

Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

**Mobilidade sócio-ocupacional no Brasil: modelos log-lineares e outras
formas de analisar**

Camilla de Oliveira

Tese apresentada para obtenção do título de
Doutora em Ciências. Área de concentração:
Economia Aplicada

Piracicaba
2024

Camilla de Oliveira
Bacharel em Ciências Econômicas

**Mobilidade sócio-ocupacional no Brasil: modelos log-lineares e outras formas de
analisar**

versão revisada de acordo com a Resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **RODOLFO HOFFMANN**

Tese apresentada para obtenção do título de
Doutora em Ciências. Área de concentração:
Economia Aplicada

Piracicaba
2024

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Oliveira, Camilla de

Mobilidade sócio-ocupacional no Brasil: modelos log-lineares e outras formas de analisar / Camilla de Oliveira. -- versão revisada de acordo com a Resolução CoPGr 6018 de 2011. -- Piracicaba, 2024.

106 p.

Tese (Doutorado) -- USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Mobilidade social 2. Modelos log-lineares 3. Intergenerational earnings elasticity 4. Great Gatsby curve I. Título

DEDICATÓRIA

A Quem me deu tudo

AGRADECIMENTOS

Agradeço a Deus por ter me permitido chegar até aqui e ter as oportunidades que tive e a Nossa Senhora do Carmo por ter me protegido neste caminho.

Ao meu orientador Rodolfo Hoffmann que sempre esteve à disposição e enriqueceu o meu trabalho com suas orientações e ensinamentos, sem ele este trabalho não seria possível.

Aos meus pais e a minha irmã que sempre me deram o necessário para chegar até o doutorado, motivaram e deram o suporte para que não desistisse.

Aos amigos que me incentivaram e não me deixaram perder de vista a importância deste trabalho. Em especial à Mariza de Almeida, que sempre esteve a postos para me apoiar e ajudar durante todo o programa de doutorado.

Ao amigo Felipe Augusto Rocha Dias, que diariamente me motivou a persistir na redação deste trabalho na reta final e aos amigos da Inteligência de Mercado do SENAI-SP.

Aos amigos das Equipes de Jovens de Nossa Senhora, em especial à minha equipe Nossa Senhora Aparecida, que me ouviu por tantos meses falar sobre a tese e rezou para que tudo corresse bem na conclusão deste trabalho. Aos amigos das paróquias São João Batista de Rio Claro e Nossa Senhora do Monte Serrate de São Paulo, em especial aos padres Antonio Maciel e Vandro Pisaneschi.

Aos amigos da Economia de Francisco que ressignificaram a economia para mim e, sem dúvida, influenciaram minhas reflexões acerca da mobilidade sócio-ocupacional expressas aqui.

Por fim, a todos os meus antigos professores, agora amigos, que me auxiliaram quando precisei de ajuda. Meu muito obrigada a todos aqueles que contribuíram de alguma maneira para a finalização deste trabalho, particularmente ao corpo docente e colaboradores da Escola Superior de Agricultura “Luiz de Queiroz”.

Ao CNPq pelo fomento desta tese de doutorado através da Cota Institucional – processo 141092/2019-0.

SUMÁRIO

RESUMO	7
ABSTRACT	8
LISTA DE FIGURAS	9
LISTA DE TABELAS	10
1. INTRODUÇÃO.....	11
2. REFERENCIAL TEÓRICO.....	15
2.1. O estudo da mobilidade social.....	15
2.2. Visões teóricas no estudo da mobilidade social	23
2.3. Conceitos de mobilidade: taxas absolutas e relativas.....	24
2.4. Estudo da mobilidade no Brasil.....	26
3. METODOLOGIA.....	31
3.1. Fonte e natureza dos dados.....	31
3.2. Hierarquização das ocupações.....	33
3.3. Tabelas de contingência	34
3.4. Modelos log-lineares	38
3.4.1. Modelos log-lineares para análises bivariadas	39
3.4.2. Modelos log-lineares para análises trivariadas.....	44
3.5. Razões de chances	46
3.5.1. Razões de chances para tabelas de contingência 2×2.....	47
3.5.2. Razões de chances para tabelas de contingência I x J.....	50
3.5.2.1. Razões de chances locais.....	50
3.5.2.2. Abordagem de células geradoras.....	52
3.6. Medidas de mobilidade a partir de tabelas de contingência.....	53
4. RESULTADOS	57
4.1. Amostra	57
4.1.1. Descrição das coortes de idade.....	58
4.1.2. Distribuição de pais e filhos por coortes de idade e matrizes de mobilidade.....	60
4.2. Medidas de mobilidade social e associação	66
4.3. Modelos log-lineares	73
4.3.1. Modelo selecionado.....	75
4.4. Uma discussão sobre a maneira de avaliar a mobilidade	80
5. CONSIDERAÇÕES FINAIS	83

REFERÊNCIAS 87
APÊNDICES 99

RESUMO

Mobilidade sócio-ocupacional no Brasil: modelos log-lineares e outras formas de analisar

Na virada do novo milênio, o fenômeno da mobilidade social, relevante na literatura sociológica desde 1950, com tímida contribuição de economistas até o fim dos anos 1990, ultrapassou a esfera acadêmica ganhando popularidade nos círculos políticos e midiáticos. Nestes círculos, porém, a tradição sociológica foi preterida em favor do Intergenerational Earnings Elasticity (IGE), medida proposta por economistas, e a The Great Gatsby Curve (GGC). Para sociólogos há, nesta adoção, falhas na conceituação do fenômeno e nas metodologias adotadas para sua mensuração, em especial na ausência de definição explícita do caráter da mobilidade abordado: absoluto ou relativo, gerando confusão no debate público. Os modelos log-lineares, utilizados por sociólogos desde os anos 70, além da pretensão de examinar a variabilidade das frequências observadas, também propõe a mensuração da mobilidade relativa através da interação entre variáveis, não se restringindo, deste modo, apenas ao caráter relativo da mobilidade. O objetivo deste estudo é investigar a mobilidade social intergeracional brasileira em sua dimensão relativa de maneira crítica, através de modelos log-lineares, como alternativa à utilização do IGE. Os dados utilizados foram obtidos a partir do suplemento de Mobilidade Sócio-ocupacional da Pesquisa Nacional por Amostra de Domicílios Anual (PNAD) de 2014. A amostra foi classificada em estratos ocupacionais, para identificar a posição do indivíduo na estrutura social, e coortes de idade, buscando captar o efeito temporal nas desigualdades de oportunidade de mobilidade. Foram estimados modelos log-lineares trivariados, considerando interações entre status de origem, status de destino e coorte de idade do indivíduo e calculadas medidas de mobilidade absoluta e de associação estatística. As medidas de associação estatística e modelos log-lineares mostraram forte associação entre status do pai e status do filho, para a amostra geral e todas as coortes de idade. O modelo log-linear de associação homogênea mostrou que a variável mais relevante na variabilidade das frequências nas tabelas de contingência é o status do pai e, entre as associações entre pares de variáveis, a mais forte é a associação entre status do pai e status do filho. Além da sua interpretação pouco trivial, os modelos log-lineares não se ajustaram bem aos dados, uma vez que a amostra usada é muito grande e leva a considerar como estatisticamente significativos os desvios de modelos não-saturados. As medidas de associação estatísticas tradicionais se mostraram intuitivas e descreveram o comportamento da mobilidade relativa satisfatoriamente. A associação entre origem e destino não é a mesma para todas as coortes de idade, apresentando redução para indivíduos mais velhos, que tiveram mais tempo para se movimentar na estrutura social. As medidas de mobilidade absoluta forneceram caracterização mais completa do fenômeno uma vez que possibilitaram a visualização da direção dos movimentos experimentados pelos indivíduos, que no caso brasileiro são predominantemente ascendentes (48,70% da população). Por fim, conclui-se que, assim como o IGE proposto por economistas, os modelos log-lineares não são a melhor alternativa para mensuração da mobilidade relativa e seu uso exclusivo é insuficiente para o entendimento da mobilidade social, evidenciando as vantagens da utilização de medidas absolutas de mobilidade para sua compreensão.

Palavras-chave: Mobilidade social, Modelos log-lineares, Intergenerational earnings elasticity, Great Gatsby curve

ABSTRACT

Socio-occupational mobility in Brazil: log-linear models and alternative analytical approaches

At the turn of the new millennium, the phenomenon of social mobility, which has been relevant in sociological literature since the 1950s, with a timid contribution from economists until the end of the 1990s, went beyond the academic sphere and gained popularity in political and media circles. In these circles, however, the sociological tradition was disregarded in favor of the Intergenerational Earnings Elasticity (IGE), a measure proposed by economists, and The Great Gatsby Curve (GGC). As far as sociologists are concerned, there are flaws in the conceptualization of the phenomenon and in the methodologies used to measure it, especially in the absence of an explicit definition of the type of mobility being addressed: absolute or relative, generating confusion in the public debate. Log-linear models, used by sociologists since the 1970s, not only claim to examine the variability of the frequencies observed, but also propose the measurement of relative mobility through the interaction between variables, thus not restricting themselves to the relative nature of mobility alone. The objective of the present study is to investigate Brazilian intergenerational social mobility in its relative dimension in a critical way, using log-linear models as an alternative to using the IGE. The data used was obtained from the Socio-occupational Mobility supplement of the 2014 Annual National Household Sample Survey (PNAD). The sample was classified into occupational strata and age cohorts to identify the individual's position in the social structure and capture the temporal effect on inequalities in mobility opportunities respectively. Trivariate log-linear models were estimated, considering interactions between origin status, destination status and the individual's age cohort, and measures of absolute mobility and statistical association were calculated. The measures of statistical association and log-linear models showed a strong association between the father's status and the son's status, for the general sample and all age cohorts. The log-linear model of homogeneous association showed that the most relevant variable in the variability of the frequencies in the contingency tables is the father's status and, among the associations between pairs of variables, the strongest is the association between the father's status and the child's status. In addition to their non-trivial interpretation, the log-linear models did not fit the data well, since the sample used is very large and leads to consider as statistically significant the deviations of non-saturated models. The traditional statistical measures of association proved intuitive and described the behavior of relative mobility satisfactorily. The association between origin and destination is not the same for all age cohorts, showing a reduction for older individuals who have had more time to move through the social structure. The absolute mobility measures provided a more complete characterization of the phenomenon since they made it possible to see the direction of the movements experienced by individuals, which in the Brazilian case are predominantly ascending (48.70% of the population). Finally, it can be concluded that, similarly to the IGE proposed by economists, log-linear models are not the best alternative for measuring relative mobility and their exclusive use is insufficient for understanding social mobility, highlighting the advantages of using absolute measures of mobility to better understand it.

Keywords: Social mobility, Log-linear models, Intergenerational earnings elasticity, Great Gatsby curve

LISTA DE FIGURAS

Figura 1. Resultados para as medidas de mobilidade tradicionais para cada coorte de idade	69
Figura 2. Resultados para as medidas do Coeficiente de Abertura (Yasuda), <i>Average Absolute Jump</i> (AAJ) e Delta, para cada coorte de idade	70
Figura 3. Resultados para as medidas de associação: Coeficiente Phi, Coeficiente de Contingência, Coeficiente de Correlação de Pearson, para cada coorte de idade.....	71
Figura 4. Resultados para a medida V de Cramer, para cada coorte de idade	72
Figura 5. Resultados para as Razões de Chances (ORs) para cada coorte de idade e amostra geral da sociedade brasileira, em 2014.....	79
Figura 6. <i>Great Gatsby Curve</i> (GGC), relação entre o <i>Intergenerational Earnings Elasticity</i> (IGE) e o coeficiente de Gini, para países selecionados.....	100
Figura 7. A relação entre o índice de Gini (G) e o valor de β (o IGE) para três valores da variância do erro da relação entre os logaritmos das rendas de pais e filhos [relação (55) com $\sigma^2 = 0,8$, $\sigma^2 = 0,5$ ou $\sigma^2 = 0,2$]	102
Figura 8. Curvas mostrando como o valor de β (o IGE) varia em função do índice de Gini para três relações entre o desvio padrão (σ) do erro do processo AR(1) entre os logaritmos das rendas de pais e filhos e o desvio padrão (S) da distribuição do logaritmo da renda dos filhos: $\sigma = 0,16 + 0,6S$, $\sigma = 0,15 + 0,65S$ ou $\sigma = 0,17 + 0,55S$	103

LISTA DE TABELAS

Tabela 1. Tabela de contingência $I \times J$	35
Tabela 2. Modelos log-lineares para tabelas de contingência bivariadas e os correspondentes graus de liberdade (g) do resíduo.....	40
Tabela 3. Tabela de contingência para o caso bivariado dicotômico.....	40
Tabela 4. Modelos log-lineares para tabelas de contingência trivariadas e graus de liberdade correspondentes	46
Tabela 5. Frequência da variável “Destino”	47
Tabela 6. Classificação cruzada entre as variáveis “Destino” e “Origem”	48
Tabela 7. Chances para a classificação cruzada numa tabela 2×2	49
Tabela 8. Classificação cruzada entre as variáveis “Destino” e “Origem” com três estratos	51
Tabela 9. Razões de chances locais para a tabela de contingência 3×3	51
Tabela 10. Exemplo de matriz de mobilidade social 3×3	53
Tabela 11. Distribuição da população nos estratos sócio-ocupacionais em 2014/milhares de pessoas	58
Tabela 12. Distribuição da frequência na amostra entre as coortes de idade/décadas.	59
Tabela 13. Distribuição dos indivíduos e seus pais entre coortes de idades e estratos sócio-ocupacionais 60	
Tabela 14. Matrizes de mobilidade sócio-ocupacional intergeracional brasileira em 2014, por coorte de idade (%)	62
Tabela 15. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014.	64
Tabela 16. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 16 e 25 anos, coorte de idade (1).....	64
Tabela 17. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 26 e 35 anos, coorte de idade (2).....	65
Tabela 18. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 36 e 45 anos, coorte de idade (3).....	65
Tabela 19. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 46 e 55 anos, coorte de idade (4).....	66
Tabela 20. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 56 e 65 anos, coorte de idade (5).....	66
Tabela 21. Padrões gerais da mobilidade intergeracional por coorte de idade	67
Tabela 22. Estatísticas de ajustamento dos modelos log-lineares estimados para a análise da mobilidade intergeracional brasileira, em 2014	74
Tabela 23. Estatísticas dos efeitos individuais e das associações consideradas no modelo de associação homogênea para a mobilidade social intergeracional brasileira, em 2014, considerando coortes de idade de 10 anos.....	76
Tabela 24. Estatísticas dos efeitos individuais e das associações consideradas no modelo saturado para a mobilidade social brasileira em 2014	77
Tabela 25. Descrição das Razões de Chances (ORs) adotadas como exemplo	78
Tabela 26. Resultados para as Razões de Chances (ORs) para cada coorte de idade e amostra geral da sociedade brasileira, em 2014.....	80
Tabela 27. Resultados estimados para o modelo de associação homogênea da mobilidade social brasileira em 2014, considerando as variáveis coorte de idade (C), status do pai (P), status do filho (F) e suas associações	104

1. INTRODUÇÃO

A busca por padrões e tendências no comportamento dos agentes sociais e no funcionamento da sociedade é parte essencial da ciência social. O debate sobre o caráter científico, o rigor metodológico e a robustez dos resultados obtidos na ciência social é suscitado pela construção de teorias sociais e modelos que estudam fenômenos envolvendo interações humanas (GOLDTHORPE, 2005). Naturalmente, no centro deste debate há uma comparação entre o corpo da ciência social e o das ciências naturais.

É necessário pontuar que, ainda que haja um esforço dos cientistas sociais em mensurar e sistematizar fenômenos sociais, há duas distinções primárias entre o objeto de estudo das duas áreas: a mutabilidade e a inserção do cientista nos fenômenos estudados (COLE, 1994). Ao contrário do que acontece em áreas como a física, em geral o estudo em disciplinas como economia e sociologia não permite a condução de experimentos e, uma vez que estes estudiosos estão inseridos no contexto social, não podem explicá-lo de forma completamente neutra, sendo levados, muitas vezes a interpretá-lo de diferentes formas (BRYANT, (1995); FLYVBJERG, (2001)).

Ao reunir sociólogos e economistas na sistematização do conhecimento a respeito da movimentação dos indivíduos na sociedade, o corpo teórico e metodológico da mobilidade social transparece o caráter interpretativo da ciência social. A literatura sobre o fenômeno apresenta abordagens diferentes, a contar: as baseadas em classe social, ocupação, status socioeconômico, status educacional e rendimentos. Usualmente as abordagens não chegam num consenso sobre a situação da fluidez social em dada sociedade, reforçando o segundo ponto destacado por Cole (1994): os estudiosos das ciências sociais são influenciados por “interesses pessoais, valores e ideologias” (GOLDTHORPE, 2005).

No estudo da mobilidade social o ponto de partida consiste na decisão de qual será o arcabouço conceitual adotado como referência das posições relativas dos indivíduos na estrutura social de interesse (ERIKSON; GOLDTHORPE, 1992). Essa questão consiste no ponto de tangência entre o estudo da estratificação social e da mobilidade onde duas correntes teóricas ganham destaque: a da análise de classes, baseada principalmente nos trabalhos dos sociólogos Karl Marx e Max Weber, e a funcionalista, norteadas pelo trabalho de John Stuart Mill (SCALON, 1999). Adota-se neste estudo a perspectiva funcionalista que vê a estrutura da sociedade através de uma

hierarquia social e atribui status aos indivíduos de acordo com suas características sócio-ocupacionais.

Com o recente interesse de economistas na área e a proposição da adoção da perspectiva de mobilidade intergeracional de renda a partir da utilização de modelos de regressão linear para mensurar a associação entre o rendimento do filho e dos pais através do coeficiente de regressão denominado *Intergenerational Earnings Elasticity* (IGE), surge a necessidade de revisitar, na literatura sociológica, a terceira geração de estudos de mobilidade, especificamente os modelos log-lineares.

Os modelos log-lineares caracterizam, assim como o IGE, um esforço metodológico na investigação da dimensão relativa da mobilidade social, ou seja, das desigualdades de oportunidades de mobilidade social experimentadas pelos indivíduos dada a sua origem. Em essência, o interesse é identificar de que forma a posição de origem na sociedade é associada à sua posição de destino.

O objetivo deste trabalho é realizar uma investigação da mobilidade sócio-ocupacional intergeracional relativa no Brasil, a fim de compreender o funcionamento da associação entre origem e destino na sociedade brasileira. Adicionalmente, é realizada uma discussão sobre o estudo da mobilidade considerando suas dimensões absolutas e relativas e, para o segundo caso, a relevância dos métodos escolhidos para a análise.

Para este fim são utilizados dados do suplemento de Mobilidade Sócio-ocupacional disponibilizado pelo Instituto Brasileira de Geografia e Estatística (IBGE) na Pesquisa por Amostra de Domicílios de 2014 (PNAD/2014), o procedimento descrito em Oliveira e Hoffmann (2021) para a estratificação das ocupações presentes na referida pesquisa e a classe de modelos log-lineares propostos por Goodman (1979a) para a análise de tabelas de contingência (caracterizadas nas matrizes de mobilidade). Complementa-se a análise de modelos log-lineares com uma contextualização do fenômeno através de medidas absolutas de mobilidade e medidas de associação estatística tradicionais.

Seguindo a sugestão de Duncan (1966) e outros autores, é realizada uma análise por coortes de idade. A análise da mobilidade sócio-ocupacional intergeracional levando em consideração coortes de idade busca mitigar limitações como a ausência de gerações perfeitas e sucessivas de pais e filhos, apontada por Behrman (2000). Além disso, as coortes de idade permitem analisar períodos menores do que o adotado por

Oliveira (2018) e Oliveira e Hoffmann (2021) a partir da utilização dos dados da PNAD/2014.

Este trabalho está estruturado em cinco capítulos, sendo esta introdução o primeiro. Em seguida, no capítulo 2, é apresentada a revisão da literatura da mobilidade social com o intuito de traçar um evolutivo histórico dos conceitos e metodologias utilizadas na área. No capítulo 3 são apresentados os modelos log-lineares utilizados para mensurar a mobilidade sócio-ocupacional intergeracional relativa da sociedade brasileira, além da análise de razões de chances e das medidas de mobilidade social absolutas tradicionais. A seção de resultados, no capítulo 4, apresenta as características da amostra, além de discutir as matrizes de mobilidade e os resultados obtidos no emprego dos modelos log-lineares. Adicionalmente, há, ao final do capítulo, a discussão sobre a utilização de modelos log-lineares em contrapartida a medidas de associação tradicionais. Por fim, no capítulo 5, são expostas as considerações finais deste trabalho.

2. REFERENCIAL TEÓRICO

2.1. O estudo da mobilidade social

O progresso das técnicas estatísticas entre o final do século XIX e início do século XX foi fortemente influenciado pela tentativa de entender e quantificar em que medida características são transmitidas entre pais e filhos, gerações e classes sociais. Entre os primeiros esforços realizados para mensurar essas relações está o trabalho de Sir Francis Galton (1886), “*Regression towards mediocrity in hereditary stature*”, em que o autor busca investigar a hereditariedade da inteligência.

Diante da dificuldade em mensurar a variável inteligência, o objetivo do estudo tornou-se identificar a possível associação na herança de características físicas entre pais e filhos. Assim, Galton descobriu o que chamou de “regressão à média” que, uma vez generalizada por Sir Ronald Aylmer Fisher, passou a ser fundamental para o desenvolvimento de estudos nas áreas das ciências sociais e da saúde (SALSBURG; GRADEL, 2009).

A suposição de Galton (1886) a respeito da transmissão da inteligência e de atributos físicos, como é o caso da estatura, de pais para filhos é algo que migra, naturalmente, para o domínio do estudo econômico (ZIMMERMAN, 1992). Ainda assim, a ideia de que a posição de um indivíduo na sociedade é influenciada pela posição de seus antepassados foi abordada primeiro por sociólogos.

O primeiro “tratado geral” sobre mobilidade social foi escrito pelo sociólogo russo Pitirim Alexander Sorokin em 1927. Intitulado “*Social mobility*” (1927), o livro inaugurou o debate acadêmico a respeito da movimentação dos indivíduos entre as camadas da estrutura social (GOLDTHORPE, 2005). A obra, que foi republicada posteriormente sob o título “*Social and Cultural Mobility*” (1959), define a mobilidade social como sendo o “fenômeno do deslocamento dos indivíduos dentro do espaço social”¹ (SOROKIN, 1959, p. 3).

A forma com que Sorokin (1959) ilustra essa definição é ainda mais intuitiva do que as propostas pela literatura recente (i.g., *The Schumpeter Hotel* (FIELDS, 2000; SCHUMPETER, 1955) e a analogia da caravana (MAJOR; MACHIN, 2018)). Ele parte da explicação do significado de ‘deslocamento no espaço social’, apresentando a

¹“(…) *the phenomenon of shifting of individuals within social space.*” (SOROKIN, 1959, p. 3)

distinção entre o espaço geográfico e o espaço social para, então, traçar um paralelo entre as ciências que os estudam (as ciências naturais e as ciências sociais, respectivamente).

O estudo de fenômenos nos dois espaços requer a adoção de um ponto de referência; no caso específico da mobilidade social, trata-se da posição do indivíduo na sociedade. Toda organização social, por menor que seja, apresenta uma hierarquia, ou seja, a estratificação social é universal e está presente em todas as sociedades, sendo considerada uma característica permanente do contexto social (SOROKIN, 1959). Tendo em vista essa característica do espaço social, o estudo da mobilidade está, necessariamente, atrelado ao estudo da estratificação social. Sorokin (1959) orienta, ainda, que o estudo deve ser iniciado definindo-se as características da ‘pirâmide’ social para, só então, ser conduzida uma investigação da sua organização interna.

Os estudos sobre mobilidade social desenvolveram-se de forma mais robusta somente após a Segunda Guerra Mundial, na década de 1950. O demógrafo Glass (1954) retomou o trabalho original de Sorokin (1927) considerando-o, até então, “o único estudo abrangente sobre mobilidade social”² (GLASS, 1954, p. 5). A análise da mobilidade social nessa época era realizada, essencialmente, a partir da análise do status ocupacional dos indivíduos. Essa escolha metodológica se justifica pelo fato de a ocupação se relacionar com o status econômico, o *background* educacional e o padrão de vida das pessoas, sendo, assim, um bom ponto de referência (GLASS, 1954).

A característica marcante desta primeira geração de estudos é a análise de matrizes de transição de status a partir de estatísticas descritivas como observamos no trabalho de Rogoff (1964)³ e a tentativa de mensurar a mobilidade em duas frentes: a mobilidade absoluta (ou total), que consiste no total de movimentos ascendentes e descendentes na estrutura social; e a relativa, que tenta entender em que medida os indivíduos estão expostos a diferentes oportunidades de mobilidade⁴.

A noção de que o fenômeno pode ser visto sob essas duas perspectivas e que há um problema de associação aleatória entre as distribuições permeia o trabalho de estatísticos italianos como Benini (1927) e Livi (1950). O chamado “Índice de Atração” desenvolvido por Benini (1927) foi aperfeiçoado por Glass (1954), resultando no conhecido “Índice de Associação”. Utilizado em grande parte das análises da primeira

² “(...) *the only comprehensive study of social mobility*” (GLASS, 1954, p. 5)

³ Originalmente publicado em 1951, ver Rogoff (1951).

⁴ Essa diferenciação será abordada na seção 2.3 deste trabalho.

geração, o índice de associação mensura a relação entre os movimentos observados na sociedade e os movimentos esperados em caso de mobilidade perfeita, ou seja, na ausência de associação entre origem e destino.

O trabalho de Glass (1954) foi pioneiro e essencial para o progresso do estudo da mobilidade social assim como outros estudos desta mesma geração, em especial: Hout (1983), que aprofunda o estudo das medidas de mobilidade a partir das matrizes de transição de status; Yasuda (1964), que propõe um índice para mensurar o grau de abertura da sociedade; Lipset e Zatterberg (1956), que iniciam um debate sobre a comparação de mobilidade entre países; Carlsson (1958) e Svalastoga (1959), que aprofundam a questão da utilização de hierarquias de prestígio relacionadas às ocupações dos indivíduos ou *status* sócio-ocupacionais (GOLDTHORPE, 2005).

A melhoria na disponibilidade de dados e o crescente interesse no estudo dos mecanismos promotores da mobilidade, até então ignorados pelas limitações das análises de tabelas de contingência, resultam numa segunda geração de trabalhos. O livro de Blau e Duncan (1967), *The American occupational structure*, propõe a utilização de *Path Analysis* (Análise de Trajetórias) para quantificar a influência de determinadas variáveis (*e.g.* educação, ocupação do pai, etc) sobre o status do indivíduo (BLAU; DUNCAN, 1967). A mobilidade, sob essa perspectiva, deixa de ser caracterizada apenas como a movimentação dos indivíduos para ser vista como um processo de obtenção de *status* que poderia ser analisado a partir de análise de regressão (GOLDTHORPE, 2005).

A utilização de tabelas de mobilidade foi retomada após o advento dos modelos log-lineares, metodologia introduzida pelos trabalhos de Hauser *et al.* (1975), Hauser (1978) e Goodman (1979b), que marcam a terceira geração de estudos sociológicos de mobilidade social. A geração seguinte combinaria as metodologias empregadas na segunda e terceira, utilizando regressões logísticas multivariadas (BREEN, 1994; DIPRETE, 1990; LOGAN, 1983).

