Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

Segregação e desigualdade: analogia na mensuração e análise da segregação por gênero em setores de atividade no Brasil

Diego Camargo Botassio

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Economia Aplicada

Piracicaba
2017

Diego Camargo Botassio
Bacharel em Ciências Econômicas

Segregação e desigualdade: analogia na mensuração e análise da segregação por gênero em setores de atividade no Brasil

versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

Dissertação apresentada para obtenção do título de Mestre em Ciências. Área de concentração: Economia Aplicada

Piracicaba
2017

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

Botassio, Diego Camargo
Segregação e desigualdade: analogia na mensuração e análise da segregação por gênero em setores de atividade no Brasil / Diego Camargo Botassio. – – versão revisada de acordo com a resolução CoPGr 6018 de 2011. – – Piracicaba, 2017 .

103 p.

Dissertação (Mestrado) – – USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Medidas de segregação 2. Medidas de desigualdade 3. Sensibilidade 4. Economia do trabalho . I. Título.

DEDICATÓRIA

Aos meus avós e padrinhos, Damalher e Maria
À maior guerreira que conheço, minha mãe
À minha irmã
Ao meu afilhado, Álvaro

AGRADECIMENTOS

Ao orientador, Rodolfo Hoffmann, pelos ensinamentos, conselhos, orientações e minuciosas leituras das versões preliminares desta dissertação. Seu comprometimento com o serviço público é admirável. Sendo sempre um exemplo ético, professor rigoroso e pesquisador em busca da verdade, ele me inspira como exemplo profissional.

Às professoras Ana Kassouf, Daniela Vaz e Lilian Maluf pelas excelentes contribuições durante o exame de qualificação. Aos membros da banca de defesa Alexandre Gori Maia, Ana Kassouf e Daniela Vaz pelas críticas e sugestões durante a defesa.

Aos professores do Programa de Pós-Graduação em Economia Aplicada da ESALQ pelos ensinamentos.

Às funcionárias exemplares do departamento, Aline e Lu, pela atenção e dedicação ao excelente trabalho que desenvolvem. A Dona Mirian por todo o carinho e atenção a todos os alunos da Pós-Graduação.

Aos amigos Augusto Seabra, Felipe Tavares, Josimar Jesus, Pedro Paiva, Gustavo Giachini, Júlia Kraemer, Cristiane Ogino, Jaquelini Gelain, João Schimidt, Taís Menezes, Luciana Lacerda, Camila Rossi e Marcos Garcias. Em especial, agradeço ao Josimar pelas discussões e comentários sobre segregação e desigualdade feitos ao longo do trabalho.

A todos os outros colegas da ESALQ com os quais compartilhei longas e frutíferas conversas durante esses dois anos. Aos colegas da EPGE, Raul e Alexandre, pelas minuciosas revisões das provas das propriedades. Agradeço a Dani (UEM) pela amizade, pelos comentários, críticas e longas discussões sobre este trabalho.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro.

Em geral, agradeço a todos que contribuíram direta e indiretamente para a conclusão deste trabalho.

SUMÁRIO

RESUMO

Segregação e desigualdade: analogia na mensuração e análise da segregação por gênero em setores de atividade no Brasil

Há algumas décadas as relações entre medidas de desigualdade e de segregação são conhecidas. Neste sentido, o objetivo principal desta dissertação é aprofundar a análise dessas relações. São descritas, pormenorizadamente, as medidas consagradas de desigualdade (índice de Gini e medida geral) e de segregação (índice de Gini para segregação e índice de Dissimilaridade). Como primeiro resultado, é demonstrado que algumas medidas de segregação conhecidas na literatura são casos particulares ou transformações da medida geral de segregação de Hutchens (2004). Advoga-se pelo uso da medida geral, em detrimento de uma transformação proposta por esse autor. Além disso, são demonstradas algumas propriedades sobre sua decomposição. Embora o efeito de transferências regressivas de renda sobre as medidas de desigualdade seja discutido há décadas, não existia análise correspondente para medidas de segregação. Para cobrir essa lacuna, são analisadas as sensibilidades da medida geral de segregação e dos índices de Gini e de Dissimilaridade a mudanças regressivas entre estratos. Para ilustrar os resultados encontrados, é analisada a evolução da segregação por gênero em grupamentos de atividade no Brasil de 1992 a 2014. Constatou-se que a segregação por gênero, conforme várias medidas, diminuiu. Além disso, de 2002 a 2014, é feita a análise da decomposição das medidas de segregação quando 53 ramos de atividade são classificados em seis grupamentos. Verifica-se que a redução da segregação por gênero entre os 53 ramos de atividade foi puxada pela segregação entre os seis grupamentos. Vários outros resultados sobre medidas de segregação são apresentados no final do trabalho.

Palavras-chave: Medidas de segregação; Medidas de desigualdade; Sensibilidade; Economia do trabalho

ABSTRACT

Segregation and inequality: analogy in the measurement and analysis of gender segregation by activity sectors in Brazil

For decades, the relations between inequality and segregation measures has been analyzed. This dissertation aims to deepen the analysis of these relations. The well known inequality (Gini index and general measure) and segregation (Gini index for segregation and Dissimilarity index) measures are described in detail. It is also demonstrated that some measures of segregation known in the literature are particular cases or transformations of Hutchens' (2004) general measure of segregation. The use of the general measure is advocated, to the detriment of a transformation proposed by the author. In addition, some properties of its decomposition are demonstrated. Although the effect of regressive income transfers on inequality measures has been discussed for decades, there is no corresponding analysis for segregation measures. To cover this gap, we analyze the sensitivity of the segregation general measure and of the Gini and Dissimilarity indices to regressive movements between strata. To illustrate the results, the evolution of gender segregation between sectors of activity in Brazil, from 1992 to 2014, is analyzed. Gender segregation, according to several measures, has decreased. In addition, from 2002 to 2014, we analyze the decomposition of segregation when 53 branches of activity are classified into 6 sectors grouping. The results show that the decline of gender segregation between activities was driven by the segregation between groups. Several other results on segregation measures are found in the last chapter.

Keywords: Segregation measures; Inequality measures; Sensitivity; Labour economics

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

CBO	Classificação Brasileira de Ocupações
CNAE	Classificação Nacional de Atividades Econômicas
IBGE	Instituto Brasileiro de Geografia e Estatística
PNAD	Pesquisa Nacional por Amostra de Domicílios
p.p.	pontos percentuais
RAIS	Relação Anual de Informações Sociais

LISTA DE SÍMBOLOS

$A(\mathbf{z})$	família de Atkinson para desigualdade
$A_S(\mathbf{X})$	família de Atkinson para segregação
$A'_S(\mathbf{X})$	família de Atkinson para segregação descrita em James e Taeuber (1985)
α	área de desigualdade
α_S	área de segregação
$B(\mathbf{z})$	medida de segregação entre os grupos (para desigualdade)
$B_S(\mathbf{X})$	medida de segregação entre os grupamentos (para segregação)
β	área entre a poligonal de Lorenz e o eixo das abscissas
β_S	área entre a curva de segregação e o eixo das abscissas
$CV(\mathbf{z})$	Coeficiente de variação para a renda
$CV_S(\mathbf{X})$	Coeficiente de variação para segregação
d_i	inclinação da poligonal de Lorenz referente à i-ésima pessoa
d_j^S	inclinação da curva de segregação referente ao j-ésimo estrato
$D(\mathbf{X})$	índice de Dissimilaridade
D_{max}	discrepância máxima da poligonal de Lorenz
D_{max}^S	discrepância máxima da curva de segregação
δ	desvio (absoluto) médio
Δ	diferença média (salvo contexto claro que indique a diferença simples)
ε	parâmetro das medidas gerais de desigualdade e segregação
$G(\mathbf{z})$	índice de Gini para desigualdade
$G_S(\mathbf{X})$	índice de Gini para segregação
γ	diferença relativa entre as participações relativas de dois estratos
$I(\mathbf{z})$	medida geral de desigualdade
$I_S(\mathbf{X})$	medida geral de segregação
$I_8(\mathbf{X})$	índice I_8 de Jahn, Schmit e Schrag (1947)
$I_p(\mathbf{X})$	índice I_p de Karmel e MacLachlan (1988)
$L(\mathbf{z})$	medida (ou índice) L de Theil para desigualdade
$L_S(\mathbf{X})$	medida L de Theil para segregação
max	valor máximo de alguma variável (ou parâmetro) de interesse
μ	renda média da população
μ_S	razão X/Y
n_j	número pessoas no j-ésimo grupo
n_g	número de estratos no g-ésimo grupamento
$O_c(\mathbf{X})$	medida geral de segregação de Hutchens (2004)
θ	proporção de uma das categorias envolvida em uma mudança entre estratos
$\Omega(\mathbf{X})$	medida genérica de segregação
P	proporção de pessoas da categoria X na população

P_j	proporção da categoria X no total do j-ésimo estrato
π_i	peso da i-ésima unidade de análise na população (para desigualdade)
Φ_i	proporção acumulada da renda até a i-ésima pessoa
p_i	proporção acumulada da população até a i-ésima pessoa
Ψ_j	proporção acumulada da categoria X até o j-ésimo estrato
$RCE(\mathbf{X})$	índice de Centralização Relativa
ρ_j	proporção acumulada da categoria Y até o j-ésimo estrato
s_j	participação do j-ésimo estrato na segregação total
t_j	participação do j-ésimo estrato na população
T	número de pessoas na população
T_j	número de pessoas no i-ésimo estrato
$T(\mathbf{z})$	medida (ou índice) T de Theil para desigualdade
$T_S(\mathbf{X})$	medida T de Theil para segregação
$w^j(\mathbf{z}^j)$	peso referente ao j-ésimo grupo na desigualdade dentro dos grupos
$w_S^g(\mathbf{X}^g)$	peso referente ao g-ésimo grupamento na segregação dentro dos grupamentos
X	número de pessoas da categoria X
X_j	número de pessoas da categoria X alocada no j-ésimo estrato
x_j	participação do j-ésimo estrato no total da categoria X
x_j/y_j	participação relativa referente ao j-ésimo estrato
$\mathbf{X}$	matriz de distribuição das pessoas das categorias X e Y nos estratos
$\mathbf{X}^*$	matriz de distribuição resultante de uma mudança regressiva na categoria X
Y	número de pessoas da categoria Y
Y_j	número de pessoas da categoria Y alocada no j-ésimo estrato
y_j	participação do j-ésimo estrato no total da categoria Y
$\mathbf{Y}^*$	matriz de distribuição resultante de uma mudança regressiva na categoria Y
z_i	renda da i-ésima pessoa
$\mathbf{z}$	vetor de distribuição de rendas
ζ	diferença absoluta entre as participações relativas de dois estratos
$\neq$	diferente
∞	infinito
$\Rightarrow$	implica
$\Leftrightarrow$	se e somente se
$\forall$	para todo
$\in$	é elemento de ($\notin$ indica não é elemento de)
$\subset$	está contido
$\mathbb{N}$	números naturais ($\mathbb{N}_0$ indica os números naturais união com o zero)
$\mathbb{R}_+$	números reais não negativos ($\mathbb{R}_{++}$ indica os números reais positivos)

1. INTRODUÇÃO

Diferentemente da ampla literatura e discussão existente sobre distribuição de renda, o conceito e as formas de medir segregação são pouco conhecidos entre os cientistas sociais. Em meados da década de 50, as discussões sobre segregação ganharam força com o trabalho de Duncan e Duncan (1955a). Desde então, os sociólogos consolidaram essa área, analisando, principalmente, a segregação racial entre negros e brancos, nos bairros das cidades norte-americanas (TAEUBER; TAEUBER, 1965). Os primeiros trabalhos com motivações econômicas (GROSS, 1968; OPPENHEIMER, 1970) introduziram a análise da segregação entre homens e mulheres no mercado de trabalho.

