



UNIVERSIDADE DE BRASÍLIA
IE – DEPARTAMENTO DE ESTATÍSTICA
ESTÁGIO SUPERVISIONADO II

Fatores associados à mobilidade espacial no Brasil:

Uma análise do perfil do migrante recente

Marília de Faria Ferreira
Pedro Henrique Monteiro Moreira

Relatório do Projeto Final
Orientadora: Ana Maria Nogales Vasconcelos
Coorientador: Herton Ellery de Araújo

Brasília,
Junho de 2011

MARÍLIA DE FARIA FERREIRA
PEDRO HENRIQUE MONTEIRO MOREIRA

Fatores associados à mobilidade espacial no Brasil:

Uma análise do perfil do migrante recente

Relatório Final de conclusão do curso de
graduação em Bacharel em Estatística,
Departamento de Estatística, Instituto de
Ciências Exatas, Universidade de Brasília.

Orientadora: ANA MARIA NOGALES VASCONCELOS

Coorientador: HERTON ELLERY DE ARAÚJO

Brasília
2011

Dedicamos este trabalho a nossos pais, irmãos e amigos que estiveram conosco e que nos ajudaram, ainda que de forma indireta, na conclusão de mais essa etapa de nossa formação profissional.

AGRADECIMENTOS

Agradecemos em primeiro lugar a Deus, que nos trouxe até onde estamos e que nos colocou juntos nisso tudo, às nossas famílias pelo carinho, preocupação e cuidado de sempre, aos amigos, ao Leandro e à Amanda que nos ajudaram e tiveram paciência com as nossas dúvidas, preocupações e estresses, relacionados ou não com esse projeto. Agradecemos também à professora e orientadora Ana Maria Nogales pelo empenho, motivação e críticas quando necessárias.

*O meu pai era paulista
Meu avô, pernambucano
O meu bisavô, mineiro
Meu tataravô, baiano
Vou na estrada há muitos anos
Sou um artista brasileiro*

Chico Buarque

SUMÁRIO

1 INTRODUÇÃO	2 -
2 OBJETIVOS	4 -
2.1 Objetivos Gerais	4 -
2.2 Objetivos Específicos	4 -
3 METODOLOGIA.....	4 -
3.1 Fonte de Dados – Pesquisa Nacional por Amostra de Domicílios	4 -
3.2 Técnicas de Dados Categorizados	7 -
3.2.1 Definição e classificação das variáveis segundo as categorias.....	7 -
3.2.2 Classificação do tipo de amostragem	7 -
3.2.3 Relação entre variáveis e Tabelas Bidimensionais	8 -
3.2.4 Técnicas de comparação entre categorias.....	8 -
3.2.5 Teste para independência.....	10 -
3.3 Regressão Logística	11 -
3.3.1 Definição do modelo	11 -
3.3.2 Construção do modelo	12 -
3.3.3 Validação do modelo	14 -
3.4 Software.....	14 -
4 RESULTADOS	15 -
4.1 Análises Descritivas	15 -
4.1.1 Comparação entre Migrantes e Não Migrantes	15 -
4.1.2 Principais Fluxos Migratórios	17 -
4.1.3 Os migrantes em relação à população residente	19 -
4.1.4 A migração intra-regional e inter-regional	20 -
4.1.5 Um perfil dos migrantes por fluxo migratório.....	21 -
4.2 Regressão Logística.....	26 -
4.2.1 Norte – Nordeste, 2001 e 2009	27 -
4.2.2 Nordeste – Sudeste, 2001 e 2009.....	30 -
4.2.3 Sudeste – Nordeste, 2001 e 2009.....	32 -
4.2.4 Sul – Sudeste, 2001 e 2009.....	35 -
4.2.5 Centro-Oeste – Sudeste, 2001 e 2009.....	37 -
4.3 Análise dos modelos.....	41 -
4.3.1 Curvas ROC.....	43 -
5 CONCLUSÃO.....	45 -
6 REFERÊNCIAS	46 -
ANEXO A – Histórico de pesquisas suplementares da PNAD.....	48 -

Fonte: IBGE – Trabalho e Rendimento: Pesquisa Nacional por Amostra de Domicílios	49 -
ANEXO B – Principais características observadas na PNAD	50 -
APÊNDICE A – Características Sócio-Demográficas - 2001, 2005 e 2009	53 -
APÊNDICE B – Características Sócio-Demográficas Médias, 2001-2009	54 -
APÊNDICE C – Características Gerais por Fluxos Intra-Regionais.....	55 -
APÊNDICE D – Características Gerais pelos Principais Fluxos Migratórios entre as Regiões	56 -

RESUMO

O estudo da mobilidade espacial é de grande relevância, pois colabora na formação de um povo em aspectos sociais e econômicos. Diante dessa importância e da ausência de publicações recentes acerca do tema, foi realizado o estudo, entre 2001 e 2009, que identificou características e variáveis associadas ao fenômeno da migração e ao perfil do migrante recente. Para uma análise mais aprofundada, foi utilizada a regressão logística para identificar as variáveis que mais se relacionavam com os principais fluxos migratórios estudados. Outra ferramenta utilizada para diferenciar os fluxos selecionados foi a razão de chances e a curva ROC. Evidenciou-se a diminuição do percentual de migrantes ao longo dos anos; o migrante, no geral, como jovem, com maior taxa alfabetização e grande parte nascida no Nordeste; a baixa interação da região Sul; o fenômeno da migração de retorno de forma significativa, ou seja, 30% dos migrantes voltaram a sua Unidade da Federação de nascimento.

Palavras-Chave: Migrante, regressão logística, razão de chances, curva ROC.

1 INTRODUÇÃO

A motivação inicial desse trabalho consistiu em estudar aplicações concretas da estatística em temas de grande relevância. Entre as diversas aplicações, percebeu-se como o fenômeno da mobilidade espacial contribui diretamente para a formação de um povo.

De acordo com o Núcleo de Estudos de População da Universidade de Campinas, o estudo da mobilidade espacial descreve movimentos migratórios e suas implicações no espaço. A partir do estudo desses movimentos, percebe-se a origem da miscigenação da população brasileira e do multiculturalismo presente no país, que se deu por meio da migração de diferentes povos para o seu território. A migração foi um fator determinante para a formação do povo brasileiro, portanto, a análise desse movimento é essencial.

Para análises desses deslocamentos, três aspectos devem ser observados: distância percorrida, mudança de residência habitual e tempo de residência no local de destino. Observando-se a distância percorrida e os limites político-administrativos ultrapassados, a movimentação é caracterizada. Se esses limites são fronteiras, tem-se a migração internacional, caso seja dentro do próprio estado, migração intra-estadual, se for uma mudança temporária, devido ao clima, agricultura e/ou estação do ano, migração sazonal. E há aqueles movimentos que não envolvem mudança de residência, como os movimentos pendulares, nos quais a pessoa sai da cidade de residência para trabalhar e, ao final do mesmo dia, volta para sua cidade, que pode ser denominada cidade dormitório.

A migração é um fenômeno presente em toda a história da humanidade. Como exemplos, têm-se as migrações devidas às causas naturais: vulcões, terremotos, furacões, tsunamis, secas, como a conhecida emigração nordestina, etc. Há ainda aqueles movimentos devidos a guerras: invasões e conquista de territórios, emigrações e busca de refúgio. No entanto, grande parte dos deslocamentos está ligada à expansão econômica.

No final do século XV, iniciou-se o processo de ocupação do continente americano por europeus e africanos. Esta migração para o novo continente foi muito intensa do século XVI até o século XX.

No Brasil, a ocupação territorial se iniciou pelo litoral, com a exploração da cana-de-açúcar e do pau-brasil nos séculos XVI e XVII. O enfraquecimento dessas

fontes de riqueza e a descoberta de jazidas de ouro nos atuais estados de Minas Gerais, Goiás e Mato Grosso foram determinantes para o povoamento do interior do país, no final do século XX. A população, que cresceu muito no interior ao longo do século XVIII, precisou, posteriormente, dispersar-se para outras regiões do país, uma vez que as jazidas se esgotavam.

Com isso, a população brasileira buscou novas fontes de riqueza, o que desencadeou um novo fluxo migratório para a região Norte, por conta da exploração da borracha, e outro em direção ao Sul e ao Sudeste, para as plantações de café.

Em 1956, a construção da nova capital do Brasil foi iniciada e, para isso, foi necessário o trabalho de milhares de pessoas, migrantes de todo o país. Ocorrendo assim, outra forte interiorização.

Muitos estudos a respeito deste tema já foram realizados, sendo a maior parte deles baseados nos Censos Demográficos. De fato, o Censo é a principal fonte de dados demográficos no Brasil, que é realizado a cada dez anos.

Uma vez que os resultados do Censo de 2010 para o quesito migração ainda não se encontravam disponíveis no início do estudo, para esta análise do perfil do migrante recente, foram utilizados os dados da Pesquisa Nacional por Amostra de Domicílios, PNAD. Além disso, os dados da PNAD permitem um acompanhamento anual da variável migração, que, por ser dinâmica, observá-la ano a ano facilita a percepção de seu comportamento. Uma limitação desse trabalho pela utilização dos dados da PNAD é seu plano amostral, que não é feito especificamente para análise da migração.

2 OBJETIVOS

2.1 Objetivos Gerais

- Identificar fatores associados à mobilidade espacial no Brasil;
- Traçar o perfil do migrante recente.

2.2 Objetivos Específicos

- Identificar os principais fluxos migratórios recentes no país;
- Caracterizar o migrante brasileiro segundo sexo, idade e unidades da federação de nascimento, origem e destino;

3 METODOLOGIA

3.1 Fonte de Dados – Pesquisa Nacional por Amostra de Domicílios

Neste estudo, foi considerado migrante o indivíduo que saiu de uma Unidade da Federação para habitar em outra, nos cinco anos antes da data da pesquisa. Dessa forma não foi considerada a migração dentro do próprio estado ou do Distrito Federal.

Os dados utilizados foram aqueles provenientes da Pesquisa Nacional por Amostra de Domicílios, PNAD. Essa escolha foi motivada por:

- A PNAD possui variáveis que caracterizam a migração, além de diversas variáveis relacionadas à saúde, educação, moradia, trabalho e renda, entre outras;
- A PNAD é realizada anualmente, o que é bom para analisar o período intercensitário;
- Desde 2004, a amostragem é feita para todo o Brasil, dessa forma, pode-se fazer uma análise da migração dos brasileiros, independente da região;
- Grande credibilidade para análises de características da população.

A Pesquisa Nacional por Amostras de Domicílios é um levantamento de dados anual, realizado por amostra probabilística de domicílios pelo Instituto Brasileiro de Geografia e Estatística, IBGE, que traz variáveis que caracterizam a população brasileira e visa produzir informações para o estudo sócio-econômico do país.

A PNAD foi criada para suprir a ausência de informações sobre a população brasileira nos períodos intercensitários e aprofundar em temas que existem pouca informação ou não existem nos censos demográficos.

De acordo com as notas técnicas da PNAD, na década de 1960, ficou evidente a necessidade de mais informações sócio-econômicas a respeito do Brasil. A pesquisa censitária fornece essas informações, mas, por ser realizada a cada dez anos, elas ficam defasadas e escassas, pois, em dez anos, o país pode mudar muito. Portanto, é interessante possuir um levantamento com periodicidade anual, como é a PNAD. Por ser uma pesquisa realizada por amostras de domicílios, permite ainda uma economia de tempo de execução e custos para o país.

O levantamento começou a ser feito a partir de 1967, de forma que, em seus primeiros anos, sua periodicidade era trimestral. A partir de 1971, ela adquiriu periodicidade anual, sendo realizada no último trimestre de cada ano, exceto em anos censitários, nos quais a pesquisa é suspensa.

As variáveis da PNAD sofreram alterações ao longo dos anos, possuindo algumas de caráter permanente, e outras que são acrescentadas para suprir a necessidade de algumas informações a mais para o país.

A PNAD possui três linhas de pesquisa:

- Pesquisa básica: É representada pelas variáveis mais importantes que indicam as características gerais da população, auxilia num acompanhamento socioeconômico do país;
- Pesquisas suplementares: Tratam de temas de interesse ligados aos temas da pesquisa básica. Fazem um aprofundamento em temas permanentes;
- Pesquisas especiais: Tratam de temas que possuem uma complexidade maior que os temas da pesquisa básica. Algumas vezes, é necessária uma amostragem distinta.

Nas primeiras décadas, informações sobre migração não eram incluídas na pesquisa básica, isso somente aconteceu a partir de 1992. Até então, era um tema recorrente em

pesquisas suplementares. No anexo I, encontra-se um histórico dos principais temas das pesquisas suplementares desde 1967.

Em 2004, a PNAD alcançou a cobertura completa do país, processo que aconteceu de forma gradativa. Em seu início, compreendia a área correspondente ao atual Estado do Rio de Janeiro e ao final da década já compreendia toda a região Sul, Sudeste, Nordeste e o Distrito Federal.

A partir de 1973, as demais regiões, Centro-Oeste e Norte, tiveram suas áreas urbanas incluídas. Em 1981 a parte rural do Centro-Oeste foi abrangida e, somente em 2004, a cobertura passou a ser de todo território nacional, com a entrada da área rural do Norte do Brasil. Cabe ressaltar que, em 1992, o Estado do Tocantins passou a ser parte da região Norte, e não da Centro-Oeste do país como até então.

A PNAD é realizada por meio de uma amostragem probabilística de domicílios feita em três estágios de seleção: municípios, setores censitários e unidades domiciliares. A seleção é feita em algumas etapas:

- Os municípios são divididos em dois grupos:
 - Aqueles que devem pertencer à amostra, auto-representativos, grupo formado pelas capitais e municípios das regiões metropolitanas;
 - Aqueles que podem pertencer à amostra, não auto-representativos.
- Entre os municípios não auto-representativos, é feita uma amostragem estratificada e, em cada estrato, são selecionados os municípios com reposição e com probabilidade proporcional à população residente obtida no último Censo Demográfico.
- Após a seleção dos municípios, selecionam-se os setores censitários da mesma forma que nos municípios não auto-representativos, com probabilidade proporcional ao número de domicílios presentes no último Censo Demográfico.
- Dentre os setores censitários, são selecionados os domicílios com igual probabilidade.

O plano amostral da PNAD é baseado nas quantidades populacionais obtidas no Censo mais recente, no caso o Censo de 2000. Isso pode trazer uma falha, uma vez que no intervalo de dez anos, municípios podem ser criados ou se desenvolverem. A PNAD não trata dessa realidade, de forma que não há uma atualização dos pesos dos setores censitários.