A partir da década de 1980 a mobilidade social começou a figurar na literatura econômica. Becker e Tomes (1986) assinalam que até então o único grande economista a abordar a mobilidade intergeracional sistematicamente havia sido Schumpeter (1955) e fazem uma crítica aos muitos estudos empíricos realizados por sociólogos apontando que careciam de modelos robustos para interpretar os resultados obtidos. Os autores sugerem um modelo baseado no “comportamento de maximização da utilidade dos

agentes, equilíbrio em diferentes mercados e forças estocásticas com incidências desiguais entre os agentes”⁵ (BECKER; TOMES, 1986, p. S3).

Becker e Tomes (1986) propõem que a dinâmica da desigualdade de renda entre as gerações de pais e filhos segue um modelo de Markov simples expresso por:

$$I_{t+1} = a + bI_t + \varepsilon_{t+1} \quad (1)$$

onde I_{t+1} é a renda atual do filho; I_t é a renda dos pais; a e b são constantes; e ε_{t+1} é o termo estocástico do modelo, representando outros fatores que afetam o rendimento do filho, independentemente da renda dos pais, sendo comum o uso do logaritmo da renda em lugar da própria renda.

A abordagem da mobilidade intergeracional de renda, inaugurada pelo trabalho de Becker e Tomes (1986), foi amplamente aceita entre os economistas. Os trabalhos de Solon (1992) e Zimmerman (1992) deram um novo passo ao propor alternativas para a mensuração dos rendimentos a partir da hipótese da renda permanente de Friedman (1957) (BLANDEN, 2013).

Os trabalhos de mobilidade desenvolvidos na literatura econômica baseiam-se, principalmente, na análise do coeficiente de regressão (b) do modelo de regressão linear do logaritmo do rendimento do filho contra o logaritmo do rendimento do pai (OLIVEIRA; HOFFMANN, 2021). O coeficiente b , denominado *Intergenerational Earnings Elasticity* (IGE), é considerado uma medida de persistência de status (SAKAMOTO; WANG, 2020). O IGE deve estar no intervalo de 0 a 1, podendo assumir esses dois valores; à medida que se aproxima de 1, expressa um aumento da imobilidade; por outro lado, quanto mais próximo de zero, mais próximo da mobilidade perfeita a sociedade se encontra.

Uma grande crítica dos sociólogos à análise baseada no IGE consiste no fato de se olhar apenas para a renda individual ou familiar no contexto da mobilidade social, não permitindo uma visão mais abrangente das condições de status dos agentes, uma vez que “de acordo com longa tradição sociológica, classes sociais deveriam ser definidas conforme as posições ocupacionais de indivíduos dentro de unidades produtivas e mercados de trabalho” (RIBEIRO, 2007, p. 103). Goldthorpe (2013) afirma

⁵ “(...) utility-maximizing behavior by all participants, equilibrium in different markets, and stochastic forces with unequal incidence among participants.” (BECKER; TOMES, 1986, p. S3)

que um dos problemas analíticos na abordagem econômica é a falta de uma distinção explícita entre mobilidade relativa e absoluta e alega que os coeficientes estimados (b), “adotados para representar as ‘elasticidades’ intergeracionais da renda dos filhos em relação à de seus pais, não são uma medida totalmente satisfatória para mensurar a mobilidade relativa” (GOLDTHORPE, 2013, p. 436)⁶.

É possível constatar um descompasso entre sociólogos e economistas no que se refere às abordagens utilizadas no estudo da desigualdade e da mobilidade. Um exemplo, colocado por Torche (2014), pode ser observado quando comparados os estudos de desigualdade de oportunidades do sociólogo Parsons (1951) e do economista Roemer (1998). Os autores distinguem duas dimensões principais no processo de desigualdade, Parsons (1951) cunha os termos características atribuídas/herdadas (“*ascription*”) e conquistadas/adquiridas (“*achievement*”), enquanto Roemer sugere a distinção entre características resultantes de circunstâncias (“*circumstances*”) e esforços (“*efforts*”). Os dois se referem ao mesmo processo, mas, como no debate da mobilidade, a terminologia e tradição sociológica foram desconsideradas pelos economistas. Essa distinção é largamente utilizada na visão funcionalista, discutida na seção seguinte e adotada na segunda geração de estudos da mobilidade para a hierarquização da sociedade.

No final dos anos noventa e início dos anos 2000, o interesse político pelo tema, principalmente no Reino Unido, se popularizou e permitiu que a discussão da mobilidade social ultrapassasse a esfera acadêmica. Nos círculos políticos e midiáticos os trabalhos desenvolvidos por economistas pautaram a forma como o assunto deveria ser abordado, enquanto a tradição sociológica foi preterida. Goldthorpe (2013) e Payne (2017) discutem essa questão salientando as falhas na conceituação do fenômeno, nas maneiras de mensurá-lo e identificar seus determinantes. Segundo os autores, a ausência de uma distinção explícita entre mobilidade relativa e absoluta na literatura econômica e a instrumentalização do fenômeno para interesses políticos resultou na confusão no debate público.

A “mobilidade social é também um dos processos mais mal compreendidos do nosso tempo”⁷ (PAYNE, 2017, p. vi), o debate público é dominado pelo discurso de que

⁶“(…) taken to represent the intergenerational ‘elasticities’ of children’s income with respect to that of their parents, are not an entirely satisfactory measure of relative mobility” (GOLDTHORPE, 2013, p. 436)

⁷ “Social mobility is also one of the most misunderstood processes of our time.” (PAYNE, 2017, p. vi)

é necessária mais mobilidade e que a mobilidade tem diminuído nas últimas décadas (GOLDTHORPE, 2013; PAYNE, 2017), mas deixa-se de lado o fato de que mais mobilidade pode significar também mais mobilidade descendente. De forma geral, há o entendimento equivocado de que a mobilidade social seja um sinônimo para ascendência social, o que não é, necessariamente, o que ocorre quando há mobilidade, tanto no caso da mobilidade intergeracional, quanto na intrageracional.

A visão econômica do fenômeno se consolidou de tal maneira que uma parcela de sociólogos passou a destacar as limitações das metodologias tradicionais na literatura sociológica e enaltecer o IGE, como é o caso do trabalho de Sakamoto e Wang (2020). Os autores listam pelo menos cinco grandes limitações das tabelas ocupacionais utilizadas em grande parte dos estudos sociológicos, a contar: (i) a natureza das ocupações, distinta em diversos contextos, caracteriza um indicador impreciso do rendimento do indivíduo; (ii) o foco em dados *cross-section*, dada a alta volatilidade do mercado de trabalho atual; (iii) a distinção dúbia entre os tipos de mobilidade estrutural e circular; (iv) a incapacidade de relacionar mobilidade e desigualdade; e (v) a insuficiência da ocupação enquanto uma variável que capte recompensas não monetárias e diferenciais compensatórios no trabalho dos indivíduos.

Sakamoto e Wang (2020) alegam que os estudos de mobilidade baseados nas ocupações e classificações de classe estão desaparecendo em decorrência das limitações mencionadas e defendem a utilização dos modelos econômicos que baseiam suas análises em distribuições de renda, mas parecem ignorar que essas abordagens também têm limitações, como a falta de confiabilidade dos dados de rendimento, que geralmente são auto declarados pelos indivíduos. Limitações dessa natureza podem não ser o caso do que encontramos em Chetty et al. (2014), mas são uma realidade no estudo de muitas sociedades que não têm disponibilidade e qualidade de dados, como é o caso do Brasil. Matrizes de mobilidade baseadas no *status* ocupacional não devem, portanto, ser descartadas, uma vez que dados a respeito da ocupação das pessoas são, em geral, mais confiáveis e disponíveis.

O modelo econômico baseado no IGE é interpretado “como uma medida de persistência intergeracional (*i.e.*, imobilidade)”⁸ e a palavra-chave neste conceito é ‘persistência’, uma vez que, medir a não persistência não é o mesmo que medir a mobilidade. Não é possível vislumbrar o fenômeno por inteiro a partir da medida de

⁸“(…) as a measure of intergenerational persistence (*i.e.*, immobility).” (SAKAMOTO; WANG, 2020)

persistência, dado que ela falha em responder se os indivíduos considerados móveis experimentaram mobilidade ascendente ou descendente. O IGE não é capaz de responder até mesmo essa questão trivial e, ainda assim, é a única sugestão utilizada para superar as limitações presentes nos modelos envolvendo a variável ocupação e matrizes de mobilidade.

Outra limitação do modelo de mobilidade, de uso comum entre economistas, consiste no fato de que, apesar de oferecer um índice interpretável e informativo⁹, ele é dependente da dispersão dos rendimentos nas gerações de origem e destino. Se a desigualdade apresentar grandes alterações de uma geração para a outra, como, por exemplo, numa elevação da desigualdade, “um coeficiente maior é necessário para dar conta do diferencial de renda maior na segunda geração”¹⁰ (BJÖRKLUND; JÄNTTI, 2011, p. 497), o que decorre da relação entre o coeficiente de correlação e o coeficiente de regressão, onde o coeficiente de correlação, ρ , é igual ao produto entre o IGE e a razão dos desvios-padrões da renda do pai, σ_O , e do filho, σ_D , como mostra a equação (2):

$$\rho = \beta \frac{\sigma_O}{\sigma_D} \quad (2)$$

Björklund e Jäntti (2011) complementam, ainda, dizendo que o coeficiente de correlação intergeracional (ρ) é preferível ao IGE enquanto medida para contornar essa limitação.

O interesse na relação entre a desigualdade e a mobilidade social permeia grande parte do debate no século XXI. Para Corak (2013)¹¹ o estudo da desigualdade entre gerações é o estudo da mobilidade intergeracional. O autor, retomando o trabalho de Solon (2004), destaca a influência de três grandes ‘instituições’ nas diferenças de mobilidade entre países: o *background* familiar, o mercado de trabalho e o Estado; e relaciona, pela primeira vez, o IGE ao Índice de Gini.

A relação entre o IGE e o Índice de Gini é uma constatação empírica, semelhante à hipótese da curva em U invertido de Kuznets (1955) que relaciona desenvolvimento econômico e desigualdade de renda. Corak (2013) observa, a partir de

⁹ O IGE.

¹⁰ “a larger coefficient is needed to account for the larger income differentials in the second generation.” (BJÖRKLUND; JÄNTTI, 2011, p. 497)

¹¹ Originalmente publicado em 2011.

evidências empíricas, que há uma associação entre os dois: num dado ponto no tempo, mais desigualdade é associada a menos mobilidade intergeracional, ou seja, um maior grau de persistência entre pais e filhos (imobilidade).

Quando há ausência de desigualdade, ou seja, com um Gini igual a zero, o IGE será indeterminado, mas havendo certo nível de desigualdade, ele poderá variar entre -1 e 1 . Krueger (2012), a partir dos dados e da relação entre IGE e Índice de Gini observada por Corak (2013), nomeia essa relação como “*The Great Gatsby Curve*” (GGC).

A correlação ilustrada na GGC suscitou debates entre os economistas. Para Mankiw (2013) o fato de haver uma correlação positiva entre imobilidade e desigualdade social não é surpreendente e não permite obter conclusões ou sugestões de política para melhora da desigualdade ou da mobilidade. Berman (2017) admite a relação proposta pela GGC e investiga a significância econômica proposta por essa correlação, chegando à conclusão de que há uma tautologia¹², uma relação mecânica entre as duas medidas: fixada a variância do termo de erro em (1), a desigualdade de renda entre filhos, medida pelo Índice de Gini, cresce com o IGE.

Berman (2017) afirma, ainda, que para que se obtenha uma evidência empírica significativa de uma relação entre mobilidade intergeracional e a desigualdade social são necessárias medidas que não possam ser deduzidas uma da outra e que a definição “da elasticidade intergeracional de rendimentos como uma medida de mobilidade é questionável e deveria ser considerada cuidadosamente”¹³ (BERMAN, 2017, p. 20), devido ao fato de que medidas de desigualdade obtidas da distribuição de renda, como é o caso do Índice de Gini, não consideram aspectos do fenômeno da mobilidade como “permutações entre a renda dos indivíduos, transferências ou mudanças de posição que não têm efeito na distribuição de renda”¹⁴ (BERMAN, 2017, p. 20).

¹² Essa questão é explorada no Apêndice A deste trabalho.

¹³ “(...) *intergenerational earnings elasticity as a measure for Mobility is questionable and should be considered cautiously.*” (BERMAN, 2017, p. 20)

¹⁴ “(...) *permutations between individual incomes, transfers or ranking changes which have no effect on the income distribution.*” (BERMAN, 2017, p. 20)

2.2. Visões teóricas no estudo da mobilidade social

No contexto do estudo da mobilidade social distinguem-se duas correntes teóricas principais: a da teoria liberal e a da análise de classes. A diferença entre as duas está, em primeira instância, na definição da importância de características individuais herdadas ou adquiridas na definição da hierarquia social utilizada.

Na teoria liberal encontramos a abordagem funcionalista da sociedade representada nos trabalhos de Davis e Moore (1945), Davis (1949) e Parsons (1954). A visão funcionalista baseia-se na suposição de que características conquistadas/adquiridas (*e.g.* educação e especialização) têm maior importância na colocação dos indivíduos nos espaços sociais em detrimento das características atribuídas/herdadas (*e.g.* histórico familiar, cor, religião)¹⁵. Essa linha teórica propõe que sociedades industriais dependem de mecanismos que incentivem a meritocracia para funcionar adequadamente, ou seja, afirmam que é ineficiente que a alocação dos agentes seja feita por critérios que não sejam o esforço individual e a capacidade.

Para os funcionalistas a alocação eficiente dos indivíduos na estrutura ocupacional resulta na diminuição de ocupações de cunho manual ou de rotina e, em conjunto com o progresso tecnológico, no aumento da mobilidade social (RIBEIRO, 2007). De acordo com Goldthorpe (2005), a perspectiva funcional consolidou nos primeiros estudos de mobilidade social uma visão ‘macro-para-macro’ do fenômeno, ou seja, a visão de que a mobilidade se dava de forma agregada e dependente do nível de desenvolvimento econômico e industrial da sociedade.

A teoria de análise de classes surge como uma antítese à teoria liberal, pois essa corrente teórica atribui o processo de hierarquização da sociedade ao contexto herdado pelos indivíduos através de características como *background* familiar, cor, religião, etc, e busca analisar de que forma os agentes enfrentam diferentes chances de mobilidade social. A maior crítica à teoria liberal nessa corrente teórica é realizada no trabalho de Erikson e Goldthorpe (1992), que adotam uma perspectiva classe-estrutural de análise onde as relações de classe desempenham um papel central na determinação das oportunidades às quais os indivíduos estão expostos.

¹⁵ Questão citada na seção anterior, ver Parsons (1951) e Roemer (1998).

2.3. Conceitos de mobilidade: taxas absolutas e relativas

Nos trabalhos seminais de Rogoff (1964) e Glass (1954) já havia a discussão da separação da mobilidade em dois componentes distintos a partir da observação das matrizes de mobilidade. Os autores sugerem a possibilidade de o fenômeno ter dois componentes diferentes: um agindo no nível individual e outro no nível agregado.

Conceitualmente esses componentes foram inicialmente caracterizados como: mobilidade circular e mobilidade estrutural. As diferenças observadas nas distribuições marginais das matrizes ocupacionais correspondiam à mobilidade estrutural, resultado de mudanças estruturais no mercado de trabalho e no caráter das ocupações entre as gerações de origem e destino. A mobilidade circular, também conhecida como de trocas, seria o resultado da subtração entre mobilidade total e a estrutural, e corresponderia à circulação de indivíduos nas vagas existentes (PASTORE, 1979).

Tal distinção foi utilizada amplamente até a terceira geração de estudos de mobilidade, onde havia uma grande preocupação dos pesquisadores na análise de quão distante ou próxima estava a sociedade do cenário de igualdade de oportunidades e quão aberta ela era. Tendo em vista as matrizes e medidas de mobilidade, foi desenvolvido o modelo de perfeita mobilidade, caracterizado pela independência estatística entre origem e destino (JONES, 1985).

Com o advento dos modelos log-lineares, porém, o interesse na utilização das matrizes de transição ganhou novo fôlego e, então, “um vigoroso debate se seguiu a respeito da relevância de conceitos tradicionais como mobilidade estrutural, de circulação ou de trocas, e sobre a especificação, interpretação e escolha entre modelos concorrentes”¹⁶ (JONES, 1985, p. 139). Sobel (1983), criticando o coeficiente de abertura desenvolvido por Yasuda (1964) e o trabalho de Hope (1982), afirma que a divisão da mobilidade social nos dois componentes, estrutural e circular, é inadequada no contexto de modelos log-lineares e deveria ser abandonada em estudos futuros. A crítica de Jones (1985) à proposição de Sobel (1983) defende a distinção macrossociológica clássica entre mobilidade circular e estrutural e alega, ainda, que a crítica feita ao coeficiente de abertura é despropositada.

¹⁶ “(...) a vigorous debate has ensued over the relevance of traditional concepts such as structural, circulation and exchange mobility, and over the specification of, interpretation of, and choice between competing models (...)” (JONES, 1985, p. 139)

A discussão sobre os componentes macrosociológicos da mobilidade ocorreu concomitantemente à validação de outras hipóteses conceituais a respeito da mobilidade. Dentro do contexto sobre padrões e tendências para mobilidade temos, inicialmente, a posição de Sorokin (1959), que acreditava que o fenômeno era caracterizado por ondas de mobilidade, ou seja, ciclos de maior ou menor mobilidade sem que houvesse, necessariamente, um padrão. Lipset e Zetterberg (1956; 1959), buscando testar as proposições de Sorokin (1959) a partir da análise da variação da mobilidade entre países ocidentais, propõem o que ficou conhecido como “a hipótese Lipset-Zetterberg”¹⁷, que afirma que países ocidentais apresentam, basicamente, os mesmos padrões de mobilidade (ERIKSON; GOLDTHORPE, 1992).

Valendo-se da melhora na disponibilidade de dados e no progresso metodológico, Featherman, Jones e Hauser (1975) reformularam a controversa hipótese Lipset-Zetterberg. A hipótese FJH afirma que as similaridades nas taxas de mobilidade entre os países não se dão no plano ‘fenotípico’, caracterizado pela mobilidade total observada, mas no plano ‘genotípico’ das chances relativas de mobilidade (ERIKSON; GOLDTHORPE, 1992; FEATHERMAN; LANCASTER JONES; HAUSER, 1975; RIBEIRO, 2007).

Erikson e Goldthorpe (1992), a partir da hipótese FJH, abordam a análise comparativa dos padrões de mobilidade social entre nações sob uma nova perspectiva. O trabalho, intitulado “*The Constant Flux*”, utiliza matrizes de mobilidade e técnicas log-lineares para propor “(...) uma distinção crucial entre taxas *absolutas* e *relativas* de mobilidade”¹⁸ (ERIKSON; GOLDTHORPE, 1992, p. 28). A partir de então houve um distanciamento da ideia original de que a mobilidade poderia ter dois componentes, um circular e outro estrutural, e a ampla aceitação de que ela pode ser expressa em termos de taxas absolutas e relativas (GOLDTHORPE, 2005).

A mobilidade social total observada corresponde ao agregado de oportunidades de mobilidade social e expressa todos os movimentos ocorridos em determinada hierarquia social, representando a taxa absoluta de mobilidade. De acordo com Ribeiro (2007), essa taxa pode ser explicada por dois efeitos distintos e complementares: o “efeito estrutural” e o “efeito da fluidez social”. O primeiro corresponde à mudança

¹⁷ Também referida como “hipótese Lipset-Bendix” em alguns trabalhos em referência ao trabalho de Lipset e Bendix (1959) (e.g. (FEATHERMAN; LANCASTER JONES; HAUSER, 1975)).

¹⁸ “(...) a crucial distinction between absolute and relative mobility rates.” (ERIKSON; GOLDTHORPE, 1992, p. 28)

entre as distribuições de origem e destino observada nas matrizes de mobilidade e é resultado de mudanças no mercado de trabalho, industrialização, mudanças sociais, etc. Já o efeito de fluidez social, expressa a desigualdade de oportunidades de mobilidade, ou seja, a diferença nas chances relativas de mobilidade enfrentada pelos indivíduos por motivos alheios às mudanças estruturais e é conhecido, na literatura, como sendo a taxa relativa de mobilidade social. Voltaremos a discutir a mobilidade relativa ao analisar os modelos log-lineares.

2.4. Estudo da mobilidade no Brasil

Assim como no resto do mundo, os estudos sobre a mobilidade social brasileira ficaram inicialmente limitados à literatura sociológica. O primeiro trabalho brasileiro relevante é o estudo dirigido por Hutchinson (1960) para a cidade de São Paulo. Neste trabalho seminal há a construção de uma escala de prestígio para as ocupações a partir da percepção dos entrevistados sobre a estrutura ocupacional e a análise da influência da educação no processo de mobilidade social na cidade de São Paulo.

Hutchinson (1960) destacou a importância do papel da educação na promoção de movimentos tanto ascendentes, quanto descendentes, mas concluiu que o maior responsável pela mobilidade econômica e ocupacional brasileira era o processo de desenvolvimento industrial pelo qual a sociedade estava passando no período. Para o autor, a mobilidade social é, em suma, uma mudança de posições dentro de uma estrutura mais ou menos estável onde, ainda que haja alterações de posições, a estrutura de status não chega a ser modificada. Esse caráter da mobilidade pode ser encontrado em sociedades que já passaram por processo de desenvolvimento e industrialização. No caso da sociedade brasileira da metade do século XX até o início do século XXI a análise é mais complexa, pois a estrutura de status passou por transformações contínuas e preservou características de estágios anteriores de desenvolvimento.

No trabalho de Foote e Hatt (1953) é apresentada pela primeira vez uma discussão sistemática sobre mobilidade social e desenvolvimento econômico. Os autores debatem o papel do estágio de desenvolvimento econômico da sociedade nas oportunidades de mobilidade. A atividade econômica passa da “produção agrícola (‘primária’) para a manufatureira (‘secundária’) e, subseqüentemente, para os serviços

“terciários” (PAYNE, 2017, p. 21)¹⁹, e isso parece criar oportunidades significativas para a mobilidade. As chances de mobilidade numa sociedade irão depender, inicialmente, das transformações nos setores econômicos e da natureza e dinâmica do mercado de trabalho. De forma que, conforme aponta Pastore (1979), a mobilidade pode ser entendida através de dois tipos de fatores: o estrutural (setores da economia, volume de empregos, caráter das ocupações) e o individual (educação, experiência, acesso à informação, características individuais como cor, situação de migrante ou não, etc).

Os estudos de mobilidade no Brasil se intensificaram pelo interesse dos sociólogos em entender os impactos das grandes transformações da estrutura social e ocupacional do país decorrentes dos processos de urbanização e industrialização que ocorreram, de maneira mais acelerada, a partir de 1940-1950. Jesus (2020) salienta que no período supracitado a população brasileira se reorganizou no território nacional através de intenso fluxo migratório, sendo identificada redução significativa da população rural, entre 1940 e 1970, passando de 69% para 44%. Nos anos 1980, 2000 e 2010 a participação da população urbana atinge 67%, 81% e 84,3%, conforme identificado pelos Censos Demográficos realizados nos respectivos anos (JESUS, 2020).

De acordo com Ribeiro (2007), a mudança na sociedade brasileira se deu rapidamente, ocorrendo em uma geração, o que dificultou que a oferta educacional acompanhasse a demanda pela educação gerada pelas novas ocupações. O autor acrescenta ao fator migratório, apresentado por Jannuzzi (1999; 2019) e Jesus (2020), outras três características relacionadas diretamente a esse processo e aos fenômenos de desigualdade e mobilidade social no país: (i) uma sociedade em que grande parte da população tem origens ligadas à ocupações rurais; (ii) um sistema educacional deficiente que se desenvolveu lentamente e levou décadas para se universalizar após o processo de industrialização e; (iii) a dicotomia presente no mercado de trabalho, que não completou sua transição para a modernidade, subsistindo setores modernos e tradicionais na economia.

Para Scalon (1999) essas características são resultado das políticas e instituições consolidadas a partir dos primeiros anos da república e do primeiro surto

¹⁹ “Economic activity transferred from agricultural (‘primary’) production to manufacturing (‘secondary’) and subsequently ‘tertiary’ services.” (PAYNE, 2017, p. 21)

industrial. Lopes (2008) enumera como fatores que contribuíram para as mudanças sociais significativas a mudança do eixo rural-urbano e constituição de mercados consumidores para produtos industriais; a abolição da escravatura em 1988 e a imigração de trabalhadores de origem europeia que resultaram numa remodelação da mão-de-obra; e as medidas de proteção da indústria nascente, apontadas por Ribeiro (2007), como uma forma de perpetuar institucionalmente desigualdades existentes na estrutura de *status* brasileira.

O trabalho de Pastore (1979), primeiro a analisar o fenômeno a nível nacional, deixa evidente a preocupação com as mudanças enfrentadas pela sociedade brasileira a partir dos processos citados. Este estudo utiliza a abordagem da primeira geração de estudos de mobilidade, valendo-se da análise de matrizes e medidas de mobilidade para os dados disponibilizados pela Pesquisa Nacional por Amostra de Domicílios (PNAD) de 1973 e faz uma tentativa de analisar o efeito de variáveis como educação do pai, educação e primeira ocupação do indivíduo no seu status atual através de análise de trajetórias.

A PNAD de 1973 também foi utilizada por Silva (1979), que chega a conclusões semelhantes às de Pastore (1979): um país marcado pela alta mobilidade social e movimentos de curta distância. Se nos trabalhos desenvolvidos nas décadas de 1970 e 1980 o emprego era a variável mais relevante, o trabalho subsequente de Pastore e Silva (2000) passa a evidenciar o ganho de importância do fator educacional no processo de mobilidade intergeracional brasileiro a partir da década de 1990. Essa mudança de importância acompanhou o processo de universalização da educação básica no país e demonstra um aumento na competitividade dentro da estrutura ocupacional da sociedade brasileira²⁰.

De meados do século XX até seu fim o Brasil foi marcado por uma desigualdade de renda persistente e a estrutura de *status* apresentou um estiramento, aumentando a distância entre ricos e pobres (OLIVEIRA, 2018). Pastore e Haller (1993) concluem que esse distanciamento entre as classes alta e baixa ocorreu em conjunto com o aumento da parcela da população alocada nas classes médias. Ainda assim, graças às mudanças na estrutura e dinâmica do mercado de trabalho da década de 1950, a sociedade brasileira apresenta alta flexibilidade.

²⁰ Para uma discussão aprofundada sobre mobilidade educacional no Brasil ver Ribeiro (2017) e Mahlmeister et al. (2019).

A mobilidade ascendente, durante todo o período entre o início da mensuração do fenômeno no país até o presente, sempre foi superior à imobilidade e à mobilidade descendente. No entanto, observa-se, após a década de 80, um aumento gradual destas medidas de mobilidade descendente e imobilidade (CAILLAUX, 1994; OLIVEIRA, 2018; OLIVEIRA; HOFFMANN, 2021; PASTORE; HALLER, 1993; PASTORE; SILVA, 2000; RIBEIRO, 2007; SCALON, 1999). O caráter da mobilidade intergeracional brasileira é de movimentos curtos, geralmente entre estratos vizinhos. Como esperado, a mobilidade estrutural, predominante até a década de 1990, passou a apresentar uma diminuição frente ao aumento dos movimentos circulares, sinalizando a maior importância de características individuais, como a educação, no processo de mudança de status (OLIVEIRA, 2018; PASTORE; SILVA, 2000; RIBEIRO, 2007; SCALON, 1999).