Sobre o conceito de segregação, destacam-se duas situações extremas. Entende-se que há segregação máxima quando, por exemplo, cada bairro de uma cidade é ocupado exclusivamente por brancos ou negros. Outro exemplo é quando não existem setores em uma economia que são ocupados, simultaneamente, por pelo menos um homem e uma mulher. Por outro lado, diz-se que não há segregação quando as proporções das duas categorias analisadas (brancos e negros ou homens e mulheres) são as mesmas em todos os estratos (bairros ou setores produtivos). Essas situações apresentadas são raras. Nas análises empíricas, os cientistas estão preocupados em avaliar como as diversas situações entre esses extremos evoluem no tempo.

Nas últimas décadas, o conceito de segregação tem sido amplamente utilizado para diversos tipos de análise. A literatura que discute e propõe novas formas de medir a segregação acompanhou o mesmo ritmo. Destaca-se, sobretudo, que a análise da segregação deixou de ser exclusiva do campo da sociologia e da economia do trabalho. Nas outras áreas da economia, por exemplo, esse tema ganhou espaço entre os economistas a partir dos modelos de redes e interações sociais (ECHENIQUE; FRYER JR., 2007; CARD; MAS; ROTHSTEIN, 2008; CURRARINI; JACKSON; PIN, 2009; BOJANOWSKI; CORTEN, 2014). Já na geografia, o estudo da segregação residencial se consolidou (WHITE, 1983; MORRILL, 1991; REARDON; O'SULLIVAN, 2004). Ainda no campo da sociologia, diversos trabalhos têm analisado a segregação racial nas escolas (THEIL; FINIZZA, 1971; ALLEN; VIGNOLES, 2007; FRANKEL; VOLIJ, 2011).

Outras análises têm discutido os conceitos e formas de mensurar segregação em casos onde há mais de duas categorias (REARDON; FIREBAUGH, 2002; FRANKEL; VOLIJ, 2011; ALONSO-VILLAR; RÍO, 2010). Nesse caso, por exemplo, ao se analisar a segregação racial nas cidades, além de branco e negro, são incluídas as categorias asiático e indígena. Este trabalho, entretanto, está focado no estudo de medidas de segregação entre duas categorias. Além disso, buscar-se-á relacionar as principais medidas de desigualdade de renda e suas análogas para segregação.

No início dos anos 80, uma série de trabalhos analisaram algumas medidas de desigualdade conhecidas e mostraram que elas eram casos particulares de uma família geral de medidas de desigualdade (COWELL; KUGA, 1981a, 1981b; COWELL, 1980; SHORROCKS, 1980, 1984). Essa família de medidas tem, como principal característica, decompor a desigualdade entre grupos de pessoas e, dentro de cada grupo, medir a desigualdade entre seus membros. A essa classe se dá o nome de medida de desigualdade aditivamente decomponível. Esses trabalhos tiveram grande impacto na discussão sobre medidas de desigualdade de renda, pois, incluíam como casos particulares, as medidas T e L desenvolvidas por Henri Theil (1967).

Por outro lado, não existe consenso sobre uma medida geral de segregação. Alguns trabalhos, por exemplo, Chakravarty e Silber (2007) e Hutchens (2004), apresentam algumas propostas. Com base nos trabalhos sobre distribuição de renda, esta dissertação apresentará uma visão alternativa ao abordar o tema, o que possibilitará discutir uma proposta de medida geral de segregação com base em uma formulação apresentada por Hutchens (2004). Mesmo que as formas de medidas gerais de segregação utilizadas neste trabalho e no de Hutchens (2004) sejam muito próximas, adianta-se que as conclusões diferem de forma substancial. Sem utilizar a transformação do autor, é possível relacionar diversas medidas de desigualdade com as de segregação.

Para mostrar a aplicação dessa medida geral de segregação, serão utilizados os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), de 1992 a 2014. A PNAD é uma pesquisa representativa para todo o território brasileiro, que coleta informações da população como renda, ocupação e condições do domicílio, dentre outras.

Um tópico que carece de análise na literatura é a sensibilidade das medidas de segregação para mudanças entre os estratos. Alguns autores (KAKWANI, 1980; SHORROCKS; FOSTER, 1987; HOFFMANN, 1992; COWELL; FLACHAIRE, 2007) analisam a sensibilidade das medidas de desigualdade a transferências de renda regressivas. Entretanto, não foram encontradas referências para essa análise no caso da segregação. Utilizando como base a literatura existente sobre desigualdade, esta dissertação apresenta uma análise similar para o caso da segregação. Outra contribuição deste trabalho é a definição de uma propriedade desejável de uma medida de segregação, que difere sutilmente da literatura, associada a uma mudança regressiva entre estratos, que é a propriedade análoga à condição de Pigou-Dalton para medidas de desigualdade, porém para segregação.

Deste modo, esta dissertação destina-se a: i) apresentar as principais medidas de desigualdade de renda e segregação consagradas na literatura, enfatizando suas relações; ii) discutir o desenvolvimento de uma medida geral de segregação aditivamente decomponível e seus casos particulares, com base em uma transformação apresentada em Hutchens (2004); iii) analisar a sensibilidade das medidas de segregação; iv) aplicar a medida aditivamente decomponível para analisar a segregação de gênero no mercado de trabalho no Brasil; e v) discutir sobre tópicos relacionados a medidas de segregação encontrados na literatura. Para cada um desses objetivos é dedicado um capítulo.

A primeira seção do Capítulo 2 apresenta a notação para medidas de segregação utilizada ao longo deste trabalho. As demais seções desse capítulo expõem, respectivamente, o índice de Gini para distribuição de renda e segregação, o índice de dissimilaridade e a medida geral de desigualdade, assim como seus casos particulares. A primeira seção do Capítulo 3 discute a proposta de uma medida geral para segregação aditivamente decomponível. A seção 3.2 deriva a sua decomposição e as seções posteriores discutem seus casos particulares, um exemplo ilustrativo para decomposição e as propriedades desejáveis de uma medida de segregação. Ressalta-se, entretanto, que a propriedade relativa a mudança entre estrato, apresentada na última seção desse capítulo, difere levemente da encontrada na literatura.

O objetivo do Capítulo 4 é fazer a análise de sensibilidade das medidas de segregação. Para isso, são utilizadas as seções 4.1, 4.2 e 4.3 para a medida geral, o índice de Gini e o índice de Dissimilaridade, respectivamente. Ressalta-se, novamente, que não foram encontrados trabalhos na literatura que analisam a sensibilidade das medidas no caso da segregação.

No Capítulo 5, é feita a análise da segregação por gênero no mercado de trabalho no Brasil, a partir do desenvolvimento da medida discutida anteriormente e pelos índices de Gini e Dissimilaridade. Este capítulo, além de uma breve introdução com revisão de literatura, tem uma seção dedicada à descrição da base de dados. Os resultados são apresentados na seção 5.2 e em suas subseções. A subseção 5.2.1 apresenta a evolução das medidas de segregação para o Brasil de 1992 a 2014, enquanto que a subseção 5.2.2 apresenta a decomposição das medidas de segregação quando os 53 ramos de atividade são classificados em seis grupamentos, de 2002 a 2014. O Capítulo 6 discorre sobre outras duas medidas de segregação e uma propriedade que diferencia os índices de Gini para desigualdade e para segregação. Por fim, são feitas as considerações finais do trabalho, apresentadas as referências bibliográficas e os apêndices.

2. PRINCIPAIS MEDIDAS DE DESIGUALDADE E SEGREGAÇÃO

2.1. Definições preliminares

Seja uma região com dois tipos de pessoas, X e Y, e k estratos. Podemos imaginar, como exemplo ilustrativo, uma cidade com brancos e negros distribuídos em k bairros ou um país com homens e mulheres ocupados em setores produtivos. O subíndice j será usado para distinguir os estratos, com $j \in \{1, 2, ..., k\}$. X e Y também representam as quantidades totais de pessoas de cada tipo. Deste modo, X_j e Y_j são as quantidades de pessoas das categorias X e Y alocadas no j-ésimo estrato, respectivamente. Denota-se T a população, de modo que $T = X + Y$. De modo semelhante, T_j é a população do j-ésimo estrato ($T_j = X_j + Y_j$). Das definições acima temos

$$X = \sum_{j=1}^{k} X_j,\ Y = \sum_{j=1}^{k} Y_j \text{ e } T = \sum_{j=1}^{k} T_j. \tag{2.1}$$

Também é interessante incluir as notações especiais

$$X_0 = 0 \text{ e } Y_0 = 0. \tag{2.2}$$

Sejam x_j e y_j as participações do j-ésimo estrato nos totais das categorias X e Y, respectivamente. Analogamente, a participação do j-ésimo estrato no total geral será t_j. Assim, temos

$$x_j = \frac{X_j}{X},\ y_j = \frac{Y_j}{Y} \text{ e } t_j = \frac{T_j}{T},\ j \in \{1, 2, ..., k\}. \tag{2.3}$$

As proporções de pessoas das categorias X e Y na população são, respectivamente,

$$P = \frac{X}{T} = \frac{\sum_{j=1}^{k} X_j}{T} \text{ e } (1 - P) = \frac{Y}{T} = \frac{\sum_{j=1}^{k} Y_j}{T}. \tag{2.4}$$

De modo semelhante, as proporções das categorias X e Y no total do j-ésimo estrato são

$$P_j = \frac{X_j}{T_j} = \frac{X_j}{X_j + Y_j} \text{ e } (1 - P_j) = \frac{Y_j}{T_j} = \frac{Y_j}{X_j + Y_j},\ j \in \{1, 2, ..., k\}\ . \tag{2.5}$$

Assim como nos trabalhos de Hutchens (1991, 2001, 2004) e Watts (1998a), pode-se definir uma matriz com as informações sobre as distribuições dos dois grupos nos k estratos. Sua definição não é essencial à teoria das medidas de segregação, porém facilita a apresentação do procedimento de cálculo das medidas. Tal matriz[1] é definida como

$$\mathbf{X} = \begin{bmatrix} X_1 & X_2 & ... & X_{k-1} & X_k \\ Y_1 & Y_2 & ... & Y_{k-1} & Y_k \end{bmatrix}_{2 \times k}.$$

[1] Em Watts (1998a, p. 489) é utilizado a transposta de $\mathbf{X}$, não acarretando diferenças para a teoria.

Portanto, sempre que citada uma medida genérica Ω, será utilizada a notação $\Omega(\mathbf{X})$ para deixar claro que $\mathbf{X}$ é o argumento da medida.

Definindo apropriadamente o conjunto dos números naturais como $\mathbb{N} = \{1, 2, ...\}$, os naturais união com o zero por $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, os reais não negativos como $\mathbb{R}_+ = [0, \infty)$ e os reais positivos por $\mathbb{R}_{++} = \mathbb{R}_+ \backslash \{0\}$, as definições de (2.1) a (2.5) implicam

$$
\begin{aligned}
X, Y, T, T_j, k &\in \mathbb{N}, \text{ com } k, T \geq 2 \\
X_j, Y_j &\in \mathbb{N}_0 \\
x_j, y_j, P_j &\in [0, 1] \subset \mathbb{R}_+ \\
t_j, P &\in (0, 1) \subset \mathbb{R}_{++}.
\end{aligned}
$$

Desse modo, uma medida de segregação é uma função $\Omega : \mathbb{R}_+^{2\times k} \rightarrow \mathbb{R}_+$, tal que Ω indica a correspondência $\mathbf{X} \mapsto \Omega(\mathbf{X})$, com $\mathbf{X} \in \mathbb{R}_+^{2\times k}$. Em seções posteriores, as notações apresentadas até aqui serão utilizadas para expor as principais medidas de segregação.

Antes de apresentar as medidas de desigualdade, é necessário fazer uma ressalva referente ao conceito de grupo nas situações de desigualdade de renda e de segregação. Quando se trabalha com desigualdade de renda, o conceito de grupos refere-se aos conjuntos de pessoas definidos a partir de algum critério como região, faixa etária, ou outros. Podemos ter, desse modo, um, dois ou mais grupos. No caso da segregação, grupos poderiam ser interpretados como dois conjuntos de pessoas mutuamente exclusivos e coletivamente exaustivos, conforme uma característica definida previamente (sexo, etc.). Seria, desse modo, inapropriado utilizar a palavra "grupo" para designar esses conjuntos, optando-se por denominá-los *categorias* ou *tipos*. O conceito análogo ao de grupo de pessoas na distribuição de renda, para o caso da segregação, é um conjunto de estratos, que é aqui designado *grupamento*.