Devido à complexidade da amostra, foram atribuídos pesos para cada indivíduo, estes pesos serão considerados para as análises dos dados.

Como o levantamento só passou a abranger à área rural da região Norte em 2004, até 2003, foi realizado o processo de expansão da amostra, utilizando a projeção da população residente de cada Unidade da Federação, considerando a evolução populacional obtida nos últimos Censos Demográficos. A partir de 2004, a pesquisa passou a fazer uma cobertura total do país.

Uma relação contendo as principais variáveis presentes na PNAD se encontra no Anexo II.

3.2 Técnicas de Dados Categorizados

Uma vez que os objetivos visam traçar um perfil do migrante brasileiro e encontrar variáveis associadas à migração e a maior parte das variáveis presentes na PNAD é categorizada, é apropriado o uso de técnicas de dados categorizados.

3.2.1 Definição e classificação das variáveis segundo as categorias

Segundo Agresti, em seu livro de introdução à análise de dados categorizados, uma variável categorizada é aquela na qual a escala de medida consiste em um conjunto de categorias. É caracterizada por possuir diferentes alternativas, nas quais, somente uma delas pode ser escolhida como resposta em um conjunto de respostas possíveis para a questão.

3.2.2 Classificação do tipo de amostragem

Os procedimentos padrões de dados categorizados assumem basicamente três modelos de amostragem probabilística: Poisson, Binomial, Multinomial.

Essas classificações são dadas de acordo com o que se define a priori, na amostragem. No modelo de Poisson, o tempo de coleta é definido a priori, ou seja, não se sabe o número de sucessos e nem o tamanho da amostra a ser coletada. Nos modelos Binomial e Multinomial, o tamanho da amostra é definido a priori, com diferença na quantidade de categorias da variável, o Binomial é para variável binária e o Multinomial é para variável com mais de duas categorias.

3.2.3 Relação entre variáveis e Tabelas Bidimensionais

Uma análise isolada de cada variável, univariada, traz menos conclusões do que uma em que diferentes variáveis são observadas ao mesmo tempo, para explicar certos fenômenos, multivariada. No entanto, em algumas situações, para se fazer análises de variáveis simultaneamente, é necessário entender a diferença entre suas classificações:

- Variável explicativa;
- Variável resposta.

Considerando duas variáveis: ‘Migrou do Nordeste para o Sudeste’ e ‘Sexo’, a variável ‘Sexo’ não pode ser influenciada pela primeira, por outro lado, ela pode influenciar a variável ‘Migrou do Nordeste para o Sudeste’. Assim podemos classificá-las como:

- ‘Sexo’: Variável explicativa;
- ‘Migrou do Nordeste para o Sudeste’: Variável resposta.

3.2.4 Técnicas de comparação entre categorias

Para comparar duas proporções de sucessos de grupos independentes p_1 e p_2 , e verificar se são estatisticamente diferentes, três medidas podem ser usadas: diferença de proporções, risco relativo e razão de chances, no entanto, nesse estudo foram utilizadas apenas as técnicas de risco relativo e razão de chances.

Risco Relativo

Esta medida consiste na razão entre as duas proporções. Este resultado varia de zero a infinito.

Seja o valor do risco relativo dado por X , se $0 < X < 1$, diz-se que o grupo do denominador tem $1/X$ vezes mais risco de apresentarem a característica em estudo. Por outro lado, se $X > 1$, a categoria do numerador possui X vezes mais risco de apresentar a característica em estudo. Caso as proporções sejam independentes, o risco é igual a um.

Para uniformizar a escala, trabalha-se com o logaritmo do risco relativo, cujo desvio padrão é dado por:

$$\hat{\sigma}(\log RR) = \sqrt{\frac{(1-p_1)}{p_1 n_{1+}} + \frac{(1-p_2)}{p_2 n_{2+}}}$$

O Intervalo de confiança para o logaritmo do risco relativo é

$$\log(RR) \pm z_{\alpha/2} \hat{\sigma}(\log RR)$$

O intervalo para o risco relativo é dado pela aplicação do exponencial nos valores encontrados na fórmula acima.

Razão de Chances

Se, por um lado, a probabilidade de sucesso (p) é dada pela razão entre o número de sucessos e o número total de eventos, por outro, a chance (odds) é dada pela razão entre a quantidade de sucessos e o número de fracassos, ou seja:

$$odds = p/(1-p)$$

Com isso, obtém-se a chance para uma linha da tabela de contingência. A razão de chance consiste na divisão das chances obtidas em duas linhas:

$$OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

A razão de chances possui algumas propriedades:

- $0 < OR < \infty$;
- Se $OR = 1$, então não há associação entre as variáveis;
- Se $1 < OR < \infty$, então a chance de sucesso é maior na primeira linha;
- Se $0 < OR < 1$, então a chance de sucesso é maior na segunda linha;
- Não é preciso diferenciar a variável resposta da variável explicativa;

Em termos amostrais também se pode calcular a razão de chance pelos produtos cruzados da tabela de contingência

$$\widehat{OR} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} = \frac{n_{11}/n_{12}}{n_{21}/n_{22}} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$

Seja o valor da razão de chances dado por X , se $0 < X < 1$, diz-se que a categoria do denominador tem $1/X$ vezes mais chance de sucesso. Por outro lado, se $X > 1$, a categoria do numerador possui X vezes mais chance de sucesso. Quando o resultado é igual a um, não existe associação entre as variáveis.

Quando a amostra é grande, como é o caso da PNAD, tem-se que o

$$\log(\widehat{OR}) \sim N\left(0, \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}\right)$$

E o intervalo de confiança é dado por:

$$\log(\widehat{OR}) \pm z_{\alpha/2} \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

Assim como no risco relativo, basta aplicar exponencial nos limites do intervalo para obtermos os limites para a razão de chance.

Quando as proporções de sucesso são muito próximas de zero, os valores da Razão de Chance e do Risco Relativo são parecidos entre si.

3.2.5 Teste para independência

Para verificação da independência entre as variáveis, podem-se usar também testes específicos, como o Qui-Quadrado.

Qui-Quadrado

São calculados os valores esperados em cada casela da tabela de contingência, e esses valores são comparados com os valores de fato observados, supondo que as proporções são iguais, pode-se verificar a independência entre as variáveis. A estatística de Qui-Quadrado é dada por:

$$\frac{\sum (n_{ij} - \mu_{ij})^2}{\mu_{ij}} \sim \chi^2_{(I-1)(J-1)}$$

onde I = número de linhas e J = número de colunas.

Com (I-1)(J-1) graus de liberdade

Esta estatística assume valores de 0 a $+\infty$ e, quanto maior o resultado maior a evidência contra a hipótese de independência.

Limitações:

- A amostra precisa ser grande;

- Indica apenas a existência de associação entre as variáveis tabeladas, e não a natureza dessa associação.

3.3 Regressão Logística

A Regressão Logística é o modelo de regressão que melhor se adequa à variável binária ‘migrante do principal fluxo’ ou ‘migrante dos demais fluxos’.

3.3.1 Definição do modelo

O modelo de regressão logística é usado quando a variável resposta é qualitativa binária. Para esse estudo, essa variável será o indivíduo ser ou não migrante do principal fluxo. A regressão logística também pode ser utilizada quando a variável resposta é qualitativa e possui mais de duas categorias, porém não será abordada no estudo.

O modelo é dado por:

$$Y_i \sim \text{Bernoulli}(\pi_i)$$

$$\log\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 X_{i,1} + \dots + \beta_{p-1} X_{i,p-1} \quad (1)$$

Y_i = Variável resposta, indivíduo ser migrante ou não, $i = 1, \dots, n$.

$X_{j,i}$ = Variável preditora j , $j = 1, 2, \dots, p - 1$, do indivíduo i .

β_k = Coeficiente relacionado à variável preditora j , $k = 0, 1, \dots, p - 1$

Como se trata de uma análise multivariada, em forma matricial nós temos:

$$\beta_{(px1)} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} \quad X_{(px1)} = \begin{bmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_{p-1} \end{bmatrix} \quad X_{i(p \times 1)} = \begin{bmatrix} 1 \\ X_{i1} \\ X_{i2} \\ \vdots \\ X_{i,(p-1)} \end{bmatrix}$$

Para o caso de apenas uma variável preditora, tem-se a seguinte fórmula para o valor esperado de Y :

$$E(Y) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)} \quad (2)$$

Expandindo esta fórmula para o caso de múltiplas variáveis preditoras:

$$E(Y) = \frac{\exp(\beta'X)}{1 + \exp(\beta'X)} = \pi \quad (3)$$

onde β' é o vetor linha dos coeficientes das variáveis preditoras.

O modelo para cada variável consiste em:

$$E(Y_i) = \frac{\exp(\beta'X_i)}{1 + \exp(\beta'X_i)} = \pi_i, \quad i = 1, \dots, n \quad (4)$$

O modelo logístico convencional exige que a variável resposta seja uma variável qualitativa binária, porém não existem restrições quanto às variáveis preditoras. Isto torna o modelo muito robusto e o torna convidativo, pois se pode explorar uma infinidade de variáveis preditoras para tentar explicar a variável binária em questão.

3.3.2 Construção do modelo

Para construir o modelo de regressão que será adotado, ou seja, selecionar quais variáveis irão compor o modelo, pode-se usar o método *stepwise*. Este método consiste em remover ou inserir variáveis explicativas no modelo uma a uma até encontrar o nível de explicação desejado.

Uma vez alocadas as variáveis no modelo, algumas delas podem perder o poder de explicação devido à existência de multicolinearidade, que é a correlação entre as variáveis explicativas. Isso acarreta em perda de precisão nas estimações do modelo. Então, para tentar sanar o problema, é necessário retirar algumas variáveis do modelo e, para fazer a escolha certa, existem testes para identificar quais variáveis estão mais propensas a deixar o modelo de regressão.

Para uma variável deixar o modelo, é necessário que seu coeficiente β seja igual a zero. Como se trata de um modelo multivariado, pode-se testar a nulidade de vários β 's de uma só vez.

Para tanto, usaremos o conceito de *deviance*, que consiste em comparar o logaritmo da verossimilhança do modelo com todas as variáveis incluídas com o logaritmo do modelo sem as variáveis que se deseja retirar do modelo.

A *deviance* de um modelo logístico é dada por:

$$DEV(X_0, X_1, \dots, X_{p-1}) = -2 \sum_{i=1}^n [Y_i \log_e(\pi_i) + (1 - Y_i) \log_e(1 - \pi_i)] \quad (5)$$

Caso a *deviance* seja pequena, podemos dizer que o modelo completo possui praticamente o mesmo poder de explicação que o modelo sem as variáveis que se deseja retirar, logo se pode usar o modelo sem as variáveis em questão, pois ele será mais simples de se trabalhar e resultará nas mesmas conclusões que o modelo completo.

Existe também o conceito de *deviance* parcial. Este conceito mostra que, para cada modelo ajustado, pode-se calcular a sua *deviance*. A diferença entre as *deviances* de dois modelos de pesquisa é denominada *deviance* parcial e, por meio dela, pode-se verificar se determinadas variáveis podem ser retiradas do modelo.

Primeiramente devem-se estimar, por meio de máxima verossimilhança, os parâmetros β 's que acompanham as variáveis X 's.

Com isso pode-se calcular a *deviance* do modelo completo. Para o modelo sem as variáveis que se deseja retirar, calcula-se a *deviance* de acordo com as variáveis remanescentes no modelo.

Com os resultados das duas *deviances* em mãos pode-se inferir se as variáveis retiradas são necessárias no modelo. Caso a diferença entre as duas seja pequena pode-se dizer que as variáveis deveriam sair o modelo, caso seja grande elas devem permanecer no modelo, porém, para não se tornar algo arbitrário e subjetivo a escolha de que valores são 'grandes' e que valores são 'pequenos', pode-se admitir para esta diferença uma distribuição Qui-Quadrado com p -z graus de liberdade.

$$H_0: \beta_0 = \beta_1$$

$$H_1: \beta_0 \neq \beta_1$$

$$\begin{aligned} & \text{DEV}(X_0, X_1, \dots, X_{p-1} | X_0, X_1, \dots, X_{z-1}) \\ &= \text{DEV}(X_z, X_{z+1}, \dots, X_{p-1}) - \text{DEV}(X_0, X_1, \dots, X_{z-1}) \quad (6) \end{aligned}$$

$$\text{DEV}(X_0, X_1, \dots, X_{p-1} | X_0, X_1, \dots, X_{z-1} \leq \chi^2(1 - \alpha; p - z) \text{ não se rejeita } H_0$$

$$\text{DEV}(X_0, X_1, \dots, X_{p-1} | X_0, X_1, \dots, X_{z-1} > \chi^2(1 - \alpha; p - z) \text{ rejeita - se } H_0$$

Usando esta técnica permite-se verificar a existência de interações entre as variáveis no modelo, basta encontrar seus respectivos coeficientes estimados e o procedimento será similar ao já proposto anteriormente.

$$B'X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

3.3.3 Validação do modelo

Deve-se tentar aplicar o mesmo modelo a diferentes amostras provenientes da mesma população ou, até mesmo, em subamostras da amostra original. Feito isso, deve-se observar se os valores dos desvios e inferências feitas baseados no modelo são similares em todos os casos testados.

3.4 Software

Para analisar e tratar os dados provenientes da PNAD, será utilizado o software SAS, nas versões 9.1 e 9.2.

4 RESULTADOS

4.1 Análises Descritivas

No início da primeira década dos anos 2000, os migrantes representavam cerca de 2,6% da população brasileira, e totalizavam aproximadamente 4,5 milhões de pessoas (Tabela 1). No entanto, no final dessa mesma década, viu-se a diminuição do volume de indivíduos que mudaram de UF de residência nos últimos cinco anos. Eles representavam, em 2009, 1,7% da população brasileira, correspondendo a 3,2 milhões de pessoas, uma redução de 1,3 milhão de pessoas.

Tabela 1. População Total, número absoluto e relativo de migrantes – Brasil, 2001 - 2009

Caracterização da População	Ano								
	2001	2002	2003	2004	2005	2006	2007	2008	2009
População	170.955.241	173.501.227	175.954.185	181.687.258	183.880.894	186.020.850	188.028.891	189.952.795	191.795.854
Migrantes	4.449.510	4.672.755	4.551.898	4.631.491	4.725.935	4.443.435	3.252.081	3.317.932	3.228.143
% de Migrantes	2,60%	2,69%	2,59%	2,55%	2,57%	2,39%	1,73%	1,75%	1,68%

4.1.1 Comparação entre Migrantes e Não Migrantes

Para uma visão geral das características dos migrantes e não migrantes, pode-se observar, na tabela do apêndice A, algumas variáveis mais relevantes em uma comparação entre migrantes e não migrantes para os anos de 2001, 2005 e 2009 e, no apêndice B, as principais variáveis utilizadas no estudo para a média de migrantes e não migrantes entre 2001 e 2009.

