boldsymbol{\eta}^{(0)} = g(\mathbf{y})$

3.4 Interpretação

Para interpretar o parâmetro β_1 , por exemplo, devemos calcular a equação 3.2 para $x_1 + 1$ e x_1 mantendo todas as outras variáveis constantes, obtendo:

$$\ln \left(\frac{\pi(x_1 + 1)}{1 - \pi(x_1 + 1)} \right) - \ln \left(\frac{\pi(x_1)}{1 - \pi(x_1)} \right) = \beta_0 + \beta_1(x_1 + 1) - (\beta_0 + \beta_1 x_1) = \beta_1 \quad (3.12)$$

Aplicando a exponencial na equação 3.8, temos:

$$\frac{\frac{\pi(x_1+1)}{1-\pi(x_1+1)}}{\frac{\pi(x_1)}{1-\pi(x_1)}} = \frac{\exp(\beta_0 + \beta_1(x_1 + 1))}{\exp(\beta_0 + \beta_1 x_1)} = \exp(\beta_1) \quad (3.13)$$

Essa expressão é conhecida como razão de chances (OR - odds ratio) e compara a probabilidade de sucesso para $x+1$ com a probabilidade de sucesso para x .

Relações:

$$\beta_1 > 0 \rightarrow OR > 1$$

$$\beta_1 < 0 \rightarrow OR < 1$$

$$\beta_1 = 0 \rightarrow OR = 1$$

Para entender melhor o que significa a razão de chances (OR), podemos exemplificá-la da seguinte forma:

Supondo que estamos interessados em analisar o desempenho dos alunos da Universidade Federal Fluminense.

$$Y_i = \begin{cases} 1, & \text{se desempenho bom} \\ 0, & \text{se desempenho ruim} \end{cases}$$

E para isso, foram escolhidas algumas variáveis para explicar o que influencia nessa questão. Uma das variáveis escolhidas foi o sexo.

$$\text{Sexo} = \begin{cases} 1, & \text{se homem} \\ 0, & \text{se mulher} \end{cases}$$

Supondo que esta variável foi significativa para o modelo, ao nível de significância de 5% e apresentou um OR = 1,36. Ou seja, a chance de um homem ter um bom desempenho é 1,36 vezes maior que uma mulher.

Caso o OR fosse menor que 1, por exemplo OR = 0,78, a interpretação ficaria da seguinte forma, a chance de uma mulher ter bom desempenho é 1,28 (=1/0,78) vezes maior que um homem.

E por fim, se OR = 1, as chances são as mesmas tanto para homens quanto para mulheres.

3.5 Inferência no Modelo

Para fazer inferência no modelo será utilizado o Teste de Wald:

- Hipóteses a ser testadas, para $j = 1, 2, \dots, p$

$$\begin{cases} H_0 : \beta_j = 0, \\ H_1 : \beta_j \neq 0. \end{cases}$$

- Estatística de Teste:

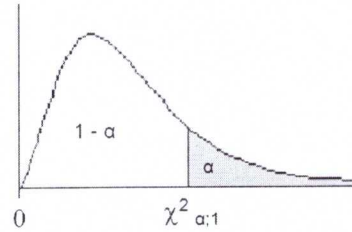
$$\text{Sob } H_0: \beta_j = 0, \text{ temos que: } Z = \frac{\hat{\beta}_j}{\sqrt{\text{VAR}(\hat{\beta}_j)}} \sim N(0,1)$$

ou alternativamente, $w_{obs} = \frac{\hat{\beta}_j^2}{V\hat{AR}(\hat{\beta}_j)} \sim \chi_1^2$

$j = 1, 2, \dots, p$

- Região Crítica:

Figura 3.1: Gráfico da Distribuição χ_1^2



- Tomada de decisão:

Se $w_{obs} \in$ região crítica, rejeita-se H_0 ao nível de significância α . Ou seja, a variável x_j é significativa para o modelo.

Se $w_{obs} \notin$ região crítica não rejeita-se H_0 ao nível de significância α . Ou seja, a variável x_j não é significativa para o modelo.

Para complementar a análise, também podemos construir o intervalo de confiança para os parâmetros. A base para a construção deste é a mesma que a utilizada para a realização de testes de hipóteses.