A próxima seção descreverá o índice de Gini para a distribuição de renda, assim como seu análogo para segregação.

2.2. Índice de Gini

Antes de definirmos o índice de Gini para segregação é interessante apresentá-lo em sua forma tradicional para a distribuição de renda. Tal apresentação permitirá comparar as duas formulações. Já se adianta que as formulações são muito semelhantes, porém deve-se atentar para alguns pormenores importantes. As definições, conceitos e notações para a forma tradicional do índice seguirão Hoffmann (1998).

Suponha uma população com n pessoas economicamente ativas. A renda é a variável z, de modo que z_i seja a renda referente à i-ésima pessoa, com $i \in \{1, 2, ..., n\}$. Considerando que os indivíduos estão ordenados das rendas menores para as maiores, ou seja,

$$z_1 \leq z_2 \leq ... \leq z_{n-1} \leq z_n, \tag{2.6}$$

pode-se definir o vetor linha com todos os valores das rendas[2] como

$$\mathbf{z} = \left[\begin{array}{ccccc} z_1 & z_2 & ... & z_{n-1} & z_n \end{array} \right].$$

Observe que a renda total da população ($\sum_{i=1}^{n} z_i$) é igual a $n\mu$, onde μ é a renda (populacional) média. Desta maneira, a proporção acumulada da renda até o i-ésimo indivíduo, obedecendo à ordenação (2.6), é

$$\Phi_i = \frac{1}{n\mu} \sum_{j=1}^{i} z_j \tag{2.7}$$

e a respectiva proporção acumulada da população é definida por

$$p_i = \sum_{j=1}^{i} \frac{1}{n} = \frac{i}{n}. \tag{2.8}$$

Também define-se a notação especial $\Phi_0 = p_0 = 0$. Essas definições permitem concluir que $\Phi_i, p_i \in [0, 1]$, para todo $i \in \{0, 1, ..., n\}$.

Pode-se, em um plano cartesiano ortogonal ($[0, 1] \times [0, 1] \subset \mathbb{R}^2_+$), unir todos os pontos dados pelas expressões (2.7), (2.8) e a notação especial. A poligonal definida pelos segmentos unindo estes pontos (p_i, Φ_i), com $i \in \{0, 1, ..., n\}$, é denominada *poligonal de Lorenz*. Quando $n \to \infty$, temos a *curva de Lorenz*. De modo geral, em trabalhos aplicados onde a população é finita, a poligonal é confundida com uma curva e recebe o nome de "curva de Lorenz". A ordenação (2.6) garante que a poligonal de Lorenz seja convexa. Nessas mesmas condições, mas com população infinita, a curva de Lorenz é convexa. Relaxando a igualdade em (2.6), temos

$$z_1 < z_2 < ... < z_{n-1} < z_n.$$

Com essa ordenação, e pressupondo população infinita, a curva de Lorenz é estritamente convexa. Não podemos afirmar o mesmo com população finita para a poligonal. Uma definição geral da curva de Lorenz é encontrada em Gastwirth (1971).

Sob a hipótese de que todos os indivíduos recebem a mesma renda ($z_i = z_j$, $\forall i, j$), temos que $\Phi_i = p_i$ e o segmento de reta que une todos os pares ordenados (p_i, Φ_i), com $i \in \{0, 1, ..., n\}$, coincidirá com a reta bissetriz do primeiro quadrante limitado pelos pontos extremos do segmento $[(0, 0)$ e $(1, 1)]$. Esse segmento recebe o nome de *linha da perfeita igualdade*. Observe que a linha da perfeita igualdade é a expressão de um caso particular da poligonal (ou curva) de Lorenz. Considerando as duas situações, a de perfeita igualdade e uma situação cuja renda varie entre os indivíduos, podemos construir um gráfico de dispersão da proporção acumulada da população contra a proporção acumulada da renda, como na Figura 2.1.

[2] As restrições impostas sobre os valores assumidos na ordenação (2.6) são: $z_i \in \mathbb{R}_+$, $i \in \{1, 2, ..., n-1\}$ e $z_n \in \mathbb{R}_{++}$. Como estamos considerando pessoas economicamente ativas (definição que inclui ocupados ou não) permitimos que todos os indivíduos até a posição $n-1$ tenham renda nula, mas necessariamente a última pessoa (n-ésima) tenha renda positiva. Em geral, poderíamos calcular o índice de Gini utilizando rendas negativas, mas não consideramos essa situação neste trabalho.

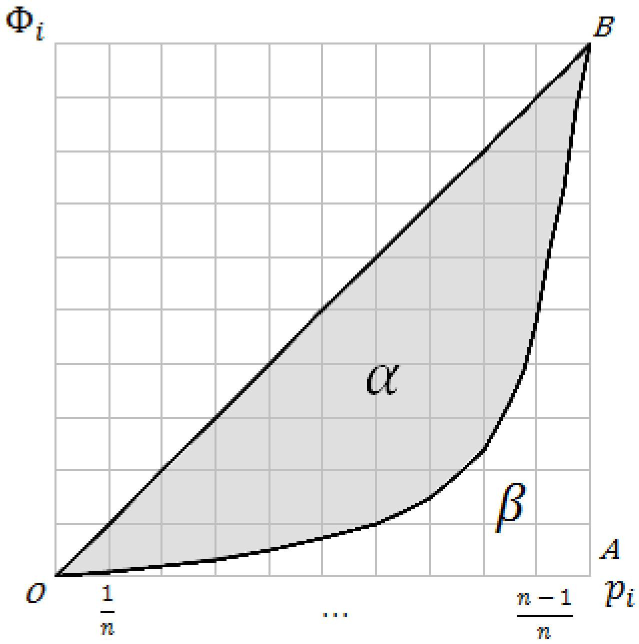

Figura 2.1. Poligonal de Lorenz e a linha da perfeita igualdade

O segmento de reta OB é a linha da perfeita igualdade. A poligonal que limita inferiormente a área α é a poligonal de Lorenz para a situação em que a renda varie entre os indivíduos. A área α é denominada *área de desigualdade*. Observe que na situação de completa igualdade temos $\alpha = 0$. Pode-se demonstrar que, para $n < \infty$, o limite superior de α é $(1 - 1/n)/2$. Assim, no limite, quando $n \to \infty$, temos que $\max \alpha = 0,5$, em que $\max \alpha$ indica o valor máximo que α pode assumir. Para população finita teremos

$$\alpha \in \left[0, \frac{1}{2}\right) \text{ ou } \alpha \in \left[0, \frac{1}{2}\left(1 - \frac{1}{n}\right)\right]. \tag{2.9}$$

Observe, na Figura 2.1, que

$$\alpha + \beta = \frac{1}{2}. \tag{2.10}$$

A área β pode ser calculada a partir da soma dos trapézios retângulos formados pela poligonal de Lorenz e o eixo das abscissas. Definindo a área do i-ésimo trapézio por S_i temos

$$S_i = \frac{1}{2}(\Phi_i + \Phi_{i-1})(p_i - p_{i-1}) = \frac{1}{2}(\Phi_i + \Phi_{i-1})\frac{1}{n} \qquad \text{ou}$$
$$S_i = \frac{1}{2n}(\Phi_i + \Phi_{i-1}).$$

Assim,

$$\beta = \sum_{i=1}^{n} S_i = \frac{1}{2n}\sum_{i=1}^{n}(\Phi_i + \Phi_{i-1}). \tag{2.11}$$

O *índice de Gini* ($G(\mathbf{z})$) é definido como a razão entre a área de desigualdade e seu valor máximo. Portanto,

$$G(\mathbf{z}) = \frac{\alpha}{\max \alpha} \Rightarrow G(\mathbf{z}) = 2\alpha. \tag{2.12}$$

Substituindo (2.10) em (2.12), temos que $G(\mathbf{z}) = 1 - 2\beta$ e, usando (2.11),

$$G(\mathbf{z}) = 1 - \frac{1}{n}\sum_{i=1}^{n}(\Phi_i + \Phi_{i-1}). \tag{2.13}$$

Da definição do índice (eq. (2.12)) e dos limites de α (eq. (2.9)), temos que

$$G(\mathbf{z}) \in [0, 1) \text{ ou } G(\mathbf{z}) \in \left[0, 1 - \frac{1}{n}\right].$$

Pode-se, também, demonstrar que a eq. (2.13) pode ser escrita como

$$G(\mathbf{z}) = \frac{\Delta}{2\mu}, \tag{2.14}$$

onde

$$\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|z_i - z_j|}{n^2}. \tag{2.15}$$

A demonstração é exaustiva e pode ser encontrada em Hoffmann (1998, p. 27–28; 40–41). A estatística Δ é denominada *diferença média*. Observe que na eq. (2.15) não é necessário que as rendas estejam ordenadas por ordem não decrescente. Portanto, a ordenação (2.6) só é necessária quando o cálculo do índice de Gini é feito pela poligonal de Lorenz.

A diferença média também pode ser obtida pela equação

$$\Delta = \sum_{i=1}^{n}\sum_{j=1}^{n}|z_i - z_j|\pi_i\pi_j, \tag{2.16}$$

onde π_i e π_j são os pesos de cada unidade de análise na população. Tal formulação é útil quando se analisa uma distribuição de renda por grupos, desde que dentro de cada grupo todos os indivíduos tenham a mesma renda. Quando as unidades de análises são individuais, temos que $\pi_i = 1/n$, para todo i, e volta-se à eq. (2.15). O valor de π_i pode ser interpretado como uma probabilidade associada a encontrar o valor de renda z_i na população.

Uma propriedade desejável em uma medida de desigualdade é o atendimento à *condição de Pigou-Dalton*. Essa condição estabelece que uma transferência de renda regressiva, ou seja, de um indivíduo com menor renda para outro com renda superior, necessariamente eleva o valor da medida de desigualdade. O índice de Gini é uma medida de desigualdade que atende a essa condição (ver demonstração em Hoffmann (1998, p. 53)).

Da Figura 2.1 observe que a inclinação do i-ésimo segmento da poligonal de Lorenz, referente à i-ésima pessoa, será

$$d_i = \frac{\Delta\Phi_i}{\Delta p_i} = \frac{\frac{1}{n\mu}\left(\sum_{j=1}^{i} z_j - \sum_{j=1}^{i-1} z_j\right)}{\frac{i}{n} - \frac{i-1}{n}} \qquad \text{ou}$$

$$d_i = \frac{z_i}{\mu}. \tag{2.17}$$

A razão z_i/μ é denominada *renda relativa* da i–ésima pessoa. Do mesmo modo, sua participação na renda total será $z_i/(n\mu)$. Nota-se, portanto, que a inclinação do i-ésimo segmento da poligonal de Lorenz (eq. (2.17)) coincide com n vezes a participação da renda da i-ésima pessoa na renda total. Substituindo (2.17) em (2.16), temos

$$\Delta = \sum_{i=1}^{n}\sum_{j=1}^{n} |d_i\mu - d_j\mu|\pi_i\pi_j. \tag{2.18}$$

Essa equação será útil posteriormente. Um outro elemento importante é a *discrepância máxima* da poligonal de Lorenz. Entende-se por discrepância máxima a maior distância vertical entre a poligonal de Lorenz e a linha da perfeita igualdade. A discrepância máxima é, também, considerada uma medida de desigualdade. Para seu desenvolvimento é fundamental o conceito de desvio médio. Seja uma variável z com média μ. O *desvio médio* é definido como

$$\delta = \frac{1}{n}\sum_{i=1}^{n} |z_i - \mu|.$$

Conforme a ordenação apresentada para a construção da poligonal de Lorenz, podemos incluir a média populacional da renda como

$$z_1 \leq z_2 \leq ... \leq z_h \leq \mu < z_{h+1} \leq ... \leq z_n. \tag{2.19}$$

Deste modo, separando em um somatório os indivíduos com desvios não positivos, em outro as pessoas com desvios positivos e considerando a definição de desvio médio teremos

$$\delta = \frac{1}{n}\left[\sum_{i=1}^{h}(\mu - z_i) + \sum_{i=h+1}^{n}(z_i - \mu)\right]. \tag{2.20}$$