Em uma análise comparativa entre migrantes e não migrantes foi observado que, com uma pequena diferença, a maior parte dos migrantes e dos não migrantes é do sexo feminino. Para os dois grupos de pessoas, as raças/cor da pele mais respondidas foram ‘branca’ e ‘parda’ (juntas, somam mais de 90% dos pesquisados em todos os anos), no entanto, ao longo dos anos, está acontecendo um suave aumento daqueles de raça ‘preta’, e um aumento na proporção de ‘pardos’ em relação a ‘brancos’, sendo que, em 2009, os migrantes de raça ‘parda’ já superaram aqueles com raça ‘branca’.

Em relação à idade, houve uma notável diferença entre aqueles que migraram e os que não migraram. No que diz respeito aos não migrantes, entre 2001 e 2009, observou-se um aumento na proporção das faixas etárias mais velhas, e uma diminuição nítida na faixa correspondente a 10 a 19 anos, que ainda diminuindo, representa mais de um quinto da população. Em relação aos migrantes, a faixa que mais se destaca é a de 20 a 29 anos, que está, ano a ano, aumentando, chegando a quase 30% dos migrantes em 2009. Os migrantes se concentram mais nas faixas etárias dos jovens, entre 20 e 29 anos, os não migrantes, por outro lado, estão melhores distribuídos ao longo das faixas que os migrantes.

A proporção do número de pessoas que sabem ler e escrever tem aumentado ao longo dos anos, no Brasil, porém, observou-se isso mais acentuado quando se analisou aquelas pessoas que são consideradas migrantes. Em 2009, 91% dos que migraram, sabem ler e escrever, e em relação aos que não são migrantes, 83%, como visto no apêndice A.

A variável ‘faixa de renda familiar mensal’ mostrou que entre migrantes e não migrantes, a partir do ano 2003, o percentual daqueles que recebiam 5 ou mais salários mínimos reduziu, em contrapartida, o percentual daqueles que recebiam entre 1 e 3 salários mínimos aumentou e ultrapassou a faixa de maior renda. De forma percentual, as demais classes se mantiveram constantes. No entanto, o comportamento dos migrantes possuiu uma maior variância em relação aos demais.

Analisando a Unidade da Federação (UF) onde nasceram, é possível notar pelo apêndice B a diferença entre os dois grupos, entre os migrantes, cerca de 42% nasceram nos estados correspondentes à região Nordeste, seguido pela região Sudeste, com 29%; por outro lado, entre os não migrantes, cerca de 40% nasceram na região Sudeste, seguidos pelos da região Nordeste (somam juntos 73% dos não migrantes). Também é interessante observar que, entre os migrantes, cerca de 30% residem na mesma UF onde nasceram, ou seja, voltaram à UF de nascimento; e cerca de 15% dos não migrantes não residem na UF de nascimento, ou seja, migraram há mais de 5 anos.

Entre 2001 e 2009, foi observado que, da mesma forma que os destinos mais frequentes dos migrantes foram as regiões Sudeste (34%) e Nordeste (27%), elas são as regiões com uma maior população, com cerca de 43% e 28% dos não migrantes,

respectivamente. No entanto, a região Centro-Oeste, com apenas 7% da população aproximadamente, foi a região mais procurada por 16% dos migrantes, o terceiro lugar mais recorrente.

Quando se trata das UF's de origem, em torno de 35% dos migrantes observados entre 2001 e 2009 saíram da região Sudeste, seguidos pelos que saíram da região Nordeste, em torno de 29%. Observa-se que a origem menos frequente foi a região Norte, que varia em torno de 10% da origem de todos os migrantes. Vale observar que a região Sudeste, seguida pela Nordeste, foi aquela mais recorrente tanto na origem quanto no destino mais procurado. A região Norte foi a de menor destino e de menor origem.

4.1.2 Principais Fluxos Migratórios

Nesta parte, o objetivo é identificar os principais fluxos migratórios do Brasil. Para tanto, será utilizada a tabela 2. Em cada célula da tabela, há quatro linhas: na primeira delas, é dado o número médio de pessoas que migraram da região de origem para a região de destino entre 2001 e 2009; na segunda linha, o percentual de migrantes que saíram daquela origem para aquele destino em relação ao total de migrantes brasileiros; na terceira linha, o percentual da linha, ou seja, o percentual de pessoas que migrou para cada região saindo de uma região específica; na quarta linha, o percentual da coluna, ou seja, o percentual de pessoas que se originou das diferentes regiões para aquela região específica.

Tabela 2. Região de origem por região de destino dos migrantes brasileiros – Brasil, 2001 - 2009

Frequência média Percentual Percentual da linha Percentual da coluna	REGIÃO DE DESTINO					Total
REGIÃO DE ORIGEM	Norte	Nordeste	Sudeste	Sul	Centro-Oeste	
Norte	155484	101908	52533	19322	83871	413117
	3,8%	2,5%	1,3%	0,5%	2,0%	10,0%
	37,6%	24,7%	12,7%	4,7%	20,3%	100,0%
	37,6%	9,2%	3,7%	3,6%	12,4%	10,0%
Nordeste	134221	362565	517814	26809	167087	1208495
	3,2%	8,8%	12,5%	0,7%	4,0%	29,2%
	11,1%	30,0%	42,9%	2,2%	13,8%	100,0%
	32,5%	32,8%	36,9%	5,0%	24,6%	29,2%
Sudeste	51324	523393	580576	190045	166189	1511529
	1,2%	12,6%	14,0%	4,6%	4,0%	36,5%
	3,4%	34,6%	38,4%	12,6%	11,0%	100,0%
	12,4%	47,4%	41,4%	35,1%	24,5%	36,5%
Sul	18318	23454	128195	256100	75696	501763
	0,4%	0,6%	3,1%	6,2%	1,8%	12,1%
	3,7%	4,7%	25,6%	51,0%	15,1%	100,0%
	4,4%	2,1%	9,1%	47,3%	11,2%	12,1%
Centro-Oeste	54126	93187	123815	49223	185548	505900
	1,3%	2,3%	3,0%	1,2%	4,5%	12,2%
	10,7%	18,4%	24,5%	9,7%	36,7%	100,0%
	13,1%	8,4%	8,8%	9,1%	27,4%	12,2%
Total	413473	1104508	1402933	541499	678391	4140804
	10,0%	26,7%	33,9%	13,1%	16,4%	100,0%
	10,0%	26,7%	33,9%	13,1%	16,4%	100,0%
	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

A tabela 2 mostra que os maiores fluxos migratórios do Brasil, em média, para o período entre 2001 e 2009, foram do Sudeste para o Nordeste, do Nordeste para o Sudeste e, principalmente, de UFs do Sudeste para outras UFs do Sudeste (todas em negrito). Por outro lado, a tabela evidencia uma baixa interação entre as UFs da região Sul com as UFs das regiões Norte e Nordeste, ou seja, os menores fluxos migratórios são aqueles do Sul para Nordeste ou Norte, e vice-versa. É importante reforçar que a tabela foi elaborada com o número médio de migrantes entre 2001 e 2009.

No que se refere aos percentuais de linha e coluna (valores da terceira e quarta linhas das células da tabela 2), é interessante notar que os percentuais da diagonal principal são os maiores, exceto o da região Nordeste, ou seja, entre aqueles que migraram de uma região específica, os destinos mais procurados, foram UFs componentes da mesma região de origem, ou, entre os migrantes que foram para uma região específica, as origens mais recorrentes são UFs da mesma região, mas para ambos os casos, quando não se trata da região Nordeste.

Para a região Nordeste, o destino mais procurado, quando ela foi a origem, foram as UFs da região Sudeste, e quando a região Nordeste foi o destino, a origem mais recorrente foram aquelas UFs da região Sudeste.

Quando se analisa a região Sul como origem, é possível perceber que mais da metade dos migrantes foram para outra UF da região Sul.

A origem mais recorrente dos migrantes foram as UFs da região Sudeste, e da mesma forma, o destino mais procurado dos migrantes brasileiros foram também da região Sudeste. Por outro lado, o destino menos buscado e com menor percentual de origem foram da região Norte.

Também é possível verificar na tabela, que embora a região Sudeste tenha sido a mais procurada, quando se analisa a quantidade de migrantes que saiu desta região, para as demais, exceto a Norte, observa-se que esse valor referente aos migrantes de saída foi maior que aqueles que vieram das outras. Ou seja, em relação às regiões Nordeste, Sul e Centro-Oeste, mais indivíduos saíram do Sudeste para estas do que destas para o Sudeste.

4.1.3 Os migrantes em relação à população residente

Já foi visto, que as regiões onde mais se encontraram migrantes foram o Sudeste e o Nordeste, no entanto, essas regiões também são aquelas onde há uma maior população. Por isso, é interessante observar o percentual de migrantes por região, para verificar se, de fato, é nessas regiões que os migrantes correspondem uma parcela mais significativa da população. A tabela 3 mostra esses valores. No entanto, vale informar que, para elaborá-la foram desconsiderados os migrantes que se originaram de outro

país com destino ao Brasil. Uma vez que o estudo é a respeito dos fluxos migratórios internos, no Brasil.

Tabela 3. Número de migrantes médio por região – Brasil, 2001 - 2009

Valores Médios de 2001 a 2009	Região					
	Total	Norte	Nordeste	Sudeste	Sul	Centro-Oeste
População média	170488649	9860931	49065909	73927325	25650178	11984306
Nº de Migrantes em média	4140804	413473	1104508	1402933	541499	678391
% de Migrantes	2,43%	4,19%	2,25%	1,90%	2,11%	5,66%

Nota-se, pela tabela 3, que as regiões onde os migrantes formaram uma parcela mais expressiva da sua população foram o Centro-Oeste e o Norte. Pela tabela 2, foi visto que mais de um terço dos migrantes foram para a região Sudeste, no entanto, esses migrantes não são tão expressivos na população como na região Centro-Oeste devido à extensa população da região Sudeste.

É interessante observar que o Sudeste, aquela região que mais recebeu os migrantes, é a região que a quantidade de migrantes foi a menos expressiva, apenas 1,9% da população é formada por esses migrantes.

4.1.4 A migração intra-regional e inter-regional

Dentro de algumas regiões, a migração entre suas Unidades da Federação é maior do que a migração para UFs das demais regiões. Essa verificação pode ser feita por meio da tabela 4, que indica para onde foram e de onde os migrantes se originam.

Tabela 4. Região de origem e destino dos migrantes brasileiros

Freq. média Percentual	Região de Destino					Total
Região de Origem	Norte	Nordeste	Sudeste	Sul	Centro- Oeste	
Norte	155484 3,75%	101908 2,46%	52533 1,27%	19322 0,47%	83871 2,03%	413117 9,98%
Nordeste	134221 3,24%	362565 8,76%	517814 12,51%	26809 0,65%	167087 4,04%	1208495 29,19%
Sudeste	51324 1,24%	523393 12,64%	580576 14,02%	190045 4,59%	166189 4,01%	1511529 36,50%
Sul	18318 0,44%	23454 0,57%	128195 3,10%	256100 6,18%	75696 1,83%	501763 12,12%
Centro-Oeste	54126 1,31%	93187 2,25%	123815 2,99%	49223 1,19%	185548 4,48%	505900 12,22%
Total	413473 9,99%	1104508 26,67%	1402933 33,88%	541499 13,08%	678391 16,38%	4140804 100,00%

Em negrito, estão os destinos mais procurados de acordo com cada origem especificada. É interessante observar que, exceto o Nordeste, o maior percentual de procura dos migrantes foi por UFs da mesma região, como já citado no item 3.1.3.

Ou seja, para as regiões Norte, Sudeste, Sul e Centro-Oeste, a migração foi maior para UFs da própria região – intra-regional – do que a migração para as demais regiões.

4.1.5 Um perfil dos migrantes por fluxo migratório

Para se analisar os perfis dos migrantes por fluxo migratório, foram escolhidos alguns fluxos mais significantes, uma vez que uma análise dos 25 fluxos seria muito extensa. Considerando cada região como origem, foram escolhidos os dois destinos com maior procura, somando dessa forma, 10 fluxos migratórios.

Como visto na seção 4.1.3, para as regiões Norte, Sudeste, Sul e Centro-Oeste, o destino mais buscado foram UFs da própria região, ou seja, intra-regional, para o Nordeste, por outro lado, o destino mais procurado foram as UFs da região Sudeste, no

entanto, em segundo lugar, foram UFs da mesma região. Logo, os fluxos migratórios detalhados no estudo serão:

1. Norte para Norte;
2. Norte para Nordeste;
3. Nordeste para Nordeste;
4. Nordeste para Sudeste;
5. Sudeste para Sudeste;
6. Sudeste para Nordeste;
7. Sul para Sul;
8. Sul para Sudeste;
9. Centro-Oeste para Centro-Oeste;
10. Centro-Oeste para Sudeste.

Os valores citados a seguir podem ser encontrados nos apêndices C e D.

Entre os migrantes que se movimentaram dentro da região Norte, é possível observar que a maior parte é formada por mulheres, embora não seja uma diferença grande, comparada aos homens. 70% desses indivíduos possuíam menos de 30 anos sendo mais de um terço dos migrantes menores de 18 anos. Em relação ao questionamento sobre a raça/cor da pele, quase 70% se classificou como ‘pardo’, as raças/cor da pele mais respondidas foram ‘parda’ e ‘branca’, as demais somam juntas, menos de 6%. Quando se observa a renda desses, verifica-se que quase a metade tem a renda entre um e três salários mínimos, menos de um quarto desses migrantes tem renda igual ou superior a três salários mínimos.

O segundo destino mais procurado pelos migrantes da região Norte é o Nordeste, desses, houve uma ligeira maioria feminina. Separando-os por faixas etárias de menores de 18 anos, entre 18 e 30 anos, entre 30 e 50 anos e maiores de 50, observa-se que a mais freqüente foi a dos mais jovens, com 33%, no entanto, o percentual entre as 3 primeiras classes foi muito próximo, apenas 8% desses migrantes tinham mais de 50 anos de idade. As respostas mais freqüentes diante da pergunta sobre a raça/cor da pele foram ‘parda’ e ‘branca’, tendo esta, metade das respostas daquela, e somando juntas 93%. Quanto à variável renda, mais de três quartos dos migrantes têm renda inferior a

três salários mínimos, e destes, um pouco menos da metade têm renda menor que um salário mínimo.