- Intervalo de confiança para o parâmetro β_j , considerando um nível de confiança de $100(1-\alpha)\%$:

$$IC_{\beta_j, 100(1-\alpha)\%} = [\hat{\beta}_j \pm z_{\alpha/2} \cdot \hat{DP}(\hat{\beta}_j)] \quad (3.14)$$

- Intervalo de confiança para OR_j , considerando um nível de confiança de $100(1-\alpha)\%$:

$$IC_{OR_j, 100(1-\alpha)\%} = [\exp(\hat{\beta}_j \pm z_{\alpha/2} \cdot \hat{DP}(\hat{\beta}_j))] \quad (3.15)$$

3.6 Qualidade do ajuste

Muitas das medidas referente à qualidade do ajuste são baseadas no log da verossimilhança encontrado quando estimamos o modelo completo com as variáveis independentes \mathbf{x} , estimando os β 's. Esta medida será chamada de $\ln L(M_{\text{completo}})$. Para medir a contribuição da variável X na verossimilhança, muitas vezes vamos comparar o modelo completo com o modelo sem a variável X , isto é, o modelo só com o parâmetro β_0 . Vamos chamar esta medida de $\ln L(M_{\text{intercepto}}) = \ln L(M_{\beta_0})$.

Existem algumas formas para medir a qualidade do ajuste, são elas:

1. Teste da Razão de Verossimilhança:

É o teste conjunto que compara se todos os parâmetros (diferente da constante) são iguais a zero (H_0), ou se existe pelo menos um parâmetro diferente de zero. É o equivalente ao teste F da regressão simples e verifica se o modelo está adequado aos dados. Sua estatística do teste segue a distribuição de qui-quadrado com o número de parâmetros sendo testados em H_0 como graus de liberdade. Essa estatística é chamada de qui-quadrado, mas na verdade ela é mensurada pela razão entre as verossimilhanças do modelo completo com o modelo somente com o intercepto:

$$\chi^2 = 2\ln \left(\frac{L(M_{\text{completo}})}{L(M_{\text{intercepto}})} \right) = 2\ln L(M_{\text{completo}}) - 2\ln L(M_{\text{intercepto}}) \quad (3.16)$$

2. Cox e Snell R^2 :

$$R_{C\&S}^2 = 1 - \frac{L(M_{\text{intercepto}})^{2/n}}{L(M_{\text{completo}})} \quad (3.17)$$

3. Nagelkerke R^2 :

$$R_{Nag}^2 = \frac{1 - \frac{L(M_{\text{intercepto}})^{2/n}}{L(M_{\text{completo}})}}{1 - L(M_{\text{intercepto}})^{2/n}} \quad (3.18)$$

Não se utiliza o R^2 e nem o $R^2_{ajustado}$ como medida de qualidade de ajuste nas regressões logísticas. Entretanto, o R^2 elaborado por McFaddens é aceito como medida de qualidade de ajuste. Ele é calculado como uma razão de verossimilhanças da seguinte forma:

4. Critério de Informação:

São medidas utilizadas quando comparamos modelos com diferentes tamanhos de amostra ou diferentes números de parâmetros.

O critério de informação de Akaike (AIC) é calculado por:

$$AIC = \frac{-2\ln L(M_{\text{completo}}) + 2p}{n} \quad (3.19)$$

onde p é o número de parâmetros do modelo e n é o número de observações.

O critério de informação bayesiano é calculado por:

$$BIC = -\chi^2 + (p - 1) \cdot \ln(n) \quad (3.20)$$

5. Valores de Y observados e Preditos Corretamente pelo Modelo

Com as probabilidades preditas do modelo, podemos construir uma tabela que compara os valores observados de Y iguais a 1 e 0 com os estimados pelo modelo, fazendo a seguinte transformação:

$$\hat{Y}_i = \begin{cases} 0, & \text{se } \hat{P}_i \leq 0,5 \\ 1, & \text{se } \hat{P}_i > 0,5 \end{cases}$$

Com essa transformação, podemos verificar a capacidade do modelo em discriminar os dois grupos $y_i=0$ e $y_i=1$.

Suponha que entre N indivíduos observados, n_0 indivíduos possuem $y_i=0$ e n_1 indivíduos possuem $y_i=1$

Tabela 3.1: Tabela de classificação

Observado	Predito		Total
	y=0	y=1	
y=0	y_0	$n_0 - y_0$	n_0
y=1	$n_1 - y_1$	y_1	n_1

Fonte: IBGE/PNAD de 2011 Tabulação do autor

O percentual de observações com $y=0$ estimado corretamente é dado por $\frac{y_0}{n_0} \cdot 100$ e o percentual de observações com $y=1$ estimado corretamente é dado por $\frac{y_1}{n_1} \cdot 100$. E o percentual de y estimado corretamente é dado por $\frac{y_0+y_1}{N} \cdot 100$, onde $N=n_0+n_1$.