Como a soma dos valores de z_i é igual à $n\mu$, temos

$$\sum_{i=1}^{n} z_i = n\mu, \text{ ou } \sum_{i=1}^{n}(z_i - \mu) = 0. \tag{2.21}$$

A eq. (2.21) diz que a média (aritmética) é a medida de tendência central que torna nula a soma dos desvios de z em relação a esta medida. Essa condição implica que a soma dos desvios não negativos é igual à soma dos desvios não positivos. Portanto, da eq. (2.20) concluímos que

$$\delta = \frac{2}{n}\sum_{i=1}^{h}(\mu - z_i) = \frac{2}{n}\sum_{i=h+1}^{n}(z_i - \mu). \tag{2.22}$$

De (2.19) e (2.17) temos

$$d_i \leq 1 \text{ para } i \in \{1, 2, ..., h\} \qquad \text{e}$$
$$d_i > 1 \text{ para } i \in \{h+1, h+2, ..., n\}.$$

Como a linha da perfeita igualdade tem inclinação igual a 1, para valores de $d_i \leq 1$ a poligonal de Lorenz se distancia da linha da perfeita igualdade. A situação oposta acontece com $d_i > 1$. Portanto, a maior distância entre a poligonal e a curva da perfeita igualdade será $D_{max} = p_h - \Phi_h$, onde D_{max} é a discrepância máxima. Usando as eqs. (2.7) e (2.8) temos

$$D_{max} = \frac{h}{n} - \frac{1}{n\mu}\sum_{i=1}^{h} z_i \qquad \text{ou}$$

$$D_{max} = \frac{1}{n\mu}\sum_{i=1}^{h}(\mu - z_i).$$

Lembrando (2.22) conclui-se que

$$D_{max} = \frac{\delta}{2\mu}. \tag{2.23}$$

O valor da discrepância máxima foi calculado a partir do ponto (p_h, Φ_h), referente à h-ésima pessoa. Entretanto, sob uma única condição, pode-se calcular seu valor a partir de qualquer ponto em um determinado conjunto. Considere o conjunto de todas as k pessoas cuja renda seja igual à renda média, ou seja, $z_j = \mu$, $j \in C = \{h-k-1, h-k-2, ..., h-1, h\}$, $k \geq 1$. Sob essa condição, $d_j = 1$, $\forall j \in C$. Deste modo, para as pessoas neste conjunto os segmentos da poligonal de Lorenz serão paralelos à linha da perfeita igualdade, de modo que $D_{max} = p_j - \Phi_j$, $\forall j \in C$. Caso nenhuma pessoa tenha a mesma renda que μ, deve-se usar, necessariamente, a posição h.

Também é importante destacar a situação na qual o valor do índice de Gini coincide com o valor da discrepância máxima. Pode-se demonstrar que, sempre que houver apenas dois valores distintos de renda, a discrepância máxima será igual ao índice de Gini. Como caso específico[3] temos que $n = 2$ é condição suficiente para $D_{max} = G(\mathbf{z})$. É óbvio que, se $n = 2$, e as duas rendas forem iguais, teremos que o índice de Gini será igual à discrepância máxima, assumindo valor zero. Essa propriedade será importante na apresentação do índice de Dissimilaridade para segregação.

Trataremos de agora em diante da análise do índice de Gini para segregação. Os trabalhos de Jahn, Schmid e Schrag (1947) e Duncan e Duncan (1955a) são os primeiros a mostrar a estreita relação entre a poligonal de Lorenz e sua análoga para segregação, refletindo, também, a estreita relação das medidas de desigualdade com as medidas de segregação. Seja X a categoria de interesse (negros ou mulheres, por exemplo). Suporemos que os estratos estão ordenados conforme a proporção de pessoas X_j com relação à população total do estrato (T_j). Deste modo, lembrando a definição (2.5), temos

$$P_1 \leq P_2 \leq ... \leq P_{k-1} \leq P_k \tag{2.24}$$

[3] Ver dedução na seção A.2, p. 93, no apêndice.

ou

$$\frac{X_1}{X_1+Y_1} \leq \frac{X_2}{X_2+Y_2} \leq ... \leq \frac{X_{k-1}}{X_{k-1}+Y_{k-1}} \leq \frac{X_k}{X_k+Y_k}.$$

Podemos definir, obedecendo a ordenação anterior, a proporção acumulada de cada categoria até o j-ésimo estrato como

$$\Psi_j = \sum_{h=0}^{j} x_h, \text{ e } \rho_j = \sum_{h=0}^{j} y_h, \; j \in \{0, 1, ..., k\}. \tag{2.25}$$

Observe que o subíndice inclui a definição especial (2.2). Essa notação implica $(\rho_0, \Psi_0) = (0, 0)$. Montando um gráfico no plano cartesiano ortogonal ($[0,1] \times [0,1]$) com os pontos dados[4] por (2.25), teremos a Figura 2.2.

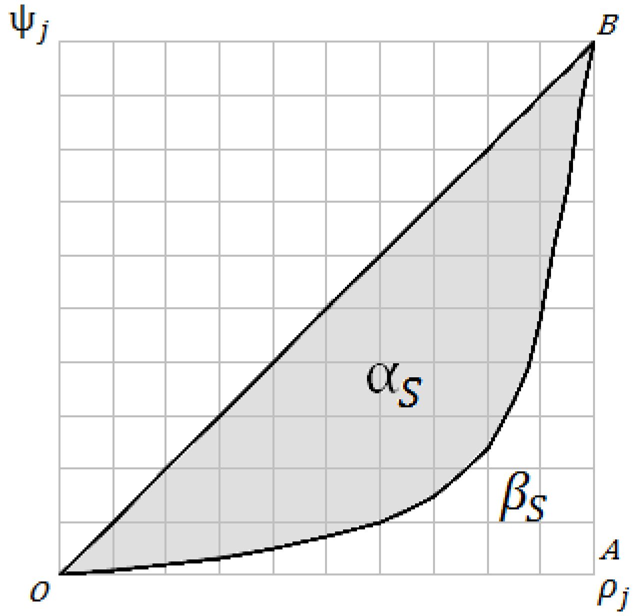

Figura 2.2. Curva de segregação

A curva convexa, que limita inferiormente a área α_S, é denominada *curva de segregação*[5]. Uma discussão detalhada e axiomática sobre as propriedades dessa curva pode ser encontrada em Hutchens (1991). A reta bissetriz do primeiro quadrante, limitada pelos pontos $O = (0, 0)$ e $B = (1, 1)$ é similar à linha da perfeita igualdade da Figura 2.1. A área α_S, limitada pela linha de perfeita igualdade da segregação e a curva de segregação, recebe o nome de *área de segregação*. Na situação de ausência de segregação teremos $\alpha_S = 0$. Por outro lado, na situação de segregação completa, temos $\max \alpha_S = 0,5$. Portanto, $0 \leq \alpha_S \leq 0,5$.

[4] Uma proposta alternativa para a construção deste gráfico é apresentada em Alonso-Villar e Río (2007). Os autores propõem o gráfico da proporção acumulada do grupo de interesse contra a proporção acumulada da população obedecendo a ordenação (2.24).

[5] A rigor deveríamos chamar de poligonal de segregação. A ideia seria a mesma da curva de Lorenz. Quando a quantidade de estratos tende a infinito a poligonal de segregação passa a se chamar curva de segregação. Por simplicidade, vamos usar em geral a expressão “curva de segregação”, mesmo quando, a rigor, se trata de uma poligonal.

A inclinação do j-ésimo segmento da curva de segregação será

$$d_j^S = \frac{\Delta \Psi_j}{\Delta \rho_j} = \frac{\Psi_j - \Psi_{j-1}}{\rho_j - \rho_{j-1}} = \frac{\sum_{h=0}^{j} x_h - \sum_{h=0}^{j-1} x_h}{\sum_{h=0}^{j} y_h - \sum_{h=0}^{j-1} y_h} = \frac{x_j}{y_j} \tag{2.26}$$

ou

$$d_j^S = \frac{X_j/X}{Y_j/Y}. \tag{2.27}$$

Assim, a inclinação de cada segmento da curva de segregação é a razão entre a participação do j-ésimo estrato no total da categoria X e a participação do mesmo estrato no total da outra categoria (Y). Essa inclinação será denominada *participação relativa* do j-ésimo estrato.

Como a curva de segregação é não decrescente, por se tratar de proporções acumuladas, temos $d_1^S \leq d_2^S \leq \cdots \leq d_{k-1}^S \leq d_k^S$, que implica

$$\frac{X_1}{Y_1} \leq \frac{X_2}{Y_2} \leq \cdots \leq \frac{X_{k-1}}{Y_{k-1}} \leq \frac{X_k}{Y_k}.$$

Portanto, a ordenação dos estratos pela eq. (2.24) ou pela ordenação acima são equivalentes. Hoffmann (1998, p. 253), ao apresentar um exemplo ilustrativo, prefere considerar os valores de X_j/Y_j para ordenar os estratos e calcular o índice de Gini para segregação.

Observe que a área de cada trapézio formado pela curva de segregação e o eixo das abscissas será

$$S_j^S = \frac{(\Psi_j + \Psi_{j-1})(\rho_j - \rho_{j-1})}{2}, \ j \in \{1, 2, ..., k\}.$$

Como β_S é a soma desses trapézios, temos

$$\beta_S = \frac{1}{2} \sum_{j=1}^{k} (\Psi_j + \Psi_{j-1})(\rho_j - \rho_{j-1}) \tag{2.28}$$

ou

$$\beta_S = \frac{1}{2} \sum_{j=1}^{k} (\Psi_j + \Psi_{j-1}) y_j. \tag{2.29}$$

A transformação de uma linha para outra é a mesma que a do denominador da eq. (2.26).

O *índice de Gini para segregação* ($G_S(\mathbf{X})$) é definido como a razão entre a área de segregação e seu respectivo valor máximo. Assim,

$$G_S(\mathbf{X}) = \frac{\alpha_S}{\max \alpha_S} \Rightarrow G_S(\mathbf{X}) = 2\alpha_S. \tag{2.30}$$

Da Figura 2.2, temos que $\alpha_S = 0,5 - \beta_S$. Substituindo esse resultado em (2.30) e utilizando (2.29), podemos escrever o índice de Gini para segregação como

$$G_S(\mathbf{X}) = 1 - \sum_{j=1}^{k} (\Psi_j + \Psi_{j-1}) y_j. \tag{2.31}$$

Analogamente à eq. (2.14), temos que

$$G_S(\mathbf{X}) = \frac{\Delta}{2\mu_S}. \tag{2.32}$$

Novamente, é importante dizer que pela formulação (2.32) não é necessário que os estratos estejam ordenados por valores não decrescentes de P_j. Observe, na eq. (2.16), que o peso de cada pessoa na população ($1/n$ ou π_i) corresponde ao incremento na abscissa do ponto correspondente à i-ésima pessoa. O valor análogo a esse na curva de segregação é y_j. Tomando a média ponderada das razões X_j/Y_j, com os fatores de ponderação y_j, temos

$$\mu_S = \sum_{j=1}^{k} \frac{X_j}{Y_j}\frac{Y_j}{Y} \qquad \text{ou}$$

$$\mu_S = \frac{X}{Y}. \tag{2.33}$$

Assim, a média (ponderada) das razões X_j/Y_j é igual a X/Y. Usando a eq. (2.18) e substituindo os valores de (2.27) e (2.33), temos

$$\Delta = \sum_{i=1}^{k}\sum_{j=1}^{k} \left|\frac{X_i}{Y_i} - \frac{X_j}{Y_j}\right| \frac{Y_i}{Y}\frac{Y_j}{Y}.$$

Desenvolvendo os termos e substituindo em (2.32), o índice de Gini para segregação pode ser escrito como

$$G_S(\mathbf{X}) = \frac{1}{2XY}\sum_{i=1}^{k}\sum_{j=1}^{k} \left|\frac{X_i}{Y_i} - \frac{X_j}{Y_j}\right| Y_iY_j.$$

Essa forma do $G_S(\mathbf{X})$ é encontrada em Hoffmann (1998, p. 256). Desenvolvendo a diferença modular, adicionando e subtraindo o termo X_iX_j dentro do módulo e utilizando a propriedade $T_h = X_h + Y_h$, para $h \in \{i, j\}$, chega-se a

$$G_S(\mathbf{X}) = \frac{1}{2XY}\sum_{i=1}^{k}\sum_{j=1}^{k} T_iT_j\left|\frac{X_i}{T_i} - \frac{X_j}{T_j}\right|. \tag{2.34}$$

Observe que

$$XY = \frac{X}{T}\frac{Y}{T}T^2 = P(1-P)T^2. \tag{2.35}$$

Substituindo (2.35) em (2.34) e lembrando (2.5) temos

$$G_S(\mathbf{X}) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} T_iT_j|P_i - P_j|}{2T^2P(1-P)}.$$

A maioria dos trabalhos na literatura utilizam essa formulação do índice de Gini para segregação. Alguns exemplos são Massey e Denton (1988), Carvalho et al. (2013) e James e Taeuber (1985).