A maior parte dos migrantes intra-regionais do Nordeste é formada por mulheres. Mais de um terço são jovens entre 18 e 30 anos, seguidos por aqueles com menos de 18 anos, menos de um quinto desses migrantes tinha mais de 50 anos. Em relação à raça/cor da pele respondida, mais de 90% disseram ser ‘pardos’ ou ‘brancos’, mas a maior parte afirmou ser ‘pardo’. O estudo da variável renda mostra que 40% dos migrantes que mudaram de UF no Nordeste têm renda de um salário mínimo ou menos, menos de um quarto dos migrantes possui renda de mais de três salários mínimos.

O destino mais procurado pelos migrantes que saíram de uma UF da região Nordeste foram as UFs da região Sudeste. Houve, entre esses migrantes, uma ligeira maioria feminina e de jovens. Metade dos migrantes tinha entre 18 e 30 anos e um quarto de menores de 18 anos. Sobre a variável raça/cor da pele, quase metade respondeu ser ‘pardo’. ‘Pardos’ e ‘brancos’ caracterizam mais de 90% dos migrantes, e esse fluxo migratório foi o que teve um percentual mais expressivo de outras raças/cor da pele (exceto as duas já citadas), aproximadamente 8% caracterizou-se como ‘negros’. Entre esses migrantes, mais da metade informou ter renda entre um e três salários mínimos, apenas um quarto afirmou ter renda superior a três salários mínimos.

Entre aqueles que mudaram de UF do Sudeste para outra de mesma região, a maior parte foram mulheres. Em relação à idade, assim como em todos os casos de migração intra-regional, mais de um terço eram jovens de 18 a 30 anos, no entanto, esse fluxo foi o que apresentou um percentual mais expressivo para os maiores de 50 anos, 12% dos migrantes desse fluxo. No que se refere à raça/cor da pele, difere dos fluxos observados anteriormente, pois mais da metade respondeu ser ‘branca’ e, apenas um terço, ‘pardos’, o percentual de ‘negros’ também se mostrou significativo. A respeito da renda, pouco mais de um terço recebe mais de três salários mínimos, e em torno de um quinto recebe um salário mínimo ou menos.

Entre os migrantes que saíram da região Sudeste, o destino mais procurado foi a região Nordeste, este foi o primeiro fluxo analisado onde foi observado uma maioria masculina, cerca de 53% dos migrantes eram homens. Neste fluxo também foi encontrada uma característica peculiar, o único onde a faixa etária com mais

observações foi a que se refere à idade entre 30 e 50 anos, no entanto, aquela de menores de 18 anos e aquela entre 18 e 30 também tiveram valores bem próximos. A raça/cor da pele mais respondida foi ‘parda’, seguida por ‘branca’, as demais somaram cerca de 5%. Em relação à renda, pode-se observar que assim como a migração intra-regional no Nordeste, a faixa de renda mais freqüente foi aquela com um ou menos salário mínimo. Entre os demais fluxos observados, esse apresenta o percentual mais crítico quanto à renda, pois aqueles que possuem mais de três salários mínimos como renda somam menos de 16%, e 47% possuem renda igual ou abaixo de um salário mínimo.

Na tabela 2, foi visto que 51% dos migrantes do Sul vão para outra UF da mesma região, é um fluxo expressivo da região sul, por isso, é interessante estudar esses 51%. Ainda que com uma diferença muito pequena, as mulheres superam o número de homens. A maior parte deles pertence à faixa etária de 18 a 30 anos, seguidos por aqueles que têm entre 30 e 50 anos. A respeito da raça/cor da pele, foi observado que quase 85% responderam ser ‘brancos’, um valor muito alto quando comparado a qualquer outro fluxo. No que se refere à renda, cerca de 46% afirmou ter renda entre um e três salários mínimos, e é o único fluxo onde o número de pessoas que afirmou ter entre três e cinco salários mínimos foi maior do que o correspondente a igual ou menor que um salário mínimo, foi também o fluxo que obteve o menor percentual de migrantes com renda de até um salário mínimo.

Entre os que migraram da região Sul para a Sudeste, o segundo destino mais procurado de quem se origina do Sul, houve uma ligeira maioria masculina. Mais de um terço tem idade entre 18 e 30 anos e 23% menos de 18 anos. Em relação à raça/cor da pele, assim como nos migrantes intra-regionais do Sul, também é observado um alto percentual de ‘brancos’, três quartos, e aproximadamente um quinto disseram ser ‘pardos’. A respeito da variável renda, é observado o menor percentual para pessoas com menos de três salários mínimos, cerca de 44% apresentou três ou mais salários mínimos como renda, valor muito maior quando comparado aos outros fluxos.

Como os demais fluxos migratórios intra-regionais, no Centro-Oeste, houve uma pequena maioria de migrantes do sexo feminino, comparada ao masculino. Mais de um terço tinham idade entre 18 e 30 anos, tendo por outro lado, um baixo percentual de

maiores de 50 anos. A raça/cor da pele mais respondida, como na região Norte e Nordeste, foi a ‘parda’, pouco mais que a metade dos migrantes, e com aqueles que responderam ‘branca’, somam mais de 90%. Em relação à variável renda, os mais respondidos foram aqueles valores que correspondem à faixa entre um e três salários mínimos, 42%, e apenas 30% têm renda acima de três salários mínimos.

Por fim, o fluxo do Centro-Oeste para o Sudeste, o segundo destino mais procurado por moradores da região Centro-Oeste. Possui uma ligeira maioria masculina e a faixa etária mais representante entre 18 e 30 anos, quase um terço desses migrantes e mais de 10% com mais de 50 anos. A raça mais respondida desses foi ‘branca’, com 61%, e ‘pardos’ com um terço. Dos fluxos observados, esse foi o que obteve um maior percentual das raças/cor da pele menos respondidas ‘indígenas’ e ‘amarelos’, sendo o percentual de ‘indígenas’ maior que o dobro de todos os demais fluxos analisados. No que se refere à variável renda, quase 40% disseram possuir entre um e três salários mínimos como renda, e mais de 40%, três ou mais, o segundo maior percentual dos fluxos analisados.

4.2 Regressão Logística

Um dos objetivos é encontrar funções que mostrem os diferentes comportamentos dos migrantes brasileiros. Para encontrar tais funções, foram estudados migrantes oriundos de cada região do país. Foram comparados os principais fluxos migratórios das regiões com seus possíveis destinos para verificar diferença entre o migrante que pertencia ao principal fluxo da região e o migrante que ia para os fluxos secundários. Por exemplo: na região Norte o maior fluxo inter-regional foi para a região Nordeste, então fez-se uma regressão para tentar modelar o comportamento daqueles que migravam em direção ao Nordeste contra os que migravam para as demais regiões.

Para todos os fluxos estudados foi usado o modelo de regressão logística com o método *stepwise/forward*, modelo este que tem por objetivo modelar o comportamento de uma variável resposta binária. No caso, a variável resposta de todos os modelos foi:

$Y_i = 0$, O indivíduo i pertence ao principal fluxo migratório da região

$Y_i = 1$, O indivíduo i não pertence ao principal fluxo migratório da região

Foram estudados os principais fluxos em 2001 e em 2009. E seu objetivo era apontar possíveis diferentes comportamentos entre os fluxos de diferentes regiões e possíveis mudanças de comportamento do ano de 2001 para 2009.

Em 2001, os principais fluxos inter-regionais de migração foram:

- Norte para Nordeste;
- Nordeste para Sudeste;
- Sudeste para Nordeste;
- Sul para Sudeste;
- Centro-Oeste para Sudeste.

A maior parte das possíveis variáveis preditoras para o modelo são categorizadas, então, para facilitar a compreensão e interpretação do modelo, foram criados modelos com no máximo quatro variáveis preditoras. As categorias de referência para o cálculo das razões de chance das variáveis ‘Idade’, ‘Renda familiar mensal’, ‘Região de nascimento’ e ‘Nasceu na UF de residência’ foram respectivamente: ‘60 anos ou mais’, ‘20 salários mínimos ou mais’, ‘País estrangeiro’ e ‘Não nasceu na UF de residência’.

Como essas quatro variáveis são categorizadas, elas serão tratadas como fatores, ou seja, cada classe das variáveis só poderá assumir 0 ou 1, então, será criado um

indicador para cada variável categorizada. Logo, se uma classe é igual a 1 automaticamente todas as demais classes da variável adquirem o valor zero. Essa variável indicadora não representa nenhum poder no teste.

A capacidade de explicação dos modelos apresentados será estudado por meio da curva ROC (*receiver operating characteristic*). Este gráfico consiste em plotar o índice de pares concordantes, ou seja, observações que foram previstas de maneira correta pelos pares de variáveis falsos positivos, que são pares previstos de maneira incorreta pelo modelo. No geral, estes gráficos possuem uma forma côncava partindo da coordenada (0,0) para a (1,1), sendo que quanto mais alto o grau de elevação inicial da curva e quanto maior a área abaixo da curva, mais poder de explicação possui o modelo. A tabela 5 mostra a capacidade de discriminação da curva ROC

Tabela 5. Área abaixo da Curva ROC por Capacidade de Discriminação

Área abaixo da curva ROC	Característica da discriminação
ROC = 0,5	Não existe discriminação
$0,7 \leq \text{ROC} < 0,8$	Discriminação aceitável
$0,8 \leq \text{ROC} < 0,9$	Excelente discriminação
$\text{ROC} \geq 0,9$	Discriminação acima do comum

Serão feitas comparações entre os anos em cada fluxo e, posteriormente, em 4.3.1, serão apresentadas as curvas ROC para cada fluxo e ano estudados (Gráficos 3 a 12) e um resumo sobre o comportamento geral entre os modelos.

4.2.1 Norte – Nordeste, 2001 e 2009

Para modelar este fluxo, no ano de 2001, foram usadas três variáveis: ‘Idade’, ‘Renda familiar mensal’ e ‘Número de pessoas na família’. Já que, com estas três variáveis, foi alcançado um poder de explicação semelhante ao modelo com quatro, para facilitar a interpretação, entraram no modelo apenas as três.

Em 2009, por outro lado, para modelar este fluxo, as três variáveis utilizadas foram: ‘Nasceu na UF de residência’, ‘Renda familiar mensal’ e ‘Número de pessoas na família’. Da mesma forma que em 2001, as três variáveis alcançaram um poder de explicação próximo ao modelo com quatro, portanto para facilitar a interpretação foram incluídas no modelo apenas estas três.

As tabelas 6 e 7 mostram as estimativas dos coeficientes da regressão logística e da razão de chances, para cada classe das variáveis para o ano de 2001 e 2009 respectivamente.

Tabela 6. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2001

Efeito	β	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-1,981	0,013			
Idade					
05 a 09 anos	-0,131	0,010	0,877	0,860	0,895
10 a 14 anos	0,552	0,011	1,737	1,702	1,774
15 a 19 anos	-0,046	0,010	0,955	0,937	0,974
20 a 24 anos	-0,241	0,010	0,786	0,770	0,802
25 a 29 anos	0,084	0,011	1,088	1,065	1,111
30 a 39 anos	-0,084	0,010	0,920	0,902	0,938
40 a 49 anos	0,864	0,010	2,372	2,325	2,420
50 a 59 anos	0,660	0,014	1,935	1,884	1,988
60 ou mais (ref.)	0,000	0,000	1,000	1,000	1,000
Nº de pessoas na família	0,128	0,002	1,136	1,132	1,141
Renda familiar mensal					
Até 1/2 sal. mín.	1,326	0,010	3,767	3,697	3,838
Entre 1/2 e 1 sal. mín.	0,358	0,010	1,431	1,402	1,460
Entre 1 e 2 sal. mín.	-0,449	0,011	0,639	0,625	0,652
Entre 2 e 3 sal. mín.	-0,072	0,013	0,930	0,907	0,954
Entre 3 e 5 sal. mín.	0,205	0,014	1,228	1,196	1,261
Entre 5 e 10 sal. mín.	0,141	0,015	1,151	1,119	1,184
Entre 10 e 20 sal. mín.	0,178	0,025	1,195	1,137	1,255
20 ou mais sal. mín.	0,000	0,000	1,000	1,000	1,000

Tabela 7. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2009

Efeito	β	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-1,368	0,011			
Nº de pessoas na família	0,136	0,002	1,146	1,140	1,151
Nasceu na UF de residência					
Sim	0,384	0,004	1,469	1,457	1,480
Não (ref.)	0,000	0,000	1,000	1,000	1,000
Renda familiar mensal					
Até 1/2 sal. mín.	-0,148	0,020	0,863	0,830	0,896
Entre 1/2 e 1 sal. mín.	0,974	0,011	2,650	2,595	2,705
Entre 1 e 2 sal. mín.	0,142	0,008	1,152	1,134	1,171
Entre 2 e 3 sal. mín.	0,341	0,009	1,406	1,382	1,431
Entre 3 e 5 sal. mín.	-0,224	0,011	0,800	0,783	0,817
Entre 5 e 10 sal. mín.	0,597	0,011	1,816	1,777	1,857
Entre 10 e 20 sal. mín.	-1,270	0,021	0,281	0,270	0,293
20 ou mais sal. mín. (ref.)	0,000	0,000	1,000	1,000	1,000

A probabilidade (π) de um indivíduo ser classificado como migrante do fluxo principal (Norte – Nordeste), em 2001, é dada por:

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_i^X + \beta_j^R + \beta N)}{1 + \exp(\beta_0 + \beta_i^X + \beta_j^R + \beta N)} = \pi$$

onde β_i^X e β_j^R são as classes observadas para as variáveis ‘Idade’ e ‘Renda familiar mensal’ respectivamente e N corresponde ao número de pessoas na família.

A probabilidade (π) de um indivíduo ser classificado como migrante do fluxo principal (Norte – Nordeste), em 2009, é dada por:

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_k^Z + \beta_j^R + \beta N)}{1 + \exp(\beta_0 + \beta_k^Z + \beta_j^R + \beta N)} = \pi$$

onde β_k^Z e β_j^R são as classes observadas para as variáveis ‘Nasceu na UF de residência’ e ‘Renda familiar mensal’ respectivamente e N corresponde ao número de pessoas na família.

Com os modelos propostos, foram feitos os diagnósticos que quantificam a eficácia do modelo, ou seja, a sua concordância com o que de fato foi observado e a área existente abaixo da curva ROC, denotada pela estatística ‘C’ (tabela 8).