Ribeiro (2017), num trabalho que examina a evolução da mobilidade intergeracional brasileira a partir dos dados de todos os suplementos de mobilidade sócio-ocupacional já disponibilizados pela PNAD/IBGE (1973, 1982, 1988, 1996 e 2014) e da utilização do esquema CASMIN²¹ de classes proposto por Erickson, Goldthorpe e Portocarrero (1979) (ERICKSON; GOLDTHORPE, 1992) e modelos log-lineares, mostra que a mobilidade total passou de 54% em 1973 para 67% em 2014. Ao considerar três grandes classes, o autor observa que a mobilidade ascendente, que era 25% em 1973, apresentou elevação, chegando a 34% em 2014; a mobilidade descendente também cresceu, passando de 6% em 1973 para 9% em 1996 e 11% em 2014.

Oliveira e Hoffmann (2021) mostram, para os dados da PNAD de 2014 e utilizando matrizes e medidas de mobilidade, que o padrão de mobilidade intergeracional brasileira se mantém alto, com a dominância de movimentos ascendentes (49,5%), mas que houve um aumento na proporção de movimentos descendentes (16,3%). Quanto à dinâmica entre mobilidade circular e estrutural, demonstram que a mobilidade circular é maior do que a estrutural, superando-a em 37 p.p.

²¹ Este esquema será discutido na seção 3.2 Hierarquização das ocupações.

3. METODOLOGIA

3.1. Fonte e natureza dos dados

Para a análise da mobilidade sócio-ocupacional intergeracional é necessária a adoção de uma hierarquia sócio-ocupacional que congregue informações sobre as ocupações e o perfil dos indivíduos que as ocupam. A partir da escolha da escala ocupacional, se faz necessária a disponibilidade de dados a respeito dos indivíduos para que sejam identificadas sua origem e destino. Por origem entende-se o status referente ao ponto de partida do indivíduo; no caso intergeracional deve ser adotado um referencial anterior ao indivíduo como, por exemplo, o status ocupacional de seu pai. O destino refere-se ao status do indivíduo no momento da análise.

A maior dificuldade consiste, justamente, na segunda etapa: encontrar dados e pesquisas de população que incluam as variáveis a respeito dos pais dos indivíduos. O cenário ideal seria o acesso a dados administrativos com informações confiáveis sobre pais e filhos. Chetty et al. (2014) utilizam registros das declarações do imposto de renda federal dos Estados Unidos para mensurar a mobilidade de renda no país e estabelecem um marco nos estudos de mobilidade justamente por utilizar dados com menos vieses de *recall data* e falta de informação.

A disponibilidade de dados, além do referencial teórico adotado, define, portanto, a abordagem utilizada para investigar, mensurar e analisar os padrões e tendências da mobilidade intergeracional brasileira. A opção mais utilizada no contexto brasileiro é a Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

A PNAD, iniciada em 1967 com periodicidade trimestral e com a periodicidade alterada para anual no ano de 1971, constitui uma pesquisa que agrega informações básicas da população brasileira como renda, idade, escolaridade, ocupação e condições de habitação. Seu questionário é composto de perguntas fixas e uma seção suplementar com perguntas temáticas. Os suplementos das PNADS contam com perguntas relacionadas a tópicos especiais a respeito da população brasileira como por exemplo: segurança alimentar, acesso à internet, acesso a serviços de saúde, etc, conforme o ano. Ao longo do período em que foi realizada houve atualizações metodológicas na aplicação da pesquisa. A PNAD tradicional foi aplicada pela última vez em 2015, sendo

que a partir de 2012 foi feita a PNAD-Contínua (IBGE, 2021), retomando a sua periodicidade trimestral.

Nos anos de 1973, 1982, 1988, 1996 e 2014, o suplemento temático da PNAD intitulado “Mobilidade Sócio-ocupacional” contou com perguntas sobre a ocupação e escolaridade dos pais quando os respondentes tinham 15 anos. Esses dados foram utilizados pela grande maioria dos trabalhos realizados no Brasil sobre a mobilidade intergeracional dos quais destacamos: Pastore (1979), Scalon (1999), Pastore e Silva (2000), Ribeiro (2007), Ribeiro (2017), Oliveira (2018), Jannuzzi et al. (2019) e Oliveira e Hoffmann (2021). A grande vantagem da utilização dos dados fornecidos pelas edições supracitadas consiste na pergunta que indica a ocupação do pai quando o indivíduo tinha 15 anos de idade, possibilitando a identificação de um ponto de “origem” para a análise de mobilidade intergeracional.

Adotou-se, para a composição da amostra geral deste estudo, um recorte de idade que seleciona todos os indivíduos com idade entre 16 e 65 anos que responderam à pergunta sobre a ocupação paterna na PNAD 2014. Essa seleção corresponde a uma análise da mobilidade intergeracional no período de 1964 até 2014, uma vez que para indivíduos com 65 anos em 2014 a origem, ou seja, a ocupação do pai quando o indivíduo tinha 15 anos, está no ano de 1964.

Visando considerar a dinâmica da mobilidade sócio-ocupacional durante as cinco décadas que compreendem as relações origem-destino na amostra e partindo da hipótese de que as associações entre origem e destino se modificam ao longo do tempo, realizou-se a divisão da amostra por coortes de idade. A seleção das coortes foi realizada considerando cinco faixas de idade de dez anos cada: 16 a 25 anos, 26 a 35 anos, 36 a 45 anos, 46 a 55 anos e 56 a 65 anos.

Deste modo, o período de análise, que compreende 50 anos, pode ser subdividido em cinco décadas, que correspondem ao período em que a ocupação do pai do indivíduo é considerada. Na primeira coorte de idade, em que são considerados os indivíduos de 16 a 25 anos, por exemplo, trabalha-se com a década de 2004 até 2013.

A presente análise da mobilidade intergeracional considera, além das variáveis referentes ao status socioeconômico de origem e destino, a variável que indica a década que inclui o ano da informação sobre a ocupação do pai: 1964 a 1973, 1974 a 1983, 1984 a 1993, 1994 a 2003 ou 2004 a 2013.

3.2. Hierarquização das ocupações

Ao realizar uma investigação sobre a mobilidade sócio-ocupacional é necessária a definição do arcabouço conceitual a ser utilizado. De acordo com Erikson e Goldthorpe (1992), as duas tradições principais são a perspectiva de estrutura de classes e a de hierarquia social. Na primeira, a mobilidade é vista como a movimentação dos indivíduos entre “posições sociais que são identificadas em termos de relações entre mercado de trabalho e unidades produtivas”²² (ERIKSON; GOLDTHORPE, 1992, p. 29), já na abordagem de hierarquia social os movimentos ocorrem entre grupos ou agregados sociais hierarquizados de acordo com algum critério como renda, prestígio, status, etc.

Uma das perspectivas mais utilizadas na construção de estratificações ocupacionais na literatura da mobilidade social baseia-se na perspectiva classe-estrutural, proposta e elaborada por Erikson, Goldthorpe e Portocarrero (1979). A perspectiva classe-estrutural postula que a estratificação ocupacional deve agregar ocupações em que os trabalhadores que as desempenham estejam envolvidos em mercados e condições de trabalho similares e tem como esquema mais utilizado o de Erikson-Goldthorpe-Portocarrero, também conhecido como esquema EGP.

Erikson e Goldthorpe (1992) argumentam que a escolha da abordagem classe-estrutural se justifica por ser a mais adequada quando o interesse é a análise de influências estruturais sobre tendências de mobilidade social. O esquema EGP diferencia as ocupações em classes de acordo com princípios derivados dos trabalhos dos sociólogos Marx e Max Weber, visando identificar aspectos do mercado de trabalho e das unidades produtivas através das relações de emprego (ERICKSON; GOLDTHORPE; PORTOCARRERO, 1979).

No presente trabalho, dado o caráter dos dados disponíveis na PNAD 2014, que não permitem mensurar aspectos subjetivos da estrutura ocupacional e a visão funcionalista adotada para permear as discussões e análises de mobilidade, optou-se pela utilização da classificação proposta por Oliveira (2018) e Oliveira e Hoffmann (2021). Adota-se, portanto, o contexto da hierarquia social, onde os movimentos são vistos a partir do deslocamento entre *status*, no caso, entre ocupações que são ordenadas a partir de *status* atribuídos.

²² “(...) *social positions that are identified in terms of relationships within labour markets and production units*” (ERIKSON; GOLDTHORPE, 1992, p. 29)

Oliveira e Hoffmann (2021) propõem um novo método para a estratificação das ocupações brasileiras com base na metodologia de Silva (1973), empregada nos trabalhos de Pastore (1979), Pastore e Silva (2000) e Ueda (2001). O procedimento utiliza critérios estatísticos objetivos para a mitigação de arbitrariedades na delimitação dos estratos sócio-ocupacionais (OLIVEIRA; HOFFMANN, 2021), evitando a discussão sobre prestígio social e análise de classes.

Essa hierarquização sócio-ocupacional relaciona renda, ocupação e escolaridade dos indivíduos controlando efeitos da idade e consiste em seis etapas: (i) seleção da amostra e padronização das variáveis referentes ao rendimento mensal da atividade principal e ao número de horas trabalhadas para 40 horas de trabalho semanais; (ii) cálculo do valor monetário do nível de escolaridade do indivíduo por meio da estimação de equações de rendimento controlando efeitos das faixas de idade e com a utilização de ajuste poligonal; (iii) definição do *Status Socioeconômico Individual* (SSEI), que consiste na média aritmética do rendimento do indivíduo na sua ocupação e do seu rendimento esperado em função da sua escolaridade; (iv) definição do *Status Socioeconômico* de cada ocupação (SSE), obtido através da média dos SSEI dos indivíduos que apresentam a mesma ocupação; (v) padronização dos escores de SSE para o intervalo de 0 a 100; e, (vi) definição dos limites dos estratos sócio-ocupacionais através do método de minimização de desigualdade intra grupos e maximização da desigualdade entre grupos (AGHEVLI; MEHRAN, 1981; DAVIES; SHORROCKS, 1989; HOFFMANN, 2005)²³.

3.3. Tabelas de contingência

As matrizes de mobilidade social são tabelas de contingência (ou classificação cruzada), que relacionam duas ou mais variáveis categóricas que apresentam escalas nominais ou ordinais e englobam todas as possíveis combinações entre status de pais e filhos, mostrando as frequências de ocorrência das categorias entre as observações da amostra. Neste estudo são adotados estratos sócio-ocupacionais que constituem seis categorias ordinais, ou seja, que apresentam um ordenamento, e vão de 1 (alto) a 6 (baixo-inferior), além de uma variável de nível, com cinco categorias nominais, representando cinco décadas distintas. As variáveis relacionadas no estudo da

²³ Para uma descrição minuciosa do método, consultar Oliveira (2018) e Oliveira e Hoffmann (2021).

mobilidade sócio-ocupacional intergeracional são Origem (ocupação do pai do indivíduo quando este tinha 15 anos) e Destino (ocupação do indivíduo no momento em que respondeu a pesquisa).

Tomemos a matriz de mobilidade que relaciona as variáveis origem e destino com categorias representando os estratos sócio-ocupacionais. A linha Origem (O_i) com $i = 1, 2, \dots, I$ estratos sócio-ocupacionais, e a coluna Destino (D_j) com $j = 1, 2, \dots, J$ estratos sócio-ocupacionais. A tabela de contingência, representando a matriz de transição de *status* (matriz de mobilidade), tem dimensão $I \times J$. No caso da mobilidade, o número de estratos sócio-ocupacionais de origem e destino é o mesmo, ou seja, $I = J$.

Tabela 1. Tabela de contingência $I \times J$

		Destino				
		1	2	...	J	Total
Origem	1	f_{11}	f_{12}	...	f_{1j}	f_{1+}
	2	f_{21}	f_{22}	...	f_{2j}	f_{2+}
	·	·	·		·	·
	·	·	·		·	·
	·	·	·		·	·
	I	f_{I1}	f_{I2}	...	f_{IJ}	f_{I+}
Total	f_{+1}	f_{+2}	...	f_{+J}	f_{++}	

Fonte: Elaborado pelo autor.

Na tabela 1 temos uma tabela de contingência mostrando uma classificação cruzada $I \times J$, onde f_{ij} indica a frequência na célula pertencente à linha (Origem) i e à coluna (Destino) j . Temos ainda as distribuições de frequências marginais obtidas através das equações (3) e (4) e o total da amostra f_{++} representado na equação (5), sendo N o total de observações da amostra.

$$f_{i+} = \sum_{j=1}^J f_{ij} \quad (3)$$

$$f_{+j} = \sum_{i=1}^I f_{ij} \quad (4)$$

$$f_{++} = \sum_{i=1}^I \sum_{j=1}^J f_{ij} = N \quad (5)$$

Silva (1990) destaca que um aspecto importante ao analisar uma tabela de contingência diz respeito a como os dados que ela apresenta foram obtidos, ou seja, ao modelo amostral utilizado. Ao utilizar uma amostra com N observações, onde os elementos são classificados a partir da combinação entre as categorias das variáveis de

interesse, supõe-se, geralmente, que ela representa uma população com distribuição multinomial. “Trata-se de uma suposição relativamente fraca, uma vez que envolve apenas a suposição de que os dados seguem uma estrutura probabilística.” (SILVA, 1990, p. 16).

O ponto de partida para a análise da tabela de contingência é o modelo nulo, onde não há qualquer influência das variáveis linha e coluna sobre as frequências observadas. A partir deste modelo, são propostos modelos onde há a influência das variáveis linha e coluna caracterizando efeitos: principais, quando tanto a variável linha quanto a variável coluna influenciam a frequência; de interação, quando há associação entre as variáveis; de covariável, em que uma terceira variável, que pode ser adotada como controle, apresenta influência; e especiais, como os advindos da escala utilizada. No caso do estudo da mobilidade social são de particular interesse os efeitos principais e de interação.

O modelo nulo, o mais parcimonioso possível, propõe não haver efeito ou associação entre as variáveis e representa, no contexto da mobilidade social, o cenário de mobilidade perfeita, onde as movimentações dos indivíduos entre os estratos são aleatórias e todos têm as mesmas probabilidades de estar em um ou outro estrato. Neste caso, a probabilidade e a frequência esperadas da célula ij são obtidas, respectivamente, através de:

$$\hat{\pi}_{ij} = \frac{1}{IJ} \quad (6)$$

$$\hat{f}_{ij} = \hat{\pi}_{ij}N = \frac{N}{IJ} \quad (7)$$

A estatística $z_{ij} = \frac{(f_{ij} - \hat{f}_{ij})}{\sqrt{\hat{f}_{ij}}}$ pode ser utilizada para comparar as frequências

observadas e estimadas e analisar a aderência do modelo aos dados. Caso a estatística z indique que as discrepâncias entre os dados e as estimativas das frequências são altas, devemos passar a considerar os efeitos de modelos hierarquicamente superiores.

Considerando os modelos que admitem efeitos e associações entre variáveis, temos, seguindo a hierarquia dos modelos, aquele que considera a existência de efeito linha ou efeito coluna. Neste modelo há o pressuposto de que o conhecimento das probabilidades marginais da linha, para o primeiro, ou da coluna, para o segundo, é suficiente para explicar a distribuição de frequências da tabela de contingência em

questão (VON EYE; MUN, 2013). As equações (8) e (9) se referem às estimativas de probabilidade e frequência para o modelo de efeito linha e as (10) e (11) para o modelo de efeito coluna.

$$\hat{\pi}_{ij} = \frac{\pi_{i+}}{J} \quad (8)$$

$$\hat{f}_{ij} = \frac{f_{i+}}{J} \quad (9)$$

$$\hat{\pi}_{ij} = \frac{\pi_{+j}}{I} \quad (10)$$

$$\hat{f}_{ij} = \frac{f_{+j}}{I} \quad (11)$$

O modelo seguinte, acima dos modelos de efeito linha e efeito coluna, na hierarquia de efeitos de uma tabela de contingência, é o modelo de efeitos principais. O modelo considera efeitos linha e coluna e caracteriza um modelo de independência, pois nenhuma hipótese é feita sobre uma possível interação entre as variáveis linha e coluna (VON EYE; MUN, 2013).

Uma das primeiras abordagens para a investigação da relação entre variáveis por meio de tabelas de contingência envolve o cálculo e comparação das distribuições percentuais ao longo das linhas e colunas. O modelo de efeitos principais baseia-se na noção de independência estatística entre duas variáveis aleatórias que diz que duas variáveis são ditas independentes se para qualquer valor de uma variável a probabilidade condicional da outra é igual à sua probabilidade marginal (SILVA, 1990). Essa definição implica que, caso seja confirmada a independência entre as duas variáveis, temos, para todo $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, J$, a equação (12) em que $\hat{\pi}_{ij}$ representa a probabilidade esperada da célula (i,j) e π_{i+} e π_{+j} são as probabilidades marginais das categorias das variáveis O_i e D_j .

$$\hat{\pi}_{ij} = \pi_{i+}\pi_{+j} \quad (12)$$

As frequências esperadas podem ser calculadas a partir de (13).

$$\hat{f}_{ij} = \frac{f_{i+}f_{+j}}{N} = N\pi_{i+}\pi_{+j} \quad (13)$$

Se esse modelo for rejeitado, as duas variáveis apresentam associação entre si, “essa interpretação é possível porque a associação (interação) da variável linha com a variável coluna é o único efeito não considerado na estimação das frequências”²⁴. (VON EYE; MUN, 2013, p. 17)

O último modelo de interesse, o de interação, leva em conta um componente adicional que corresponde à interação entre a variável linha e a variável coluna e será apresentado na próxima subseção.

3.4. Modelos log-lineares

Uma vez que a mobilidade total considera tanto o efeito de mudanças estruturais quanto o efeito de mudanças nas chances relativas de mobilidade, a mobilidade relativa será, então, caracterizada pela associação estatística entre um ponto de origem e um ponto de destino quando desprezamos o componente estrutural da mobilidade (caracterizado pelas distribuições marginais das tabelas de contingência).

A classe de modelos log-lineares proposta por Goodman (1979b) pode ser utilizada para a mensuração da mobilidade relativa. Os modelos log-lineares fornecem como medida de associação as razões de chances (*odds ratios*)²⁵ que, no contexto de matrizes de mobilidade, avaliam as possibilidades de um indivíduo se movimentar na estrutura social, entre os diferentes estratos, a partir do seu ponto de origem. “Quanto menor a associação entre origem e destino, menos a origem determina a chance de ascensão” (RIBEIRO, 2007, p. 189).

Orientados para a análise de associação em classificações cruzadas de variáveis categóricas ordenadas, geralmente representadas por tabelas de contingência, os modelos log-lineares incluem “o modelo de associação nula, o modelo de associação uniforme e modelos que descrevem possíveis efeitos de linhas e/ou colunas na associação.” (GOODMAN, 1979b, p. 537)²⁶ e permitem uma análise da associação

²⁴ “This interpretation is possible because the association (interaction) of the row variable with the column variable is the only effect that is not taken into account when the expected cell frequencies are estimated.” (VON EYE; MUN, 2013, p. 17)

²⁵ Usaremos “razão de chances” como tradução de odds ratio. Se p_0 é a probabilidade de ocorrência de um evento em dada situação e p_1 é a probabilidade associada ao mesmo evento em uma situação distinta, e definindo $q_0 = 1 - p_0$ e $q_1 = 1 - p_1$, a razão de chances é $\frac{p_1/q_1}{p_0/q_0}$.

²⁶ “the null association model, the uniform association model, and models that describe the possible effects of the rows and/or columns on the association.” (GOODMAN, 1979b, p. 537)

entre as variáveis em suas diferentes categorias, semelhante à análise de variância com dois fatores. A especificação das categorias afeta os resultados obtidos na utilização destes modelos, que podem se alterar devido, por exemplo, a alterações como redução ou aumento de categorias.

A partir das relações probabilísticas possíveis dentro de uma tabela de contingência, descritas na subseção anterior, podemos adotar a notação log-linear que oferece, de acordo com von Eye e Mun (2013), algumas vantagens como: (i) facilidade de identificação dos modelos como pertencentes à classe de modelos lineares generalizados; (ii) semelhança dos modelos aos modelos de análise de variância; (iii) a presença de modelos aditivos que facilitam a interpretação dos coeficientes; e, (iv) a relação dos modelos log-lineares com as razões de chances.

3.4.1. Modelos log-lineares para análises bivariadas

Considerando uma tabela de contingência que relaciona duas variáveis, O e D , e classifica uma amostra de N indivíduos em I e J categorias, parte-se de um modelo que considera apenas os efeitos individuais de cada variável, o modelo de efeitos principais (ou de independência). A partir do logaritmo natural da equação (13), temos:

$$\ln \hat{f}_{ij} = \ln N + \ln \pi_{i+} + \ln \pi_{+j} \quad (14)$$

A equação (15), uma reparametrização da equação (14), representa a forma log-linear do modelo de independência, onde o logaritmo natural da frequência esperada da célula (mobilidade) é igual à soma do logaritmo natural do tamanho da amostra (λ), o logaritmo natural da probabilidade de apresentar origem na i -ésima categoria (λ_i^O) e o logaritmo natural da probabilidade de ter como destino a j -ésima categoria (λ_j^D).

$$\ln \hat{f}_{ij} = \lambda + \lambda_i^O + \lambda_j^D \quad (15)$$

Na tabela 2 são apresentados os possíveis modelos de classificação cruzada na forma log-linear para o caso bivariado e seus respectivos graus de liberdade do resíduo.

Tabela 2. Modelos log-lineares para tabelas de contingência bivariadas e os correspondentes graus de liberdade (g) do resíduo

Modelo	Equação do modelo	Graus de liberdade (g)
Nulo	$\ln \hat{f}_{ij} = \ln N = \lambda$	$g = IJ - 1$
Efeito linha	$\ln \hat{f}_{ij} = \ln N + \ln \pi_{i+} = \lambda + \lambda_i^O$	$g = IJ - 1 - (I - 1)$
Efeito coluna	$\ln \hat{f}_{ij} = \ln N + \ln \pi_{+j} = \lambda + \lambda_j^D$	$g = IJ - 1 - (J - 1)$
Efeitos principais	$\ln \hat{f}_{ij} = \ln N + \ln \pi_{i+} + \ln \pi_{+j} = \lambda + \lambda_i^O + \lambda_j^D$	$g = IJ - 1 - (I - 1) - (J - 1)$
Modelo saturado	$\ln \hat{f}_{ij} = \ln N + \ln \pi_{i+} + \ln \pi_{+j} + \ln \pi_{ij}$ $= \lambda + \lambda_i^O + \lambda_j^D + \lambda_{ij}^{OD}$	$g = IJ - 1 - (I - 1) - (J - 1)$ $- [(I - 1)(J - 1)] = 0$

Fonte: Elaborado pelo autor.

Nota: IJ representa o número de células, I o número de linhas e J o número de colunas.

O modelo saturado leva em consideração a interação entre as variáveis linha (Origem) e coluna (Destino) e, segundo Silva (1990), tem a propriedade de se ajustar perfeitamente aos dados. O modelo saturado é assim denominado por apresentar zero graus de liberdade no resíduo.

Tomando como exemplo a Tabela 3, a seguir, que representa a tabela de contingência 2×2 para duas variáveis com duas categorias cada, já na forma log-linear, examinemos a aplicação do modelo saturado, considerando $\ln \hat{f}_{ij} = \mu_{ij}$.

Tabela 3. Tabela de contingência para o caso bivariado dicotômico

	Variável Coluna (D)			Média
	$j = 1$	$j = 2$		
Variável	$i = 1$	$\mu_{11} = \lambda + \lambda_1^O + \lambda_1^D + \lambda_{11}^{OD}$	$\mu_{12} = \lambda + \lambda_1^O + \lambda_2^D + \lambda_{12}^{OD}$	$\mu_{1+} = \lambda + \lambda_1^O$
Linha (O)	$i = 2$	$\mu_{21} = \lambda + \lambda_2^O + \lambda_1^D + \lambda_{21}^{OD}$	$\mu_{22} = \lambda + \lambda_2^O + \lambda_2^D + \lambda_{22}^{OD}$	$\mu_{2+} = \lambda + \lambda_2^O$
	Média	$\mu_{+1} = \lambda + \lambda_1^D$	$\mu_{+2} = \lambda + \lambda_2^D$	$\mu_{++} = \lambda$

Fonte: Elaborado pelo autor.

Temos neste caso quatro equações e nove incógnitas. Para a estimação das frequências esperadas μ_{ij} é necessário impor restrições aos parâmetros. A abordagem mais comum é a de soma zero, discutida em von Eye e Mun (2013), onde são impostas restrições estabelecendo que os parâmetros referentes a um tipo de efeito tenham soma igual a zero²⁷. As restrições necessárias podem ser expressas pela equação (16), onde $I = 2$ e $J = 2$, no caso da tabela 2×2 .

²⁷ Para detalhamento da abordagem citada e discussão sobre outras abordagens possíveis consultar von Eye e Mun (2013).

$$\sum_{i=1}^I \lambda_i^O = \sum_{j=1}^J \lambda_j^D = \sum_i \lambda_{ij}^{OD} = \sum_j \lambda_{ij}^{OD} = 0 \quad (16)$$

As restrições descritas pela equação (16) implicam, na tabela de contingência bivariada com duas categorias, que:

$$\lambda_1^O = -\lambda_2^O \quad (17)$$

$$\lambda_1^D = -\lambda_2^D \quad (18)$$

$$\lambda_{11}^{OD} = \lambda_{22}^{OD} = -\lambda_{12}^{OD} = -\lambda_{21}^{OD} \quad (19)$$

Os valores marginais μ_{i+} e μ_{+j} serão as médias aritméticas das linhas e colunas. Por exemplo, para μ_{1+} temos que:

$$\mu_{1+} = \frac{(\lambda + \lambda_1^O + \lambda_1^D + \lambda_{11}^{OD}) + (\lambda + \lambda_1^O + \lambda_2^D + \lambda_{12}^{OD})}{2} \quad (20)$$

Levando em consideração as restrições descritas em (18) e (19), reescrevemos (20):

$$\mu_{1+} = \frac{2\lambda + 2\lambda_1^O + \lambda_2^D - \lambda_2^D + \lambda_{12}^{OD} - \lambda_{12}^{OD}}{2} = \frac{2(\lambda + \lambda_1^O)}{2} = \lambda + \lambda_1^O \quad (21)$$

Generalizando, temos:

$$\mu_{i+} = \sum_{j=1}^J \frac{\mu_{ij}}{J} \quad (22)$$

$$\mu_{+j} = \sum_{i=1}^I \frac{\mu_{ij}}{I} \quad (23)$$

$$\mu_{++} = \sum_{i=1}^I \sum_{j=1}^J \frac{\mu_{ij}}{IJ} \quad (24)$$

Os valores estimados dos parâmetros podem ser calculados diretamente através das equações de (25) a (28) e mostram que os parâmetros λ representam, segundo Silva (1990), “acréscimos (ou decréscimos) sobre a média geral atribuíveis ao fato de se estar localizado numa dada linha i ou coluna j ”, ou seja, o efeito das categorias sobre a frequência de cada célula.

$$\hat{\lambda} = \mu_{++} \quad (25)$$

$$\hat{\lambda}_i^O = \mu_{i+} - \lambda \quad (26)$$

$$\hat{\lambda}_j^D = \mu_{+j} - \lambda \quad (27)$$

$$\hat{\lambda}_{ij}^{OD} = \mu_{ij} - \mu_{i+} - \mu_{+j} + \mu_{++} = \mu_{ij} - (\hat{\lambda} + \hat{\lambda}_i^O + \hat{\lambda}_j^D) \quad (28)$$

O modelo saturado, por se ajustar perfeitamente aos dados, é pouco interessante quando desejamos analisar a influência da origem sobre as chances relativas de mobilidade, mas é importante como um ponto de partida para identificação de modelos mais parcimoniosos que descrevam as relações existentes entre as variáveis da tabela de contingência.