De modo semelhante à eq. (2.23), e lembrando que $\mu_S = X/Y$, teremos que a discrepância máxima do índice de Gini para segregação será

$$D^S_{max} = \frac{\delta}{2}\frac{Y}{X}. \tag{2.36}$$

Essa medida será discutida após a apresentação do índice de Dissimilaridade.

A Tabela 2.1 apresenta um exemplo numérico hipotético para ilustrar o cálculo das medidas de segregação. As duas categorias de análise (X e Y) são distribuídas em quatro estratos.

Tabela 2.1. Distribuição das categorias X e Y em quatro estratos hipotéticos

Estrato	X_j	Y_j	X_j/Y_j	x_j	y_j	x_j/y_j	ρ_j	Ψ_j	$\lvert x_j - y_j \rvert$
1	1	5	0,2	0,0909	0,3846	0,2364	0,0909	0,3846	0,2937
2	2	4	0,5	0,1818	0,3077	0,5909	0,2727	0,6923	0,1259
3	3	3	1	0,2727	0,2308	1,1818	0,5455	0,9231	0,0420
4	5	1	5	0,4545	0,0769	5,9091	1	1	0,3776
Total	11	13							0,8392

Fonte: elaboração própria.

Para os dados da tabela, o valor do índice de Gini para segregação é 0,5454. Uma característica bastante difundida a favor desse índice é o atendimento a uma propriedade similar à condição de Pigou-Dalton para distribuição de renda (BUTLER, 1987; HUTCHENS, 1991; SILBER, 1989). Essa propriedade, no caso para segregação, será definida formalmente na seção 3.5, p. 46. A próxima seção apresenta, de forma pormenorizada, o índice de Dissimilaridade para segregação.

2.3. Índice de Dissimilaridade

O índice de Dissimilaridade é a medida mais utilizada na literatura sobre segregação e foi proposto por Jahn, Schmid e Schrag (1947). Os autores definem um *escore* como a diferença, em valor absoluto, entre o número observado de pessoas do grupo de interesse e o valor esperado desse valor, sob a hipótese de que a proporção de pessoas de cada categoria seja a mesma em todos os estratos. Admitindo que a categoria de interesse é a X, temos o *escore* do j-ésimo estrato definido por

$$escore_j = \left| X_j - \frac{X}{T} T_j \right|.$$

Cada *escore* terá valor máximo quando só existirem pessoas de um tipo em cada estrato. Desse modo, considerando essa suposição e ordenando os estratos de modo que até o h-ésimo estrato temos somente pessoas da categoria X e do $(h+1)$-ésimo estrato até o último só pessoas da categoria Y, ou seja,

$$\begin{aligned} X_1, ..., X_h \neq 0, \qquad & X_{h+1} = ... = X_k = 0 \qquad \text{e} \\ Y_1 = ... = Y_h = 0, \qquad & Y_{h+1}, ..., Y_k \neq 0, \end{aligned}$$

o valor máximo da soma dos *escores* será

$$\max \sum_{j=1}^{k} escore_j = \sum_{j=1}^{h} \left| X_j - \frac{X}{X+Y} X_j \right| + \sum_{j=h+1}^{k} \left| 0 - \frac{X}{X+Y} Y_j \right| \quad \text{ou}$$

$$\max \sum_{j=1}^{k} escore_j = 2\left(X - \frac{X^2}{X+Y} \right).$$

O *índice de Dissimilaridade* ($D(\mathbf{X})$), segundo Jahn, Schmid e Schrag (1947), é definido como a razão entre a soma dos *escores* e seu valor máximo. Assim, temos que

$$D(\mathbf{X}) = \frac{\sum_{j=1}^{k} escore_j}{\max \sum_{j=1}^{k} escore_j} \Rightarrow D(\mathbf{X}) = \frac{\sum_{j=1}^{k} \left| X_j - \frac{X}{T} T_j \right|}{2\left(X - \frac{X^2}{T} \right)}. \tag{2.37}$$

Desenvolvendo a eq. (2.37) podemos encontrar a expressão

$$D(\mathbf{X}) = \frac{1}{2} \sum_{j=1}^{k} \left| \frac{X_j}{X} - \frac{Y_j}{Y} \right| = \frac{1}{2} \sum_{j=1}^{k} |x_j - y_j|. \tag{2.38}$$

Essa é a formulação mais difundida do índice, pois foi amplamente utilizada após o trabalho de Duncan e Duncan (1955a). Como $D(\mathbf{X})$ é uma soma de valores absolutos, o menor valor de $D(\mathbf{X})$ será zero, se e somente se, $x_j = y_j$, para todo j. Por outro lado, para cada j, teremos que a diferença modular será máxima quando uma das duas parcelas da diferença é zero. Desse modo, o valor máximo de $D(\mathbf{X})$ será 1. É desnecessário fazer a prova formal desse limite superior, pois o índice foi definido como a soma dos *escores* dividido por seu valor máximo. Portanto, temos que $0 \leq D(\mathbf{X}) \leq 1$. Para os dados hipotéticos da Tabela 2.1, o índice de Dissimilaridade assume valor 0,4196.

Desenvolvendo a eq. (2.38) podemos encontrar

$$D(\mathbf{X}) = \frac{1}{2XY} \sum_{j=1}^{k} |X_j T - X T_j|. \tag{2.39}$$

Substituindo (2.35) em (2.39) e desenvolvendo os termos, temos

$$D(\mathbf{X}) = \sum_{j=1}^{k} \frac{T_j |P_j - P|}{2TP(1-P)}. \tag{2.40}$$

A eq. (2.40) é outra forma muito comum de expressar o índice de Dissimilaridade, como encontrado em James e Taeuber (1985), Massey e Denton (1988) e Harrison (2001), por exemplo. De modo geral, quando os trabalhos são relacionados a segregação geográfica a forma (2.40) é a mais comum. Quando a análise é de segregação no mercado de trabalho é mais frequente a formulação (2.38).

Da eq. (2.33) podemos calcular o desvio (absoluto) médio das razões X_j/Y_j ponderadas por y_j como

$$\delta = \sum_{j=1}^{k} \left| \frac{X_j}{Y_j} - \mu_S \right| y_j. \tag{2.41}$$

Substituindo $\mu_S = X/Y$ e $y_j = Y_j/Y$ em (2.41), e desenvolvendo os termos, temos

$$\delta = \sum_{j=1}^{k} \left| \frac{X_j}{Y} - \frac{XY_j}{Y^2} \right|.$$

Lembrando que a discrepância máxima do índice de Gini para segregação é $\delta Y/(2X)$ (eq. (2.36)), tem-se

$$D_{max}^{S} = \frac{1}{2} \sum_{j=1}^{k} \left| \frac{X_j}{X} - \frac{Y_j}{Y} \right|. \tag{2.42}$$

Comparando a eq. (2.42) com a eq. (2.38), verifica-se que o índice de Dissimilaridade é igual à discrepância máxima da curva de segregação. Essa propriedade é bem retratada na literatura, porém existe uma relação entre o índice de Gini para segregação e o índice de Dissimilaridade que raramente é citada.

Do resultado anterior, temos que a discrepância máxima é o índice de Dissimilaridade. Para o índice de Gini temos que, se $n = 2$, o valor do índice coincidirá com a discrepância máxima. Assim, quando analisamos as duas categorias divididas em dois estratos teremos que $D(\mathbf{X}_{2\times 2}) = G_S(\mathbf{X}_{2\times 2})$. De modo geral, desde que haja apenas dois valores distintos para a proporção P_j, independentemente da quantidade de estratos, o índice de Gini para segregação será igual ao índice de Dissimilaridade. Na próxima seção, será apresentada a medida geral de desigualdade e seus casos particulares.

2.4. Medida geral de desigualdade

O artigo *The class of additively decomposable inequality measures*, de Shorrocks (1980), apresenta detalhadamente as propriedades de uma medida geral de desigualdade proposta por vários trabalhos no início da década de 80. De acordo com o autor, uma medida de desigualdade é dita *aditivamente decomponível* quando pode ser expressa como a soma de dois termos, o primeiro sendo a soma ponderada das medidas de desigualdade *dentro dos grupos* e, o segundo, referente à medida da desigualdade *entre grupos*.

Seja z a variável renda, de modo que z_i represente a renda da i-ésima pessoa com $i \in \{1, 2, ..., n\}$. Esses indivíduos podem ser divididos em k grupos conforme alguma característica de interesse. Nota-se que esses grupos não precisam, necessariamente, ser criados a partir de faixas de renda. Podem ser definidos com base em regiões, setores, diferenças raciais ou religiosas, por exemplo. Assim, temos que o subíndice para o grupo j assume valores no conjunto $\{1, 2, ..., k\}$, com $n \geq k$. O número de pessoas no j-ésimo grupo será denotado por n_j.

Das definições apresentadas, temos que uma medida de desigualdade $I(\mathbf{z}; n)$ é dita aditivamente decomponível se pode ser escrita como

$$I(\mathbf{z}; n) = I(\mathbf{z}^1, \mathbf{z}^2, ..., \mathbf{z}^k; n) \qquad \text{ou}$$
$$I(\mathbf{z}; n) = \sum_{j=1}^{k} w^j(\boldsymbol{\mu}; \mathbf{n}) I^j(\mathbf{z}^j; n_j) + B(\boldsymbol{\mu}; \mathbf{n}) \tag{2.43}$$

onde $\mathbf{z}^j$ é um vetor com os n_j valores de renda das pessoas do grupo j. Na eq. (2.43), a expressão $I^j(\mathbf{z}^j; n_j)$ é a medida de desigualdade entre os indivíduos do j-ésimo grupo, que é ponderado por $w^j(\boldsymbol{\mu}; \mathbf{n})$, e $B(\boldsymbol{\mu}; \mathbf{n})$ é o termo referente à desigualdade entre grupos. $\boldsymbol{\mu}$ é o vetor com as médias por grupo $[\mu_1\, \mu_2\, ...\, \mu_k]$ e $\mathbf{n}$ o vetor com os valores do número de indivíduos em cada grupo $[n_1\, n_2\, ...\, n_k]$. A *medida geral de desigualdade* apresentada por Cowell (1980), Shorrocks (1980, 1984) e Cowell e Kuga (1981a, 1981b) é[6]

$$I(\mathbf{z}; n) = \frac{1}{\varepsilon(1-\varepsilon)} \left[1 - \frac{1}{n} \sum_{i=1}^{n} \left(\frac{z_i}{\mu} \right)^{1-\varepsilon} \right], \tag{2.44}$$

onde z_i/μ é a renda relativa da i-ésima pessoa, que é, também, o valor da declividade do segmento da poligonal de Lorenz referente a esse indivíduo. ε é um parâmetro da medida que assume valores reais, considerando-se, normalmente, valores de -1 a 1.