Tabela 8. Associação entre probabilidades preditas e respostas observadas – 2001 e 2009

2001		2009	
Discriminação	Resultado	Discriminação	Resultado
Percentual concordante	67,1%	Percentual concordante	62,0%
Percentual discordante	32,3%	Percentual discordante	35,9%
C	0,67	C	0,63

Observa-se que o percentual de pares classificados de maneira correta pelo modelo reduziu de 67,1% para 62,0%, assim como o valor da estatística C de 0,67 para 0,63, evidenciando assim que o modelo perdeu um pouco de sua capacidade de explicação.

Após o cálculo das razões de chance, em 2001, pode-se concluir que os migrantes que saem do Norte para o Nordeste, no geral, têm maior probabilidade de serem mais velhos do que os que se destinam às demais regiões, tendem também a ter famílias com número de pessoas ligeiramente maiores. Em relação à renda, a chance de receber até 1 salário mínimo é maior entre os migrantes que vão para o Nordeste do que entre aqueles que vão para as outras regiões.

Para 2009, é possível observar que a probabilidade de terem feito uma migração de retorno à UF de nascimento é maior entre os migrantes que voltaram ao Nordeste do que os provenientes de outras regiões. Em 2001, a faixa de renda mais expressiva eram as mais baixas, já em 2009, são as intermediárias. E da mesma forma que em 2001, os

migrantes que vão para o Nordeste têm maior chance de ter famílias maiores do que os migrantes que vão para outros destinos.

4.2.2 Nordeste – Sudeste, 2001 e 2009

Para modelar este fluxo, no ano de 2001 e, da mesma forma, no ano de 2009, foram usadas três variáveis: ‘Idade’, ‘Renda familiar mensal’ e ‘Região de Nascimento’. Já que, com estas três variáveis foi alcançado um poder de explicação semelhante ao modelo com quatro, para facilitar a interpretação, entraram no modelo apenas as três.

As tabelas 9 e 10 mostram as estimativas dos coeficientes da regressão logística e da razão de chances, para cada classe das variáveis para o ano de 2001 e 2009 respectivamente.

Tabela 9. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2001

Efeito	β	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-1,683	0,00669			
Faixa etária					
05 a 09 anos	0,2	0,00373	1,221	1,213	1,230
10 a 14 anos	0,2087	0,00388	1,232	1,223	1,241
15 a 19 anos	0,1135	0,00352	1,120	1,112	1,128
20 a 24 anos	-0,1318	0,00327	0,877	0,871	0,882
25 a 29 anos	0,1754	0,00333	1,192	1,184	1,200
30 a 39 anos	-0,0947	0,00306	0,910	0,904	0,915
40 a 49 anos	0,1744	0,00379	1,191	1,182	1,199
50 a 59 anos	-0,1616	0,0055	0,851	0,842	0,860
60 ou mais (ref.)	0,000	0,000	1,000	1,000	1,000
Região de nascimento					
Norte	-0,2321	0,00781	0,793	0,781	0,805
Nordeste	1,8231	0,00664	6,191	6,111	6,272
Sudeste	-0,0649	0,00687	0,937	0,925	0,950
Sul	-0,9798	0,0111	0,375	0,367	0,384
Centro-Oeste	-1,694	0,0174	0,184	0,178	0,190
País estrangeiro (ref.)	0,000	0,000	1,000	1,000	1,000
Renda familiar mensal					
Até 1/2 sal. mín.	0,2908	0,00317	1,337	1,329	1,346
Entre 1/2 e 1 sal. mín.	0,1937	0,00304	1,214	1,207	1,221
Entre 1 e 2 sal. mín.	0,1624	0,00311	1,176	1,169	1,184
Entre 2 e 3 sal. mín.	0,2779	0,00409	1,320	1,310	1,331
Entre 3 e 5 sal. mín.	-0,2951	0,00479	0,744	0,738	0,751
Entre 5 e 10 sal. mín.	-0,4756	0,00597	0,622	0,614	0,629
Entre 10 e 20 sal. mín.	-0,3728	0,00919	0,689	0,677	0,701
20 ou mais sal. mín. (ref.)	0,000	0,000	1,000	1,000	1,000

Tabela 10. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2009

Efeito	B	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-1,246	0,010			
Faixa etária					
05 a 09 anos	0,176	0,006	1,192	1,179	1,205
10 a 14 anos	0,323	0,006	1,381	1,366	1,396
15 a 19 anos	0,017	0,005	1,017	1,007	1,028
20 a 24 anos	0,193	0,004	1,213	1,203	1,224
25 a 29 anos	0,208	0,005	1,231	1,220	1,243
30 a 39 anos	0,105	0,005	1,110	1,101	1,120
40 a 49 anos	-0,230	0,006	0,795	0,786	0,804
50 a 59 anos	-0,372	0,008	0,690	0,678	0,701
60 ou mais (ref.)	0,000	0,000	1,000	1,000	1,000
Região de nascimento					
Norte	-0,296	0,011	0,744	0,729	0,760
Nordeste	0,869	0,010	2,385	2,340	2,431
Sudeste	1,412	0,011	4,105	4,017	4,194
Sul	-0,300	0,018	0,741	0,716	0,767
Centro-Oeste	-1,219	0,017	0,296	0,286	0,306
País estrangeiro (ref.)	0,000	0,000	1,000	1,000	1,000
Renda familiar mensal					
Até 1/2 sal. mín.	-0,491	0,010	0,612	0,601	0,624
Entre 1/2 e 1 sal. mín.	-0,797	0,006	0,451	0,445	0,456
Entre 1 e 2 sal. mín.	0,176	0,004	1,192	1,183	1,201
Entre 2 e 3 sal. mín.	0,369	0,004	1,446	1,434	1,458
Entre 3 e 5 sal. mín.	0,196	0,005	1,216	1,206	1,227
Entre 5 e 10 sal. mín.	-0,070	0,006	0,932	0,922	0,943
Entre 10 e 20 sal. mín.	0,120	0,007	1,128	1,112	1,144
20 ou mais sal. mín. (ref.)	0,000	0,000	1,000	1,000	1,000

A probabilidade (π) de um indivíduo ser classificado como migrante do fluxo principal (Nordeste – Sudeste), em 2001 e em 2009, é dada por:

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)}{1 + \exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)} = \pi$$

onde β_i^X , β_j^R e β_k^N são as classes observadas para as variáveis ‘Idade’, ‘Renda familiar mensal’ e ‘Região de Nascimento’ respectivamente.

Com os modelos propostos, foram feitos os diagnósticos que quantificam a eficácia do modelo, ou seja, a sua concordância com o que de fato foi observado e a área existente abaixo da curva ROC, denotada pela estatística ‘C’ (tabela 11).

Tabela 11. Associação entre probabilidades preditas e respostas observadas – 2001 e 2009

2001		2009	
Discriminação	Resultado	Discriminação	Resultado
Percentual concordante	71,3%	Percentual concordante	65,6%
Percentual discordante	27,0%	Percentual discordante	32,8%
C	0,721	C	0,664

Observa-se que o percentual de pares classificados de maneira correta pelo modelo reduziu de 71,3% para 65,6%, assim como o valor da estatística C de 0,721 para 0,664, evidenciando assim que o modelo perdeu um pouco de sua capacidade de explicação.

Com as razões de chance, foi observado que, para ambos os anos analisados, o percentual de jovens que migram para o Sudeste é maior do que o percentual de jovens que migram para outras regiões, sendo em 2009 de forma mais acentuada que em 2001. Em uma análise geral, os migrantes desse fluxo têm uma maior chance do que os que migram para outros destinos de terem nascido no Nordeste, em 2001, e no Sudeste, em 2009, Ou seja, a migração observada em 2009 tem uma característica de migração de retorno.

Os indivíduos que migraram para o Sudeste, em 2001, tiveram uma maior chance de possuírem baixa renda do que quando comparados aos demais migrantes. Já em 2009, essa maior chance esteve relacionada com as faixas de renda mais intermediárias.

4.2.3 Sudeste – Nordeste, 2001 e 2009

Para modelar este fluxo, no ano de 2001, foram usadas três variáveis: ‘Idade’, ‘Região de nascimento’ e ‘Nasceu na UF de residência’.

Em 2009, por outro lado, para modelar este fluxo, as três variáveis utilizadas foram: ‘Renda familiar mensal’, ‘Região de nascimento’ e ‘Nasceu na UF de residência’. Da mesma forma que em 2001, as três variáveis alcançaram um poder de explicação próximo ao modelo com quatro, portanto para facilitar a interpretação entraram no modelo apenas estas três.

As tabelas 12 e 13 mostram as estimativas dos coeficientes da regressão logística e da razão de chances, para cada classe das variáveis para o ano de 2001 e 2009 respectivamente.

Tabela 12. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2001

Efeito	B	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-1,580	0,011			
Faixa etária					
5 a 9 anos	1,086	0,006	2,962	2,926	2,999
10 a 14 anos	0,741	0,007	2,097	2,069	2,126
15 a 19 anos	0,383	0,007	1,467	1,446	1,488
20 a 24 anos	-0,295	0,007	0,745	0,734	0,755
25 a 29 anos	-0,062	0,007	0,940	0,928	0,952
30 a 39 anos	-0,454	0,006	0,635	0,628	0,642
40 a 49 anos	-0,251	0,007	0,778	0,768	0,788
50 a 59 anos	-0,700	0,010	0,497	0,487	0,506
60 ou mais (ref.)	0,000	0,000	1,000	1,000	1,000
Região de nascimento					
Norte	-0,135	0,019	0,874	0,842	0,907
Nordeste	3,127	0,011	22,805	22,306	23,316
Sudeste	-0,182	0,011	0,834	0,816	0,852
Sul	-1,519	0,017	0,219	0,212	0,226
Centro-Oeste	-3,124	0,039	0,044	0,041	0,047
País estrangeiro (ref.)	0,000	0,000	1,000	1,000	1,000
Nasceu na UF de residência					
Sim	0,460	0,003	1,583	1,575	1,591
Não (ref.)	0,000	0,000	1,000	1,000	1,000

Tabela 13. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2009

Efeito	β	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-1,3743	0,006			
Nasceu na UF de residência					
Sim	0,314	0,002	1,368	1,364	1,373
Não (ref.)	0,000	0,000	1,000	1,000	1,000
Região de nascimento					
Norte	0,096	0,008	1,100	1,084	1,117
Nordeste	1,624	0,006	5,075	5,013	5,138
Sudeste	0,118	0,006	1,126	1,111	1,140
Sul	-0,956	0,012	0,384	0,375	0,394
Centro-Oeste	-1,316	0,014	0,268	0,261	0,276
País estrangeiro (ref.)	0,000	0,000	1,000	1,000	1,000
Renda familiar mensal					
Até 1/2 sal. mín.	0,327	0,006	1,387	1,370	1,403
Entre 1/2 e 1 sal. mín.	-0,161	0,004	0,852	0,844	0,859
Entre 1 e 2 sal. mín.	0,117	0,003	1,124	1,117	1,131
Entre 2 e 3 sal. mín.	0,167	0,003	1,181	1,174	1,189
Entre 3 e 5 sal. mín.	-0,052	0,004	0,949	0,943	0,956
Entre 5 e 10 sal. mín.	-0,374	0,004	0,688	0,682	0,694
Entre 10 e 20 sal. mín.	-0,118	0,006	0,888	0,879	0,898
20 ou mais sal. mín. (ref.)	0,000	0,000	1,000	1,000	1,000

A probabilidade (π) de um indivíduo ser classificado como migrante do fluxo principal (Sudeste – Nordeste), em 2001, é dada por:

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)}{1 + \exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)} = \pi$$

onde β_i^X , β_j^R e β_k^N são as classes observadas para as variáveis ‘Idade’ e ‘Região de nascimento’ e ‘Nasceu na UF de residência’ respectivamente.

A probabilidade (π) de um indivíduo ser classificado como migrante do fluxo principal (Sudeste – Nordeste), em 2009, é dada por:

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)}{1 + \exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)} = \pi$$

onde β_i^X , β_j^R e β_k^N são as classes observadas para as variáveis ‘Nasceu na UF de residência’, ‘Renda familiar mensal’ e ‘Região de nascimento’ respectivamente.

Com os modelos propostos, foram feitos os diagnósticos que quantificam a eficácia do modelo, ou seja, a sua concordância com o que de fato foi observado e a área existente abaixo da curva ROC, denotada pela estatística ‘C’ (tabela 14).

Tabela 14. Associação entre probabilidades preditas e respostas observadas – 2001 e 2009

2001		2009	
Discriminação	Resultado	Discriminação	Resultado
Percentual concordante	84,9%	Percentual concordante	69,9%
Percentual discordante	12,8%	Percentual discordante	26,5%
C	0,861	C	0,717

Observa-se que o percentual de pares classificados de maneira correta pelo modelo reduziu de 84,9% para 69,9%, assim como o valor da estatística C de 0,86 para 0,72, mostrando assim que o modelo perdeu um pouco de sua capacidade de explicação.

Pelas razões de chance, em 2001, foi visto que os indivíduos do fluxo Sudeste para o Nordeste têm uma maior chance de terem menos de 20 anos do que os que vão para os demais destinos.

Nos dois anos estudados, viu-se que os migrantes que foram para o Nordeste tiveram uma chance expressivamente maior de terem nascido em UFs do Nordeste do que os que se destinaram a demais localidades, e essa chance também é maior quando se observa a variável ‘Nasceu na UF de residência’, isso reitera o fenômeno da migração de retorno. Em 2009, aconteceu de forma mais branda do que em 2001.

Os indivíduos que pertencem a esse fluxo têm menos chance de terem rendas mais elevadas do que os migrantes que se destinam às demais regiões.

4.2.4 Sul – Sudeste, 2001 e 2009

Para modelar este fluxo, em 2001, foram usadas quatro variáveis: ‘Idade’, ‘Renda familiar mensal’ e ‘Região de nascimento’ e ‘Nasceu na UF de residência’.

Em 2009, por outro lado, para modelar este fluxo, três variáveis foram utilizadas: ‘Nasceu na UF de residência’, ‘Renda familiar mensal’ e ‘Região de nascimento’. Diferentemente do ocorrido em 2001, essas três variáveis alcançaram um poder de explicação próximo ao modelo com quatro, portanto para facilitar a interpretação entraram no modelo apenas estas três.

As tabelas 15 e 16 mostram estimativas dos coeficientes da regressão logística e da razão de chances, para cada classe das variáveis para 2001 e 2009 respectivamente.