3.7 Técnicas de diagnóstico

Estudos de simulação têm sugerido o resíduo t_{Di} para as análises de diagnósticos em MLG's. Em particular, para modelos binomiais, esse resíduo é expresso:

$$t_{Di} = \pm \sqrt{\frac{2}{1 - \hat{h}_{ii}}} \left\{ y_i \log \left(\frac{y_i}{n_i \pi_i} \right) + (n_i - y_i) \log \left(\frac{n_i - y_i}{n_i - n_i \hat{\pi}_i} \right) \right\}^{1/2}, \quad (3.21)$$

para $0 < y_i < n_i$. Quando $y_i=0$ ou $y_i=n_i$, o componente do desvio padronizado fica da seguinte maneira:

$$t_{Di} = \frac{\{2n_i |\log(1 - \hat{\pi}_i)|\}^{1/2}}{\sqrt{1 - \hat{h}_{ii}}} \quad \text{e} \quad t_{Di} = \frac{\{2n_i |\log \hat{\pi}_i|\}^{1/2}}{\sqrt{1 - \hat{h}_{ii}}} \quad (3.22)$$

respectivamente. O resíduo Studentizado t_{Si} também usado para detectar a presença de pontos aberrantes mesmo tendo, em geral, distribuição assimétrica acentuada, tem a seguinte forma:

$$t_{Si} = \frac{1}{\sqrt{1 - \hat{h}_{ii}}} \frac{(y_i - n_i \hat{\pi}_i)}{\{n_i \hat{\pi}_i (1 - \hat{\pi}_i)\}^{1/2}} \quad (3.23)$$

Para medir a influência das observações nas estimativas dos coeficientes, utiliza-se a distância de Cook aproximada, que é dada por:

$$LD_i = \frac{\hat{h}_{ii}}{(1 - \hat{h}_{ii})^2} \frac{(y_i - n_i \hat{\pi}_i)^2}{n_i \hat{\pi}_i (1 - \hat{\pi}_i)} \quad (3.24)$$

Hosmer e Lemeshow (1989) observam que $\hat{\pi}_i$ depende das probabilidades ajustadas $\hat{\pi}_i$, $i=1, \dots, p$, e com isso, os resíduos t_{Si} e t_{Di} e a medida de influência LD_i também dependem. Então,

$$h_{ii} = n_i \pi_i (1 - \pi_i) \mathbf{x}_i^T (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{x}_i$$

,

com $\mathbf{V} = \text{diag}\{n_1 \pi_1 (1 - \pi_1), \dots, n_p \pi_p (1 - \pi_p)\}$. Hosmer e Lemeshow mostram através de um estudo numérico que o comportamento de \hat{h}_{ii} numa regressão logística pode ser muito diferente do comportamento de \hat{h}_{ii} na regressão linear para uma mesma matriz modelo \mathbf{X} .

A Tabela 3.2 mostra os possíveis valores de t_{Si} , LD_i e \hat{h}_{ii} em função das probabilidades ajustadas. A quantidade \hat{h}_{ii} pode ser interpretada de forma semelhante á h_{ii} da regressão normal linear para $0,1 \leq \hat{\pi}_i \leq 0,9$. Entretanto, quando $\hat{\pi}_i$ é pequeno ou alto, \hat{h}_{ii} fica em geral pequeno, o que pode dificultar a avaliação de pontos que estejam mais afastados no subespaço gerado pelas colunas da matriz \mathbf{X} . Sendo assim, a proposta alternativa são os gráficos de t_{Si} , t_{Di} e LD_i contra as probabilidades ajustadas $\hat{\pi}_i$. Esses gráficos informam sobre os pontos aberrantes e influentes com relação às probabilidades ajustadas.

Tabela 3.2: Possíveis valores para t_{Si} , t_{Di} e LD_i segundo as probabilidades ajustadas

Medida	Probabilidade ajustada				
	0,0-0,1	0,2-0,3	0,3-0,7	0,7-0,9	0,9-1,0
t_{Si}	grande ou pequeno	moderado	moderado ou pequeno	moderado	grande ou pequeno
LD_i	pequeno	grande	moderado	grande	pequeno
\hat{h}_{ii}	pequeno	grande	moderado ou pequeno	grande	pequeno

Fonte: Modelos de regressão com apoio computacional, Gilberto A. Paula, 2004, p.219

4 Resultados

4.1 Descrição das variáveis

A variável resposta que será utilizada no modelo foi definida da seguinte maneira:

$$Y_i = \begin{cases} 1, & \text{se imóvel próprio} \\ 0, & \text{caso contrário} \end{cases}$$

onde imóvel próprio foi classificado como quitado e estão pagando.