De (2.44), a desigualdade dentro do grupo j com n_j membros é

$$I^j(\mathbf{z}^j; n_j) = \frac{1}{\varepsilon(1-\varepsilon)} \left[1 - \frac{1}{n_j} \sum_{i=1}^{n_j} \left(\frac{z_{ji}}{\mu_j} \right)^{1-\varepsilon} \right]. \tag{2.45}$$

em que z_{ji} indica que a i-ésima pessoa pertence ao grupo j e μ_j é a renda média desse grupo. Também tem-se de (2.44) que a desigualdade entre todos os n membros será

$$I(\mathbf{z}; n) = \frac{1}{\varepsilon(1-\varepsilon)} \left[1 - \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(\frac{z_{ji}}{\mu} \right)^{1-\varepsilon} \right]. \tag{2.46}$$

Adicionando e subtraindo

$$\frac{1}{\varepsilon(1-\varepsilon)} \sum_{j=1}^{k} \frac{n_j}{n} \left(\frac{\mu_j}{\mu} \right)^{1-\varepsilon}$$

em (2.46) e usando (2.45), temos que a decomposição de $I(\mathbf{z}; n)$ é

$$I(\mathbf{z}; n) = \sum_{j=1}^{k} \frac{n_j}{n} \left(\frac{\mu_j}{\mu} \right)^{1-\varepsilon} I^j(\mathbf{z}^j; n_j) + \frac{1}{\varepsilon(1-\varepsilon)} \left[1 - \sum_{j=1}^{k} \frac{n_j}{n} \left(\frac{\mu_j}{\mu} \right)^{1-\varepsilon} \right]. \tag{2.47}$$

[6] Existem pequenas diferenças de uma formulação para outra. Para os trabalhos de Shorrocks o parâmetro de análise é $c = 1 - \varepsilon$. Nos trabalhos de Cowell temos $\beta = -\varepsilon$. A notação apresentada segue Hoffmann (1998).

Comparando a expressão (2.47) com (2.43), temos que

$$B(\boldsymbol{\mu};\mathbf{n}) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1-\sum_{j=1}^{k}\frac{n_j}{n}\left(\frac{\mu_j}{\mu}\right)^{1-\varepsilon}\right]. \tag{2.48}$$

A eq. (2.48) é a medida de desigualdade *entre grupos*. Observe que os pesos[7] das medidas dentro dos grupos são dados por

$$w^j(\boldsymbol{\mu};\mathbf{n}) = \frac{n_j}{n}\left(\frac{\mu_j}{\mu}\right)^{1-\varepsilon}.$$

Para simplificar as expressões, suprimiremos os argumentos referentes aos tamanhos da população ou dos grupos. Segundo Shorrocks (1980, p. 623), o limite inferior da medida (2.44) é sempre zero, enquanto que seu limite superior, para $\varepsilon < 1$, é

$$\max I(\mathbf{z};\varepsilon<1) = \frac{1-n^{-\varepsilon}}{\varepsilon(1-\varepsilon)}.$$

Caso $\varepsilon = 0$, é fácil verificar, pelo procedimento de L'Hôpital, que $\max I(\mathbf{z};\varepsilon = 0) = \ln(n)$. Ainda segundo o autor, se $\varepsilon \geq 1$, então a medida geral de desigualdade não é limitada superiormente.

Note que essa medida não está definida para $\varepsilon = 1$ ou $\varepsilon = 0$, por se tratar de uma indefinição matemática. Conforme a regra de L'Hôpital, derivando o numerador e o denominador em relação a ε, temos

$$\lim_{\varepsilon\to d} I(\mathbf{z}) = \lim_{\varepsilon\to d}\frac{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{z_i}{\mu}\right)^{1-\varepsilon}\ln\left(\frac{z_i}{\mu}\right)}{1-2\varepsilon} \tag{2.49}$$

onde d é igual a 0 ou 1. Para $d = 1$ teremos

$$\lim_{\varepsilon\to 1} I(\mathbf{z}) = -\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{z_i}{\mu}\right) \Rightarrow I(\mathbf{z};\varepsilon=1) = L(\mathbf{z}). \tag{2.50}$$

Temos, portanto, que no limite, quando ε tende a um, a medida geral de desigualdade é o índice L de Theil (1967, p. 126). Esse também é conhecido na literatura como a segunda medida de desigualdade de Theil. Se todos os indivíduos possuem a mesma renda, temos que $z_i = \mu$, para todo i, que implica $L(\mathbf{z}) = 0$. Por outro lado, quando pelo menos um indivíduo tem renda nula, sua contribuição para o índice tende a infinito. Deste modo, temos que $L(\mathbf{z}) \in [0,\infty)$.

Para o índice L de Theil, o coeficiente de decomposição referente ao j-ésimo grupo é $w^j(\boldsymbol{\mu};\varepsilon = 1) = n_j/n$. Observe que esse peso corresponde à participação do grupo na

[7] Em Shorrocks (1980), por exemplo, os argumentos dos pesos são dados por $w^j(\boldsymbol{\mu};\mathbf{n})$. Estes argumentos são suficientes para o cálculo do peso de cada grupo, mas não necessários. Poderíamos indicar os argumentos necessários escrevendo $w^j(\mu_j,\mu;n_j,n)$.

população. Por isso, diz-se que o índice L de Theil é uma medida "democrática". Nota-se também que a soma desses pesos é igual a um.

Da eq. (2.49), aplicando o limite quando $\varepsilon \to 0$, temos

$$\lim_{\varepsilon \to 0} I(\mathbf{z}) = \sum_{i=1}^{n} \left(\frac{z_i}{n\mu}\right) \ln\left(\frac{z_i}{\mu}\right) \Rightarrow I(\mathbf{z}; \varepsilon = 0) = T(\mathbf{z}) \tag{2.51}$$

que é o índice T de Theil (1967, p. 91), ou a primeira medida de desigualdade de Theil. De modo semelhante à medida L de Theil, quando todas as pessoas possuem a mesma renda o índice assume valor zero. Por outro lado, o valor máximo de $T(\mathbf{z})$ será $\ln(n)$ quando um indivíduo possuir toda a renda e os demais possuírem renda nula. Portanto, temos que $T(\mathbf{z}) \in [0, \ln(n)]$. A prova formal desse resultado, a partir da teoria da informação, envolve o multiplicador de Lagrange e pode ser encontrada em Shannon (1948) ou Theil (1967, 1972). O trabalho de Shannon (1948) é a principal obra sobre Teoria da Informação e também a referência para Henri Theil desenvolver as medidas L e T utilizando os conceitos de informação e entropia.

Para o índice T de Theil, o coeficiente de decomposição referente ao j-ésimo grupo é igual a

$$w^j(\boldsymbol{\mu}; \varepsilon = 0) = \frac{\mu_j n_j}{\mu n}. \tag{2.52}$$

Observe que o numerador dessa expressão é a renda total no grupo j, enquanto que o denominador é a renda total da economia. Desse modo, esse índice pondera a desigualdade dentro dos grupos pela sua participação na renda na população. Assim, diferentemente da medida $L(\mathbf{z})$, o índice T de Theil é uma medida "não-democrática". Entretanto, Hoffmann (1998, p. 114) ressalta que ponderações não-democráticas são usuais em economia.

Conforme a eq. (2.52), a soma desses pesos também é igual a um. Somente para o T e o L de Theil, como casos da medida geral de desigualdade, a soma desses coeficientes é igual a um. Portanto, nesses casos específicos, a *soma ponderada* das medidas de desigualdade dentro dos grupos torna-se uma *média ponderada*. Além disso, pode ser demonstrado que, se $\varepsilon \notin \{0, 1\}$, então $1 - \sum_j w^j(\boldsymbol{\mu})$ é proporcional à parcela entre grupos $B(\boldsymbol{\mu})$ da eq. (2.44) (SHORROCKS, 1980, p. 624).

Sendo z a variável renda, temos que sua média será μ e seu desvio-padrão igual a

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^{n} z_i^2 - \mu^2}.$$

Assim, o *coeficiente de variação* da renda, definido por $CV(\mathbf{z}) = \sigma/\mu$, será

$$CV(\mathbf{z}) = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n} z_i^2 - \mu^2}{\mu^2}} \qquad \text{ou} \qquad CV(\mathbf{z}) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{z_i}{\mu}\right)^2 - 1}. \tag{2.53}$$

Observe que, utilizando $\varepsilon = -1$ na eq. (2.44) da medida geral de desigualdade, temos

$$I(\mathbf{z}; \varepsilon = -1) = \frac{1}{2}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{z_i}{\mu}\right)^2 - 1\right]. \tag{2.54}$$

Substituindo (2.54) em (2.53), temos que o coeficiente de variação é uma transformação monotônica crescente de $I(\mathbf{z}; \varepsilon = -1)$, dada por $CV(\mathbf{z}) = \sqrt{2I(\mathbf{z}; \varepsilon = -1)}$.

Um outro grupo de medidas que se relaciona com a medida geral é a família de medidas de desigualdade de Atkinson (1970). Pressupondo que $\varepsilon > 0$ e $\varepsilon \neq 1$, essa família de medidas, denotada por $A(\mathbf{z})$, será

$$A(\mathbf{z}) = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{z_i}{\mu}\right)^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}. \tag{2.55}$$

Desde que pelo menos uma pessoa tenha renda positiva, o termo entre colchetes da equação anterior é sempre positivo. De acordo com a equação da medida geral, este termo é igual a $1 - \varepsilon(1-\varepsilon)I(\mathbf{z})$. Assim,

$$A(\mathbf{z}) = 1 - [1 - \varepsilon(1-\varepsilon)I(\mathbf{z})]^{\frac{1}{1-\varepsilon}}. \tag{2.56}$$

Como $\varepsilon > 0$ e $\varepsilon \neq 1$, temos que

$$\frac{\partial A(\mathbf{z})}{\partial I(\mathbf{z})} = \varepsilon\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{z_i}{\mu}\right)^{1-\varepsilon}\right]^{\frac{\varepsilon}{1-\varepsilon}} > 0.$$

Logo, a família de medidas de desigualdade de Atkinson é uma transformação monotônica crescente da medida $I(\mathbf{z})$, dada pela eq. (2.56). É fácil verificar na eq. (2.55) que, se $0 < \varepsilon < 1$, e um único indivíduo se apropria de toda a renda, então o valor da medida será $\max A(\mathbf{z}) = 1 - n^{\frac{-\varepsilon}{1-\varepsilon}}$. Por outro lado, se $\varepsilon \geq 1$, basta que um indivíduo tenha renda nula para que a medida assuma valor 1. O valor para ε igual a um pode ser obtido pelo procedimento de L'Hôpital. Para uma distribuição perfeitamente igualitária ($z_i = \mu$, $\forall i$), a medida sempre assumirá valor zero, independentemente do valor de ε. Além disso, a família de medidas de Atkinson, por ser uma transformação estritamente crescente da medida geral de segregação, atende à condição de Pigou-Dalton.

Também existe uma medida de Atkinson análoga para segregação. Conforme James e Taeuber (1985)[8], a família de medidas de Atkinson para segregação ($A'_S(\mathbf{X})$) é

$$A'_S(\mathbf{X}) = 1 - \frac{P}{1-P}\left[\frac{\sum_{j=1}^{k}(1-P_j)^{1-\varepsilon}\, P_j^{\varepsilon}\, T_j}{PT}\right]^{\frac{1}{1-\varepsilon}}, \ \varepsilon \in (0,1). \tag{2.57}$$

No capítulo posterior será deduzida essa expressão a partir de uma medida geral de segregação desenvolvida neste trabalho.

[8] Ver a errata do artigo de James e Taeuber (1985), Errata... (1986), para a expressão correta dessa família de medidas.

É curioso como é mais comum encontrar referências à família da medida de Atkinson para segregação do que a alguma medida similar ao T e L de Theil para segregação. Com exceção da formulação do T de Theil para segregação em Hutchens (1991), que raramente é citada na literatura, não há outras menções a medidas de segregação análogas ao L de Theil para segregação. Talvez esse fato ocorra devido à construção de uma medida de entropia racial desenvolvida pelo próprio Theil e Finizza (1971) e aperfeiçoada posteriormente em Theil (1972, seções 1.4, 2.2 e 3.2).

No próximo capítulo é apresentada e discutida a proposta de uma medida geral para segregação, assim como a análise de suas propriedades, sua decomposição e seus casos particulares.