Tabela 15. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2001

Efeito	β	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-1,220	0,005			
Faixa etária					
05 a 09 anos	0,236	0,003	1,266	1,258	1,275
10 a 14 anos	0,210	0,004	1,233	1,225	1,242
15 a 19 anos	0,110	0,003	1,116	1,109	1,123
20 a 24 anos	-0,049	0,003	0,952	0,946	0,957
25 a 29 anos	0,102	0,003	1,107	1,101	1,114
30 a 39 anos	-0,121	0,003	0,886	0,881	0,891
40 a 49 anos	0,062	0,003	1,064	1,056	1,071
50 a 59 anos	-0,171	0,005	0,843	0,835	0,851
60 ou mais (ref.)	0,000	0,000	1,000	1,000	1,000
Região de nascimento					
Norte	-0,542	0,007	0,582	0,574	0,590
Nordeste	1,515	0,005	4,549	4,501	4,596
Sudeste	-0,153	0,006	0,858	0,849	0,867
Sul	-0,131	0,006	0,877	0,867	0,887
Centro-Oeste	-1,845	0,014	0,158	0,154	0,162
País estrangeiro (ref.)	0,000	0,000	1,000	1,000	1,000
Renda familiar mensal					
Até 1/2 sal. mín.	0,181	0,003	1,198	1,191	1,205
Entre 1/2 e 1 sal. mín.	0,043	0,003	1,044	1,038	1,050
Entre 1 e 2 sal. mín.	0,124	0,003	1,132	1,126	1,138
Entre 2 e 3 sal. mín.	0,321	0,004	1,378	1,369	1,388
Entre 3 e 5 sal. mín.	-0,083	0,004	0,920	0,913	0,927
Entre 5 e 10 sal. mín.	-0,374	0,005	0,688	0,682	0,695
Entre 10 e 20 sal. mín.	-0,258	0,008	0,772	0,761	0,784
20 ou mais sal. mín. (ref.)	0,000	0,000	1,000	1,000	1,000
Nasceu na UF de residência					
Sim	0,216	0,001	1,540	1,532	1,548
Não (ref.)	0,000	0,000	1,000	1,000	1,000

Tabela 16. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2009

Efeito	β	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-1,291	0,00563			
Nasceu na UF de residência					
Sim	0,2555	0,00154	1,667	1,657	1,677
Não (ref.)	0,000	0,000	1,000	1,000	1,000
Região de Nascimento					
Norte	-0,0131	0,00711	0,987	0,973	1,001
Nordeste	1,5202	0,00578	4,573	4,522	4,625
Sudeste	0,1423	0,00597	1,153	1,140	1,166
Sul	-0,374	0,00674	0,688	0,679	0,697
Centro-Oeste	-1,3358	0,0131	0,263	0,256	0,270
País estrangeiro (ref.)	0,000	0,000	1,000	1,000	1,000
Renda familiar mensal					
Até 1/2 sal. mín.	0,4252	0,0057	1,530	1,513	1,547
Entre 1/2 e 1 sal. mín.	-0,0857	0,00419	0,918	0,910	0,925
Entre 1 e 2 sal. mín.	0,0745	0,00287	1,077	1,071	1,083
Entre 2 e 3 sal. mín.	0,1145	0,00313	1,121	1,114	1,128
Entre 3 e 5 sal. mín.	-0,0997	0,00332	0,905	0,899	0,911
Entre 5 e 10 sal. mín.	-0,3823	0,00403	0,682	0,677	0,688
Entre 10 e 20 sal. mín.	-0,1623	0,00512	0,850	0,842	0,859
20 ou mais sal. mín. (ref.)	0,000	0,000	1,000	1,000	1,000

A probabilidade (π) de um indivíduo ser classificado como migrante do fluxo principal (Sul – Sudeste), em 2001, é dada por:

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N + \beta_l^F)}{1 + \exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N + \beta_l^F)} = \pi$$

onde β_i^X , β_j^R , β_k^N e β_l^F são as classes observadas para as variáveis ‘Idade’ e ‘Renda familiar mensal’, ‘Nasceu na UF de residência’ e ‘Região de nascimento’ respectivamente.

A probabilidade (π) de um indivíduo ser classificado como migrante do fluxo principal (Sul – Sudeste), em 2009, é dada por:

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)}{1 + \exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)} = \pi$$

onde β_i^X , β_j^R e β_k^N são as classes observadas para as variáveis ‘Nasceu na UF de residência’ e ‘Renda familiar mensal’ e ‘Região de Nascimento’ respectivamente.

Com os modelos propostos, foram feitos os diagnósticos que quantificam a eficácia do modelo, ou seja, a sua concordância com o que de fato foi observado e a área existente abaixo da curva ROC, denotada pela estatística ‘C’ (tabela 17).

Tabela 17. Associação entre probabilidades preditas e respostas observadas – 2001 e 2009

2001		2009	
Discriminação	Resultado	Discriminação	Resultado
Percentual concordante	81,2%	Percentual concordante	70,4%
Percentual discordante	18,2%	Percentual discordante	26,1%
C	0,815	C	0,721

Observa-se que o percentual de pares classificados de maneira correta pelo modelo reduziu de 81,2% para 70,4%, assim como o valor da estatística C de 0,82 para 0,72, evidenciando assim que o modelo perdeu um pouco de sua capacidade de explicação.

Com as razões de chance, foi visto que, em 2001, os migrantes do Sul para o Sudeste têm uma suave tendência de serem mais jovens, comparados àqueles que migraram aos demais lugares.

Os indivíduos que, saindo do Sul, migraram para o Sudeste, tiveram maior chance de serem nativos do Nordeste do que os que se destinaram para as demais regiões, em 2001 e em 2009. Da mesma forma que nos demais fluxos, percebe-se uma maior evidência da migração de retorno nesse fluxo do que nos demais.

Em 2001, os migrantes pertencentes a este fluxo tiveram uma menor chance de terem rendas elevadas quando comparados aos pertencentes aos demais fluxos. Em 2009, este comportamento se acentuou.

4.2.5 Centro-Oeste – Sudeste, 2001 e 2009

Para modelar este fluxo, no ano de 2001, foram usadas três variáveis: ‘Idade’, ‘Região de nascimento’ e ‘Nasceu na UF de residência’. Com estas três

variáveis foi alcançado um poder de explicação semelhante ao modelo com quatro, logo, para facilitar a interpretação, entraram no modelo apenas as três.

Em 2009, por outro lado, para modelar este fluxo, quatro variáveis foram utilizadas: ‘Idade’, ‘Região de Nascimento’, ‘Nasceu na UF de residência’ e ‘Renda familiar mensal’.

As tabelas 18 e 19 mostram as estimativas dos coeficientes da regressão logística e da razão de chances, para cada classe das variáveis para o ano de 2001 e 2009 respectivamente.

Tabela 18. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2001

Efeito	B	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-0,918	0,00527			
Faixa etária					
05 a 09 anos	0,299	0,00415	1,348	1,337	1,359
10 a 14 anos	0,189	0,00428	1,208	1,198	1,218
15 a 19 anos	0,072	0,00425	1,074	1,065	1,083
20 a 24 anos	-0,060	0,00404	0,942	0,934	0,949
25 a 29 anos	-0,068	0,00395	0,934	0,927	0,942
30 a 39 anos	-0,174	0,00333	0,840	0,835	0,846
40 a 49 anos	-0,056	0,00404	0,945	0,938	0,953
50 a 59 anos	-0,215	0,00572	0,807	0,798	0,816
60 ou mais (ref.)	0,000	0,000	1,000	1,000	1,000
Região de nascimento					
Norte	-1,265	0,013	0,282	0,275	0,290
Nordeste	1,544	0,00578	4,683	4,631	4,737
Sudeste	-0,236	0,00558	0,790	0,781	0,798
Sul	-0,347	0,00589	0,707	0,699	0,715
Centro-Oeste	-0,216	0,00633	0,806	0,796	0,816
País estrangeiro (ref.)	0,000	0,000	1,000	1,000	1,000
Nasceu na UF de residência					
Sim	0,276	0,00157	1,317	1,313	1,321
Não (ref.)	0,000	0,000	1,000	1,000	1,000

Tabela 19. Análise dos estimadores de Máxima Verossimilhança e Razão de Chances Estimada, 2009

Efeito	β	Erro Padrão	Razão de Chances	Intervalo de confiança (95%) de Wald	
Intercepto	-1,1925	0,00511			
Faixa etária					
05 a 09 anos	0,2066	0,00391	1,229	1,220	1,239
10 a 14 anos	0,1831	0,00384	1,201	1,192	1,210
15 a 19 anos	0,1268	0,00379	1,135	1,127	1,144
20 a 24 anos	0,0637	0,00327	1,066	1,059	1,073
25 a 29 anos	0,0348	0,00323	1,035	1,029	1,042
30 a 39 anos	-0,0406	0,00292	0,960	0,955	0,966
40 a 49 anos	-0,0074	0,00376	0,993	0,985	1,000
50 a 59 anos	-0,2679	0,00526	0,765	0,757	0,773
60 ou mais (ref.)	0,000	0,000	1,000	1,000	1,000
Região de nascimento					
Norte	-0,2894	0,00664	0,749	0,739	0,759
Nordeste	1,2643	0,00524	3,541	3,504	3,577
Sudeste	0,1678	0,0054	1,183	1,170	1,195
Sul	-0,5877	0,00625	0,556	0,549	0,562
Centro-Oeste	-0,3391	0,00662	0,712	0,703	0,722
País estrangeiro (ref.)	0,000	0,000	1,000	1,000	1,000
Nasceu na UF de residência					
Sim	0,2216	0,00141	1,248	1,245	1,252
Não (ref.)	0,000	0,000	1,000	1,000	1,000
Renda familiar mensal					
Até 1/2 sal. mín.	0,369	0,00528	1,446	1,431	1,461
Entre 1/2 e 1 sal. mín.	-0,1787	0,00389	0,836	0,830	0,843
Entre 1 e 2 sal. mín.	0,0804	0,00268	1,084	1,078	1,089
Entre 2 e 3 sal. mín.	0,0566	0,00294	1,058	1,052	1,064
Entre 3 e 5 sal. mín.	-0,0428	0,00308	0,958	0,952	0,964
Entre 5 e 10 sal. mín.	-0,3175	0,00365	0,728	0,723	0,733
Entre 10 e 20 sal. mín.	-0,1074	0,00468	0,898	0,890	0,906
20 ou mais sal. mín. (ref.)	0,000	0,000	1,000	1,000	1,000

A probabilidade (π) de um indivíduo ser classificado como migrante do fluxo principal (Centro-Oeste – Sudeste), em 2001, é dada por:

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)}{1 + \exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N)} = \pi$$

onde β_i^X , β_j^R e β_k^N são as classes observadas para as variáveis ‘Idade’ e ‘Região de nascimento’ e ‘Nasceu na UF de residência’ respectivamente.

A probabilidade (π) de um indivíduo ser classificado como migrante do fluxo principal (Centro-Oeste – Sudeste), em 2009, é dada por:

$$E(Y_i) = \frac{\exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N + \beta_l^F)}{1 + \exp(\beta_0 + \beta_i^X + \beta_j^R + \beta_k^N + \beta_l^F)} = \pi$$

onde β_i^X , β_j^R , β_k^N e β_l^F são as classes observadas para as variáveis ‘Idade’ e ‘Renda familiar mensal’, ‘Nasceu na UF de residência’ e ‘região de nascimento’ respectivamente.

Com os modelos propostos, foram feitos os diagnósticos que quantificam a eficácia do modelo, ou seja, a sua concordância com o que de fato foi observado e a área existente abaixo da curva ROC, denotada pela estatística ‘C’ (tabela 20).

Tabela 20. Associação entre probabilidades previstas e respostas observadas – 2001 e 2009

2001		2009	
Discriminação	Resultado	Discriminação	Resultado
Percentual concordante	72,6%	Percentual concordante	70,6%
Percentual discordante	24,9%	Percentual discordante	28,8%
C	0,738	C	0,709

Observa-se que o percentual de pares classificados de maneira correta pelo modelo reduziu de 72,6% para 70,6%, assim como o valor da estatística C de 0,74 para 0,71, evidenciando assim que o modelo perdeu um pouco de sua capacidade de explicação.

Pelo estudo das razões de chance, foi observado que, para ambos os anos analisados, o percentual de jovens que migram para o Sudeste é maior do que o percentual de jovens que migram para outras regiões, sendo em 2009 de forma mais acentuada que em 2001. Em uma análise geral, os migrantes desse fluxo têm uma maior chance do que os que migram para outros destinos de terem nascido no Nordeste.

Os indivíduos que migraram para o Sudeste, em 2009, tiveram uma maior chance de possuírem baixa renda do que quando comparados aos demais migrantes.

4.3 Análise dos modelos

Os modelos foram feitos seguindo o método *stepwise forward* e a tabela 21 apresenta um resumo das variáveis que compuseram os modelos criados por fluxo e por ano.

Tabela 21:Quadro-resumo dos modelos criados.

Ano	Fluxo	Idade	Região de Nascimento	Nasceu na UF de Residência	Renda	Número de pessoas na família
2001	N – Ne	x			x	x
	Ne – Se	x	x		x	
	Se – Ne	x	x	x		
	S – Se	x	x	x	x	
	CO – Se	x	x	x		
2009	N – Ne			x	x	x
	Ne – Se	x	x		x	
	Se – Ne		x	x	x	
	S – Se		x	x	x	
	CO – Se	x	x	x	x	

As variáveis ‘Região de nascimento’ e ‘Renda familiar mensal’ entraram em oito dos dez modelos obtidos, sendo que em todas as regiões no ano de 2001, a variável ‘Idade’ fez parte do modelo, e o mesmo aconteceu com a variável ‘Renda’ no ano de 2009. A variável ‘Região de Nascimento’ entrou em todos os modelos, exceto os da região Norte, modelo este que a trocou pela variável ‘Número de pessoas na família’. A variável ‘Nasceu na UF de residência’ entrou em todos os modelos, exceto nos da região Nordeste e em 2001 da região Norte. Ao contrário dos demais, os modelos para as regiões Nordeste e Sul se mantiveram os mesmos do ano de 2001 para o ano de 2009. A região Sul e a região Centro-Oeste precisaram de quatro variáveis preditoras nos anos de 2001 e 2009, respectivamente. Todos os demais modelos foram compostos por apenas três variáveis.