As variáveis explicativas usadas no modelo serão apresentadas na Tabela 4.1:

Tabela 4.1: Descrição das variáveis

Variável	Descrição	Valores
X_1	Sexo do chefe	1: homem 0: mulher
X_2	Idade do chefe	Menos de 30 30 a 49 anos 50 anos ou mais
X_3	Escolaridade do chefe	Ensino Fundamental Ensino Médio Ensino Superior
X_4	Classe social pelo critério econômico	Baixa Média Alta
X_4	Classe social pelo critério sociológico	Baixa Média Alta
X_5	Raça	Branco ou Amarelo Negro, pardo, indígena
X_6	Tipo de família	casal com ou sem filho Mãe sozinha Outros tipos de família

Fonte: IBGE/PNAD de 2011 Tabulação do autor

A escolha dessas variáveis se deve às características sociais e econômicas do chefe do domicílio, que segundo a teoria econômica, estão intimamente ligadas à capacidade do consumo de um bem.

Para observar como as variáveis explicativas se distribuem em relação à riqueza imobiliária (variável resposta), foi feito um cruzamento entre estas, que será apresentado a seguir na Tabela 4.2. É possível notar que a proporção de domicílios com riqueza imobiliária aumenta com a idade do chefe e diminui com o nível de instrução.

Tabela 4.2: Cruzamento das variáveis explicativas com a variável resposta

Variáveis	Y		Total
	0	1	
X_1			
Mulher	26,2	73,8	100,0
Homem	26,2	73,8	100,0
X_2			
15 a 19 anos	59,7	40,3	100,0
20 a 29 anos	51,4	48,6	100,0
30 a 39 anos	34,7	65,3	100,0
40 a 49 anos	24,0	76,0	100,0
50 a 59 anos	18,0	82,0	100,0
60 a 64 anos	14,4	85,6	100,0
65 anos ou mais	12,8	87,2	100,0
X_3			
Sem instrução	19,8	80,2	100,0
Ensino fundamental incompleto	23,2	76,8	100,0
Ensino fundamental completo	28,9	71,1	100,0
Ensino médio incompleto	36,8	63,2	100,0
Ensino médio completo	30,3	69,7	100,0

Ensino superior incompleto	36,6	63,4	100,0
Ensino superior completo	23,6	76,4	100,0
X_4-Econômico			
Baixa	29,5	70,5	100,0
Média	26,2	73,8	100,0
Alta	23,9	76,1	100,0
X_4-Social			
Baixa	29,2	70,8	100,0
Média	28,5	71,5	100,0
Alto	25,1	74,9	100,0
X_5			
Negro, pardo, mulato e indígena	28,0	72,0	100,0
Branco e amarelo	24,5	75,5	100,0
X_6			
Casal sem filho	24,5	75,5	100,0
Casal com filho	24,8	75,2	100,0
Mãe sozinha	26,4	73,6	100,0
Outro tipo de família	31,1	68,9	100,0

Fonte: IBGE/PNAD de 2011 Tabulação do autor

As variáveis categóricas que possuem mais de duas categorias sofreram transformações para se tornarem do tipo binária. São elas: X_2 , X_3 , X_4 -Eco, X_4 -Soc, X_6 .

$$X_2 - 30a49 = \begin{cases} 1, \text{ se idade entre 30 e 49 anos} \\ 0, \text{ caso contrário} \end{cases}$$

$$X_2 - 50oumais = \begin{cases} 1, \text{ se idade maior ou igual a 50 anos} \\ 0, \text{ caso contrário} \end{cases}$$

$$X_3 - \text{Emédio} = \begin{cases} 1, \text{ se ensino médio completo ou incompleto} \\ 0, \text{ caso contrário} \end{cases}$$

$$X_3 - \text{Esuperior} = \begin{cases} 1, \text{ se ensino superior completo ou incompleto} \\ 0, \text{ caso contrário} \end{cases}$$

$$X_4 - \text{Cmediaeco} = \begin{cases} 1, \text{ se classe baixa} \\ 0, \text{ caso contrário} \end{cases}$$

$$X_4 - \text{Caltaeco} = \begin{cases} 1, \text{ se classe alta} \\ 0, \text{ caso contrário} \end{cases}$$

$$X_4 - \text{Cmediasoc} = \begin{cases} 1, \text{ se classe baixa} \\ 0, \text{ caso contrário} \end{cases}$$

$$X_4 - \text{Caltasoc} = \begin{cases} 1, \text{ se classe alta} \\ 0, \text{ caso contrário} \end{cases}$$

$$X_6 - \text{casal} = \begin{cases} 1, \text{ se casal com ou sem filhos} \\ 0, \text{ caso contrário} \end{cases}$$