Os modelos log-lineares descritos na tabela 3 representam, no caso bivariado, a classe de modelos log-lineares hierárquicos. Essa classe de modelos consiste nos modelos log-lineares onde a presença de “termos de ordem superior implicam todos os possíveis termos de ordens inferiores das variáveis incluídas nos termos de ordem superior.” (VON EYE; MUN, 2013, p. 56)²⁸. Isso equivale a dizer que no caso de uma análise trivariada, com variáveis A, B e C, por exemplo, ao incluir no modelo o termo λ^{ABC} , para que seja um modelo hierárquico é necessário incluir todos os termos de ordens inferiores: λ^{AB} , λ^{AC} , λ^{BC} , λ^A , λ^B e λ^C .

O ajustamento de modelos log-lineares pode ser avaliado a partir da estatística de divergência de potência expressa pela equação a seguir:

$$I(\delta) = \frac{2}{\delta(\delta+1)} \sum_i \sum_j f_{ij} \left[\left(\frac{f_{ij}}{\hat{f}_{ij}} \right)^\delta - 1 \right] \quad (29)$$

onde o parâmetro δ é um número real ($-\infty < \delta < \infty$); f_{ij} é a frequência observada da célula ij ; e \hat{f}_{ij} é a frequência esperada da célula ij .

A estatística $I(\delta)$ é importante porque, tomando valores específicos de δ , é possível demonstrar que algumas medidas de ajustamento muito utilizadas são casos especiais dela. Dos cinco casos especiais demonstrados por von Eye e Mun (2013),

²⁸ “higher order terms imply all possible lower order terms of the variables included in the higher order terms.” (VON EYE; MUN, 2013, p. 56)

destacamos dois, quando δ é igual a 1 e quando tende a 0. Para o caso em que $\delta = 1$, a estatística $I(\delta)$ é igual ao χ^2 de Pearson e, no caso particular em que $\delta \rightarrow 0$, a estatística tende à razão de verossimilhança (L_m^2). As duas relações são expressas nas equações (30) e (31).

$$I(1) = \chi^2 = \sum_i \sum_j \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}} \quad (30)$$

$$I(\delta \rightarrow 0) \rightarrow L_m^2 = 2 \sum_i \sum_j f_{ij} \ln \left(\frac{f_{ij}}{\hat{f}_{ij}} \right) \quad (31)$$

As duas medidas, o χ^2 de Pearson e a Razão de Verossimilhança, assim como a estatística $I(\delta)$, são, sob as condições de regularidade, assintótica e identicamente distribuídas como uma distribuição qui-quadrado (χ^2) com g graus de liberdade, média g , variância $2g$ e assimetria $\sqrt{\frac{8}{g}}$, sendo g o número de graus de liberdade do resíduo do modelo adotado.

Apesar de o teste de qui-quadrado de Pearson (χ^2 de Pearson) geralmente se aproximar mais da distribuição de qui-quadrado do que a razão de máxima verossimilhança, essa segunda apresenta melhores características de decomposição. Segundo Scalon (1999), L_m^2 apresenta propriedades aditivas que possibilitam seu uso para a comparação do ajustamento entre modelos sucessivos, sendo, portanto, mais utilizado para a análise de ajustamento de modelos log-lineares.

Como coloca Scalon (1999), testes de significância para amostras muito grandes tendem a apresentar valores elevados, indicando a rejeição da hipótese nula mesmo que a divergência seja irrelevante. Por esta razão, são utilizados outros testes como os critérios de informação, a contar: o critério de informação de Akaike (AIC) e o critério de informação Bayesiano (BIC).

A estatística BIC pode ser obtida através da estatística de qui-quadrado da razão de verossimilhança (L_m^2), como expresso na equação (32), a seguir, onde N é o tamanho da amostra e g os graus de liberdade do modelo.

$$BIC = L_m^2 - [g \times \ln N] \quad (32)$$

3.4.2. Modelos log-lineares para análises trivariadas

Os modelos log-lineares para análises envolvendo três variáveis ($I \times J \times K$) representam extensões dos modelos apresentados na subseção anterior, podendo ser generalizados para um número maior de variáveis. Esses modelos, no entanto, apresentam padrões de associação e aspectos adicionais que necessitam de atenção (SILVA, 1990).

Na exposição do modelo log-linear para o caso bivariado considerou-se duas variáveis qualitativas ordinais: origem e destino; para o caso em que consideramos três variáveis adicionamos uma variável para distinguir as coortes de idade, isto é, as faixas de idade dentre os indivíduos da amostra.

Partindo-se do modelo que descreve todas as associações possíveis entre as três variáveis: origem (O), destino (D) e coorte de idade (C), com categorias $i = 1, 2, \dots, I$, $j = 1, 2, \dots, J$ e $k = 1, 2, \dots, K$, representado na equação (33), a seguir, temos o modelo saturado (ODC), ou de terceira ordem. Assim como no caso bivariado, este modelo se ajusta perfeitamente aos dados e as frequências esperadas são iguais às observadas.

$$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ij}^{OD} + \lambda_{ik}^{OC} + \lambda_{jk}^{DC} + \lambda_{ijk}^{ODC} \quad (33)$$

Usualmente, modelos de ordens menores são preferíveis ao modelo saturado porque, como coloca Agresti (2018), suavizam os dados da amostra e são mais fáceis de interpretar. Partindo para os modelos de ordens inferiores na classe de modelos log-lineares hierárquicos, temos o modelo de associação homogênea (OD , OC , DC), ou modelo de associação 2 a 2, representado a seguir.

$$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ij}^{OD} + \lambda_{ik}^{OC} + \lambda_{jk}^{DC} \quad (34)$$

O modelo representado pela equação (34) também é conhecido como modelo condicionalmente dependente e descreve associações entre todas as variáveis onde a associação entre pares de variáveis será independente dos níveis de uma terceira variável, aquela não considerada na associação. Tomemos, por exemplo, a associação entre as variáveis origem (O) e destino (D) representada, na equação (34), pelo termo λ_{ij}^{OD} ; neste caso a associação entre as duas variáveis será a mesma em todas as faixas de

idade representadas pela variável C , ou seja, a associação entre os estratos sócio-ocupacionais de origem e destino será a mesma, independente da década observada.

Um outro modelo possível é o de independência condicional (OC, DC), no qual se admite que, quando controlamos por uma terceira variável, não há associação entre as outras duas.

$$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ik}^{OC} + \lambda_{jk}^{DC} \quad (35)$$

No modelo de independência condicional descrito pela equação (35) não há um termo de associação entre origem e destino, indicando que essas duas variáveis são condicionalmente independentes dada a coorte de idade. O mesmo raciocínio poderia ser aplicado para o modelo (OD, OC) ou (OD, DC) que representam a independência condicional entre destino e coorte de idade e origem e coorte de idade, respectivamente.

No caso em que apenas um termo de segunda ordem, λ_{jk}^{DC} por exemplo, é diferente de zero, temos o modelo de independência marginal, ou parcial. Este modelo é representado pela equação (36), a seguir.

$$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{jk}^{DC} \quad (36)$$

O modelo de independência marginal descrito pela equação (36) estabelece que a variável origem é independente das variáveis destino e coorte de idade, mas essas duas últimas possuem associação, ou seja, são dependentes uma da outra. Podemos ter, para o caso trivariado, os seguintes modelos de independência marginal: (O, DC), (D, OC) e (C, OD).

Por fim, há, ainda, o caso em que todos os parâmetros de associação são iguais a zero e, portanto, configura-se a independência entre as variáveis analisadas. Este modelo, expresso pela equação (37), não possui termos de associação e é chamado de modelo de independência mútua ou múltipla (O, D, C). Neste modelo, cada par de variáveis é mutuamente e condicionalmente independente.

$$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C \quad (37)$$

O número de graus de liberdade (g) para um modelo trivariado é obtido de forma similar ao modelo bivariado. No quadro a seguir são representados nove modelos considerando os tipos de associação descritos anteriormente e o cálculo dos seus respectivos graus de liberdade.

Tabela 4. Modelos log-lineares para tabelas de contingência trivariadas e graus de liberdade correspondentes

Modelo	Equação do modelo	Graus de liberdade (g)
Independência Mútua	$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C$	$g = IJK - I - J - K + 2$
Independência Marginal (D)	$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ik}^{OC}$	$g = (J - 1)(IK - 1)$
Independência Marginal (O)	$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{jk}^{DC}$	$g = (I - 1)(JK - 1)$
Independência Marginal (C)	$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ij}^{OD}$	$g = (K - 1)(IJ - 1)$
Independência Condicional (D)	$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ij}^{OD} + \lambda_{jk}^{DC}$	$g = J(I - 1)(K - 1)$
Independência Condicional (O)	$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ij}^{OD} + \lambda_{ik}^{OC}$	$g = I(J - 1)(K - 1)$
Independência Condicional (C)	$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ik}^{OC} + \lambda_{jk}^{DC}$	$g = K(I - 1)(J - 1)$
Associação Homogênea	$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ij}^{OD} + \lambda_{ik}^{OC} + \lambda_{jk}^{DC}$	$g = (I - 1)(J - 1)(K - 1)$
Modelo saturado	$\ln \hat{f}_{ijk} = \lambda + \lambda_i^O + \lambda_j^D + \lambda_k^C + \lambda_{ij}^{OD} + \lambda_{ik}^{OC} + \lambda_{jk}^{DC} + \lambda_{ijk}^{ODC}$	$g = 0$

Fonte: Elaborado pelo autor.

Nota: IJK representa o número de células, I o número de linhas, J o número de colunas e K o número de níveis.

A seleção dos modelos é realizada, como no caso bivariado, a partir do resultado da estatística de qui-quadrado de Wald, ou razão de verossimilhança (L_m^2) e do critério de informação Bayesiano (BIC).

3.5. Razões de chances

Uma outra maneira de expressar e analisar as relações entre as variáveis e frequências de uma tabela de contingência é através do estudo das chances e razões de chances. Esse tipo de análise envolve a especificação de efeitos que, por hipótese,

podem explicar a distribuição de frequência em determinada classificação cruzada, podendo essa ser bivariada ou multivariada (VON EYE; MUN, 2013).

De acordo com von Eye e Mun (2013), há diversas maneiras de expressar esses efeitos, mas a mais intuitiva e de fácil interpretação é a de efeitos do tipo contrastes, que são aqueles que envolvem, na análise de cada variável, a comparação entre duas de suas categorias. As chances envolvem a análise de apenas uma variável, enquanto as razões de chances avaliam a interação entre duas variáveis.

Rudas (1998) destaca que a utilização de razões de chances como medidas de associação entre variáveis é preferível a outras devido a sua importante propriedade de ser independente de variações nas frequências marginais da tabela de contingência, sendo dependente apenas da associação entre as variáveis envolvidas.

3.5.1. Razões de chances para tabelas de contingência 2×2

Consideremos inicialmente a variável destino com duas categorias ordinais representando a posição atual do indivíduo numa sociedade com dois estratos: Baixo (1) e Alto (2). Na Tabela 5 é proposto um exemplo numérico onde temos uma amostra com 10.000 indivíduos dos quais 7.000 pertencem ao estrato Baixo e 3.000 ao estrato Alto.

Tabela 5. Frequência da variável “Destino”

Destino	
1-Baixo	2-Alto
7.000	3.000

Fonte: Elaborado pelo autor.

A partir das informações contidas na tabela 5, temos que a probabilidade de um indivíduo pertencer ao estrato Alto é de 30%, enquanto para o estrato Baixo é 70%. Podemos calcular, levando em consideração que estamos analisando apenas uma variável, a chance de selecionarmos aleatoriamente um indivíduo que pertença ao estrato Baixo em contrapartida ao Alto. Essa chance pode ser calculada através da equação (38) que expressa que a chance do evento v ocorrer (Ω_v) é igual à razão entre a sua probabilidade, p_v , e a probabilidade de todos os outros eventos possíveis, $1 - p_v$.

$$\Omega_v = \frac{p_v}{1-p_v} \quad (38)$$

A chance de pertencer ao estrato Baixo, será, portanto, $\Omega_1 = 2,33$. Na análise de variáveis qualitativas e no escopo das ciências sociais essa informação não exaure as possibilidades de estudo. A relação expressa pela chance se mantém frente à consideração de outras variáveis como, por exemplo, o estrato de origem do indivíduo? Essa questão sugere que é necessário avaliar possíveis associações entre uma variável e outras que podem estar relacionadas a ela.

De maneira geral, se p_0 é a probabilidade de ocorrência de um evento em uma situação de referência e p_1 é a probabilidade de ocorrência do mesmo evento em uma situação alternativa, a razão de chances (odds ratio) de ocorrência do evento devido à mudança da situação é

$$\theta = \frac{\frac{p_1}{1-p_1}}{\frac{p_0}{1-p_0}} \quad (39)$$

Para fins de exemplo, olhemos para o caso bivariado adicionando às informações contidas na Tabela 3 a variável Origem, que representa a posição do pai do indivíduo na sociedade quando este tinha 15 anos, contendo as mesmas duas categorias ordinais da variável “Destino”: Baixo (1) e Alto (2)

Tabela 6. Classificação cruzada entre as variáveis “Destino” e “Origem”

Origem	Destino	
	1-Baixo	2-Alto
1-Baixo	$n_{11} = 6.600$ $(p_{11} = 0,66)$	$n_{12} = 1.400$ $(p_{12} = 0,14)$
2-Alto	$n_{21} = 400$ $(p_{21} = 0,04)$	$n_{22} = 1.600$ $(p_{22} = 0,16)$

Fonte: Elaborado pelo autor.

Se considerarmos a Tabela 6 e o evento de interesse for “ter destino no estrato Alto”, a situação de referência for “origem no estrato Baixo” e a situação alternativa for “origem no estrato Alto”, verifica-se, com um pouco de álgebra, que

$$\theta = \frac{p_{11}p_{22}}{p_{12}p_{21}} = \frac{n_{11}n_{22}}{n_{12}n_{21}} \quad (40)$$

Verifica-se, ainda, que, alterando o evento de interesse para “ter destino no estrato Baixo” chega-se exatamente à mesma expressão.

Considerando a classificação cruzada representada na Tabela 6, temos quatro chances e apenas uma razão de chances. A interpretação das chances, representadas na tabela 7, pode ser feita a partir das probabilidades apresentadas entre parênteses na Tabela 6, da seguinte forma: $\Omega_{1|1}$ nos diz qual a chance de um indivíduo pertencer ao estrato Baixo, dado que ele tem origem no estrato Baixo; $\Omega_{1|2}$, qual é a chance de um indivíduo estar no estrato Baixo se ele tem origem no estrato Alto; $\Omega_{2|1}$, qual a chance de um indivíduo pertencer ao estrato Alto se ele tem origem no estrato Baixo; e, por fim, $\Omega_{2|2}$, qual a chance de um indivíduo estar no estrato Alto, dado que ele tem origem no estrato Alto.

Tabela 7. Chances para a classificação cruzada numa tabela 2×2

$\Omega_{1 1} = \frac{p_{11}}{p_{12}} = \frac{0,66}{0,14} = 4,714$	$\Omega_{2 1} = \frac{p_{12}}{p_{11}} = \frac{0,14}{0,66} = 0,212$
$\Omega_{1 2} = \frac{p_{21}}{p_{22}} = \frac{0,04}{0,16} = 0,25$	$\Omega_{2 2} = \frac{p_{22}}{p_{21}} = \frac{0,16}{0,04} = 4$

Fonte: Adaptado de von Eye e Mun (2013).

Esses resultados sugerem que há forte efeito da origem do indivíduo sobre sua posição de destino. Para analisar a força desse efeito, dessa associação presumida, uma comparação entre as chances pode ser realizada através do cálculo da razão entre elas, chamada Razão de Chances (*Odds Ratio*).

A equação (41) expressa a razão de chances entre as chances $\Omega_{1|1}$ e $\Omega_{1|2}$, isto é, a razão entre a chance de um indivíduo com origem no estrato Baixo estar no estrato Baixo e a chance de um indivíduo com origem no estrato Alto estar no estrato Baixo.

$$\theta_1 = \frac{\Omega_{1|1}}{\Omega_{1|2}} \quad (41)$$

Para o exemplo proposto obtemos que a chance de um indivíduo com origem no estrato Baixo estar no estrato Baixo é superior a quatro vezes a chance de um indivíduo com origem no estrato Alto estar no estrato Baixo.

Pelo exemplo da Tabela 6 podemos obter quatro chances e uma única razão de chances, porque uma vez que todas as razões de chances possíveis utilizam a mesma informação, elas terão a mesma interpretação. Desta forma, conclui-se que a razão de chances não varia com mudanças na ordenação das categorias das variáveis na tabela 2×2 , o que é uma das consequências da importante propriedade de independência das razões de chances com relação às distribuições marginais da tabela de contingência.

Caso as frequências f_{12} e/ou f_{21} sejam iguais a zero, a razão de chances não pode ser calculada. Quando a razão de chances é igual à unidade, não há associação entre as variáveis e quanto mais distante da unidade estiver o valor da razão de chances, mais forte será a associação entre elas (RUDAS, 1998; VON EYE; MUN, 2013). De acordo com Rudas (1998), se as variáveis envolvidas na análise forem ordinais e tiverem as mesmas categorias, como é o caso nos estudos de mobilidade, a direção da razão de chances, se for maior ou menor que a unidade, é informativa quanto à natureza da associação presente.

Tanto para o caso de tabelas 2×2 quanto para tabelas $I \times J$ com $I > 2$ e $J > 2$, a robustez dos resultados pode ser testada através de testes assintóticos como estatísticas de erro-padrão, o teste z e intervalos de confiança.

3.5.2. Razões de chances para tabelas de contingência $I \times J$

Para tabelas de contingência $I \times J$, onde $I > 2$ e $J > 2$, ou seja, em que as variáveis possuem mais de duas categorias, múltiplas razões de chances podem ser calculadas. Há diversos métodos disponíveis para organizar as razões de chances, sendo os mais utilizados o de Razões de Chances Locais e a Abordagem de Células Geradoras.

3.5.2.1. Razões de chances locais

O método de Razões de Chances Locais compara probabilidades de linhas e colunas adjacentes a partir de tabelas 2×2 retiradas da tabela $I \times J$ original. Podem ser obtidas $(I - 1)(J - 1)$ sub-tabelas 2×2 que darão origem a $(I - 1)(J - 1)$ razões de

chances. Este método é aplicável para variáveis nominais e ordinais, mas é mais útil no segundo caso (VON EYE; MUN, 2013).

Tomando o exemplo da Tabela 6 e adicionando um estrato intermediário, passamos a ter uma tabela de contingência 3×3 , representada na Tabela 8, em que tanto a variável linha (Origem), quanto a variável coluna (Destino), apresentam três categorias: Baixo (1), Médio (2) e Alto (3).

Tabela 8. Classificação cruzada entre as variáveis “Destino” e “Origem” com três estratos

Origem	Destino		
	1-Baixo	2-Médio	3-Alto
1-Baixo	p_{11}	p_{12}	p_{13}
2-Médio	p_{21}	p_{22}	p_{23}
3-Alto	p_{31}	p_{32}	p_{33}

Fonte: Elaborado pelo autor.

Para a tabela de contingência 3×3 representada na Tabela 8 podem ser obtidas $(I - 1)(J - 1) = (3 - 1)(3 - 1) = 4$ sub tabelas e razões de chances. Trata-se das 4 tabelas 2×2 obtidas combinando as linhas 1 e 2 ou 2 e 3 com as colunas 1 e 2 ou 2 e 3.

Na tabela 9 são apresentadas as possíveis razões de chances.

Tabela 9. Razões de chances locais para a tabela de contingência 3×3

$\theta = \frac{p_{11}p_{22}}{p_{12}p_{21}}$	$\theta = \frac{p_{12}p_{23}}{p_{13}p_{22}}$
$\theta = \frac{p_{21}p_{32}}{p_{22}p_{31}}$	$\theta = \frac{p_{22}p_{33}}{p_{23}p_{32}}$

Fonte: Adaptado de von Eye e Mun (2013).

O conjunto das razões de chances representadas na tabela 9 descrevem a estrutura de associação presente na matriz 3×3 . É possível realizar, de acordo com von Eye e Mun (2013), uma generalização para essa abordagem a partir do cálculo de razões de chances para células não adjacentes. Assumindo $i' > i$ e $j' > j$ a expressão da razão de chances é:

$$\theta = \frac{p_{ij}p_{i'j'}}{p_{ij'}p_{i'j}} \quad (42)$$

Se $i' = i + 1$ e $j' = j + 1$:

$$\theta = \frac{p_{ij}p_{i+1,j+1}}{p_{i,j+1}p_{i+1,j}} \quad (43)$$

A equação (43) é o equivalente à razão de chances locais obtidas através de linhas e colunas adjacentes, utilizando generalização para quando, por exemplo, $i' = i + 2$ e $j' = j + 2$, obtemos uma quinta razão de chances, obtida por células não adjacentes e expressa na equação (44).

$$\theta = \frac{p_{11}p_{33}}{p_{13}p_{31}} \quad (44)$$

3.5.2.2. Abordagem de células geradoras

A abordagem de células geradoras consiste na escolha de uma categoria de referência. Num primeiro momento é selecionada, para cada variável representada na tabela, uma categoria de referência e, feito isso, as demais categorias são comparadas a ela.

Utilizando a Tabela 8 como exemplo, temos uma tabela de contingência 3×3 representando as variáveis Destino e Origem, ambas com três categorias ordinais representando estratos sociais: Baixo (1), Médio (2) e Alto (3). Ao selecionar a categoria Baixo da variável Destino, ou seja, $I=1$, e a categoria Baixo da variável Origem, $J=1$, temos a célula de referência a_{11} .

A seleção de qualquer outra célula que não esteja na primeira coluna (adotada como coluna de referência), nem na primeira linha (adotada como linha de referência) irá definir uma célula geradora que, em conjunto com a célula de referência gera uma sub tabela 2×2 . Leva-se em consideração que as células da matriz são representadas por a_{ij} em que os índices i e j representam a linha e a coluna, respectivamente. Então a célula geradora é caracterizada como a_{ij} , com $i > 1$ e $j > 1$. Desta forma, as quatro células da sub-tabela 2×2 resultante serão: a_{11} , a_{1j} , a_{i1} e a_{ij} .

Há uma razão de chances para cada célula geradora numa tabela $I \times J$, portanto, temos, assim como no método de razões de chances locais, $(I - 1)(J - 1)$ razões de chances que descrevem a estrutura de associação entre as variáveis da tabela.

De acordo com von Eye e Mun (2013), as razões de chances obtidas pelos dois métodos fornecem a mesma informação e a escolha entre um ou outro procedimento pode ser feita com base na interpretação desejada e nas hipóteses pré-estabelecidas a respeito das razões de chances.

3.6. Medidas de mobilidade a partir de tabelas de contingência

Entre as formas de mensurar a mobilidade social está a utilização de medidas de mobilidade calculadas a partir de tabelas de contingência classificando duas ou mais variáveis categóricas. Essas medidas, características da primeira geração de estudos de mobilidade, fornecem medidas de fácil compreensão e visualização dos movimentos entre diferentes estratos sociais (OLIVEIRA, 2018; OLIVEIRA; HOFFMANN, 2021).

Considerando uma tabela de contingência (M) classificando origem (O) e destino (D) com I, J estratos sociais, tem-se uma matriz de mobilidade social $I \times J$. Os indivíduos são distribuídos nas contagens das células de acordo com seu ponto de origem, identificado neste trabalho como o status ocupacional do seu pai, e o seu ponto de destino, seu próprio status ocupacional em 2014.

Os valores correspondentes à proporção de indivíduos em cada célula da matriz de mobilidade social apresentada na Tabela 10 são resultado da equação (45),

$$p_{ij} = \frac{n_{ij}}{N} \quad (45)$$

com $N = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$.

Tabela 10. Exemplo de matriz de mobilidade social 3x3

STATUS DE ORIGEM	STATUS DE DESTINO			TOTAL
	BAIXO	MÉDIO	ALTO	
BAIXO	p_{11}	p_{12}	p_{13}	$p_{1\cdot}$
MÉDIO	p_{21}	p_{22}	p_{23}	$p_{2\cdot}$
ALTO	p_{31}	p_{32}	p_{33}	$p_{3\cdot}$
TOTAL	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	1

Fonte: Elaborado pelo autor.

Na diagonal principal, em destaque, são caracterizados todos aqueles indivíduos que se encontravam, em 2014, no mesmo estrato de origem e são, portanto, considerados imóveis. Abaixo da diagonal principal estão os indivíduos que apresentaram mobilidade descendente, em estrato inferior ao de origem, e, acima, os que experimentaram mobilidade ascendente e estão numa situação superior à de seus pais. Para autores como Pastore (1979) e Behrman (2000), as matrizes de mobilidade oferecem uma alternativa didática de visualização e mensuração da mobilidade social e, permitem, ainda, uma categorização mais flexível da mobilidade e de tipos de mobilidade do que análises considerando variáveis contínuas (BEHRMAN, 2000).

Neste estudo, para a mensuração da mobilidade sócio-ocupacional a partir de matrizes de transição, são utilizadas várias medidas, incluindo as elaboradas por Yasuda (1964), que consideram as dimensões da mobilidade total, imobilidade, mobilidade ascendente e descendente, mobilidade circular e mobilidade estrutural. Considerando uma amostra com N indivíduos, distribuídos numa matriz de mobilidade com k linhas e k colunas, são imóveis todos os indivíduos que estão na diagonal principal $n_{ij} = n_{ii}$ em que $i = j$. Deste modo, a mobilidade total pode ser representada pela proporção de indivíduos fora da diagonal principal, como descreve a equação (46).

$$\text{Mobilidade total: } \frac{N - \sum_{i=1}^k n_{ii}}{N} \quad (46)$$

A mobilidade estrutural é aquela associada às diferenças entre as distribuições marginais de origem e de destino e pode ser obtida através da equação (47), em que \bar{n}_{ii} representa, para cada categoria de destino ou origem, o mínimo entre o total da linha $n_{i\cdot}$ e o total da coluna $n_{\cdot i}$.

$$\text{Mobilidade estrutural: } \frac{1}{N} (N - \sum_{i=1}^k \bar{n}_{ii}) \quad (47)$$

Note que $N - \sum_{i=1}^k \bar{n}_{ii} = \sum_{i=1}^k (n_{i\cdot} - \bar{n}_{ii}) = \sum_{i=1}^k (n_{i\cdot} - \bar{n}_{ii})$ é o número mínimo de pessoas que precisa mudar de status para que a distribuição marginal no destino seja compatível com a distribuição marginal na origem.