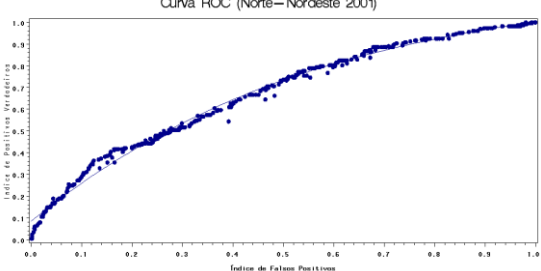
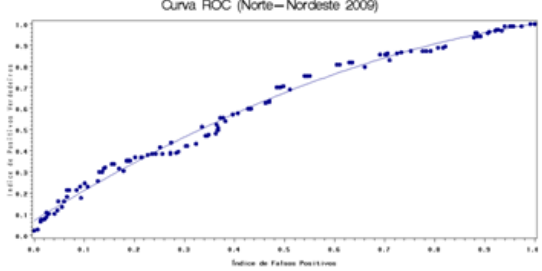
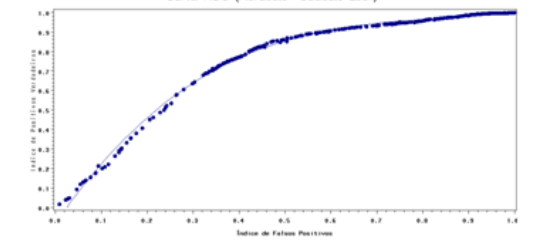
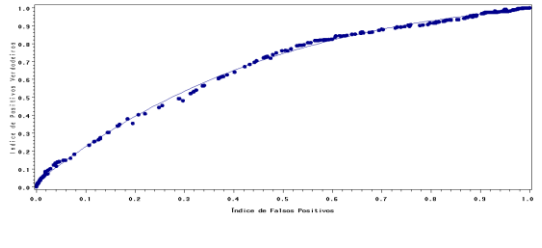
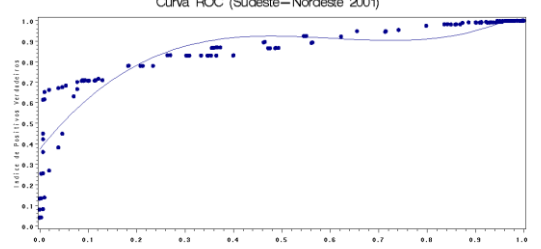
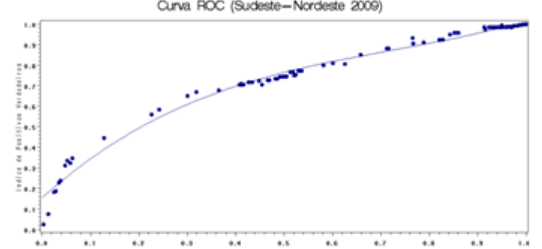
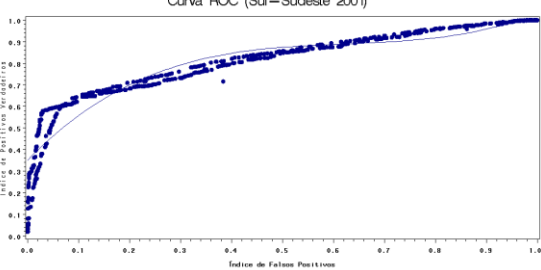
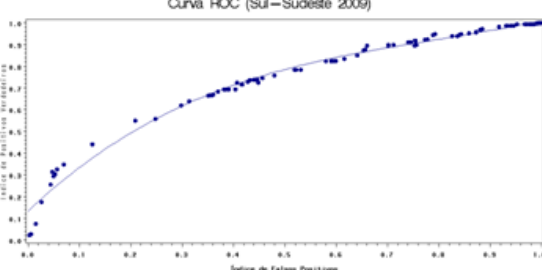
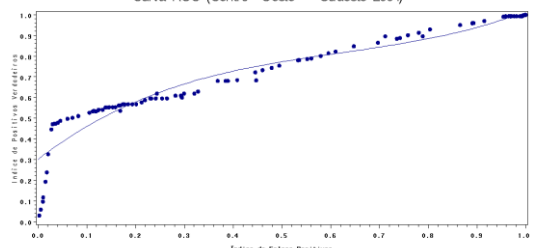
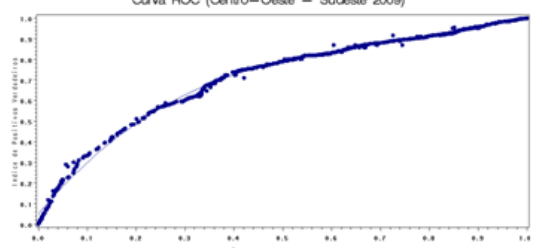
Como citado anteriormente, a tabela 22 apresenta as áreas abaixo das curvas ROC, para cada fluxo por ano, e um resumo sobre o comportamento geral entre os modelos.

Tabela 22: Capacidade de explicação dos modelos – 2001 e 2009.

Ano	Região	Percentual de Concordantes	Área abaixo da curva ROC
2001	N – Ne	67,10%	0,674
	Ne – Se	71,30%	0,721
	Se – Ne	84,90%	0,861
	S – Se	81,20%	0,815
	CO – Se	72,60%	0,738
2009	N – Ne	62,00%	0,63
	Ne – Se	65,60%	0,664
	Se – Ne	69,90%	0,717
	S – Se	70,40%	0,721
	CO – Se	70,60%	0,709

Como exposto anteriormente, a curva ROC indica a capacidade de explicação dos modelos, quanto mais alto o grau de elevação inicial da curva e maior a área abaixo da curva, mais capacidade de explicação possui o modelo. As curvas ROC dos modelos estudados se encontram nos gráficos de 1 a 10.

4.3.1 Curvas ROC

Norte para Nordeste	<p>Gráfico 1. Curva ROC: Norte – Nordeste, 2001</p> 	<p>Gráfico 2. Curva ROC: Norte – Nordeste, 2009</p> 
Nordeste para Sudeste	<p>Gráfico 3. Curva ROC: Nordeste – Sudeste, 2001</p> 	<p>Gráfico 4. Curva ROC: Nordeste – Sudeste, 2009</p> 
Sudeste para Nordeste	<p>Gráfico 5. Curva ROC: Sudeste – Nordeste, 2001</p> 	<p>Gráfico 6. Curva ROC: Sudeste – Nordeste, 2009</p> 
Sul para Sudeste	<p>Gráfico 7. Curva ROC: Sul – Sudeste, 2001</p> 	<p>Gráfico 8. Curva ROC: Sul – Sudeste, 2009</p> 
Centro- Oeste para Sudeste	<p>Gráfico 9. Curva ROC: Centro-Oeste – Sudeste, 2001</p> 	<p>Gráfico 10. Curva ROC: Centro-Oeste – Sudeste, 2009</p> 

Os modelos para o ano de 2009 diminuíram sua capacidade de explicação em relação a todos respectivos modelos de 2001. Dessas perdas de capacidade, vale destacar o Sul e o Sudeste, que perderam 10% e 15% de capacidade de explicação do modelo, respectivamente. Isto mostra uma mudança de comportamento no perfil do migrante, quando se estuda seu destino, pois no ano de 2001, houve uma média de 75,4% de pares concordantes nos modelos e, em 2009, esta média caiu para 67,7% e o mesmo aconteceu com a área abaixo da curva ROC que caiu de 0,761 para 0,688.

As possíveis variáveis que eram candidatas a entrar nos modelos eram ‘Sabe ler e escrever’, ‘Idade’, ‘Região de nascimento’, ‘Sexo’, ‘Raça/Cor da pele’, ‘Número de pessoas na família’, ‘Nasceu na UF de residência’ e ‘Renda familiar mensal’. Essas variáveis modelaram bem o comportamento no ano de 2001, porém, no ano de 2009, devido a uma possível mudança de comportamento, alguma variável relevante não deve ter sido considerada, o que ocasionou uma perda de capacidade dos modelos. Entre as regiões, nos dois anos, a diferença também mudou, pois no ano de 2001, a diferença do melhor modelo para o pior modelo foi de 17,8% enquanto, no ano de 2009, esta diferença caiu para 8,6%, ou seja, os modelos diminuíram a capacidade de explicação de 2001 para 2009, em contrapartida, a variância entre a explicação dos modelos também diminuiu consideravelmente.

5 CONCLUSÃO

O fenômeno da migração tem grande importância em diversos aspectos sociais e econômicos do país, a partir disso, percebeu-se a importância de um estudo mais detalhado sobre este fenômeno.

Com o estudo, observou-se que o número de migrantes recentes no país está diminuindo, apresentando uma redução de 1,3 milhão de pessoas entre 2001 e 2009.

Foram encontradas variáveis que estão associadas à mobilidade espacial no Brasil e, dessa forma, comparou-se o perfil do migrante brasileiro com o do não migrante e seu comportamento ao longo da primeira década do século XXI. Entre os migrantes, a proporção de pessoas que dizem ser de raça/cor da pele branca tem diminuído, há um alto percentual de jovens entre 20 e 29 anos, possuem taxa de alfabetização maior do que os não migrantes e um pouco menos da metade nasceu na região Nordeste.

Analisando possíveis origens e destinos, foi visto que os maiores fluxos migratórios entre regiões foram do Nordeste ao Sudeste e do Sudeste ao Nordeste, e a região que menos se interagiu com as demais foi a região Sul, entre os migrantes que se originam de uma UF da região Sul, mais da metade se destina a outra UF da região Sul.

As migrações intra-regionais se mostraram intensas e, exceto na região Nordeste, foram os maiores fluxos dadas as regiões de origem. Da mesma forma que as quantidades mais expressivas de migrantes foram nas regiões Sudeste e Nordeste, essas também têm a maior população, no entanto, quando observadas suas respectivas populações, em percentual, as regiões com um maior percentual de migrantes foram o Centro-Oeste e o Norte, sendo o Sudeste a região com um menor percentual de migrantes.

O fenômeno da migração de retorno, ou seja, a migração do indivíduo que saiu da UF de nascimento e retornou à UF natal, foi verificado em todos os fluxos analisados, representando, em média, 30% do total de movimento migratório.

Também foi possível notar que, além da diminuição do número de migrantes, houve uma mudança de comportamento, pois os modelos usados em 2001 explicaram bem a realidade, já os mesmos modelos aplicados a 2009 não obtiveram níveis de explicação tão satisfatórios, evidenciando assim que as variáveis usadas para explicar o modelo em 2001 já não tiveram o mesmo efeito em 2009.

6 REFERÊNCIAS

[1] Instituto migração e direito humanos. **Glossário**. Definições sobre conceitos de mobilidade espacial. Disponível em: <<http://www.migrante.org.br/glossario.htm>>. Acesso em: 02 Novembro 2010

[2] Superintendência de Estudos Econômicos e Sociais da Bahia. **Notas Metodológicas**. Conceitos e definições de termos a respeito de mobilidade espacial. Disponível em: <http://www.sei.ba.gov.br/index.php?option=com_content&view=article&id=150&Itemid=238>. Acesso em 02 Novembro 2010

[3] ELLERY, Herton. Migração Interna no Brasil. **Comunicados do IPEA**. Brasília, n.61, Agosto 2010. Disponível em:<http://www.ipea.gov.br/portal/images/stories/PDFs/comunicado/100817_comunicadoipea61.pdf>. Acesso em 2 Novembro 2010

[4] Núcleo de Estudos da População. **Mobilidade espacial da População**. Apresenta uma definição sobre mobilidade espacial da população. Disponível em: <<http://www.nepo.unicamp.br/pesquisa/linhaPesq6.html>>. Acesso em 12 Novembro 2010

[5] Instituto Brasileiro de Geografia e Estatística. **Trabalho e Rendimento: Pesquisa Nacional por Amostra de Domicílios**. Disponível em:<http://www.ibge.gov.br/projetos/comite_estatisticas_sociais/?page_id=157>. Acesso em 24 Janeiro 2011

[6] CABRAL, Gabriela. **Povoamento Brasileiro**. Equipe Brasil Escola. Disponível em: <<http://www.brasilecola.com/historiab/povoamento-brasileiro.htm>>. Acesso em: 12 Janeiro 2011

[7] FERNANDES, Duval.; VASCONCELLOS, Idamila. Proposta para a inserção da variável migração em sistemas de indicadores municipais. **São Paulo em Perspectiva**, São Paulo, v. 19, n. 3, 2005. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-88392005000300011&lng=en&nrm=iso>. Acesso em: 24 janeiro 2011

[8] OGLIARI, P. J. **Regressão Logística**, 8 de Dez. de 2004. Notas de aula.

[9] NETER, J.; Wasserman, W.; KUTNER, M.H. **Applied linear statistical models**, 4.ed. 1996

[10] SILVA, Eduardo Freitas. **Dados Categorizados**, 2010. Notas de Aula.

[11] AGRETI, A. An **Introduction to Categorical Data Analysis**, 2.ed. 2007

[12] Instituto Brasileiro de Geografia e Estatística. **Pesquisa Nacional por Amostra de Domicílios**. Brasil, 2006. 27.

ANEXO A – Histórico de pesquisas suplementares da PNAD

- 1969 – Migração interna.
- 1970 – Migração interna.
- 1972 – Rendimento Familiar (pesquisa especial) e Fecundidade.
- 1973 – Migração, Fecundidade e Mão-de-obra.
- 1974/1975 – Estudo Nacional da Despesa Familiar – ENDEF (pesquisa especial sobre consumo alimentar e orçamento familiar).
- 1976 – Mobilidade Social e Cor das Pessoas (pesquisa especial realizada por meio de uma subamostra), Migração e Fecundidade.
- 1977 – Habitação (pesquisa suplementar restrita ao Distrito Federal e às Regiões Metropolitanas de Recife, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre), Migração e Fecundidade
- 1978 – Fecundidade e Mortalidade.
- 1979 – Consumo de Energia.
- 1981 – Saúde.
- 1982 – Educação.
- 1983 – Mão-de-obra e previdência.
- 1984 – Fecundidade feminina.
- 1985 – Situação do menor (pesquisa suplementar restrita ao Distrito Federal e às nove Regiões Metropolitanas existentes à época: Belém, Fortaleza, Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo, Curitiba e Porto Alegre).
- 1986 – Anticoncepção (pesquisa especial realizada por meio de uma subamostra e restrita às mulheres de 15 a 54 anos de idade), acesso a serviços de saúde, Suplementação alimentar e Associativismo.
- 1988 – Participação político-social e Estoque de aparelhos utilizadores de energia.
- 1989 – Trabalho.
- 1990 – Trabalho.
- 1992 – Ensino supletivo, Nupcialidade e Trabalho das crianças de 5 a 9 anos de idade.