Como X_1 e X_5 já eram do tipo binário, estas não sofreram mudanças.

$$X_1 = \begin{cases} 1, \text{ se homem} \\ 0, \text{ se mulher} \end{cases}$$

$$X_5 = \begin{cases} 1, \text{ se branco ou amarelo} \\ 0, \text{ se negro, pardo, mulato ou indígena} \end{cases}$$

4.2 Especificação do modelo

O modelo de regressão logística, é dado pela equação 4.1 em que $\pi(\mathbf{x})$ é a probabilidade do domicílio ser próprio dado \mathbf{x} , que contém os valores observados das variáveis explicativas $\mathbf{x}=(1, x_1, x_2, \dots, x_p)^T$

$$\log \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (4.1)$$

No nosso estudo, foram utilizados dois modelos, um para analisarmos a classe média definida pelo critério econômico e outro para o estudo da classe média sociológica. Os modelos se apresentam da seguinte maneira:

Modelo 1 - Critério econômico

$$\log \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 - 30a49 + \beta_3 X_2 - 50oumais + \beta_4 X_3 - \text{Emédio} + \beta_5 X_3 - \text{Esuperior} \\ + \beta_6 X_4 - \text{Cmédiaeco} + \beta_7 X_4 - \text{Caltaeco} + \beta_8 X_5 + \beta_9 X_6 - \text{casal} \quad (4.2)$$

Modelo 2 - Critério sociológico

$$\log \left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 - 30a49 + \beta_3 X_2 - 50oumais + \beta_4 X_3 - \text{Emédio} + \beta_5 X_3 - \text{Esuperior} \\ + \beta_6 X_4 - \text{Cmédiassoc} + \beta_7 X_4 - \text{Caltasoc} + \beta_8 X_5 + \beta_9 X_6 - \text{casal} \quad (4.3)$$

4.3 Modelos estimados

Modelo 1 - Critério econômico

Com auxílio da Tabela 4.3, podemos notar que apenas duas variáveis não foram significativas para o modelo, ao nível de significância de 5%: X_3 -Emédio e X_4 -Caltaeco, pois seus p-valores foram 0,111 e 0,116 respectivamente.

Também é possível analisar a razão de chances de cada uma das variáveis significativas.

- A chance de uma mulher possuir riqueza imobiliária é 1,17 ($= 1/0,852$) vezes maior do que um homem;
- A chance de uma pessoa que tem de 30 a 49 anos possuir riqueza imobiliária é 2,629 vezes maior do que uma pessoa que está fora dessa faixa etária;
- A chance de uma pessoa que tem 50 anos ou mais possuir riqueza imobiliária é aproximadamente 11,065 vezes maior do que uma pessoa que está fora dessa faixa etária;
- A chance de uma pessoa que cursou o ensino superior possuir riqueza imobiliária é 1,081 vezes maior do que uma pessoa que cursou o ensino fundamental;
- A chance de uma pessoa que pertence à classe baixa pelo critério econômico possuir riqueza imobiliária é 1,14 ($= 1/0,878$) vezes maior que uma pessoa de classe média pelo mesmo critério;
- A chance de uma pessoa de cor branca ou amarela possuir riqueza imobiliária é 1,170 vezes maior que uma pessoa de negra, parda, mulata ou indígena;
- A chance de um casal com ou sem filhos possuir riqueza imobiliária é 2,323 vezes maior que uma mãe ou outro tipo de família.

Tabela 4.3: Estimativa dos parâmetros referentes ao modelo logístico utilizando as classes econômicas

Variável	β	DP	P-valor	Exp(β)	IC(Exp(β), 95%)
X_1 -Homem	-0,160	0,009	<0,001	0,852	[0,837 ; 0,867]
X_2 -30a49	0,967	0,009	<0,001	2,629	[2,584 ; 2,677]
X_2 -50oumais	2,404	0,012	<0,001	11,065	[10,810 ; 11,331]
X_3 -Emédio	-0,013	0,008	0,111	0,987	[0,972 ; 1,002]
X_3 -Esuperior	0,078	0,012	<0,001	1,081	[1,056 ; 1,107]
X_4 -Cmédiaeco	-0,130	0,009	<0,001	0,878	[0,863 ; 0,894]
X_4 -Caltaeco	-0,019	0,012	0,116	0,981	[0,958 ; 1,004]
X_5 -Branco	0,157	0,007	<0,001	1,170	[1,154 ; 1,186]
X_6 -Casal	0,843	0,009	<0,001	2,323	[2,283 ; 2,365]
Constante	-0,573	0,012	<0,001	0,564	[0,551 ; 0,577]