A mobilidade circular é calculada a partir da subtração da mobilidade estrutural da mobilidade total, conforme equação (48).

$$\text{Mobilidade circular: } \frac{1}{N} (\sum_{i=1}^k \bar{n}_{ii} - \sum_{i=1}^k n_{ii}) \quad (48)$$

O coeficiente de abertura proposto por Yasuda (1964) busca identificar o grau de flexibilidade da sociedade a partir de uma comparação entre a situação de mobilidade circular perfeita, em que não há limitações de movimentação na estrutura social, e o caso da sociedade objeto de estudo. Neste sentido, tem-se um índice, calculado pela equação (49), que ao atingir o valor unitário indica uma sociedade com perfeita mobilidade.

$$Y = \frac{\sum_{i=1}^k \bar{n}_{ii} - \sum_{i=1}^k n_{ii}}{\sum_{i=1}^k \bar{n}_{ii} - \sum_{i=1}^k \frac{n_{i \cdot} \times n_{\cdot i}}{N}} \quad (49)$$

Há, ainda, outras duas medidas, a exemplo do Coeficiente de Abertura de Yasuda, que podem ser utilizadas na mensuração da mobilidade social: o Índice de Bartholomeu e a medida Delta.

O Índice de Bartholomeu, também conhecido como *Average Absolute Jump* (AAJ), é uma medida de mobilidade voltada para a análise dos movimentos entre os estratos sociais e é muito utilizada pela sua simplicidade (BARTHOLOMEW, 1967; BERMAN, 2022; CHATTOPADHYAY; DEWANJI; ROY, 2019). Bartholomeu (1967) descreve seu índice como uma medida de movimento, o AAJ²⁹, que é igual a 0 no caso de completa imobilidade e assume valores superiores a zero quando ocorre movimentação entre os estratos da sociedade.

$$\text{AAJ} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^k n_{ij} |j - i| \quad (50)$$

A medida Delta, por sua vez, corresponde à razão entre a imobilidade e a imobilidade esperada no caso de independência entre as distribuições de pais e filhos, vezes cem, representando uma forma de mensurar a associação entre as duas variáveis. Quanto maior o Delta, maior a imobilidade, ou seja, maior a força da associação entre origem e destino. A medida pode ser descrita por:

²⁹ Para descrição completa da medida, consultar Berman (2022) e Chattopadhyay, Dewanji e Roy (2019).

$$D = \frac{\sum_{i=1}^k n_{ii}}{\frac{1}{N} \sum_{i=1}^k n_{i.} \times n_{.i}} \quad (51)$$

A análise das matrizes de mobilidade relacionando *status* do pai, nas linhas, e *status* do filho, nas colunas, a partir de medidas de associação pode ser realizada levando-se em conta diversas medidas de associação como o Coeficiente de Correlação de Pearson, o Coeficiente Phi, o Coeficiente de Contingência e a medida V de Cramer. Essas medidas são utilizadas para contrapor a análise de associação através da utilização de modelos log-lineares e, por sua popularidade e disseminação, não são apresentadas pormenorizadamente neste trabalho. Para o detalhamento da aplicação destas medidas há a sugestão dos trabalhos relacionados. Para o Coeficiente de Correlação de Pearson, Snedecor e Cochran (1989) e Brown e Benedetti (1977); Coeficiente Phi, Liebetrau (1983) e Fleiss, Levin e Paik (2003); Coeficiente de Contingência, Liebetrau (1983) e Kendall e Stuart (1979); V de Cramer, Kendall e Stuart (1979).

4. RESULTADOS

4.1. Amostra

A partir da utilização dos dados da PNAD/2014 foram selecionadas para composição da amostra pessoas em idade ativa, com idade de 16 até 65 anos, e ocupadas na semana de referência da pesquisa. Assim como em Oliveira (2018), a restrição de idade máxima de 65 anos faz com que a análise da mobilidade abranja o período de 1964 até 2014, pois, uma vez que o suplemento sobre mobilidade sócio-ocupacional elenca a ocupação do pai quando o respondente tinha 15 anos, para os indivíduos com 65 anos em 2014 é equivalente dizer que a ocupação do pai remonta ao ano 1964.

Foram excluídos da análise indivíduos que não responderam sobre sua escolaridade, em anos de educação formal, e também aqueles que declararam ter trabalhado menos de 6 horas ou mais de 98 horas semanais, essa decisão visou mitigar vieses nos resultados. Também foram retirados da amostra aqueles indivíduos que não responderam ao suplemento de mobilidade sócio-ocupacional pela ausência de informação a respeito da ocupação do pai que impossibilitaria a análise da mobilidade intergeracional para esses indivíduos.

A amostra final para a análise da mobilidade intergeracional brasileira contém 22.918 observações que, considerando os fatores de expansão da subamostra referente ao suplemento, representam uma população de 59.673 milhares de pessoas das quais 59% são do sexo masculino e 41% do sexo feminino; aproximadamente 49% brancos e 50% pretos e pardos; metade dos indivíduos com idade entre 21 e 40 anos.

O nível de escolaridade mais recorrente entre os indivíduos da amostra, cerca de 30%, é o Ensino Médio completo, equivalente a 11 anos de escolaridade formal, enquanto 10% concluíram apenas o Ensino Fundamental (8 anos de estudo), 5% declararam não ter instrução e 16% possuem pelo menos o Ensino Superior completo (15 anos ou mais). Quanto à condição censitária, mais de 88% dos indivíduos se encontram na área urbana. Há predominância de indivíduos que residem no sudeste do país, 45%, e que residem na região nordeste, 24%.

Considerando o salário mínimo de 2014, R\$ 724 em valor corrente, cerca de 40% dos indivíduos recebiam como rendimento pela atividade principal entre um e dois salários mínimos, 24% recebia menos que um salário mínimo e 10% recebia mais que

cinco salários mínimos. Com relação às horas semanais de trabalho, 7% trabalhavam até 20 horas semanais, enquanto 45% trabalhavam entre 20 e 40 horas semanais.

Adotando a classificação das ocupações da PNAD/2014 proposta por Oliveira (2018) e Oliveira e Hoffmann (2021), que hierarquiza as 483 ocupações presentes nessa edição da pesquisa, temos a distribuição dos indivíduos da amostra na tabela 11.

Tabela 11. Distribuição da população nos estratos sócio-ocupacionais em 2014/milhares de pessoas

Estrato ocupacional	Frequência	Frequência cumulativa	Porcentagem %	Porcentagem cumulativa %
Alto	3.853	3.853	6,46	6,46
Médio-Superior	7.504	11.357	12,58	19,04
Médio-Médio	8.353	19.710	14,0	33,04
Médio-Inferior	9.037	28.747	15,14	48,18
Baixo-Superior	15.721	44.468	26,35	74,53
Baixo-Inferior	15.205	59.673	25,48	100

Fonte: Elaborado pelo autor, com base nos microdados do IBGE.

A presença nos estratos alto e médio-superior é muito limitada e há predominância na base da estrutura social adotada. Mais de 51% dos indivíduos encontram-se nos estratos baixos, enquanto no estrato alto há apenas 6% dos indivíduos da amostra.

4.1.1. Descrição das coortes de idade

A amostra foi subdividida em cinco coortes de idade, de dez anos cada, visando a inclusão de uma variável referente à década em que o pai do indivíduo estava no mercado de trabalho. Considerando que a resposta sobre o *status* ocupacional do pai remonta ao período temporal em que o filho tinha 15 anos, as décadas correspondentes e suas respectivas coortes de idade são: 1964 até 1973, para os indivíduos com idade entre 56 e 65 anos em 2014; 1974 até 1983, indivíduos com idade entre 46 e 55 anos; 1984 até 1993, 36 a 45 anos; 1994 até 2003, 26 a 35 anos; e 2004 até 2013, 16 a 25 anos.

A distribuição dos indivíduos entre as coortes, apresentada na tabela 12, evidencia a maior representatividade nas décadas mais recentes, quase 70% da amostra é composta por indivíduos com idade de até 45 anos. A década com maior número de pais-filhos é a que compreende 1994-2003, 26,57% dos indivíduos, seguida pela terceira década, 1984-1993, juntas essas duas décadas sucessivas representam 51,45%

da população. A menor representatividade é observada entre indivíduos com idade entre 56 e 65 anos, em que a resposta sobre a ocupação dos pais remonta ao período entre 1964 e 1973, 10,39% da amostra.

Tabela 12. Distribuição da frequência na amostra entre as coortes de idade/décadas.

Identificador	Coorte	Frequência na população (milhares)	%	Frequência cumulativa	% cumulativa
1	2004-2013	10.486	17,57	10.486	17,57
2	1994-2003	15.854	26,57	26.340	44,14
3	1984-1993	14.850	24,89	41.190	69,03
4	1974-1983	12.283	20,58	53.473	89,61
5	1964-1973	6.199	10,39	59.673	100,00
Total		59.673	100,00		

Fonte: Elaborado pelo autor, com base nos microdados do IBGE.

A primeira década, 2004 a 2013, possui 10.486 milhares de pessoas, na qual 58,77% são do sexo masculino. Cerca de 77,35% têm até 11 anos de escolaridade, sendo a escolaridade mais recorrente o ensino médio completo (39,48%), seguida pelo ensino fundamental completo (9,87%) e ensino superior completo (incluindo pós-graduação) (8,48%).

A segunda década, 1994 a 2003, possui 15.854 milhares de pessoas, sendo 57,77% homens. A escolaridade mais frequente é o ensino médio completo 37,10%, seguida pelo ensino superior completo (incluindo pós-graduação), 20,02%, e o fundamental completo, 8,20%. Aproximadamente 67,8% possuem até 11 anos de escolaridade.

Na terceira coorte, que compreende os anos entre 1984 e 1993, há 14.850 milhares de pessoas, das quais 57,15% são homens. Cerca de 71,41% possuem até 11 anos de escolaridade, sendo mais frequente aqueles com o ensino médio completo, 29,19%, e ensino superior completo, 16,52%.

Na década entre 1974 e 1983, há 12.283 milhares de pessoas, sendo 60,13% homens. Os indivíduos representados nesta década possuem, mais frequentemente, o ensino médio completo, 24,21% e o ensino superior completo, 15,88%, 70,81% possuem até 11 anos de escolaridade.

Por fim, na quinta coorte, representando indivíduos com idade entre 56 e 65 anos, temos 6.199 milhares de pessoas, sendo 64,02% homens. Os níveis de escolaridade mais frequentes são, novamente, o ensino médio completo, 16,52%, e o ensino superior completo, 13,71%. Um aspecto relevante desta coorte é o número de analfabetos, superior ao observado nas coortes anteriores – 1,20% na primeira coorte, 2,28% na segunda, 5,15% na terceira e 7,84% na quarta – chegando a 12,21% dos indivíduos representados no período (1973-1964).

Há, portanto, duas características observadas na composição das coortes: uma gradual diminuição da baixa escolaridade nas coortes mais contemporâneas a 2014 e a alta representatividade de indivíduos do sexo masculino em todas elas.

4.1.2. Distribuição de pais e filhos por coortes de idade e matrizes de mobilidade

A variável coorte de idade é utilizada para a identificação da existência de associação da estrutura ocupacional em diferentes pontos no tempo e a mobilidade intergeracional experimentada pelos indivíduos brasileiros. Deste modo, é interessante analisar a distribuição dos indivíduos e seus pais ao longo destas cinco décadas para identificar se há mudanças nas distribuições da estrutura sócio-ocupacional.

Neste sentido, a tabela 13, abaixo, representa a distribuição dos indivíduos e seus pais entre as coortes e os estratos sócio-ocupacionais da sociedade brasileira.

Tabela 13. Distribuição dos indivíduos e seus pais entre coortes de idades e estratos sócio-ocupacionais

Estrato	Coorte 1 2004-2013 (%)		Coorte 2 1994-2003 (%)		Coorte 3 1984-1993 (%)		Coorte 4 1974-1983 (%)		Coorte 5 1964-1973 (%)	
	Pai	Filho	Pai	Filho	Pai	Filho	Pai	Filho	Pai	Filho
1 - Baixo-Inferior	29,0	26,0	33,2	28,5	40,3	26,7	46,4	25,7	52,6	22,8
2 - Baixo-Superior	29,9	31,6	26,7	25,8	24,7	26,5	22,8	24,1	16,5	25,1
3 - Médio-Inferior	21,1	11,9	18,3	15,5	16,7	15,8	14,0	16,9	12,2	15,0
4 - Médio-Médio	8,8	18,9	8,7	12,1	7,9	11,8	7,6	12,5	9,0	14,6
5 - Médio-Superior	8,2	7,8	8,7	10,1	6,8	12,3	6,4	14,1	6,2	15,5
6 – Alto	3,0	3,8	4,3	8,0	3,7	6,9	2,9	6,7	3,4	7,1

Fonte: Elaborado pelo autor, com base nos microdados do IBGE.

As diferentes distribuições de pais e filhos entre as coortes de idade permitem verificar que a presença de filhos nos estratos baixos, com relação a pais em estratos baixos, é expressiva nas coortes de 3, 4 e 5. Na coorte 5, por exemplo, representando pais em ocupações na década de 1964 a 1973, 69,1% dos pais estão no estrato baixo, enquanto os filhos, considerando como destino suas ocupações em 2014, somam 47,8%, resultando uma diferença de 21,3 pontos percentuais, a mais alta observada entre as cinco coortes. Para a coorte 1, essa diferença é de 1,3 pontos percentuais, a distância entre representatividade de pais e filhos no estrato baixo é pequena.

Com relação às diferenças nas distribuições, o maior ‘esvaziamento’ no estrato baixo é observado nas coortes 4 e 5. Esse movimento parece ter refletido um aumento na participação dos filhos no estrato médio. Com relação ao estrato alto, em todas as coortes a participação dos filhos é superior à dos pais, a maior diferença é encontrada na coorte 4, onde a diferença entre a participação dos pais e dos filhos neste estrato é de 3,8 p.p., indicando que os indivíduos desta coorte conseguiram atingir *status* sócio-ocupacional no topo da estratificação superior ao experimentado pelos seus pais no topo da estratificação. A presença de filhos no estrato alto, superior à dos pais em todas as coortes, é mais expressiva nas mais antigas.

Na tabela 14, a seguir, são apresentadas as matrizes de mobilidade sócio-ocupacional para a sociedade brasileira, em 2014, por coorte de idade, na forma de porcentagens totalizando 100% em cada linha (origem). É possível, assim como na tabela 13, observar as diferenças nas distribuições, representadas pelos totais marginais nas matrizes. Um aspecto relevante, oferecido pela visualização das matrizes, é o comportamento da diagonal principal, que caracteriza a imobilidade, e as participações acima e abaixo da diagonal, representando a mobilidade ascendente e descendente, respectivamente.

Em todas as coortes a maior parte dos filhos encontra-se na diagonal principal. A coorte de idade com mais exceções, na maior predominância nas diagonais, é a primeira, com indivíduos entre 16 e 25 anos. Nesta coorte há quatro linhas (origens) nas quais predomina a maior concentração de indivíduos abaixo da diagonal principal e não na diagonal principal, indicando a maior mobilidade descendente em detrimento da imobilidade. Para os pais do estrato médio-inferior (3), médio-médio (4), médio-superior (5) e alto (6), a manutenção dos filhos estrato no paterno ou, ainda, a mobilidade ascendente para estratos mais altos, não se realiza com a mesma intensidade que a mobilidade descendente. A maior parte dos indivíduos com origem nos três

primeiros estratos citados encontra-se no estrato baixo-superior (2), enquanto a parcela mais expressiva de filhos de pais do estrato alto (6) está no estrato médio-médio (4).

O mesmo pode ser observado, com menor recorrência, nas demais coortes. Na coorte 2, para os filhos de pais do estrato médio-inferior, há mais recorrência do destino baixo-superior (2), o mesmo que ocorre nas coortes 3 e 4. Na coorte 4, para os filhos de pais do estrato médio-médio (4) há, também, maior recorrência de filhos no estrato baixo-superior (2). Na coorte 5, os filhos de pais do estrato médio-inferior (3) e médio-médio (4), possuem maior presença nos estratos baixo-inferior (1) e baixo-superior (2), respectivamente.

A única exceção em que há maior recorrência, fora da diagonal principal, localizada em alguma posição que representa mobilidade ascendente, ocorre na coorte 2, para filhos de pais do estrato médio-médio (4) que estão no estrato médio-superior (5).

Tabela 14. Matrizes de mobilidade sócio-ocupacional intergeracional brasileira em 2014, por coorte de idade (%)

	STATUS DO PAI	STATUS DO FILHO						Total	Total dos Pais
		1	2	3	4	5	6		
Coorte 1	1 - Baixo-Inferior	46,4	26,7	11,2	11,4	3,2	1,1	100	29,0
	2 - Baixo-Superior	21,2	39,0	11,6	19,0	6,6	2,6	100	29,9
	3 - Médio-Inferior	18,8	33,1	16,0	21,6	6,7	3,7	100	21,1
	4 - Médio-Médio	13,5	27,4	11,5	29,2	14,6	3,7	100	8,8
	5 - Médio-Superior	12,6	24,6	9,7	21,1	23,7	8,3	100	8,2
	6 - Alto	0,6	23,8	1,9	33,9	10,0	29,9	100	3,0
	Total dos Filhos N = 10.486 mil	26,0	31,6	11,9	18,9	7,8	3,8	100	100
<hr/>									
	STATUS DO PAI	STATUS DO FILHO						Total	Total dos Pais
		1	2	3	4	5	6		
Coorte 2	1 - Baixo-Inferior	41,2	26,8	14,4	9,3	7,2	1,1	100	33,2
	2 - Baixo-Superior	19,8	31,5	15,1	16,3	12,3	5,0	100	26,7
	3 - Médio-Inferior	13,3	22,2	21,4	19,6	17,2	6,4	100	18,3
	4 - Médio-Médio	9,7	21,8	13,4	18,1	27,2	9,7	100	8,7
	5 - Médio-Superior	5,1	16,0	9,2	18,0	32,2	19,5	100	8,7
	6 - Alto	1,3	8,7	6,7	11,0	33,8	38,5	100	4,3
	Total dos Filhos N = 15.854 mil	22,8	25,1	15,0	14,6	15,5	7,1	100	100

(continua)

(continuação)

	STATUS DO PAI	STATUS DO FILHO						Total	Total dos Pais
		1	2	3	4	5	6		
Coorte 3	1 - Baixo-Inferior	39,3	24,6	15,6	8,7	9,5	2,2	100	40,3
	2 - Baixo-Superior	20,0	30,5	17,7	13,9	13,5	4,4	100	24,7
	3 - Médio-Inferior	19,0	23,5	23,2	14,1	14,7	5,4	100	16,7
	4 - Médio-Médio	12,0	19,6	18,0	21,6	18,7	10,1	100	7,9
	5 - Médio-Superior	9,4	12,8	11,3	14,6	29,2	22,6	100	6,8
	6 - Alto	5,3	9,9	5,1	13,2	26,9	39,6	100	3,7
	Total dos Filhos N = 14.850 mil	25,7	24,1	16,9	12,5	14,1	6,7	100	100
<hr/>									
	STATUS DO PAI	STATUS DO FILHO						Total	Total dos Pais
		1	2	3	4	5	6		
Coorte 4	1 - Baixo-Inferior	39,3	27,9	13,4	9,0	7,8	2,7	100	46,4
	2 - Baixo-Superior	19,9	31,7	19,2	11,5	12,2	5,6	100	22,8
	3 - Médio-Inferior	15,3	25,0	22,2	14,8	15,6	7,1	100	14,0
	4 - Médio-Médio	16,6	19,6	15,0	19,2	16,9	12,7	100	7,6
	5 - Médio-Superior	8,2	16,7	13,1	16,6	26,7	18,7	100	6,4
	6 - Alto	1,5	10,7	3,9	15,5	26,6	41,9	100	2,9
	Total dos Filhos N = 12.283 mil	26,7	26,5	15,8	11,8	12,3	6,9	100	100
<hr/>									
	STATUS DO PAI	STATUS DO FILHO						Total	Total dos Pais
		1	2	3	4	5	6		
Coorte 5	1 - Baixo-Inferior	37,9	28,0	15,8	10,1	5,2	3,0	100	52,6
	2 - Baixo-Superior	21,3	30,2	12,7	16,6	12,7	6,5	100	16,5
	3 - Médio-Inferior	22,8	21,6	19,9	9,6	13,1	13,1	100	12,2
	4 - Médio-Médio	17,5	20,6	16,5	16,8	13,2	15,4	100	9,0
	5 - Médio-Superior	8,0	16,5	12,4	18,1	25,0	20,0	100	6,2
	6 - Alto	6,9	16,8	10,5	6,1	25,6	34,1	100	3,4
	Total dos Filhos N = 6.199 mil	28,5	25,8	15,5	12,1	10,1	8,0	100	100

Fonte: Elaborado pelo autor, com base nos microdados do IBGE.

De modo geral, a menor diferença entre as distribuições de pais e filhos nas coortes mais recentes é esperada. Para os indivíduos da coorte 1, em especial, que têm idade entre 16 e 25 anos, a ocupação do pai e a sua própria distam, no máximo, 10 anos entre si, houve menos tempo para mobilidade intergeracional. E a mobilidade descendente relativamente elevada nessa coorte se explica pela juventude dos filhos,

ainda no início de eventuais carreiras e/ou pouca experiência no trabalho. Nas tabelas 15 a 20 são apresentadas as matrizes de mobilidade intergeracional em número de pessoas, para o ano de 2014, e por coorte de idade.

Tabela 15. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014.

STATUS DO PAI	STATUS DO FILHO EM 2014						TOTAL DOS PAIS
	1	2	3	4	5	6	
1. Baixo-Inferior	9.408	6.200	3.313	2.197	1.659	475	23.252
2. Baixo-Superior	3.008	4.867	2.317	2.287	1.697	674	14.848
3. Médio-Inferior	1.707	2.551	2.081	1.721	1.379	624	10.064
4. Médio- Médio	653	1.083	735	1.047	963	492	4.973
5. Médio-Superior	369	759	476	780	1.250	794	4.428
6. Alto	59	262	115	322	557	793	2.108
TOTAL DOS INDIVÍDUOS N = 59.673	15.205	15.721	9.037	8.353	7.504	3.853	59.673

Fonte: Elaborado pelo autor.

Tabela 16. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 16 e 25 anos, coorte de idade (1).

STATUS DO PAI	STATUS DO FILHO EM 2014						TOTAL DOS PAIS
	1	2	3	4	5	6	
1. Baixo-Inferior	1.410	813	340	348	98	32	3.041
2. Baixo-Superior	663	1.222	362	596	206	80	3.129
3. Médio-Inferior	415	733	355	479	149	82	2.213
4. Médio- Médio	125	254	106	270	135	35	925
5. Médio-Superior	108	212	83	182	204	72	861
6. Alto	2	75	6	107	31	95	316
TOTAL DOS INDIVÍDUOS N = 10.485	2.723	3.309	1.252	1.982	823	396	10.485

Fonte: Elaborado pelo autor.

Tabela 17. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 26 e 35 anos, coorte de idade (2).

STATUS DO PAI	STATUS DO FILHO EM 2014						TOTAL DOS PAIS
	1	2	3	4	5	6	
1. Baixo-Inferior	2.172	1.414	758	491	380	57	5.271
2. Baixo-Superior	838	1.335	640	690	523	212	4.237
3. Médio-Inferior	385	643	618	566	497	186	2.895
4. Médio- Médio	134	302	185	251	377	135	1.384
5. Médio-Superior	71	222	127	250	446	270	1.387
6. Alto	9	59	46	75	230	262	680
TOTAL DOS INDIVÍDUOS N = 15.854	3.609	3.975	2.374	2.322	2.452	1.122	15.854

Fonte: Elaborado pelo autor.

Tabela 18. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 36 e 45 anos, coorte de idade (3).

STATUS DO PAI	STATUS DO FILHO EM 2014						TOTAL DOS PAIS
	1	2	3	4	5	6	
1. Baixo-Inferior	2.353	1.472	935	518	570	133	5.981
2. Baixo-Superior	733	1.116	648	510	496	160	3.662
3. Médio-Inferior	473	584	578	350	366	135	2.485
4. Médio- Médio	140	229	211	252	219	118	1.170
5. Médio-Superior	95	130	114	148	294	228	1.009
6. Alto	29	54	28	72	146	216	544
TOTAL DOS INDIVÍDUOS N = 14.850	3.822	3.583	2.513	1.851	2.091	990	14.850

Fonte: Elaborado pelo autor.

Tabela 19. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 46 e 55 anos, coorte de idade (4).

STATUS DO PAI	STATUS DO FILHO EM 2014						TOTAL DOS PAIS
	1	2	3	4	5	6	
1. Baixo-Inferior	2.236	1.587	764	511	442	157	5.696
2. Baixo-Superior	556	886	537	321	342	156	2.797
3. Médio-Inferior	263	428	381	253	268	122	1.714
4. Médio- Médio	156	184	140	180	158	119	936
5. Médio-Superior	65	131	103	130	209	147	785
6. Alto	5	38	14	55	94	149	355
TOTAL DOS INDIVÍDUOS N = 12.283	3.280	3.253	1.939	1.449	1.514	848	12.283

Fonte: Elaborado pelo autor.

Tabela 20. Matriz de mobilidade intergeracional da sociedade brasileira (milhares), em 2014, indivíduos com idade entre 56 e 65 anos, coorte de idade (5).

STATUS DO PAI	STATUS DO FILHO EM 2014						TOTAL DOS PAIS
	1	2	3	4	5	6	
1. Baixo-Inferior	1.237	914	517	330	170	97	3.263
2. Baixo-Superior	218	309	130	169	130	66	1.022
3. Médio-Inferior	172	163	150	72	99	99	756
4. Médio-Médio	98	115	92	94	74	86	559
5. Médio-Superior	31	64	48	70	97	77	387
6. Alto	15	36	22	13	55	73	213
TOTAL DOS INDIVÍDUOS N = 6.199	1.770	1.600	959	748.451	624	498	6.199

Fonte: Elaborado pelo autor.

4.2. Medidas de mobilidade social e associação

Esta seção apresenta e analisa medidas de mobilidade e associação estatística calculadas a partir das tabelas de contingência apresentadas na seção anterior. A tabela 21, a seguir, apresenta os padrões de mobilidade sócio-ocupacional para a amostra total

e cada uma das coortes consideradas, utilizando medidas de mobilidade social tradicionais e de associação estatística³⁰. Percebe-se claramente o aumento da predominância da mobilidade ascendente conforme aumenta a distância da coorte ao ano de 2014, ou seja, conforme aumenta a idade do indivíduo, aumenta também a mobilidade ascendente.