- 1993 – Ensino supletivo, Nupcialidade e Trabalho das crianças de 5 a 9 anos de idade.
- 1995 – Ensino supletivo, Nupcialidade e Trabalho das crianças de 5 a 9 anos de idade.
- 1996 – Mobilidade Social.
- 1998 – Saúde e Trabalho das crianças de 5 a 9 anos de idade.
- 1999 – Trabalho das crianças de 5 a 9 anos de idade.
- 2001 – Trabalho infantil.
- 2002 – Trabalho das crianças de 5 a 9 anos de idade.
- 2003 – Saúde, Participação em programas voltados para educação e Trabalho das crianças de 5 a 9 anos de idade.
- 2004 – Aspectos complementares de educação, Acesso a transferências de renda de programas sociais e Segurança alimentar, e Trabalho das crianças de 5 a 9 anos de idade.
- 2005 – Acesso à Internet e Posse de telefone móvel celular para uso pessoal e trabalho das crianças de 5 a 9 anos de idade.
- 2006 – Aspectos complementares de educação, acesso a transferências de renda de programas sociais, Trabalho Infantil.
- 2007 – Aspectos complementares da educação de jovens e adultos e Educação profissional e Trabalho das crianças de 5 a 9 anos de idade.
- 2008 – Tabagismo (pesquisa especial), Saúde, Acesso à Internet e Posse de telefone móvel celular para uso pessoal.

Fonte: IBGE – Trabalho e Rendimento: Pesquisa Nacional por Amostra de Domicílios

ANEXO B – Principais características observadas na PNAD

A) Características da habitação

- Espécie do domicílio;

Para os domicílios particulares permanentes:

- Material das paredes e cobertura do prédio;
- Número de cômodos e de dormitórios;
- Condição de ocupação e valor do aluguel ou da prestação mensal;
- Forma de abastecimento de água e esgotamento sanitário;
- Destino do lixo;
- Combustível utilizado para cozinhar;
- Forma de iluminação;
- Existência de linha telefônica;
- Existência de bens duráveis – filtro, fogão, geladeira, freezer, máquina de lavar roupa, rádio, microcomputador, dvd e televisão.

B) Características demográficas gerais:

- Sexo;
- Idade;
- Cor ou raça;
- Condição no domicílio e na família;

C) Características de migração:

- Naturalidade em relação ao município e à Unidade da Federação;
- Unidade da Federação de nascimento;
- Tempo de residência no município e na Unidade da Federação dos migrantes, inclusive os de retorno;
- Unidade da Federação de residência anterior e há cinco anos da data de referência da pesquisa.

D) Características de fecundidade feminina:

- Número de filhos nascidos vivos, por sexo;
- Número de filhos nascidos mortos, por sexo;

- Número de filhos que estavam vivos, por sexo;
- Data de nascimento e sexo do último filho nascido vivo.

E) Características de educação:

- Alfabetização;
- Escolarização;
- Série e grau freqüentados pelos estudantes;
- Última série concluída, grau correspondente e conclusão do curso para pessoas que não são estudantes.

F) Características de trabalho e rendimento:

- Condição de atividade e de ocupação na semana de referência e no período de referência de 365 dias;
- Ocupação, atividade, posição na ocupação, categoria do emprego nos trabalhos principal e secundário da semana de referência, no trabalho principal do período de 365 dias ou no último trabalho do período de referência de cinco anos;
- Rendimento e horas trabalhadas nos trabalhos principal e secundário e em outros trabalhos da semana de referência;
- Contribuição para instituto de previdência nos trabalhos principal e secundário e em outros trabalhos da semana de referência e no trabalho principal dos 365 dias;
- Tempo de permanência no trabalho principal da semana de referência e no trabalho principal do período de referência de 365 dias;
- Setor do emprego e área do emprego público nos trabalhos principal e secundário da semana de referência e no trabalho principal dos 365 dias;
- Tempo gasto do domicílio para o trabalho, local de estabelecimento do empreendimento e número de pessoas ocupadas no empreendimento do trabalho principal da semana de referência;
- Relação com o empreendimento dos conta-própria e empregadores em atividade agrícola no trabalho principal da semana de referência e no trabalho principal do período de referência de 365 dias;

- Forma de remuneração, recebimento de auxílios (alimentação, transporte, educação e saúde) e jornada diurna ou noturno do emprego no trabalho principal da semana de referência;
- Associação a sindicato, contribuição para previdência privada e idade com que começou a trabalhar das pessoas ocupadas no período de referência de 365 dias;
- Recebimento do seguro-desemprego das pessoas sem trabalho na semana de referência que saíram do último emprego no período de referência de 365 dias;
- Condição de aposentado, pensionista e cuidar dos afazeres domésticos, independentemente da condição de atividade; e rendimentos não oriundos de trabalho, independentemente da condição de atividade.

Fonte: IBGE – Trabalho e Rendimento: Pesquisa Nacional por Amostra de Domicílios

APÊNDICE A – Características Sócio-Demográficas - 2001, 2005 e 2009

Características Sócio-Demográficas	2001				2005				2009			
	Migrantes		Não Migrantes		Migrantes		Não Migrantes		Migrantes		Não Migrantes	
	Número	%	Número	%	Número	%	Número	%	Número	%	Número	%
Total	4449510	100	170955241	100	4725935	100	183880894	100	3228143	100	191795854	100
Sexo												
Homens	2198271	49,4	83231162	48,7	2338036	49,5	89608501	48,7	1626946	50,4	93356489	48,7
Mulheres	2251239	50,6	87724079	51,3	2387899	50,5	94272393	51,3	1601197	49,6	98439365	51,3
Cor da pele ou Raça												
Branca	2272746	51,1	88783926	53,4	2255579	47,7	89083665	49,8	1466775	45,4	90874817	48,3
Preta	206410	4,6	9436646	5,7	239777	5,1	11302657	6,3	223005	6,9	13020403	6,9
Parda	1955372	43,9	67136145	40,4	2204799	46,7	77415971	43,3	1515300	46,9	83144010	44,2
Amarela	9637	0,2	752357	0,5	18651	0,4	835608	0,5	10590	0,3	858719	0,5
Indígena	5345	0,1	217082	0,1	7129	0,2	343982	0,2	12473	0,4	402205	0,2
Saber ler e escrever												
Sim	3911314	87,9	133157966	77,9	4219538	89,3	147344917	80,1	2942067	91,1	155396642	82,5
Não	538196	12,1	37797275	22,1	506397	10,7	36535977	19,9	286076	8,9	32962576	17,5
Idade												
Menos de 10 anos	491904	11,1	31484047	18,9	557575	11,8	31025221	17,3	305801	9,5	28672442	15,2
10 a 14 anos	450830	10,1	16366588	9,8	444405	9,4	16715669	9,3	315549	9,8	17091785	9,1
15 a 19 anos	553499	12,4	17024278	10,2	508446	10,8	17183673	9,6	320276	9,9	16602139	8,8
20 a 24 anos	666105	15,0	15292624	9,2	689231	14,6	16580142	9,3	483123	15,0	15998135	8,5
25 a 29 anos	600116	13,5	13056765	7,8	647137	13,7	14753620	8,2	481693	14,9	15962666	8,5
30 a 39 anos	821157	18,5	24661656	14,8	916453	19,4	25966830	14,5	639900	19,8	28190358	15,0
40 a 49 anos	472063	10,6	19892059	12,0	493925	10,5	22750226	12,7	341968	10,6	25062276	13,3
50 a 59 anos	220174	4,9	13267827	8,0	269380	5,7	16038260	9,0	174060	5,4	19217044	10,2
60 anos ou mais	173662	3,9	15302001	9,2	199383	4,2	17980956	10,0	165773	5,1	21562373	11,4
Naturalidade em relação à UF												
Naturais	1273422	28,6	146682833	85,8	1464165	31,0	157306806	85,6	933863	28,9	163903924	85,5
Não Naturais	3176088	71,4	24266413	14,2	3261770	69,0	26566177	14,4	2294280	71,1	27891930	14,5
Número de pessoas na família												
1	208709	4,7	4665816	2,8	197697	4,2	5744037	3,2	157348	5,0	7023974	3,7
2	704364	16	23081397	14	848163	18,2	28070714	15,7	677980	21,4	32643823	17,4
3	1055487	24	37443284	23	1162381	24,9	42963575	24,1	832068	26,2	49507294	26,3
4	1125058	25	45122402	27	1178070	25,3	49176536	27,5	840212	26,5	51468932	27,4
5	750544	17	29132472	18	747627	16,0	28188572	15,8	418190	13,2	27180555	14,5
6 ou mais	605348	14	26902474	16	525689	11,3	24392906	13,7	249411	7,9	20133157	10,7
Renda familiar												
Até 1/2 salário mín.	855459	19,6	32962500	19,9	313054	6,8	12504272	7,0	135367	4,3	8217594	4,4
Mais de 1/2 a 1 sal. mín.	1072364	24,5	41064910	24,8	905073	19,6	34936159	19,6	300825	9,7	19665366	10,5
Mais de 1 a 2 sal. mín.	1080167	24,7	43021733	26,0	1398984	30,3	54164113	30,3	751999	24,1	45583357	24,3
Mais de 2 a 3 sal. mín.	499521	11,4	18101010	10,9	693001	15,0	28029375	15,7	585994	18,8	34497829	18,4
Mais de 3 a 5 sal. mín.	389663	8,9	13852158	8,4	618494	13,4	23471699	13,2	572143	18,4	36059520	19,2
Mais de 5 a 10 sal. mín.	283447	6,5	9258631	5,6	410729	8,9	15062200	8,4	429847	13,8	25377105	13,5
Mais de 10 a 20 sal. mín.	102363	2,3	2951236	1,8	137945	3,0	5471093	3,1	170728	5,5	9210102	4,9
Mais de 20 sal. mín.	91387	2,1	4541736	2,7	135724	2,9	4849132	2,7	167243	5,4	9294405	4,9

APÊNDICE B – Características Sócio-Demográficas Médias, 2001-2009

Características Sócio-Demográficas	Migrantes (%)	Não Migrantes (%)
Sexo		
Masculino	49,8	48,7
Feminino	50,2	51,3
Raça/Cor da pele		
Branca	48,3	50,6
Preta	5,6	6,4
Parda	45,4	42,3
Outras	0,7	0,7
Sabe ler e escrever		
Sim	89,2	79,9
Não	10,8	20,1
Idade		
05 a 09 anos	13,1	13,9
10 a 19 anos	21,6	22,3
20 a 29 anos	28,4	18,7
30 a 39 anos	17,9	14,6
40 a 49 anos	10,0	12,4
50 anos ou mais	9,0	18,2
Nasceu na UF de residência		
Sim	30,0	85,3
Não	70,0	14,7
Região de Nascimento		
Norte	7,4	6,5
Nordeste	42,2	33,1
Sudeste	28,6	39,7
Sul	13,8	15,3
Centro-Oeste	8,0	5,5
Número de pessoas na família		
01	4,5	3,2
02	18,5	15,5
03	24,7	24,3
04	25,5	27,4
05	15,2	15,9
06 ou mais	11,7	13,6
Renda familiar mensal		
Até 1/2 salário mín.	9,3	10,0
Mais de 1/2 a 1 sal. mín.	16,6	18,4
Mais de 1 a 2 sal. mín.	25,6	27,4
Mais de 2 a 3 sal. mín.	16,6	15,2
Mais de 3 a 5 sal. mín.	13,3	13,4
Mais de 5 a 10 sal. mín.	10,0	9,0
Mais de 10 a 20 sal. mín.	5,3	3,2
Mais de 20 sal. mín.	3,4	3,4

APÊNDICE C – Características Gerais por Fluxos Intra-Regionais

Características Sócio-Demográficas	Fluxos Migratórios				
	N - N	Ne - Ne	Se - Se	S - S	CO - CO
Sexo					
Homens	48,7%	47,6%	49,1%	49,1%	48,2%
Mulheres	51,3%	52,4%	50,9%	50,9%	51,8%
Idade					
Menos de 18 anos	34,3%	28,8%	23,7%	26,2%	30,5%
18 a 30 anos	35,7%	35,6%	34,8%	34,7%	34,1%
30 a 50 anos	22,9%	25,8%	29,5%	28,9%	27,3%
50 anos ou mais	7,1%	9,8%	11,9%	10,3%	8,0%
Raça/Cor da Pele					
Branca	26,2%	36,2%	57,9%	84,2%	40,7%
Preta	4,6%	4,2%	7,2%	2,3%	5,2%
Parda	68,6%	59,1%	34,4%	13,0%	53,2%
Amarela	0,4%	0,3%	0,4%	0,3%	0,5%
Indígena	0,1%	0,2%	0,2%	0,2%	0,4%
Renda familiar mensal					
Até 1 salário min.	29,4%	39,6%	21,1%	14,4%	27,8%
Entre 1 e 3 salários min.	47,3%	36,8%	43,2%	45,9%	42,3%
Entre 3 e 5 salários min.	12,1%	9,3%	14,1%	18,8%	12,8%
Entre 5 e 10 salários min.	6,5%	8,2%	10,9%	13,1%	9,2%
Entre 10 e 20 salários min.	3,1%	3,6%	6,1%	5,7%	5,6%
Mais de 20 salários mín.	1,6%	2,5%	4,6%	2,1%	2,3%

APÊNDICE D – Características Gerais pelos Principais Fluxos Migratórios entre as Regiões

Características Sócio-Demográficas	Fluxos Migratórios				
	N - Ne	Ne - Se	Se - Ne	S - Se	CO - Se
Sexo					
Homens	49,2%	49,3%	52,8%	50,3%	50,6%
Mulheres	50,8%	50,7%	47,2%	49,7%	49,4%
Idade					
Menos de 18 anos	32,5%	24,5%	29,3%	23,2%	25,7%
18 a 30 anos	30,7%	49,7%	28,1%	35,9%	32,8%
30 a 50 anos	28,4%	20,5%	32,3%	31,1%	31,1%
50 anos ou mais	8,4%	5,2%	10,2%	9,7%	10,4%
Raça/Cor da Pele					
Branca	31,4%	43,2%	36,2%	75,4%	60,7%
Preta	6,0%	7,8%	4,7%	4,1%	4,5%
Parda	62,0%	48,5%	58,6%	19,6%	33,1%
Amarela	0,2%	0,2%	0,2%	0,7%	0,8%
Indígena	0,3%	0,3%	0,2%	0,1%	0,9%
Renda familiar mensal					
Até 1 salário min.	36,7%	22,3%	46,9%	16,8%	20,2%
Entre 1 e 3 salários min.	39,9%	51,4%	37,2%	39,5%	39,2%
Entre 3 e 5 salários min.	10,6%	13,3%	6,8%	15,1%	15,0%
Entre 5 e 10 salários min.	9,0%	6,6%	4,9%	14,2%	13,2%
Entre 10 e 20 salários min.	2,2%	3,5%	1,7%	8,7%	7,5%
Mais de 20 salários mín.	1,5%	2,9%	2,5%	5,6%	5,0%