3. ENSAIO SOBRE UMA MEDIDA GERAL DE SEGREGAÇÃO

3.1. Uma medida geral de segregação

Na poligonal de Lorenz definida na seção 2.2, temos na abscissa do plano cartesiano a proporção acumulada da população. A distância, nesse eixo, entre dois indivíduos consecutivos, é a participação relativa de cada um na população ($1/n$). Observe que esse valor é o peso referente a qualquer indivíduo na equação da medida geral de desigualdade. O valor similar a esse na curva de segregação é y_j, que é uma característica do j-ésimo estrato.

Temos, também, que a declividade do i-ésimo segmento da poligonal de Lorenz é igual a z_i/μ. Na medida geral de desigualdade (eq. (2.44)), esse é o valor que aparece dentro do somatório. Na curva de segregação, a declividade do j-ésimo estrato é x_j/y_j.

Utilizando os valores da declividade dos segmentos da curva de segregação (x_j/y_j) e a participação do j-ésimo estrato no total da categoria Y (y_j), podemos definir uma medida de segregação similar à medida geral de desigualdade. Esta medida é

$$I_S(\mathbf{X}) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{j=1}^{k} y_j \left(\frac{x_j}{y_j}\right)^{1-\varepsilon}\right]. \tag{3.1}$$

Essa medida será denominada *medida geral de segregação*. É interessante notar como o conceito de poligonal de Lorenz não é relacionado à construção da medida geral de desigualdade, porém, a utilização desse recurso facilita o desenvolvimento de uma medida análoga para segregação.

Uma versão modificada da eq. (3.1) aparece na prova do primeiro teorema em Hutchens (2004, p. 574), porém, o autor utiliza uma transformação dessa formulação para todas as propriedades e teoremas em seu artigo. Neste capítulo, vamos demonstrar as implicações dessa transformação, defendendo o uso da medida em sua forma original. É importante destacar que a eq. (3.1) foi deduzida aqui de forma independente do trabalho de Hutchens, pois o referido autor não utiliza a analogia entre a curva de segregação e a poligonal de Lorenz. Entretanto, os dois raciocínios convergiram para a mesma expressão.

Lembramos que a partir da medida geral de desigualdade é possível encontrar as medidas T e L de Theil, uma transformação para o coeficiente de variação e outra para a família de Atkinson. Assim, é razoável supor que existam medidas análogas a essas para o caso da segregação, a partir da medida geral de segregação (eq. (3.1)). Além disso, poder-se-ia imaginar que diversas propriedades da medida geral de desigualdade também são atendidas para sua análoga de segregação, sendo uma delas a característica de ser aditivamente decomponível. Vamos mostrar que essas analogias são possíveis, porém, deve-se utilizar a eq. (3.1) ao invés da transformação apresentada em Hutchens (2004).

Essa analogia das medidas gerais de segregação e desigualdade, sem utilizar nenhuma tipo de transformação, é a principal contribuição teórica desse capítulo.

Seja $O_c(\mathbf{X})$ a medida geral de segregação de Hutchens (2004, p. 563), utilizada ao longo de todo seu artigo. Assim, as medidas $I_S(\mathbf{X})$ e $O_c(\mathbf{X})$ se relacionam da seguinte forma: $O_c(\mathbf{X}) = \varepsilon(1-\varepsilon)I_S(\mathbf{X})$, para $\varepsilon \in (0,1)$. Desse modo,

$$O_c(\mathbf{X}) = 1 - \sum_{j=1}^{k} y_j \left(\frac{x_j}{y_j}\right)^{1-\varepsilon}, \ \varepsilon \in (0,1). \tag{3.2}$$

É importante ressaltar que, nas palavras do autor (HUTCHENS, 2004, p. 563), "deve-se referir a $O_c(\mathbf{X})$ como uma medida generalizada de entropia de segregação"[1]. Nota-se, ainda, que o $I_S(\mathbf{X})$ inclui os casos em que $\varepsilon \notin (0,1)$.

A primeira implicação ao utilizar a transformação proposta pelo autor é que limitaria a medida geral de segregação para valores de ε no intervalo $(0,1)$. Por exemplo, para ε igual a zero ou um, a medida de Hutchens necessariamente seria igual a zero, independentemente de como as categorias se distribuam nos estratos. Ao utilizar a medida $I_S(\mathbf{X})$, permitiremos que ε assuma qualquer valor real.

Analisaremos a contribuição de cada estrato para o total do valor da medida. Da eq. (3.1), a medida geral de segregação pode ser reescrita como

$$I_S(\mathbf{X}) = \frac{1}{\varepsilon(1-\varepsilon)} \sum_{j=1}^{k} \left[\frac{1}{k} - y_j \left(\frac{x_j}{y_j}\right)^{1-\varepsilon}\right].$$

Logo, considerando ε diferente de zero ou um, a contribuição do j-ésimo estrato no total da medida (s_j) será

$$s_j = \frac{1/k - y_j^{\varepsilon} x_j^{1-\varepsilon}}{\varepsilon(1-\varepsilon)}.$$

Se ε está entre zero e um, no limite quando $x_j \to 0$ ou $y_j \to 0$, a contribuição desse estrato no total é $1/[k\varepsilon(1-\varepsilon)]$. Desse modo, quando todos os estratos são ocupados exclusivamente por uma categoria, e $\varepsilon \in (0,1)$, o valor máximo da medida será $1/[\varepsilon(1-\varepsilon)]$.

Se $\varepsilon > 1$, no limite quando $x_j \to 0$, a participação s_j tenderá a infinito. Por outro lado, se existe um estrato h tal que $y_h \to 0$, então $s_h = 1/[k\varepsilon(1-\varepsilon)]$. Desse modo, se $\varepsilon > 1$ e pelo menos um estrato seja ocupado exclusivamente por pessoas da categoria Y, a medida geral de segregação tenderá a infinito. Será demonstrado posteriormente que o mesmo resultado é válido para $\varepsilon = 1$.

Se $\varepsilon < 0$ e existe pelo menos um estrato tal que $x_j \to 0$, então temos que $s_j = 1/[k\varepsilon(1-\varepsilon)]$. De modo semelhante, se existe outro estrato tal que $y_h \to 0$, então $s_h \to \infty$. Logo, se existe pelo menos um estrato que seja ocupado exclusivamente pela categoria X e $\varepsilon < 0$, então a medida geral de segregação tenderá a infinito. Tal resultado também é válido para $\varepsilon = 0$ (veja p. 39).

[1] $O_c(\mathbf{X})$ *shall be referred to as a generalized entropy measure of segregation.*

Pode-se demonstrar que a medida geral de segregação não assume valores negativos[2], independentemente do valor de ε. Assim, utilizando esse resultado com os dos parágrafos anteriores, concluímos que a medida geral de segregação assume valores de zero a $1/[\varepsilon(1-\varepsilon)]$ se $\varepsilon \in (0,1)$, e poderá tender a infinito caso ε não pertença a esse intervalo.

Outra medida de segregação encontrada na literatura é o índice Raiz Quadrada de Hutchens (2001). Essa medida é o caso particular de $O_c(\mathbf{X})$ para $\varepsilon = 0,5$ ou $O_c(\mathbf{X}; \varepsilon = 0,5)$. Verifica-se que $I_S(\mathbf{X}; \varepsilon = 0,5) = 4O_c(\mathbf{X}; \varepsilon = 0,5)$ ou

$$I_S(\mathbf{X}; \varepsilon = 0,5) = 4\Big(1 - \sum_{j=1}^{k} \sqrt{x_j y_j}\Big).$$

De agora em diante denominaremos $I_S(\mathbf{X}; \varepsilon = 0,5)$ de *índice Raiz Quadrada* e $O_c(\mathbf{X}; \varepsilon = 0,5)$ de *índice Raiz Quadrada de Hutchens*. Para os dados da Tabela 2.1, é possível verificar que o índice Raiz Quadrada é igual a $0,5545$ e o índice Raiz Quadrada de Hutchens igual a $0,1386$. Hutchens (2004) argumenta que o $O_c(\mathbf{X}; \varepsilon = 0,5)$ é a melhor medida de segregação, pois obedece a diversas propriedades desejáveis de uma medida, conforme julgamento do autor.

Tendo em vista que $I_S(\mathbf{X})$ foi adaptada a partir de uma medida de desigualdade aditivamente decomponível, a próxima seção apresenta a decomposição da medida geral de segregação.

3.2. Decomposição da medida geral de segregação

A contribuição de Cowell (1980), Shorrocks (1980, 1984) e Cowell e Kuga (1981a, 1981b) foi a formulação de uma classe de medidas de desigualdade aditivamente decomponíveis. Tendo em vista a semelhança entre a medida dos autores e a desenvolvida neste trabalho, esta seção buscará decompor a medida $I_S(\mathbf{X})$ de modo semelhante à medida de desigualdade. O objetivo é apresentar uma medida de segregação aditivamente decomponível, isto é, decompor seu valor em duas parcelas. A primeira deve ser referente à segregação dentro dos grupamentos de estratos e, a segunda, à segregação entre os grupamentos.

O subíndice g será utilizado para denotar um grupamento de estratos que toma valores em um conjunto $G = \{1, 2, ..., G\}$. Claramente, em uma economia com k estratos, devemos ter que a cardinalidade do conjunto G deve ser menor ou igual a k. É importante destacar que cada estrato deve pertencer a um e somente um grupamento. Deste modo, da eq. (3.1), temos que a medida de segregação será

$$I_S(\mathbf{X}) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{g \in G}\sum_{j=1}^{n_g} y_{gj}\left(\frac{x_{gj}}{y_{gj}}\right)^{1-\varepsilon}\right], \tag{3.3}$$

[2] Veja Propriedade 7, p. 47, e a respectiva demonstração na página 93.

com $y_{gj} = Y_{gj}/Y$ e $x_{gj} = X_{gj}/X$, sendo Y_{gj} e X_{gj} os números de pessoas das categorias Y e X, respectivamente, no j-ésimo estrato. O valor n_g indica a quantidade de estratos no g-ésimo grupamento.

A segregação dentro do grupamento g, de acordo com (3.1), será

$$I_S^g(\mathbf{X}^g) = \frac{1}{\varepsilon(1-\varepsilon)} \left[1 - \sum_{j=1}^{n_g} \frac{Y_{gj}}{Y_g} \left(\frac{X_{gj}/X_g}{Y_{gj}/Y_g} \right)^{1-\varepsilon} \right],$$

As quantidades Y_g e X_g indicam, respectivamente, o número de pessoas das categorias Y e X no g-ésimo grupamento. Também é importante acrescentar a notação $x_g = X_g/X$ e $y_g = Y_g/Y$, que são as participações de cada grupamento nas duas categorias.

Adicionando e subtraindo o termo

$$\sum_{g \in G} y_g \left(\frac{x_g}{y_g} \right)^{1-\varepsilon}$$

dentro dos colchetes da eq. (3.3), e desenvolvendo o resultado, teremos

$$I_S(\mathbf{X}) = \sum_{g \in G} y_g \left(\frac{x_g}{y_g} \right)^{1-\varepsilon} \frac{1}{\varepsilon(1-\varepsilon)} \left[1 - \sum_{j=1}^{n_g} \frac{Y_{gj}}{Y_g} \left(\frac{X_{gj}/X_g}{Y_{gj}/Y_g} \right)^{1-\varepsilon} \right] + \frac{1}{\varepsilon(1-\varepsilon)} \left[1 - \sum_{g \in G} y_g \left(\frac{x_g}{y_g} \right)^{1-\varepsilon} \right],$$

ou

$$I_S(\mathbf{X}) = \sum_{g \in G} w_S^g(\mathbf{X}^g) I_S^g(\mathbf{X}^g) + B_S(\mathbf{X}),$$

onde $\mathbf{X}^g$ é uma submatriz da matriz particionada $\mathbf{X}$ dada por

$$\mathbf{X} = \left[\begin{array}{ccccccc} \mathbf{X}^1 & \mathbf{X}^2 & \ldots & \mathbf{X}^g & \ldots & \mathbf{X}^G \end{array} \right], \text{ com } \mathbf{X}^g_{2 \times n_g}, \forall g,$$

e

$$w_S^g(\mathbf{X}^g) = y_g \left(\frac{x_g}{y_g} \right)^{1-\varepsilon}. \tag{3.4}$$

Destaca-se que $\mathbf{X}^g$ é argumento suficiente para o cálculo dos pesos w_S^g. Os argumentos necessários são $w_S^g(X_g, Y_g, X, Y)$.