Fonte: IBGE/PNAD de 2011 Tabulação do autor

Para calcular a probabilidade de um indivíduo ter riqueza imobiliária, dado suas características, usando o critério econômico de classe social, usamos a seguinte equação:

$$P(Y = 1) = \frac{\exp(-0,573 - 0,160X_1Homem + \dots + 0,157X_5Branco + 0,843X_6Casal)}{1 + \exp(-0,573 - 0,160X_1Homem + \dots + 0,157X_5Branco + 0,843X_6Casal)} \quad (4.4)$$

A probabilidade de um homem, de 50 anos ou mais de idade, com ensino superior, pertencente à classe média econômica, branco e pertencente a uma família formada por um casal é igual a 0,9321.

Modelo 2 - Critério sociológico

Com auxílio da Tabela 4.4, é possível notar que todas as variáveis foram significativas para o modelo, pois seus p-valores são menores que o nível de significância adotado, 5%. Também podemos analisar a razão de chances de cada uma das variáveis.

- A chance de uma mulher possuir riqueza imobiliária é 1,227 (= 1/0,815) vezes maior

do que um homem;

- A chance de uma pessoa que tem de 30 a 49 anos possuir riqueza imobiliária é 2,476 vezes maior do que uma pessoa que está fora dessa faixa etária;
- A chance de uma pessoa que tem 50 anos ou mais possuir riqueza imobiliária é aproximadamente 5,837 vezes maior do que uma pessoa que está fora dessa faixa etária;
- A chance de uma pessoa que não cursou o ensino médio possuir riqueza imobiliária é 1,06 (= 1/0,943) vezes maior do que uma pessoa que cursou o ensino médio;
- A chance de uma pessoa que cursou o ensino superior possuir riqueza imobiliária é 1,124 vezes maior do que uma pessoa que cursou o ensino fundamental ou não tem instrução;
- A chance de uma pessoa que pertence à classe média pelo critério sociológico possuir riqueza imobiliária é 1,082 vezes maior que uma pessoa de classe baixa pelo mesmo critério;
- A chance de uma pessoa que pertence à classe alta pelo critério sociológico possuir riqueza imobiliária é 1,189 vezes maior que uma pessoa de classe baixa pelo mesmo critério;
- A chance de uma pessoa de cor branca ou amarela possuir riqueza imobiliária é 1,09 vezes maior que uma pessoa de negra, parda, mulata ou indígena;
- A chance de um casal com ou sem filhos possuir riqueza imobiliária é 1,905 vezes maior que uma mãe ou outro tipo de família.

Como todas as variáveis explicativas do modelo são categóricas, a magnitude das estimativas dos parâmetros mostram quais variáveis explicativas têm um peso maior na probabilidade de ter riqueza imobiliária. Analisando os dois modelos estudados observa-se que as variáveis que aumentam a probabilidade de se obter uma casa própria é a idade (mais velho) do chefe do domicílio e o tipo de família (casal).

Tabela 4.4: Estimativa dos parâmetros referentes ao modelo logístico utilizando as classes sociológicas

Variável	β	DP	P-valor	Exp(β)	IC(Exp(β), 95%)
X_1 -Homem	-0,205	0,010	<0,001	0,815	[0,799 ; 0,831]
X_2 -30a49	0,907	0,010	<0,001	2,476	[2,429 ; 2,526]
X_2 -50oumais	1,764	0,012	<0,001	5,837	[5,700 ; 5,975]
X_3 -Emédio	-0,059	0,009	<0,001	0,943	[0,926 ; 0,959]
X_3 -Esuperior	0,117	0,012	<0,001	1,124	[1,098 ; 1,151]
X_4 -Cmédiasoc	0,079	0,008	<0,001	1,082	[1,065 ; 1,099]
X_4 -Caltasoc	0,173	0,018	<0,001	1,189	[1,147 ; 1,232]
X_5 -Branco	0,086	0,007	<0,001	1,090	[1,061 ; 1,091]
X_6 -Casal	0,644	0,009	<0,001	1,905	[1,871 ; 1,938]
Constante	-0,427	0,013	<0,001	0,653	[0,636 ; 0669]

Fonte: IBGE/PNAD de 2011 Tabulação do autor

Utilizando o critério sociológico, a equação é dada por:

$$P(Y = 1) = \frac{\exp(-0,427 - 0,205X_1Homem + \dots + 0,086X_5Branco + 0,644X_6Casal)}{1 + \exp(-0,427 - 0,205X_1Homem + \dots + 0,086X_5Branco + 0,644X_6Casal)} \quad (4.5)$$

Um indivíduo com as mesmas características descritas acima, terá probabilidade de 0,9232 de ter riqueza imobiliária.