Tabela 21. Padrões gerais da mobilidade intergeracional por coorte de idade

MEDIDAS	COORTE					2014 (GERAL)
	1	2	3	4	5	
Medidas de mobilidade						
Ascendente (%)	36,5	45,4	46,2	47,9	48,9	44,9
Imobilidade (%)	33,9	32,1	32,4	32,9	31,6	32,6
Descendente (%)	29,6	22,6	21,4	19,2	19,5	22,5
Estrutural (%)	12,6	15,4	15,1	19,7	24,1	15,2
Circular (%)	53,5	52,5	52,5	47,4	44,3	52,2
Coefficiente de Abertura (Y)	0,817	0,812	0,826	0,820	0,840	0,823
Delta	1,55	1,61	1,52	1,46	1,36	1,53
Average Absolute Jump (AAJ)	1,15	1,22	1,28	1,27	1,34	1,25
Medidas de Associação						
Coefficiente de Correlação de Pearson	0,351	0,424	0,378	0,379	0,363	0,391
Coefficiente de Contingência	0,418	0,459	0,416	0,400	0,376	0,413
Coefficiente Phi	0,460	0,517	0,457	0,436	0,406	0,453
V de Cramer	0,206	0,231	0,204	0,195	0,181	0,203
Medidas de posição e dispersão do status do filho						
Média	13,0	15,1	14,7	14,4	14,4	14,4
Desvio-padrão	11,7	13,7	13,8	14,4	14,7	13,8
Coefficiente de variação	90,0	91,0	93,8	99,6	102,2	95,5

Fonte: Elaborado pelo autor, com base nos microdados do IBGE.

As coortes de 2 a 5, cuja distância temporal do ponto de origem é superior à dos indivíduos da coorte 1, apresentam maior mobilidade ascendente em contrapartida a uma mobilidade descendente e imobilidade menores. A maior mobilidade ascendente

³⁰ Apresentadas na seção 3.6.

observada entre as cinco coortes é encontrada justamente na coorte 5, indicando que a distância temporal entre a origem e o destino do indivíduo é relevante nas oportunidades de mobilidade ascendente experimentadas pela indivíduos. Há mais tempo-oportunidade de mobilidade para os filhos conforme o intervalo de tempo entre sua medida de *status* e a do seu pai aumenta.

As medidas de mobilidade estrutural e mobilidade circular apresentam, no contexto das coortes de idade, comportamento divergente. A mobilidade estrutural, partindo da coorte 1 em direção à coorte 5, aumenta, indicando aumento na distância entre a estrutura sócio-ocupacional do indivíduo de referência e a do pai deste indivíduo.

Por exemplo, quando olhamos para a coorte 5, estamos observando indivíduos com idade entre 56 e 65 anos, atuando no mercado de trabalho em 2014 e cuja ocupação dos pais remonta o período entre 1964 e 1973. Nota-se, neste caso, um resultado de mobilidade estrutural superior ao das demais coortes e ao observado para a amostra geral, sem divisão por coortes, indicando maior diferença entre a distribuição dos pais e a dos filhos entre os estratos sócio-ocupacionais e caracterizando-a como sendo mais relevante para a mobilidade observada nos filhos do que na coorte 1, por exemplo.

O mesmo não acontece para a medida da mobilidade circular, que apresenta resultado superior nas coortes de idade mais recentes, ou seja, entre os indivíduos mais jovens, indicando que para aqueles cuja distância temporal dos pais é, ainda, pequena, a mobilidade apresenta como característica mais relevante a mudança de estrato pela substituição de posições na estrutura social. O Coeficiente de Yasuda, medida que busca avaliar a abertura da sociedade, apresenta crescimento partindo da coorte 1 em direção à coorte 5. Neste sentido, observa-se, portanto, que há maior flexibilidade para pontos menos contemporâneos.

As medidas de posição e dispersão dos *status* dos filhos mostram que a média, o desvio padrão e o coeficiente de variação são mais baixos para a coorte 1, ou seja, aqueles com idade entre 16 e 25 anos. Isso se explica pela juventude das pessoas dessa coorte. Para muitas carreiras de maior status, 16 a 25 anos ainda é uma idade em que a pessoa está estudando. A juventude dessa primeira coorte também condiciona as medidas de mobilidade socioeconômica, como será discutido adiante.

A figura 1 apresenta, de forma gráfica, os resultados da tabela 21, ilustrando as medidas tradicionais de mobilidade que constroem o padrão de mobilidade intergeracional brasileira para as cinco coortes de idade da amostra. Os resultados, já

discutidos, evidenciam a relevância da distância temporal entre origem e destino na mobilidade experimentada pelos indivíduos. Ressalta-se, ainda, o fato de que a coorte 1 apresenta resultados em nível bastante diferente para as medidas de mobilidade ascendente, descendente e imobilidade num comparativo com as demais.

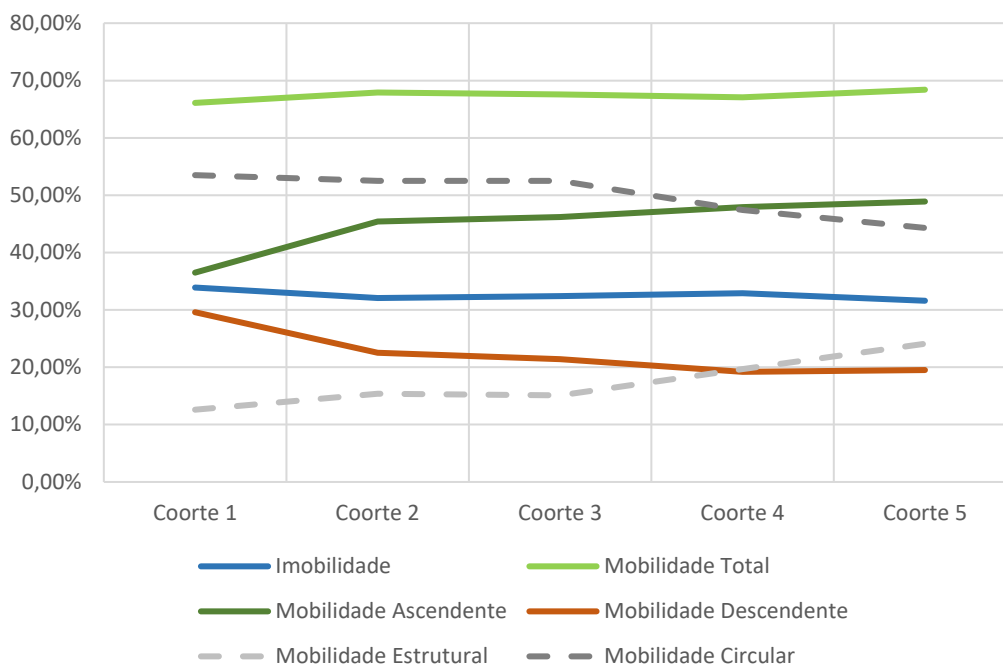


Figura 1. Resultados para as medidas de mobilidade tradicionais para cada coorte de idade

Fonte: Elaborado pelo autor.

A figura 2 apresenta, em conjunto com o Coeficiente de Abertura de Yasuda, as medidas do Índice de Bartholomew e a medida Delta, medidas de mobilidade expostas para complementar a análise das medidas da figura 1.

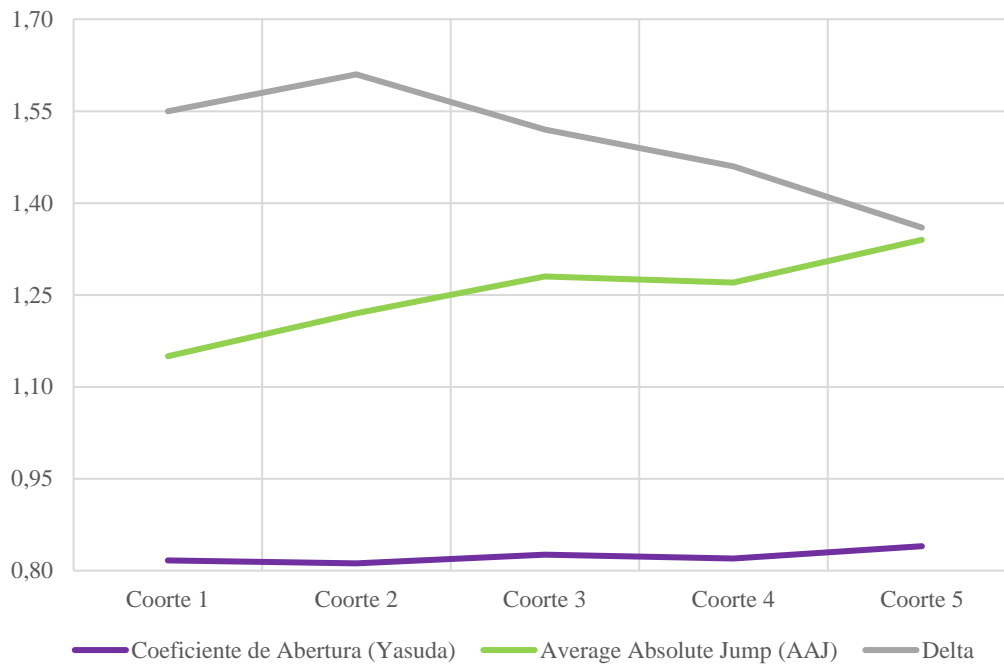


Figura 2. Resultados para as medidas do Coeficiente de Abertura (Yasuda), *Average Absolute Jump* (AAJ) e Delta, para cada coorte de idade

Fonte: Elaborado pelo autor.

O coeficiente de abertura, num comparativo entre as cinco coortes, é bastante semelhante, mas apresenta valor levemente superior na quinta coorte, reforçando que coortes mais antigas apresentam maior mobilidade, resultado esperado, dado o espaço temporal para experimentar tanto a mobilidade consequência de mudanças estruturais no mercado de trabalho, quanto a de trocas de posição, decorrente de mais tempo de experiência, por exemplo.

A medida *Average Absolute Jump* (AAJ) apresenta uma tendência crescente na direção da coorte 1 à coorte 5, indicando que, conforme aumenta a distância temporal entre o ponto de origem e o ponto de destino, a distância média percorrida na estrutura sócio-ocupacional também aumenta. O Delta apresenta uma tendência decrescente após a coorte 2, em direção à coorte 5, indicando diminuição da imobilidade observada em relação à mobilidade esperada em caso de independência entre origem e destino e, por consequência, da força da associação entre o *status* do pai e o *status* do filho.

A figura 3 ilustra os resultados para várias medidas de associação: Coeficiente de Contingência, Coeficiente de Correlação de Pearson, Coeficiente Phi, considerando as cinco tabelas de contingência relacionando *status* de origem e destino para as coortes de idade na análise da mobilidade intergeracional brasileira com base em dados da PNAD 2014.

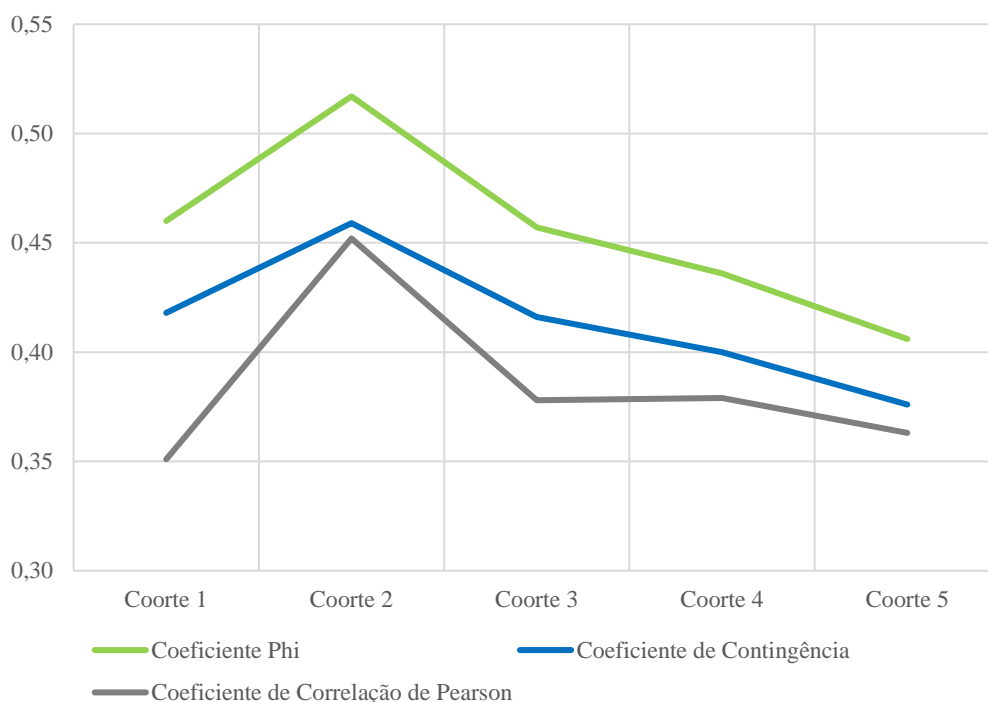


Figura 3. Resultados para as medidas de associação: Coeficiente Phi, Coeficiente de Contingência, Coeficiente de Correlação de Pearson, para cada coorte de idade

Fonte: Elaborado pelo autor.

Para coortes de idade mais antigas, como as coorte 4 e 5, a força da associação entre *status* do filho e *status* do pai é menor. Os resultados observados na figura 3 indicam que a associação entre *status* do pai e do filho existe e não é constante ao longo do tempo (nas coortes de idade).

Um aspecto importante é o salto na associação entre origem e destino observado entre a primeira e a segunda coorte de idade. A associação nesta última é superior ao resultado de todas as outras coortes. A questão torna-se, neste ponto, não o resultado observado para a coorte 2 e sim a baixa associação na coorte 1, para os indivíduos com idade entre 16 e 25 anos. A baixa associação entre *status* do pai e *status* do filho neste caso pode se dar devido à condição de entrada recente no mercado de trabalho e, ainda, à maior chance de contemporaneidade de pais e dos filhos no mercado.

Para a coorte 1, em início de carreira, ainda não houve tempo para que a pessoa se firmasse na carreira; são muito raras as atividades profissionais nas quais a pessoa possa obter sucesso antes dos 25 anos de idade. Assim, nessa coorte ainda não houve tempo para que se concretizasse a alta associação entre origem e destino. De modo geral, a coorte 1 é aquela que apresenta números mais baixos de permanência, no

comparativo entre as cinco coortes, em origem e destino no estrato Alto e origem e destino no estrato Médio Superior. Entre os filhos de pais do estrato alto, 29,9% estão neste estrato na coorte 1, enquanto na coorte 3, por exemplo, este número chega a 39,6%. Ou, ainda, dos filhos de pais do estrato médio-superior, 23,7% estão neste estrato na coorte 1, enquanto na coorte 2, temos 32,2%. Nos dois casos, a coorte 1 apresenta a menor porcentagem de manutenção de *status* nestes estratos.

A medida V de Cramer, ilustrada graficamente na figura 4, mostra, também, que a associação entre o *status* do pai e do filho na coorte 1 é inferior ao observado para a coorte 2, após a qual o poder da associação perde força.

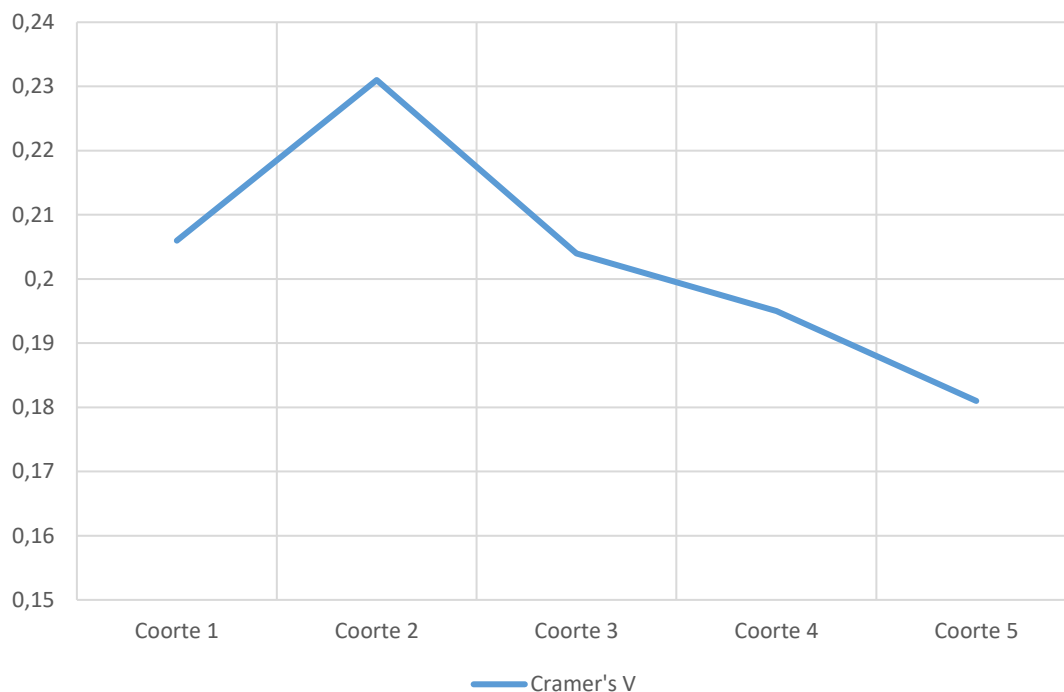


Figura 4. Resultados para a medida V de Cramer, para cada coorte de idade

Fonte: Elaborado pelo autor.

Ressalta-se que as medidas de associação representam uma forma de mensurar imobilidade: quanto maior a associação entre as variáveis de origem e destino, maior será a imobilidade observada na tabela de contingência. Apesar deste aspecto, a medida de associação é tratada como uma dimensão relativa da mobilidade, porque, para os indivíduos móveis, não permite identificar a direção da mobilidade experimentada.

4.3. Modelos log-lineares

Na análise da mobilidade intergeracional sob a perspectiva da associação entre origem, caracterizada pelo *status* sócio-ocupacional do pai, destino, caracterizado pelo *status* sócio-ocupacional do filho, e a dimensão temporal, representada pela década em que o pai do indivíduo estava no mercado de trabalho, realizada através da aplicação de modelos log-lineares, é possível avaliar nove cenários.

Entre os nove modelos possíveis estão: o modelo de independência mútua, onde nenhuma das três variáveis apresenta associação estatística entre si; três modelos de independência marginal, nos quais uma das variáveis é adotada como controle tendo em vista a associação entre as outras duas variáveis; três de independência condicional em que há associação entre uma variável e as demais, separadamente; o de associação homogênea, onde há todas as associações possíveis entre duas variáveis, com exceção da associação tripla; e o modelo saturado, que consiste no modelo que relaciona todas as associações possíveis e presentes nos modelos anteriores, com a adição da interação tripla.

A aplicação dos modelos log-lineares foi realizada a partir do *software* estatístico SAS³¹, por meio da função PROC CATMOD, que realiza a análise de dados categóricos através da transformação das probabilidades das células das tabelas de contingência por meio de uma função-resposta. A função-resposta escolhida depende das premissas teóricas adotadas na construção do modelo, que pode ser um modelo linear ou log-linear.

Entre as técnicas utilizadas para a estimação está o método de máxima verossimilhança, procedimento adotado como padrão da função PROC CATMOD que, para evitar a super parametrização na combinação linear das funções-resposta, adota a restrição de soma zero. Deste modo, o número de funções resposta será sempre menor, em uma unidade, que o número das categorias consideradas pelo modelo. Considerando a relação entre três variáveis com $5 \times 6 \times 6$ categorias, há, portanto, $180 - 1 = 179$ funções-resposta.

Dado o elevado número de observações, 59.673 milhares, e seu impacto sobre as estimativas e testes de significância dos modelos, adotou-se uma estratégia para reponderar as observações sem que os graus de liberdade fossem afetados. Neste

³¹ SAS 9.4.

sentido, as frequências observadas, de acordo com as coortes, *status* do pai e do filho, foram multiplicadas pela razão entre a frequência amostral observada, antes da expansão da amostra, e a frequência populacional observada. Deste modo, todas as observações foram normalizadas ponderando-se o tamanho da amostra.

Na tabela 22 são apresentados os resultados da razão de verossimilhança (L_m^2) e do critério de informação Bayesiano (*BIC*) para cada um dos nove modelos log-lineares estimados. As duas medidas visam subsidiar a análise da qualidade de ajustamento dos modelos de associação entre *status* do pai, *status* do filho e coorte de idade no estudo da mobilidade sócio-ocupacional intergeracional brasileira.

Tabela 22. Estatísticas de ajustamento dos modelos log-lineares estimados para a análise da mobilidade intergeracional brasileira, em 2014

Modelo	Associações	Razão de Verossimilhança (L_m^2)	Graus de Liberdade (<i>gl</i>)	<i>p</i> -valor	<i>BIC</i>
Independência mútua	C P F	5354,45	165	<0.0001	3.697,90
Independência marginal (F)	C*P F	1226,55	145	<0.0001	-179,00
Independência marginal (P)	C*F P	4963,77	145	<0.0001	3.508,02
Independência marginal (C)	P*F C	4746,96	140	<0.0001	3.291,21
Independência condicional (F)	F*C F*P	835,87	120	<0.0001	-368,89
Independência condicional (P)	P*C P*F	619,06	120	<0.0001	-585,70
Independência condicional (C)	C*P C*F	4.356,29	125	<0.0001	3.101,33
Associação Homogênea	C*P C*F P*F	232,19	100	<0.0001	-771,78
Modelo Saturado	C*P*F	-	0	-	-

Fonte: Elaborado pelo autor, com base nos microdados do IBGE.

O modelo saturado, representado na última linha da tabela 22, considera todas as interações contidas nos oito modelos anteriores. Da mesma forma, todos os modelos, partindo do modelo de independência mútua, consideram as associações do modelo anterior, caracterizando, assim, modelos log-lineares hierarquizados. Por considerar todas as associações possíveis, o modelo saturado possui zero graus de liberdade e se ajusta perfeitamente aos dados, esse modelo não é interessante para a análise da associação entre as variáveis.

Todos os modelos apresentaram alta significância para a respectiva razão de verossimilhança. Idealmente, conforme von Eye e Mun (2012), em modelos log-lineares a estatística da razão de verossimilhança deveria ser não significativa, indicando bom

ajustamento do modelo. Esse requisito não é atendido por nenhum dos modelos propostos.

O resultado, no entanto, é esperado dada a natureza dos dados utilizados. Para amostras muito grandes, como a PNAD, é necessário levar em consideração que as estatísticas de qualidade de ajustamento são diretamente influenciadas pelo tamanho da amostra (SILVA, 1990). A seleção do melhor modelo, neste caso, deve ser a escolha daquele com o menor resultado para a razão de verossimilhança possível, ainda que este seja significativo.

Uma medida adicional, indicada por Pastore e Silva (2000), é a utilização do critério de informação Bayesiano (*BIC*), que permite comparar diferentes modelos. A seleção do modelo pela estatística *BIC* deve considerar que seu valor cresce com a soma de quadrados residual e que, portanto, quanto mais baixo seu valor, melhor será o modelo numa comparação com os demais. Neste caso, conforme a tabela comparativa das razões de verossimilhança e *BICs*, o modelo de associação homogênea, com três interações duplas, é o melhor entre os nove estimados.

4.3.1. Modelo selecionado

O modelo de associação homogênea implica que a associação condicional entre *status* do pai (origem) e *status* do filho (destino) independe dos níveis da variável de coorte de idade (década), assim como a associação entre *status* do pai e coorte de idade independe dos estratos da variável *status* do filho e a associação entre *status* do filho e coortes de idade independe dos estratos do *status* do pai.

Para Pastore e Silva (2000), este modelo estabelece que as distribuições ocupacionais se alteram ao longo do tempo e a associação entre os *status* dos pais e dos filhos independe de uma terceira categorização, representada na análise dos autores pelos anos da pesquisa da PNAD.

Na tabela 23, a seguir, são apresentados os resultados para os efeitos individuais de cada variável e par de variáveis considerando o modelo de associação homogênea selecionado na seção anterior, aplicado à amostra da PNAD de 2014 descrita anteriormente.

Tabela 23. Estatísticas dos efeitos individuais e das associações consideradas no modelo de associação homogênea para a mobilidade social intergeracional brasileira, em 2014, considerando coortes de idade de 10 anos

Variável/Par de associação	Graus de liberdade	χ^2 de Wald	Pr > χ^2
Coorte de idade (C)	4	1.695,45	<0,0001
Status do pai (P)	5	3.928,91	<0,0001
Status do filho (F)	5	680,59	<0,0001
Associação entre Coorte de idade e Status do pai (C*P)	20	581,18	<0,0001
Associação entre Coorte de idade e Status do filho (C*F)	20	357,19	<0,0001
Associação entre Status do filho e Status do pai (P*F)	25	3.588,99	<0,0001
Razão de Verossimilhança	100	232,19	<0,0001

Fonte: Elaborado pelo autor, com base nos microdados do IBGE.

Além da estatística de ajustamento, abordada na seção anterior, são apresentadas as estatísticas para a influência das variáveis de forma isolada e das associações entre os três pares de variáveis. Um maior valor, absoluto, da estatística de teste χ^2 de Wald indica maior intensidade da associação. A partir dos resultados apresentados na tabela 23, observa-se que as associações mais fortes na determinação das frequências são o *status* do pai e a interação entre *status* do pai e *status* do filho.

O modelo aponta, portanto, que a força da associação entre *status* do filho e *status* do pai é forte, só não superando a influência do *status* do pai. Para avaliar se essa associação é observada na amostra como um todo foi estimado um modelo log-linear bivariado, considerando apenas o *status* do pai e o *status* do filho como variáveis. Na tabela 24, são apresentados os resultados para este exercício, considerando todas as associações possíveis entre essas duas variáveis, ou seja, caracterizando o modelo log-linear em sua forma saturada.

Tabela 24. Estatísticas dos efeitos individuais e das associações consideradas no modelo saturado para a mobilidade social brasileira em 2014

Variável/Par de associação	Graus de liberdade	χ^2 de Wald	Pr > χ^2
Status do pai (P)	5	4.460,83	<0,0001
Status do filho (F)	5	782,58	<0,0001
Associação entre Status do filho e Status do pai (P*F)	25	3.591,17	<0,0001
Razão de Verossimilhança	0	.	.

Fonte: Elaborado pelo autor, com base nos microdados do IBGE.

O modelo saturado, como colocado anteriormente, se ajusta perfeitamente aos dados e não constitui interesse na análise de forma isolada. No entanto, em uma comparação direta entre ele e o modelo trivariado de associação homogênea, é possível observar a força das associações descritas. Neste sentido, observa-se que, novamente, há forte influência do *status* do pai e da associação entre o *status* de origem e o *status* de destino, sendo estes dois superiores ao efeito do *status* do filho.

Propõe-se uma análise de razões de chances (ORs) extraídas de cada uma das matrizes de mobilidade intergeracional, descritas nas tabelas de 15 a 20, para avaliar se há diferenças substanciais nas desigualdades de oportunidade de mobilidade social entre a amostra geral e as coortes de idade. Cabe sinalizar que uma razão de chances igual a 1 significa ausência de desigualdade de oportunidades de mobilidade social, ou seja, não há associação. Para razões de chances superiores a 1, a probabilidade de atingir determinada posição de destino é maior para a origem considerada do que para a origem adotada como referência.

A escolha das razões de chances propostas busca identificar se há, primeiro, divergência entre as desigualdades de oportunidades de mobilidade entre as diferentes coortes de idade, e, depois, de que forma as desigualdades se caracterizam para posições na estrutura sócio-ocupacional brasileira quando se alteram os pontos de origem.

A partir do resultado dos modelos log-lineares com relação à força da associação entre as variáveis consideradas – *status* do filho, *status* do pai e coorte de idade –, que revelou uma forte associação entre *status* do filho e do pai e, ainda, um efeito ainda mais forte do *status* do pai, foram escolhidos seis exemplos de razões de chances, calculados a partir das matrizes de mobilidade observadas e apresentadas anteriormente.

Os exemplos de razões de chances mantêm inalterado o *status* dos filhos nas comparações, enquanto é alterado o status dos pais, ou seja, estamos comparando as chances de estar em um determinado estrato sócio-ocupacional dado o estrato de origem do indivíduo. A tabela 25 descreve os exemplos que parte do destino no estrato baixo-inferior para filhos de pais do estrato baixo-inferior e filhos de pais do estrato alto.

Tabela 25. Descrição das Razões de Chances (ORs) adotadas como exemplo

Razão de Chances (OR)	<i>Chance no numerador</i> $\left(\frac{p_1}{1-p_1}\right)$	<i>Chance no denominador</i> $\left(\frac{p_0}{1-p_0}\right)$
OR1	Filho no estrato baixo-inferior, com origem no estrato baixo-inferior	Filho no estrato baixo-inferior, com origem no estrato alto
OR2	Filho no estrato baixo-superior, com origem no estrato baixo-inferior	Filho no estrato baixo-superior, com origem no estrato alto
OR3	Filho no estrato médio-inferior, com origem no estrato baixo-inferior	Filho no estrato médio-inferior, com origem no estrato alto
OR4	Filho no estrato médio-médio, com origem no estrato baixo-inferior	Filho no estrato médio-médio, com origem no estrato alto
OR5	Filho no estrato médio-superior, com origem no estrato baixo-inferior	Filho no estrato médio-superior, com origem no estrato alto
OR6	Filho no estrato alto, com origem no estrato baixo-inferior	Filho no estrato alto, com origem no estrato alto

Fonte: Elaborado pelo autor.

A figura 5 apresenta, graficamente, os resultados de cada uma das ORs para cada uma das cinco coortes de idade adotadas neste estudo e as razões de chances para a amostra geral.

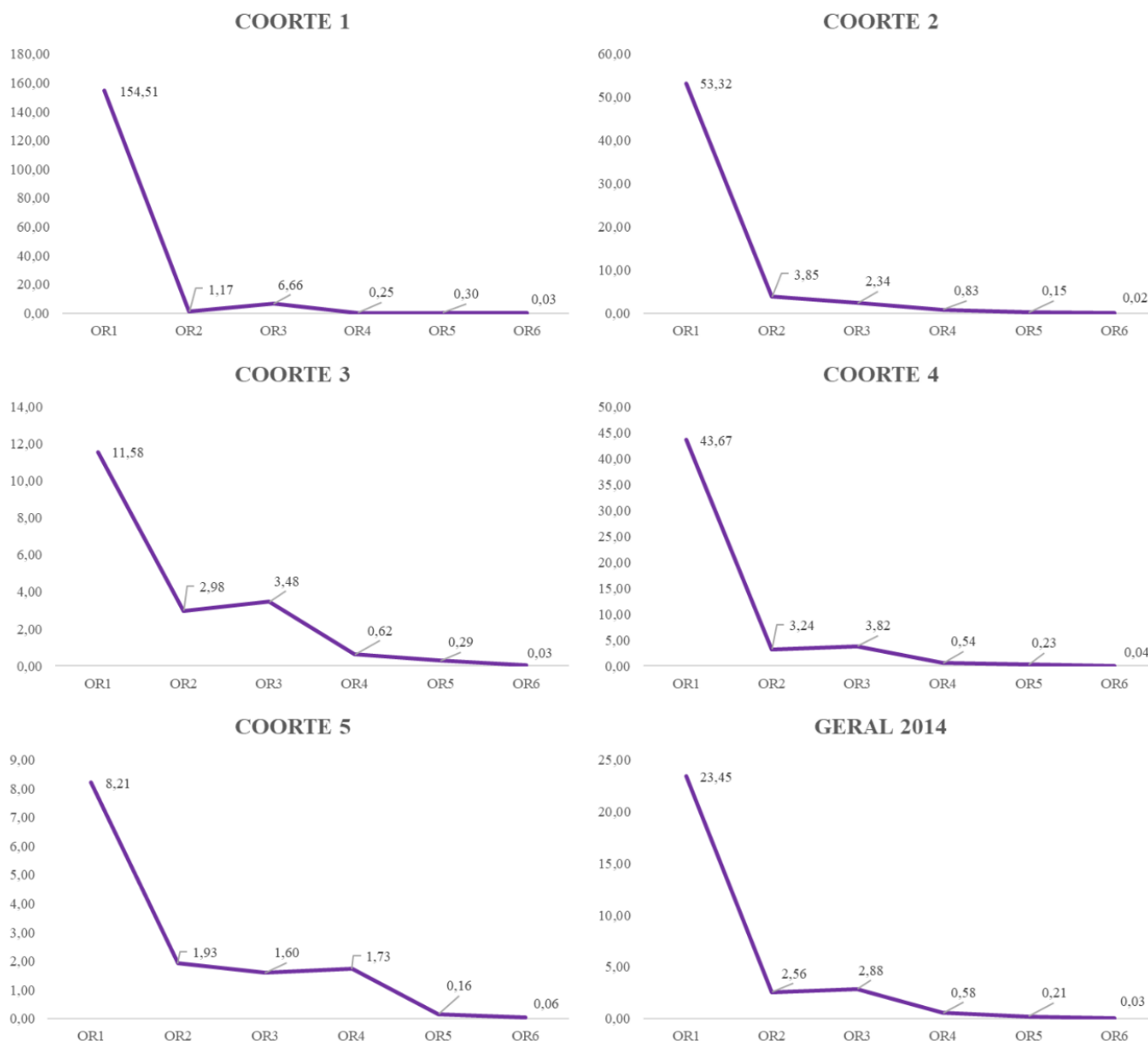


Figura 5. Resultados para as Razões de Chances (ORs) para cada coorte de idade e amostra geral da sociedade brasileira, em 2014

Fonte: Elaborado pelo autor.

A trajetória das ORs ilustradas na figura 5 revela que a razão entre chances de estar em um estrato, dependendo da origem do pai, diminui conforme o estrato de origem se aproxima do topo da estrutura sócio-ocupacional. Na OR 1, são comparados indivíduos que estão no estrato baixo-inferior e possuem origem no estrato baixo-inferior, configurando a imobilidade, com indivíduos do mesmo estrato baixo-inferior com origem no estrato alto, configurando uma grande mobilidade descendente. Neste cenário, observa-se que os filhos de pais do estrato baixo-inferior apresentam chance muito elevada de permanecer neste estrato quando comparada à chance de que filhos de pais do estrato alto apresentem mobilidade descendente de longa distância.

Dada a magnitude muito dispare entre os resultados das razões de chances do gráfico 1, o que dificulta uma visualização dos resultados a partir da OR 2, as ORs são apresentadas também na tabela 26.

Tabela 26. Resultados para as Razões de Chances (ORs) para cada coorte de idade e amostra geral da sociedade brasileira, em 2014.

COORTE-	Razão de Chances (OR)					
	OR1	OR2	OR3	OR4	OR5	OR6
COORTE 1	154,51	1,17	6,66	0,25	0,30	0,03
COORTE 2	53,32	3,85	2,34	0,83	0,15	0,02
COORTE 3	11,58	2,98	3,48	0,62	0,29	0,03
COORTE 4	43,67	3,24	3,82	0,54	0,23	0,04
COORTE 5	8,21	1,93	1,60	1,73	0,16	0,06
GERAL 2014	23,45	2,56	2,88	0,58	0,21	0,03

Fonte: Elaborado pelo autor.

Observa-se que, conforme o status de destino aumenta – da OR 1, que representa o destino no estrato baixo-inferior, para a OR 6, que representa o destino no estrato alto – as razões de chances, tanto nas coortes quanto na amostra geral, diminuem.

Note-se, por exemplo, que, de acordo com os valores da OR 6, em todos os recortes, a razão de chances entre um filho estar no estrato alto tendo pai no estrato baixo-inferior e um filho estar no mesmo estrato alto tendo pai no estrato alto é muito inferior a 1, indicando que a chance de imobilidade para filhos do estrato alto é muito maior do que a chance de mobilidade ascendente de longa distância para filhos do estrato baixo-inferior.

Como é possível observar, existe mudança entre as razões de chances das diferentes coortes de idade, indicando que há influência da coorte de idade sobre a associação pai-filho.

4.4. Uma discussão sobre a maneira de avaliar a mobilidade

A proposição e utilização de modelos log-lineares no estudo da mobilidade social atingiu seu auge entre os anos 1980 e 1990. Uma das justificativas para sua adoção é a de que o modelo, além de se adequar bem à análise de tabelas de

classificação cruzada, permite um estudo apurado da associação entre variáveis selecionadas no contexto da mobilidade, bem como a observação de mudanças nos padrões de mobilidade entre diferentes amostras ao se considerar, por exemplo, uma variável categórica distinguindo diferentes períodos.

Os modelos log-lineares, como coloca Kelley (1990), ganharam atenção no estudo da mobilidade social com o trabalho de Goldthorpe e seus associados entre as décadas de 1970 e 1980 (ERIKSON; GOLDTHORPE, 1985; ERIKSON; GOLDTHORPE; PORTOCARERO, 1983; GOLDTHORPE, 1971; 1985; GOLDTHORPE; LLEWELLYN; PAYNE, 1980) e, a seu modo, despertaram grande interesse em detrimento dos métodos propostos por Blau e Duncan (1967).

Entre os aspectos determinantes para o advento dos modelos log-lineares no estudo da mobilidade, em oposição às medidas tradicionais propostas por Yasuda (1964) e análises de trajetórias, estão a maior disponibilidade de dados de qualidade e amostras maiores. Neste sentido, os ganhos ocasionados pelo uso da nova metodologia não superam suas limitações que incluem: poder estatístico limitado do método em lidar com questões de hierarquia e medida da associação; a impossibilidade de trabalhar com medidas contínuas em contraposição a categorias e estratos discretos; e a dificuldade de incorporar variáveis relevantes para a mobilidade social no modelo, como por exemplo a educação, rendimento, entre outras (KELLEY, 1990).

Ainda que seja possível mitigar algumas dessas questões, como é o caso da incorporação de variáveis como educação e rendimento na determinação de estruturas sócio-ocupacionais, há outras como a influência do número de categorias adotadas sobre os resultados e o fato de que em amostras muito grandes é muito difícil obter um modelo, que não seja o saturado, que se ajuste bem aos dados, como costuma ser o caso na utilização de *surveys* como a PNAD, o que torna menos justificável a sua utilização.

Além dos obstáculos para a aplicação de modelos log-lineares, representados pelas questões apresentadas anteriormente, estes são menos intuitivos que os modelos disseminados pela primeira e segunda geração de metodologias aplicadas ao estudo da mobilidade e apresentam, ainda, diferentes interpretações e perspectivas para a sua interpretação, como pode ser observado na contraposição entre a análise conduzida por Pastore e Silva (2000), que procuram interpretar modelos log-lineares a partir da utilização de medidas clássicas e os conceitos de mobilidade circular e estrutural, e a posição de Ribeiro (2007), que refuta a validade destes conceitos na análise da mobilidade social.

Os modelos log-lineares possuem relevância para o caso de amostras pequenas, assim como a análise de variância para dados experimentais, mas para amostras muito grandes estão longe de ser fundamentais para analisar corretamente a mobilidade social como sugere Ribeiro (2007). Uma vez que os modelos log-lineares se propõem a explorar a contribuição de variáveis sobre a variabilidade nas frequências de tabelas de classificação cruzada e a interação delas entre si, aponta-se, como alternativa para a análise das interações entre variáveis, a existência de medidas mais disseminadas e de fácil entendimento que cumprem essa função satisfatoriamente.

O comportamento das medidas de associação para as diferentes coortes de idade corrobora os resultados observados para os modelos log-lineares, que demonstram haver associação entre as duas variáveis, e para os resultados das razões de chances propostas para cada coorte de idade, que demonstraram que essa associação não é constante ao longo das coortes de idade. Essas medidas, portanto, suprem, de forma satisfatória, a necessidade de entender de que forma se relacionam a origem e o destino do indivíduo nas suas oportunidades de mobilidade, além de serem medidas de fácil aplicação em tabelas de contingência e não oferecerem dificuldades de ajustamento ou interpretação de resultados para amostras muito grandes, como verificou-se na aplicação de modelos log-lineares no contexto da mobilidade sócio-ocupacional intergeracional brasileira em 2014.

5. CONSIDERAÇÕES FINAIS

O estudo da mobilidade social tem constituído grande interesse na literatura sociológica e econômica ao longo das últimas três décadas sem, no entanto, oferecer um arcabouço metodológico consolidado para a mensuração do fenômeno. Ao longo dos anos, as medidas de mobilidade absoluta propostas na primeira geração metodológica de estudos foram substituídas por metodologias que traduzem o grande interesse na dimensão relativa da mobilidade, caracterizada pelas desigualdades de oportunidades de mobilidade social experimentadas pelos indivíduos.

Neste sentido, por muitos anos os modelos log-lineares representaram para os sociólogos o método fundamental para o estudo da influência do *status* de origem do indivíduo sobre as suas chances de mobilidade dentro da estrutura social (RIBEIRO, 2007). Com o interesse de economistas pela área de estudo houve, então, a proposição de uma nova metodologia para a mensuração desta relação. Trata-se da abordagem da mobilidade intergeracional de renda, que se baseia, essencialmente, na análise do coeficiente de regressão do modelo de regressão linear do logaritmo do rendimento do filho contra o logaritmo do rendimento do pai – o *Intergenerational Earnings Elasticity* (IGE) (BECKER; TOMES, 1986; SAKAMOTO; WANG, 2020; SOLON, 1992; ZIMMERMAN, 1992).

Partindo do pressuposto de que os modelos log-lineares oferecem uma análise minuciosa da mobilidade sócio-ocupacional, este estudo se propôs a aplicá-lo, em detrimento dos modelos de regressão utilizados pela literatura econômica para o estudo da mobilidade, e verificar se os resultados obtidos são satisfatórios no entendimento do fenômeno.

Os dados da PNAD 2014, utilizados para esse exercício, fornecem informações sobre a ocupação do indivíduo e, também, a variável de maior relevância no estudo da mobilidade sócio-ocupacional intergeracional, a ocupação do pai quando o indivíduo tinha 15 anos. Deste modo, utilizando a escala sócio-ocupacional desenvolvida por Oliveira (2018) e Oliveira e Hoffmann (2021), foram obtidas as matrizes de mobilidade sócio-ocupacional para o Brasil em 2014 para cinco coortes de idade de dez anos cada.

Os modelos log-lineares estimados apontam para a existência e força das associações entre *status* do pai e *status* do filho, *status* do filho e coorte de idade, e *status* do pai e coorte de idade. A associação mais relevante no contexto das matrizes de

mobilidade observadas é a do *status* do pai, seguida pela associação entre as variáveis de origem e destino.

A aplicação do modelo log-linear na mobilidade social revelou-se um esforço metodológico que traz pouco ganho, em comparação com a adoção de medidas de associação tradicionais, dado seu caráter pouco intuitivo e de difícil interpretação no contexto apresentado. De modo geral, é possível concordar com Kelley (1990): a utilização de modelos log-lineares configura “um preço alto a pagar pela elegância estatística”³² (KELLEY, 1990, p. 320).

Soma-se a isso o fato de que a análise da dimensão relativa da mobilidade revela-se insuficiente para o entendimento do fenômeno no seu aspecto mais relevante: a direção dos movimentos experimentados pelos indivíduos móveis. Neste sentido, observa-se que a utilização dos modelos de regressão proposta pela literatura econômica da mobilidade social é insuficiente não apenas pela observação exclusiva do caráter relativo do fenômeno, como também pela simplificação (ignorância) da posição do indivíduo na sociedade, que deixa de lado toda a literatura sociológica construída através das teorias da estratificação social e proposição de conceitos acerca da posição do indivíduo na estrutura social, que supera o conceito simplista da renda.

Já a utilização de modelos log-lineares, defendida por parte da literatura sociológica no estudo da mobilidade social como o auge metodológico na área, apesar de considerar a teoria da estratificação na definição do status do indivíduo na estrutura social, oferece pouco à análise do fenômeno, pois, além de não oferecer a dimensão absoluta do fenômeno, configura uma análise influenciada pela definição do número de estratos e de difícil interpretação em comparação direta às medidas tradicionais de associação e, ainda, às medidas tradicionais de mobilidade social da primeira geração metodológica da área.

Por fim, a utilização de medidas de mobilidade tradicionais, que mensuram o aspecto absoluto do fenômeno, em conjunto com medidas de associação estatísticas tradicionais suprem, de modo satisfatório, a análise do fenômeno considerando as duas dimensões da mobilidade: absoluta e relativa. Ambas apresentam a vantagem de possuir fácil aplicação, entendimento e interpretação, preservando seu caráter informativo e não oferecem os desafios de ajustamento e interpretação como se verificou na aplicação de

³² “*It is too high a price to pay for statistical elegance.*” (KELLEY, 1990, p. 320)

modelos log-lineares no contexto da mobilidade sócio-ocupacional intergeracional brasileira em 2014.

REFERÊNCIAS

- AGHEVLI, B.; MEHRAN, F. Optimal grouping of income distribution data. **Journal of the American Statistical Association**, 76, n. 373, p. 22-26, 1981.
- AGRESTI, A. **An introduction to categorical data analysis**. John Wiley & Sons, 2018. 1119405262.
- BARTHOLOMEW, D. J. **Stochastic Models for Social Processes**. Reino Unido: John Wiley & Sons, 1967.
- BECKER, G. S.; TOMES, N. Human Capital and the Rise and Fall of Families. **Journal of Labor Economics**, 4, n. 3, p. S1-S39, 1986.
- BEHRMAN, J. R. Social mobility: concepts and measurement. *In*: BIRDSALL, N. e GRAHAM, C. (Ed.). **New markets, new opportunities?: Economic and social mobility in a changing world**: Brookings Institution Press, 2000.
- BENINI, R. **Gruppi chiusi e gruppi aperti in alcuni fatti collettivi di combinazioni**. Imprimerie Nationale, 1927.
- BERMAN, Y. Understanding the Mechanical Relationship between Inequality and Intergenerational Mobility. **SSRN Electronic Journal**, 2017.
- BERMAN, Y. The Long-Run Evolution of Absolute Intergenerational Mobility. **American Economic Journal: Applied Economics**, 14, n. 3, p. 61-83, 2022.
- BJÖRKLUND, A.; JÄNTTI, M. Intergenerational income mobility and the role of family background. *In*: SALVERDA, W.; NOLAN, B., *et al* (Ed.). **Oxford handbook of Economic Inequality**. Oxford: Oxford University Press, 2011.
- BLANDEN, J. Cross-country rankings in intergenerational mobility: a comparison of approaches from economics and sociology. **Journal of Economic Surveys**, 27, n. 1, p. 38-73, 2013-02-01 2013.

- BLAU, P. M.; DUNCAN, O. D. **The American occupational structure**. New York: The Free Press, 1967.
- BREEN, R. Individual Level Models for Mobility Tables and other Cross-Classifications. **Sociological Methods & Research**, 23, n. 2, p. 147-173, 1994.
- BROWN, M. B.; BENEDETTI, J. K. Sampling behavior of tests for correlation in two-way contingency tables. **Journal of the American Statistical Association**, n. 72, p. 309-315, 1977.
- BRYANT, C. G. A. **Practical Sociology**. Cambridge: Polity Press, 1995.
- CAILLAUX, E. L. Cor e mobilidade social no Brasil. **Estudos Afro-Asiáticos**, 26, p. 53-66, 1994.
- CARLSSON, G. **Social Mobility and Class Structure**. C. W. K. Gleerup, 1958.
- CHATTOPADHYAY, N.; DEWANJI, A.; ROY, B. K. Re-examining Bartholomew's Index of Social Mobility: Some New Derivations with Statistical Inference. **Calcutta Statistical Association Bulletin**, 71, n. 2, p. 99-112, 2019.
- CHETTY, N.; HENDREN, N.; KLINE, P.; SAEZ, E. Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States. 129, n. 4, 2014.
- COLE, S. Why Sociology Doesn't Make Progress Like the Natural Sciences. **Sociological Forum**, 9, n. 2, p. 133-154, 1994.
- CORAK, M. Inequality from generation to generation: the United States in comparison. *In*: RYCROFT, R. (Ed.). **The economics of inequality, poverty, and discrimination in the 21st Century**. Santa Barbara, CA: ABC-CLIO, 2013. p. 613.
- DAVIES, J. B.; SHORROCKS, A. F. Optimal grouping of income and wealth data. **Journal of econometrics**, 42, n. 1, p. 97-108, 1989.

DAVIS, K. **Human society**. New York: Macmillan, 1949.

DAVIS, K.; MOORE, W. E. Some principles of stratification. **American sociological review**, 10, n. 2, p. 242-249, 1945.

DIPRETE, T. A. Adding Covariates to Loglinear Models for the Study of Social Mobility. **American Sociological Review**, 55, n. 5, p. 757-773, 1990.

DUNCAN, O. D. Methodological issues in the analysis of social mobility. *In*: SMELSER, N. J. e LIPSET, S. M. (Ed.). **Social structure and mobility in economic development**. Chicago, Illinois: Aldine Publishing Company, 1966. cap. 2, p. 51-97.

ERICKSON, R.; GOLDTHORPE, J. H. The CASMIN Project and the American dream. **European Sociological Review**, n. n. 8, p. 283-306, 1992.

ERICKSON, R.; GOLDTHORPE, J. H.; PORTOCARRERO, L. Intergenerational mobility in Three Western European industrial societies. **British Journal of Sociology**, n. n.30, 1979.

ERIKSON, R.; GOLDTHORPE, J. H. Are American Rates of Social Mobility Exceptionally High? New Evidence on an Old Issue. **European Sociological Review**, 1, n. 1, p. 1-22, 1985.

ERIKSON, R.; GOLDTHORPE, J. H. **The Constant Flux: A Study of Class Mobility in Industrial Societies**. Clarendon Press, 1992.

ERIKSON, R.; GOLDTHORPE, J. H.; PORTOCARERO, L. Intergenerational class mobility and the convergence thesis: England, France and Sweden. **The British Journal of Sociology**, 34, n. 3, p. 303-343, 1983.

FEATHERMAN, D. L.; LANCASTER JONES, F.; HAUSER, R. M. Assumptions of social mobility research in the U.S.: The case of occupational status. **Social Science Research**, 4, n. 4, p. 329-360, 1975.

- FIELDS, G. S. Income mobility: concepts and measures: patterns and underlying causes. <https://digitalcommons.ilr.cornell.edu/articles/1106/>: Cornell University, ILR School 2000.
- FLEISS, J. L.; LEVIN, B.; PAIK, M. C. **Statistical methods for rates and proportions**. 3rd ed. Hoboken, Nova Jersey: John Wiley & Sons, 2003.
- FLYVBJERG, B. Is theory possible in social science? *In*: FLYVBJERG, B. e SAMPSON, S. (Ed.). **Making Social Science Matter: Why Social Inquiry Fails and How it Can Succeed Again**. Cambridge: Cambridge University Press, 2001. p. 25-37.
- FOOTE, N. N.; HATT, P. K. Social Mobility and Economic Advancement. **The American Economic Review**, 43, n. 2, p. 364-378, 1953.
- FRIEDMAN, M.; NATIONAL BUREAU OF ECONOMIC RESEARCH. **A theory of the consumption function**. Princeton,: Princeton University Press, 1957. 243 p. p. (National Bureau of Economic Research General series,, v. no 63).
- GALTON, F. Regression Towards Mediocrity in Hereditary Stature. **The Journal of the Anthropological Institute of Great Britain and Ireland**, 15, p. 246-263, 1886.
- GLASS, D. V. **Social Mobility in Britain**. California: Free Press, 1954.
- GOLDTHORPE, J. H. THEORIES OF INDUSTRIAL SOCIETY: REFLECTIONS ON THE RECRUDESCENCE OF HISTORICISM AND THE FUTURE OF FUTUROLOGY. **European Journal of Sociology / Archives Européennes de Sociologie / Europäisches Archiv für Soziologie**, 12, n. 2, p. 263-288, 1971.
- GOLDTHORPE, J. H. On Economic Development and Social Mobility. **The British Journal of Sociology**, 36, n. 4, p. 549-573, 1985.

- GOLDTHORPE, J. H. Progress in sociology: The case of social mobility research. *In*: SVALLFORS, S. (Ed.). **Analyzing inequality: Life chances and social mobility in comparative perspective**. Stanford, California: Stanford University Press, 2005. v. 71, cap. 3, p. 56-82.
- GOLDTHORPE, J. H. Understanding – and Misunderstanding – Social Mobility in Britain: The Entry of the Economists, the Confusion of Politicians and the Limits of Educational Policy. **Journal of Social Policy**, 42, n. 3, p. 431-450, 2013.
- GOLDTHORPE, J. H.; LLEWELLYN, C.; PAYNE, C. **Social Mobility and Class Structure in Modern Britain**. Clarendon Press, 1980.
- GOODMAN, L. A. Multiplicative models for the analysis of occupational mobility tables and other kinds of cross-classification tables. **American Journal of Sociology**, 84, n. 4, p. 804-819, 1979a.
- GOODMAN, L. A. Simple Models for the Analysis of Association in Cross-Classifications having Ordered Categories. **Journal of the American Statistical Association**, 74, n. 367, p. 537-552, 1979b.
- HAUSER, R. M. A Structural Model of the Mobility Table. **Social Forces**, 56, n. 3, p. 919-953, 1978.
- HAUSER, R. M.; KOFFEL, J. N.; TRAVIS, H. P.; DICKINSON, P. J. Temporal Change in Occupational Mobility: Evidence for Men in the United States. **American Sociological Review**, 40, n. 3, p. 279-297, 1975.
- HOFFMANN, R. Desigualdade de renda no Brasil em 2005 e sua decomposição. **Centro de Gestão e Estudos Estratégicos. II Seminário de Análise dos Resultados da Pesquisa Nacional por Amostra de Domicílios**, 2005.
- HOFFMANN, R.; BOTASSIO, D. C.; JESUS, J. G. Distribuição de renda: medidas de desigualdade, pobreza, concentração, segregação e polarização. São Paulo, Brasil: Edusp 2019.

HOPE, K. Vertical and Nonvertical Class Mobility in Three Countries. **American Sociological Review**, 47, n. 1, p. 99-113, 1982.

HOUT, M. **Mobility tables**. Beverly Hills: Sage Publications, 1983.

HUTCHINSON, B. A. **Mobilidade e trabalho: um estudo na cidade de São Paulo**. Centro Brasileiro de Pesquisas Educacionais, INEP, Ministério da Educação e Cultura, 1960.

JANNUZZI, P. D. M. Construção de uma escala sócio-ocupacional: notas metodológicas. **Revista Brasileira de Estatística**, 60, n. 2, 1999.

JESUS, J. G. D. **Negros em movimento: migração e desigualdade racial no Brasil**. 2020. (Doutorado em Economia Aplicada) -, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. Disponível em: <https://dx.doi.org/10.11606/t.11.2020.tde-18052020-102639> Acesso em: 5 de outubro de 2021.

JONES, F. L. STRUCTURAL AND CIRCULATION MOBILITY ARE ALIVE AND WELL: A CRITICAL COMMENT ON SOBEL'S PREMATURE OBITUARY NOTICE. **Sociology**, 19, n. 1, p. 139-142, 1985.