Observando os intervalos de confiança para Exp(β) da classe média nos dois critérios, é possível notar que estes são disjuntos e também, no critério econômico ele é menor que 1 (indicando que a chance diminui) e no critério sociológico é maior que 1 (indicando que a chance aumenta). Com isso, podemos dizer que existe diferença na probabilidade de possuir riqueza imobiliária comparando os dois critérios utilizados.

4.4 Qualidade do ajuste

Com as Tabelas 4.5, 4.6 e ?? podemos observar que apesar das estatísticas Cox e Snell R Square e Nagelkerke R Square estarem baixas, o percentual de $Y=1$ estimados corretamente e o percentual de Y estimado total foram considerados muito bons. O desvio do modelo 1 foi de 61315,46 com p-valor $<0,001$ e o do modelo 2 foi de 30165,53 com p-valor $<0,001$, indicando um ajustamento adequado. Comparando os dois modelos quanto a predição, é possível notar que $y=0$ e o y total estimados corretamente são maiores no modelo 1 e o $y=1$ estimado corretamente é semelhante nos dois modelos.

Tabela 4.5: Qualidade do ajuste

Estatísticas	Modelo 1	Modelo 2
$-2L_{final}$	480.771,68	448.332,02
Qui-Quadrado	61.315,46	30.165,53
graus de liberdade	9	9
p-valor	$<0,001$	$<0,001$
Cox e Snell R Square	11,80%	7,20%
Nagelkerke R Square	17,60%	10,40%

Fonte: IBGE/PNAD de 2011 Tabulação do autor

Podemos observar pela Tabela 4.6 que o percentual de $Y=1$ e Y total estimados corretamente está elevado, que mostra que os modelos estão sendo estimados de forma adequada.

Tabela 4.6: Tabela de classificação

Estatística	Modelo 1	Modelo 2
$Y=0$ estimado corretamente	13,8%	10,5%
$Y=1$ estimado corretamente	96,6%	96,8%
Y total estimado corretamente	76,3%	72,4%

Fonte: IBGE/PNAD de 2011 Tabulação do autor

4.5 Técnicas de Diagnóstico

Para analisar os resíduos, primeiramente fizemos um gráfico dos resíduos padronizados contra as probabilidades ajustadas com o banco de dados todo. Com isso, percebemos que esses Gráficos (Figuras 4.2 e 4.3) mostram que existem observações menores do que -2, indicando um grande resíduo.

A partir desses gráficos de resíduos dos dois modelos, foi escolhido um ponto de corte que elimine todos os valores menores do que -2,5. Com isso, excluimos os dados aberrantes da nossa amostra e assim estimamos os modelos escolhidos.

Figura 4.1: Gráficos dos resíduos padronizados

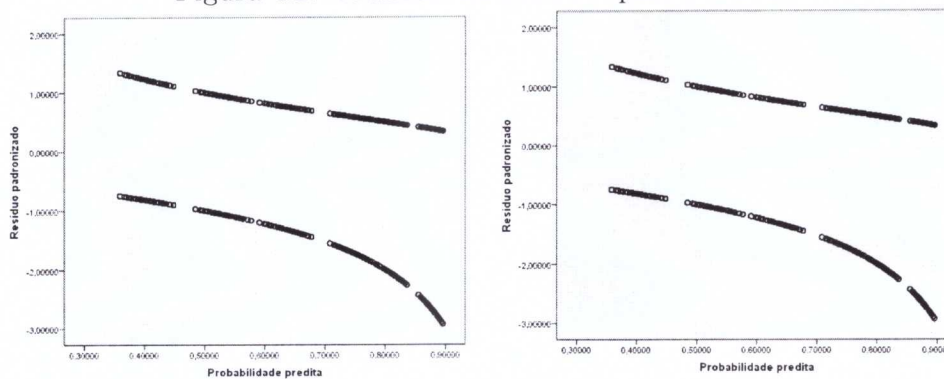


Figura 4.2: Modelo 1

Figura 4.3: Modelo 2

Como as observações estão próximas, nenhum dado foi excluído pela análise dos gráficos de Cook's. (Gráfico 4.5 e 4.6)