KELLEY, J. The failure of a Paradigm: Log-linear models of social mobility. *In*: CLARK, J. (Ed.). **John Goldthorpe: Consensus And Controversy**. Londres: Routledge, 1990. cap. 20, p. 319-346.

KENDALL, M. G.; STUART, A. **The advanced theory of statistics**. 4 ed. nova York: Macmillan, 1979.

KRUEGER, A. B. **The rise and consequences of inequality**, 2012. Disponível em: https://obamawhitehouse.archives.gov/sites/default/files/krueger_cap_speech_final_remarks.pdf Acesso em: 3 de outubro de 2021.

KUZNETS, S. Economic Growth and Income Inequality. **The American Economic Review**, 45, n. 1, p. 1-28, 1955.

LIEBETRAU, A. **Measures of Association**. Thousand Oaks, California: 1983. DOI 10.4135/9781412984942. Disponível em: <https://methods.sagepub.com/book/measures-of-association> Acesso em: 19 de outubro de 2023.

LIPSET, S. M.; BENDIX, R. **Social mobility in industrial society**. Berkeley: University of California Press, 1959.

LIPSET, S. M.; ZETTERBERG, H. L. A theory of social mobility. *In*: ASSOCIATION, I. S. (Ed.). **Transactions of the third world congress of sociology**. London: International Sociological Association, 1956. v. iii.

LIPSET, S. M.; ZETTERBERG, H. L. Social mobility in industrial societies *In*: LIPSET, S. M. e BENDIX, R. (Ed.). **Social mobility in industrial society**. Berkeley: University of California, 1959.

LIVI, L. Sur la mesure de la mobilité sociale. Résultats d'un sondage sur la population italienne. **Population**, p. 65-76, 1950.

LOGAN, J. A. A Multivariate Model for Mobility Tables. **American Journal of Sociology**, 89, n. 2, p. 324-349, 1983.

LOPES, J. R. B. **Desenvolvimento e mudança social: formação da sociedade urbano-industrial no Brasil**. Rio de Janeiro: Centro Edelstein de Pesquisas Sociais, 2008. 233 p.

MAHLMEISTER, R.; FERREIRA, S. G.; VELOSO, F.; MENEZES-FILHO, N. *et al.* Revisitando a mobilidade intergeracional de educação no Brasil. **Revista Brasileira de Economia**, 73, n. 2, p. 159-180, 2019.

MAJOR, L. E.; MACHIN, S. **Social Mobility: And Its Enemies**. Penguin Books Limited, 2018.

MANKIW, G. Observations on the Great Gatsby Curve. Greg Mankiw's Blog. MANKIW, G. 2013. Disponível em: <http://gregmankiw.blogspot.com/2013/07/some-observations-on-great-gatsby-curve.html>. Acesso: 18 de junho de 2021.

MONTAGNER, P.; TAIRA, E.; JANNUZZI, P. Migração e Mobilidade Social: revisitando a discussão a partir da Pnad 2014. **Revista Ciências do Trabalho**, n. 13, 2019.

OLIVEIRA, C. **Mobilidade sócio-ocupacional no Brasil no novo milênio**. 2018. (Mestrado em Economia Aplicada) -, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba.

OLIVEIRA, C.; HOFFMANN, R. Mobilidade socio-ocupacional no Brasil: novo procedimento para delimitação dos estratos ocupacionais e análise dos dados da PNAD de 2014. **Sociologias**, 23, n. 57, p. 208-238, 2021.

PARSONS, T. **The social system**. New York: Free Press, 1951.

PARSONS, T. An Analytical Approach to the Theory of Social Stratification'in Essays in Sociological Theory, Glencoe. Free Press, rev. ed 1954.

PASTORE, J. **Desigualdade e mobilidade social no Brasil**. TA Queiroz, 1979.

PASTORE, J.; HALLER, A. O. O que está acontecendo com a mobilidade social no Brasil? *In*: VELLOSO, J. P. D. R. e ALBUQUERQUE, R. C. D. (Ed.). **Pobreza e mobilidade social**. São Paulo: Nobel, 1993. cap. 1.

PASTORE, J.; SILVA, N. D. V. **Mobilidade social no Brasil**. Makron Books, 2000.

- PAYNE, G. **The new social mobility. How the politicians got it wrong.** 1 ed. Bristol University Press, 2017.
- RIBEIRO, C. A. C. **Estrutura de classe e mobilidade social no Brasil.** Edusc São Paulo, 2007.
- RIBEIRO, C. A. C. Tendências da desigualdade de oportunidades no Brasil: mobilidade social e estratificação educacional. **Mercado de trabalho: conjuntura e análise (IPEA)**, 62, p. 49-65, 2017.
- ROEMER, J. E. **Equality of opportunity.** Cambridge, MA: Harvard University Press, 1998.
- ROGOFF, N. Recent trends in urban occupational mobility. *In*: HATT, P. K. e REISS, A. J. J. (Ed.). **Reader in urban sociology.** Illinois: Glencoe, 1951. p. 406-420.
- ROGOFF, N. Recent trends in urban occupational mobility. *In*: HATT, P. K. e REISS, A. J. J. (Ed.). **Cities and society: the revised reader in urban sociology.** 5 ed. Glencoe: Free Press, 1964. p. 432 - 445.
- RUDAS, T. **Odds Ratios in the Analysis of Contingency Tables.** Thousand Oaks, California: SAGE, 1998.
- SAKAMOTO, A.; WANG, S. X. The declining significance of occupation in research on intergenerational mobility. **Research in Social Stratification and Mobility**, p. 100521, 2020.
- SALSBURG, D. S.; GRADEL, J. M. **Uma senhora toma chá ... como a estatística revolucionou a ciência no século XX.** Zahar, 2009.
- SCALON, M. C. **Mobilidade social no Brasil: padrões e tendências.** Editora Revan, 1999.

- SCHUMPETER, J. A. **Imperialism and Social Classes: Two Essays**. Ludwig von Mises Institute, 1955.
- SILVA, N. D. V. Posição social das ocupações. **Rio de Janeiro: IBGE, Centro de Informática**, 1973.
- SILVA, N. D. V. As duas faces da mobilidade. **Dados: Revista de Ciências Sociais**, n. 21, p. 46-67, 1979.
- SILVA, N. D. V. **Introdução à análise de dados qualitativos**. São Paulo: Vértice, Editora Revistas dos Tribunais, 1990.
- SNEDECOR, G. W.; COCHRAN, W. G. **Statistical Methods**. 8th ed. Ames: Iowa State University Press, 1989.
- SOBEL, M. E. Structural Mobility, Circulation Mobility and the Analysis of Occupational Mobility: A Conceptual Mismatch. **American Sociological Review**, 48, n. 5, p. 721-727, 1983.
- SOLON, G. Intergenerational income mobility in the United States. **The American Economic Review**, p. 393-408, 1992.
- SOLON, G. A model of intergenerational mobility variation over time and place. *In*: CORAK, M. (Ed.). **Generational Income Mobility in North America and Europe**. Cambridge: Cambridge University Press, 2004. p. 38-47.
- SOROKIN, P. A. **Social mobility**. New York: Harper, 1927.
- SOROKIN, P. A. **Social and cultural mobility**. 1 ed. New York: Free Press, 1959.
- SVALASTOGA, K. **Prestige, Class, and Mobility**. Scandinavian University Books, 1959.

TORCHE, F. Intergenerational Mobility and Inequality: The Latin American Case. **Annual Review of Sociology**, 40, p. 619-642, 2014.

UEDA, E. M. **Educação e rendimentos: uma abordagem econométrica**. 2001. Dissertação - Instituto de Economia, UNICAMP, Campinas.

VON EYE, A.; MUN, E. Y. **Log-Linear Modeling: Concepts, Interpretation, and Application**. First edition ed. Hoboken, New Jersey: Wiley, 2013.

VON EYE, A.; MUN, E. Y.; MAIR, P. Log-linear modeling. **Wiley Interdisciplinary Reviews: Computational Statistics**, 4, n. 2, p. 218-223, 2012.

YASUDA, S. A methodological inquiry into social mobility. **American Sociological Review**, p. 16-23, 1964.

ZIMMERMAN, D. J. Regression Toward Mediocrity in Economic Stature. **The American Economic Review**, 82, n. 3, p. 409-429, 1992.

APÊNDICES

APÊNDICE A. Estendendo a interpretação de Berman (2017) sobre a *Great Gatsby Curve*

A *Great Gatsby Curve* (GGC), curva empírica que ilustra a associação entre a mobilidade, mensurada pelo *Intergenerational Earnings Elasticity* (IGE), e a desigualdade de renda, mensurada pelo Coeficiente de Gini, amplamente discutida em círculos políticos e midiáticos, configura uma relação que é, pelo menos parcialmente, de natureza mecânica.

Berman (2017), em seu artigo “*Understanding the mechanical relationship between inequality and intergenerational mobility*”, discorre sobre a relação existente entre as duas medidas, que torna a GGC uma tautologia, chegando à conclusão de que medir a desigualdade através do Índice de Gini e a mobilidade intergeracional através do IGE é, sob certas condições, equivalente. Neste apêndice é explorada a relação entre as duas medidas, a partir da apresentação dos passos fundamentais da dedução desta relação.

A Figura 6, a seguir, mostra a *Great Gatsby Curve* (GGC), com a associação positiva representada por uma linha pontilhada, que é o melhor ajuste linear de mínimos quadrados ($R^2 = 0,6$) para o conjunto de dados de indivíduos que nasceram entre o início e metade da década de 1960 e tiveram suas rendas mensuradas no final da década de 1990 no recorte de países presentes no gráfico.

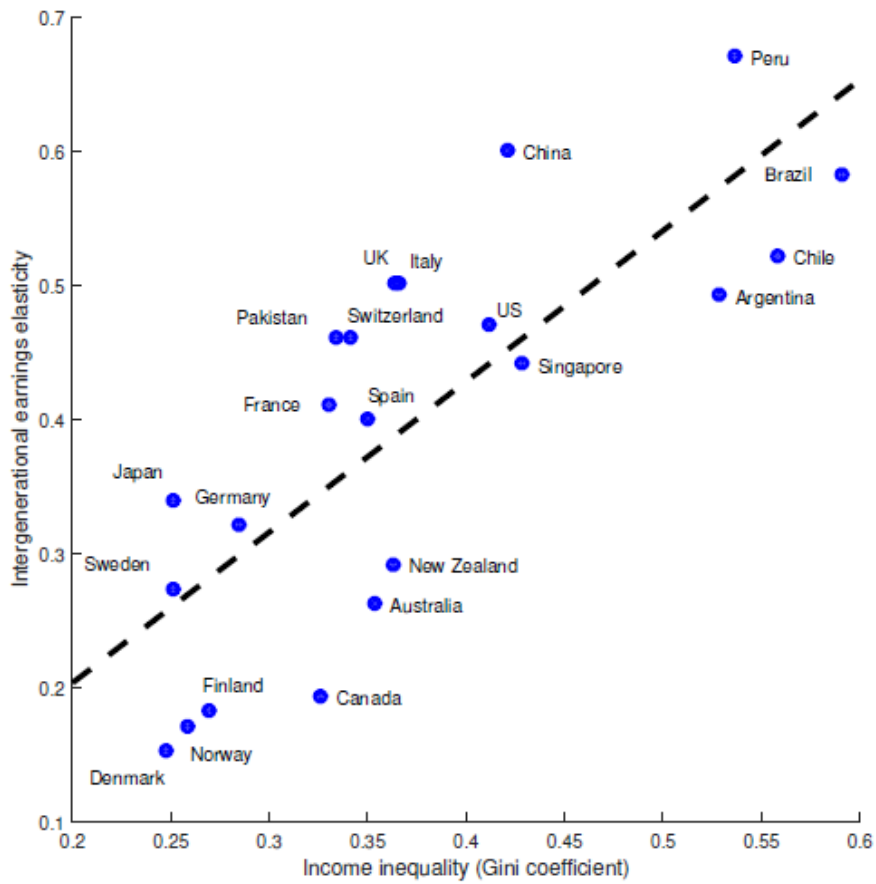


Figura 6. *Great Gatsby Curve* (GGC), relação entre o *Intergenerational Earnings Elasticity* (IGE) e o coeficiente de Gini, para países selecionados

Fonte: Extraído de Berman (2017, p. 8)

Sendo Y_i a renda do i -ésimo indivíduo e N o tamanho da população, o índice de Gini pode ser definido como

$$G = \frac{\sum_{i=1}^N \sum_{j=1}^N |Y_i - Y_j|}{2N \sum_{j=1}^N Y_j} \quad (52)$$

Define-se $X_{i,t} = \ln(Y_{i,t})$, sendo $Y_{i,t}$ a renda do i -ésimo indivíduo na geração t , e $X_{i,t-1} = \ln(Y_{i,t-1})$ como o valor correspondente na geração anterior. Sendo $\varepsilon_{i,t}$ um ruído branco com variância σ^2 , podemos escrever

$$X_{i,t} = \alpha + \beta X_{i,t-1} + \varepsilon_{i,t} \quad (53)$$

onde β corresponde ao IGE usado como medida da imobilidade ou persistência de situação econômica.

Quando $\beta^2 < 1$ tem-se um modelo autorregressivo de ordem 1, ou, AR(1), estacionário com média $\frac{\alpha}{(1-\beta)}$ e variância $\frac{\sigma^2}{(1-\beta^2)}$. Neste caso, a distribuição da renda $Y_{i,t}$ converge para uma distribuição log-normal com média $\frac{\alpha}{(1-\beta)}$ e variância:

$$S^2 = \frac{\sigma^2}{(1-\beta^2)} \quad (54)$$

De acordo com Hoffmann, Botassio e Jesus (2019, p. 178), o índice de Gini dessa distribuição é dado por

$$G = 2\Phi\left(\sqrt{\frac{\sigma^2}{2(1-\beta^2)}}\right) - 1 \quad (55)$$

onde Φ indica a função de distribuição de uma distribuição normal reduzida. Como essa função é sempre crescente, a expressão mostra que o índice de Gini é uma função crescente de β para $\beta^2 < 1$. Ressalte-se que até aqui apenas apresentamos, de maneira um pouco diferente, o argumento básico de Berman (2017). A expressão (55) é equivalente à expressão (3.2) em Berman (2017, p. 9).

A figura (7) mostra a forma da relação entre G e β para três valores da variância σ^2 .

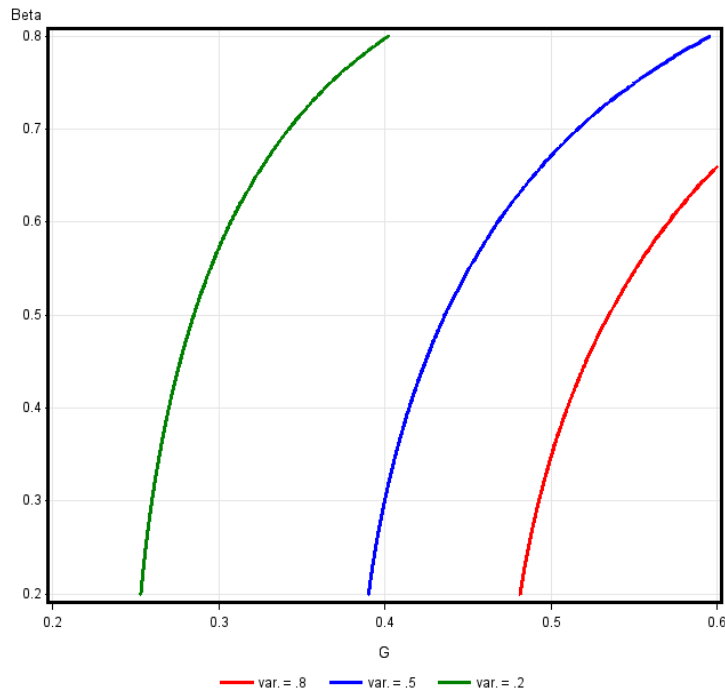


Figura 7. A relação entre o índice de Gini (G) e o valor de β (o IGE) para três valores da variância do erro da relação entre os logaritmos das rendas de pais e filhos [relação (55) com $\sigma^2 = 0,8$, $\sigma^2 = 0,5$ ou $\sigma^2 = 0,2$]

Fonte: Elaborado pelo autor.

Quando se compara a forma das curvas da Figura 7 com a dispersão de pontos para valores do índice de Gini e do IGE de vários países, Figura 6, verifica-se que as observações empíricas indicam uma curva mais afastada da vertical, inclinando-se mais suavemente para a direita. Berman explora várias alternativas para obter curvas mais semelhantes à sugerida pelos dados empíricos, considerando, por exemplo, um processo AR(2) em lugar do AR(1). Aqui vamos considerar uma nova alternativa, que consiste em admitir que a variância do erro (σ^2) tenda a crescer com a variância (S^2) de $X_{i,t}$.

Pode-se argumentar que não é razoável admitir que σ^2 (a variância do elemento aleatório em (53)) seja fixo. É plausível estabelecer que essa variância cresce com a variância do log da renda que, por sua vez, cresce com G .

De (54) e (55) segue-se que:

$$G = 2\Phi\left(\sqrt{\frac{S^2}{2}}\right) - 1 \quad (56)$$

e

$$\frac{1+G}{2} = \Phi\left(\sqrt{\frac{S^2}{2}}\right) \quad (57)$$

Vamos pressupor que σ cresça com a maior variabilidade da renda e do log da renda no país:

$$\sigma = \theta_0 + \theta_1 S \quad (58)$$

com $\theta_1 > 0$.

Escolhido um valor qualquer para S , o correspondente valor de G é obtido por meio de (56) e, dados θ_1 e θ_2 , o valor de σ é obtido de (58). Finalmente, tendo S e σ , a expressão (54) permite determinar β . Por meio desse procedimento são obtidas as curvas apresentadas na Figura 8. Verifica-se que essas curvas têm a inclinação semelhante à indicada pela dispersão de pontos empiricamente observada.

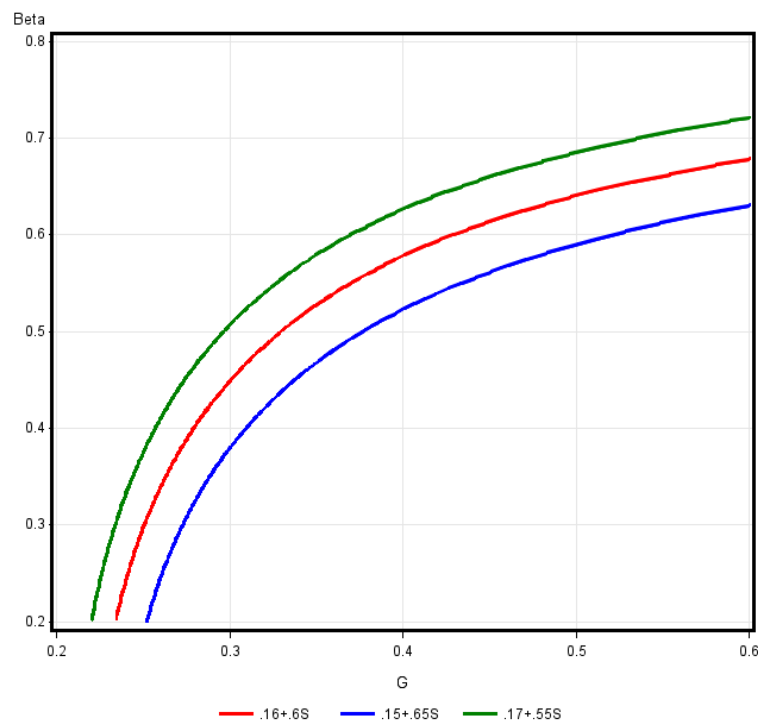


Figura 8. Curvas mostrando como o valor de β (o IGE) varia em função do índice de Gini para três relações entre o desvio padrão (σ) do erro do processo AR(1) entre os logaritmos das rendas de pais e filhos e o desvio padrão (S) da distribuição do logaritmo da renda dos filhos: $\sigma = 0,16 + 0,6S$, $\sigma = 0,15 + 0,65S$ ou $\sigma = 0,17 + 0,55S$.

APÊNDICE B. Resultados adicionais do modelo log-linear de associação homogênea

Na tabela a seguir, são apresentadas as estimativas obtidas para o modelo de associação homogênea considerando status do pai, status do filho e coorte de idade para a amostra da PNAD 2014.

Analisando os parâmetros estimados, observa-se que há alguns não estatisticamente significativos, tendo como base a comparação com as categorias de base. É o caso, por exemplo, do efeito da influência de pertencer à coorte 4, correspondente à década de 1994-2003.

Os efeitos principais, representados pelas variáveis coorte de idade (C), status do pai (P) e status do filho (F), são estatisticamente significativos em sua maioria, com exceção da quarta coorte de idade (década entre 1994-2003, idade entre 26 e 35 anos) e status do filho baixo-inferior.

Tabela 27. Resultados estimados para o modelo de associação homogênea da mobilidade social brasileira em 2014, considerando as variáveis coorte de idade (C), status do pai (P), status do filho (F) e suas associações

Parâmetro	Categoria(s)	Estimativa	Erro-padrão	χ^2	Pr > χ^2
C	1	-0,305	0,0227	180,72	<,0001
	2	0,5252	0,016	1.083,05	<,0001
	3	0,3548	0,0171	429,07	<,0001
	4	0,0213	0,0196	1,18	0,2774
P	Baixo-Inferior	0,9577	0,0184	2.721,13	<,0001
	Baixo-Superior	0,692	0,0186	1.379,11	<,0001
	Médio-Inferior	0,3879	0,0199	378,28	<,0001
	Médio-Médio	-0,222	0,0237	87,92	<,0001
	Médio-Superior	-0,4099	0,0261	246,92	<,0001
F	Baixo-Inferior	-0,0638	0,0346	3,4	0,0652
	Baixo-Superior	0,4557	0,0222	422,39	<,0001
	Médio-Inferior	-0,0877	0,0284	9,52	0,002
	Médio-Médio	0,1065	0,0227	22,04	<,0001
	Médio-Superior	0,0796	0,0222	12,89	0,0003

(continua)

(continuação)

C×P	1*Baixo-Inferior	-0,4785	0,0332	207,3	<,0001
	1*Baixo-Superior	0,0976	0,0327	8,92	0,0028
	1*Médio-Inferior	0,178	0,0359	24,55	<,0001
	1*Médio-Médio	-0,0125	0,0477	0,07	0,7925
	1*Médio-Superior	0,1737	0,0501	12,02	0,0005
	2*Baixo-Inferior	-0,2157	0,0246	77	<,0001
	2*Baixo-Superior	0,0553	0,0256	4,66	0,0309
	2*Médio-Inferior	0,0444	0,0286	2,42	0,1199
	2*Médio-Médio	-0,057	0,0367	2,42	0,1199
	2*Médio-Superior	0,0724	0,0381	3,61	0,0573
	3*Baixo-Inferior	0,0409	0,0252	2,63	0,1049
	3*Baixo-Superior	0,0577	0,0273	4,48	0,0344
	3*Médio-Inferior	0,0334	0,0304	1,21	0,2718
	3*Médio-Médio	-0,0696	0,0392	3,15	0,0759
	3*Médio-Superior	-0,0899	0,0422	4,53	0,0333
	C×F	4*Baixo-Inferior	0,2628	0,0277	89,73
4*Baixo-Superior		0,0528	0,0308	2,93	0,087
4*Médio-Inferior		-0,0642	0,0352	3,32	0,0685
4*Médio-Médio		-0,0194	0,0437	0,2	0,6582
4*Médio-Superior		-0,0719	0,0474	2,3	0,1298
1*Baixo-Inferior		0,2726	0,0333	67,06	<,0001
1*Baixo-Superior		0,3176	0,0305	108,19	<,0001
1*Médio-Inferior		-0,1084	0,0409	7,02	0,0081
1*Médio-Médio		0,3902	0,0359	118,21	<,0001
1*Médio-Superior		-0,3523	0,0479	54,06	<,0001
2*Baixo-Inferior		-0,0987	0,0254	15,12	0,0001
2*Baixo-Superior		-0,0818	0,0238	11,83	0,0006
2*Médio-Inferior		-0,032	0,0288	1,23	0,267
2*Médio-Médio		0,0042	0,0291	0,02	0,8851
2*Médio-Superior		0,2077	0,0301	47,58	<,0001
3*Baixo-Inferior		-0,0488	0,0256	3,64	0,0564
3*Baixo-Superior	-0,1312	0,0248	28,06	<,0001	
3*Médio-Inferior	0,0913	0,0288	10,03	0,0015	
3*Médio-Médio	-0,1263	0,0313	16,31	<,0001	
3*Médio-Superior	0,1732	0,0316	29,98	<,0001	
4*Baixo-Inferior	-0,0737	0,0278	7,03	0,008	
4*Baixo-Superior	-0,0462	0,0266	3,01	0,0827	
4*Médio-Inferior	0,0276	0,0321	0,74	0,3898	
4*Médio-Médio	-0,1551	0,0349	19,74	<,0001	
4*Médio-Superior	0,0818	0,0357	5,25	0,0219	

(continua)

(continuação)

P×F	Baixo-Inferior*Baixo-Inferior	13.749	0,0386	1.271,68	<,0001
	Baixo-Inferior*Baixo-Superior	0,4479	0,029	237,85	<,0001
	Baixo-Inferior*Médio-Inferior	0,3114	0,036	74,87	<,0001
	Baixo-Inferior*Médio-Médio	-0,2442	0,0345	50,01	<,0001
	Baixo-Inferior*Médio-Superior	-0,6062	0,0357	288,54	<,0001
	Baixo-Superior*Baixo-Inferior	0,4387	0,0416	111,43	<,0001
	Baixo-Superior*Baixo-Superior	0,3953	0,0296	178,08	<,0001
	Baixo-Superior*Médio-Inferior	0,1904	0,0377	25,45	<,0001
	Baixo-Superior*Médio-Médio	-0,0294	0,034	0,75	0,3868
	Baixo-Superior*Médio-Superior	-0,3554	0,0352	102	<,0001
	Médio-Inferior*Baixo-Inferior	0,1794	0,0455	15,54	<,0001
	Médio-Inferior*Baixo-Superior	0,0524	0,0338	2,4	0,1216
	Médio-Inferior*Médio-Inferior	0,4037	0,039	106,95	<,0001
	Médio-Inferior*Médio-Médio	-0,00896	0,037	0,06	0,8088
	Médio-Inferior*Médio-Superior	-0,2341	0,0379	38,18	<,0001
	Médio-Médio*Baixo-Inferior	-0,1161	0,0578	4,04	0,0445
	Médio-Médio*Baixo-Superior	-0,1299	0,0439	8,75	0,0031
	Médio-Médio*Médio-Inferior	0,0167	0,0519	0,1	0,7476
	Médio-Médio*Médio-Médio	0,1621	0,0445	13,26	0,0003
	Médio-Médio*Médio-Superior	0,0585	0,0443	1,74	0,1871
	Médio-Superior*Baixo-Inferior	-0,5532	0,0697	62,98	<,0001
	Médio-Superior*Baixo-Superior	-0,3476	0,05	48,4	<,0001
	Médio-Superior*Médio-Inferior	-0,2612	0,0605	18,66	<,0001
	Médio-Superior*Médio-Médio	0,0189	0,0494	0,15	0,7028
	Médio-Superior*Médio-Superior	0,4729	0,0423	124,93	<,0001

Fonte: Elaborado pelo autor, com base nos microdados do IBGE.