Figura 4.4: Gráficos da influência de Cook's

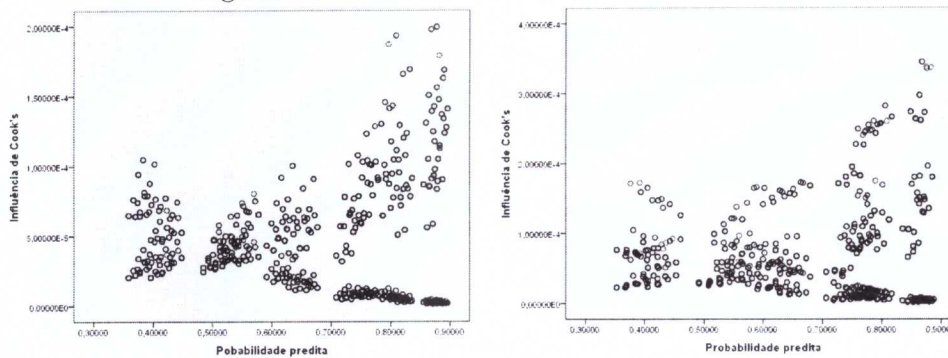


Figura 4.5: Modelo 1

Figura 4.6: Modelo 2

5 Conclusão

A riqueza imobiliária discutida nesse trabalho, levou em consideração apenas os imóveis residenciais urbanos no Brasil no ano de 2011. Foi possível observar que a grande maioria dos imóveis eram próprios (73,8%), ou seja grande parte da população possui riqueza imobiliária.

Foi possível notar que a riqueza imobiliária da população brasileira em 2011 quanto ao sexo, estava distribuída de forma igual, enquanto na variável idade, quanto mais velho, mais possui essa riqueza. Em relação ao nível de instrução, algo que chama muito a atenção, as pessoas sem instrução ou com ensino fundamental incompleto são as que possuem mais riqueza imobiliária seguido das pessoas com ensino superior completo.

No critério de classes econômicas a classe alta é a que possui mais riqueza imobiliária, já na classificação sociológica, a classe baixa e médio inferior tem mais riqueza imobiliária que as outras classes médias. Os brancos/amarelos possuem mais riqueza imobiliária e as mães solteiras também.

A distribuição da população quanto as classes ainda é um ponto muito discutido, pois como vimos, existem diversas maneiras delas serem classificadas. Nesse trabalho escolhemos tratar de dois critérios com visões bem distintas do que é a classe média. O primeiro critério usado pelos economistas que está interessado na renda domiciliar per capita e o outro utilizado pelos sociólogos, que acham que a ocupação é o fator que mais importa para a classificação das classes.

Observando os intervalos de confiança para $\text{Exp}(\beta)$ da classe média nos dois critérios, é possível notar que estes são disjuntos e também, no critério econômico ele é menor que 1 (indicando que a chance diminui) e no critério sociológico é maior que 1 (indicando que a chance aumenta). Com isso, podemos dizer que existe diferença na probabilidade de possuir riqueza imobiliária comparando os dois critérios utilizados.

A partir dos critérios estabelecidos, foram estimados dois modelos de regressão logística, onde o principal objetivo era entender como se comportavam as classes quando relacionadas com variável de interesse (Riqueza imobiliária). Com auxílio dos modelos verificamos que a classe média pelo critério econômico tem menos chances de possuir riqueza imobiliária quando comparada às classes baixa. Já no modelo sociológico, a classe média tem mais chances de possuir riqueza imobiliária se comparada com a classe baixa.

Referências Bibliográficas

- ANNETTE, J. (1945) *An Introduction to Generalized Linear Models*. Chapman Hall/CRC texts in statistical science series, 2 edn.
- DAVISON, A. (2003) *Statistical Models*. Cambridge University Press.
- DE SOUZA, A.; LAMOUNIER, B. (2010) *A classe média brasileira - Ambições, valores e projetos de sociedade*. Elsevier.
- HOSMER, D.W.; LEMESHOW, S. (1989) *Applied Logistic Regression. Series in Probability and Mathematical Statistics*. Wiley-Interscience publication.
- KLEINBAUM, D. (1994) *Logistic Regression - A Self Learning Text*. Springer.
- MEDEIROS, J. B. (2010) *Redação científica: a prática de fichamentos, resumos, resenhas*. São Paulo: Atlas.
- SOUZA, J. (2012) *Os batalhadores Brasileiros: nova classe média ou nova classe trabalhadora?* Belo Horizonte: Editora UFMG.
- TAFNER, P.; CARVALHO, M. (2007) *Evolução da distribuição familiar da riqueza imobiliária no Brasil: 1995-2004*.
- UCHÔA, C.; KERSTENETZKY, C. L. (2012) *É “nova classe média” classe média? o que diz a pesquisa de orçamentos familiares (2008-2009)*.