Assim, temos que a medida de segregação entre os estratos ($I_S(\mathbf{X})$) pode ser decomposta em uma parcela referente à segregação entre grupamentos ($B_S(\mathbf{X})$) e outra parcela que é uma soma ponderada das medidas de segregação dentro de cada grupamento.

Observe que a soma dos pesos é igual a

$$\sum_{g=1}^{G} w_S^g(\mathbf{X}^g) = 1 - \left[1 - \sum_{g=1}^{G} y_g \left(\frac{x_g}{y_g} \right)^{1-\varepsilon} \right]$$

ou

$$\sum_{g=1}^{G} w_S^g(\mathbf{X}^g) = 1 - \varepsilon(1-\varepsilon) B_S(\mathbf{X}).$$

São duas as principais implicações da equação anterior. A primeira é que somente para ε igual a zero ou um, os pesos são independentes da parcela referente à segregação entre os grupamentos. Além disso, somente para esses valores, podemos garantir que a soma dos pesos será igual a um.

Pode ser demostrado[3] que $B_S(\mathbf{X}) \geq 0$. Desse modo, se $B_S(\mathbf{X}) = 0$, os pesos também somam um. Nessa situação, a segregação considerando todos os estratos corresponde à soma ponderada das medidas de segregação dentro dos grupamentos. Assim, ε igual a zero ou um é suficiente para que os pesos tenham soma igual a um e sejam independentes da parcela entre grupamentos. Caso $B_S(\mathbf{X})$ seja maior que zero, somente com ε igual a um ou zero teremos a soma dos pesos igual à unidade. De fato, empiricamente, é extremamente rara a situação em que a parcela entre grupamentos seja igual a zero. Seria necessário o valor X_g/Y_g ser o mesmo para todos os grupamentos ($X_g/Y_g = X/Y$, $\forall g$).

A segunda implicação é sobre a soma dos pesos para valores de ε diferentes de zero ou um. Como $\varepsilon(1-\varepsilon)$ é maior que zero se $0 < \varepsilon < 1$ e considerando $B_S(\mathbf{X}) > 0$, então a soma dos componentes de decomposição será menor que um. Por outro lado, se $\varepsilon < 0$ ou $\varepsilon > 1$, a soma dos pesos excederá a unidade. Theil (1967, p. 125) havia notado esse fato especificamente para o quadrado do coeficiente variação, pois esse corresponde à medida geral de desigualdade para $\varepsilon = -1$ multiplicada por dois. A seguir, são apresentados os casos particulares para a medida geral de segregação conforme valore de ε.

3.3. Casos particulares

Analisaremos os casos em que $\varepsilon \notin (0, 1)$. Nota-se que, no limite quando $\varepsilon \to 0$ ou $\varepsilon \to 1$, o valor de $I_S(\mathbf{X})$ é indeterminado. Aplicando a regra de L'Hôpital e tomando o limite quando ε tende a zero temos

$$I_S(\mathbf{X}; \varepsilon = 0) = \lim_{\varepsilon \to 0} \frac{\sum_{j=1}^{k} y_j \left(\frac{x_j}{y_j}\right)^{1-\varepsilon} \ln\left(\frac{x_j}{y_j}\right)}{1 - 2\varepsilon} \quad \text{ou} \tag{3.5}$$

$$I_S(\mathbf{X}; \varepsilon = 0) = \sum_{j=1}^{k} x_j \ln\left(\frac{x_j}{y_j}\right). \tag{3.6}$$

Comparando a expressão acima com a eq. (2.51), denominaremos a eq. (3.6) de *T de Theil para segregação* ($T_S(\mathbf{X})$). Essa medida já havia sido proposta anteriormente em Hutchens (1991, p. 48), porém não havia sido derivada de uma medida mais geral.

É possível que um estrato seja ocupado somente por pessoas de uma categoria. Considerando que o j-ésimo estrato seja ocupado exclusivamente por pessoas da categoria Y e utilizando a regra de L'Hôpital temos

$$\lim_{x_j \to 0} x_j \ln\left(\frac{x_j}{y_j}\right) = 0.$$

[3] Ver prova da Propriedade 7 (enunciada na página 47), que se encontra no apêndice (p. 93).

Portanto, sob estas circunstâncias, a contribuição deste estrato para a medida de segregação total será zero. Por outro lado, no limite quando $y_i \to 0$ tem-se que $x_j \ln(x_j/y_j) \to \infty$ e então $T_S(\mathbf{X}) \to \infty$. Assim, quando pelo menos um estrato for ocupado exclusivamente por pessoas da categoria X, o índice T de Theil para segregação tenderá a infinito.

Considerando a situação de ausência de segregação, quando, para cada categoria, a respectiva proporção é a mesma em todos os estratos, temos que $x_j = y_j$ para todo j. Logo, de (3.6), temos que o valor mínimo de $T_S(\mathbf{X})$ é zero. Assim, considerando os resultados anteriores, concluímos que $T_S(\mathbf{X}) \in [0, \infty)$. É importante destacar que, para que este índice tenda a infinito, é condição suficiente que pelo menos um estrato seja ocupado exclusivamente pela categoria X. Para os índices de Gini para segregação e Dissimilaridade, era condição tanto necessária quanto suficiente que todos os estratos fossem ocupados, exclusivamente, por uma categoria para que os índices assumissem valor máximo. Para ilustrar a aplicação deste índice, para os valores da Tabela 2.1, temos que $T_S(\mathbf{X}) = 0,6263$.

Na decomposição da medida, temos que a medida de segregação dentro de um grupamento é ponderada pela sua participação na categoria X, pois, de acordo com (3.4), $w_S^g(\mathbf{X}^g; \varepsilon = 0) = x_g$. Da mesma maneira que para a medida T de Theil de desigualdade, a soma desses pesos é igual a um. Esse resultado já havia sido derivado na seção anterior.

Lembrando a eq. (3.5) e considerando, agora, o limite quando ε tende a 1,

$$I_S(\mathbf{X}; \varepsilon = 1) = \sum_{j=1}^{k} y_j \ln\left(\frac{y_j}{x_j}\right). \tag{3.7}$$

Comparando a eq. (2.50) com a eq. (3.7), denominaremos esta última expressão de índice *L de Theil para segregação* ($L_S(\mathbf{X})$). Não é difícil mostrar que, no limite quando $y_j \to 0$ (um estrato "sem" pessoas do grupo Y), a contribuição deste estrato para o índice será zero. Por outro lado, quando $x_j \to 0$, então $L_S(\mathbf{X}) \to \infty$. Portanto, se existir um estrato sem pessoas da categoria X o índice crescerá indefinidamente. As discussões sobre condições necessárias e suficientes para máximo e mínimo são iguais ao índice $T_S(\mathbf{X})$. Assim, concluímos que $L_S(\mathbf{X}) \in [0, \infty)$. Pode-se verificar que $L_S(\mathbf{X}) = 0,5414$ para os dados apresentados na Tabela 2.1.

No sentido contrário ao $T_S(\mathbf{X})$, a parcela dentro dos grupamentos para o índice L de Theil para segregação é ponderada pela participação relativa da categoria Y, pois, segundo a eq. (3.4), $w_S^g(\mathbf{X}^g; \varepsilon = 1) = y_g$. Desse modo, para os dois casos onde os pesos somam um, as parcelas dentro dos grupamentos são ponderadas pela sua participação na categoria X ou sua participação na categoria Y.

Assim, tanto no caso da segregação quanto no caso da desigualdade, os coeficientes de decomposição das medidas T e L de Theil somam um. No caso da desigualdade, a medida L de Theil é dita "democrática", pois a ponderação é feita pela participação dos grupos na população, enquanto que a medida análoga para segregação pondera a parcela dentro dos grupamentos pelas suas participações na categoria Y. Por outro lado, a medida T de Theil para segregação utiliza como ponderação a participação de cada grupamento na categoria X, enquanto que a medida análoga para desigualdade utiliza a participação do grupo na renda total, caracterizando-a como uma medida "não-democrática".

Desse modo, não é aplicável o conceito de "medida democrática" para o índice L de Theil para segregação e tampouco para o T, pois ambos consideram como casos extremos de ponderação a participação de cada grupamento em uma das duas categorias. Assim, poder-se-ia argumentar que uma medida "democrática" de segregação seria o índice Raiz Quadrada, uma vez que os coeficientes de decomposição são da forma $w_S^g(\mathbf{X}^g; \varepsilon = 0,5) = \sqrt{y_g x_g}$, pois consideram com iguais pesos as participações do grupamento nas duas categorias. Entretanto, ressalta-se, conforme demonstrado na seção anterior, que para ε diferente de zero ou um, a soma dos coeficiente de decomposição não são necessariamente iguais a um e poderão estar relacionados com a parcela entre os grupamentos.

A variância das razões X_j/Y_j, ponderada por y_j, é

$$\sigma_S^2 = \sum_{j=1}^{k} y_j \left(\frac{X_j}{Y_j} \right)^2 - \mu_S^2.$$

Utilizando $\mu_S = X/Y$ (eq. (2.33)), o *coeficiente de variação para segregação*, definido por $CV_S(\mathbf{X}) = \sigma_S/\mu_S$, será

$$CV_S(\mathbf{X}) = \sqrt{\sum_{j=1}^{k} y_j \left(\frac{x_j}{y_j} \right)^2 - 1}. \tag{3.8}$$

Como $I_S(\mathbf{X}; \varepsilon = -1) = [\sum_{j=1}^{k} y_j (x_j/y_j)^2 - 1]/2$, temos que o coeficiente de variação para segregação é uma transformação monotônica crescente da medida geral de segregação dada por $CV_S(\mathbf{X}) = \sqrt{2 I_S(\mathbf{X}; \varepsilon = -1)}$. Para os dados da Tabela 2.1, o coeficiente de variação para segregação assume valor 1,4619.

Usando a eq. (3.8), chega-se facilmente à expressão

$$CV_S^2(\mathbf{X}) = \sum_{j=1}^{k} y_j \left(\frac{x_j}{y_j} - 1 \right)^2,$$

que é exatamente o coeficiente de variação para segregação definido em Hutchens (1991, p. 48). Nota-se, entretanto, que o coeficiente de variação de Hutchens é o quadrado do coeficiente deduzido neste trabalho. De fato, Hutchens define o coeficiente de variação da renda como sua variância dividida pela renda média ao quadrado. Como consequência, a medida similar para segregação corresponderia ao quadrado do medida correta. Caso ele tivesse definido, corretamente, o coeficiente de variação como o desvio padrão dividido pela média, o coeficiente de Hutchens e o deduzido neste trabalho seriam iguais.

Outra medida que pode ser deduzida a partir da medida geral de segregação é a família de Atkinson para segregação. Por analogia com a eq. (2.56), a partir da medida geral de segregação (eq. (3.1)), temos

$$A_S(\mathbf{X}) = 1 - [1 - \varepsilon(1-\varepsilon)I_S(\mathbf{X})]^{\frac{1}{1-\varepsilon}} \qquad \text{ou} \tag{3.9}$$

$$A_S(\mathbf{X}) = 1 - \left[\sum_{j=1}^{k} y_j \left(\frac{x_j}{y_j}\right)^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}. \tag{3.10}$$

Substituindo as definições (2.3), (2.4) e (2.5), e desenvolvendo os termos na equação anterior, temos

$$A_S(\mathbf{X}) = 1 - \frac{1-P}{P}\left[\frac{\sum_{j=1}^{k} P_j^{1-\varepsilon}(1-P_j)^{\varepsilon} T_j}{(1-P)T}\right]^{\frac{1}{1-\varepsilon}}. \tag{3.11}$$

Comparando a eq. (2.57) com a eq. (3.11), temos que a medida de Atkinson para segregação encontrada na literatura é similar à transformação acima da medida geral de segregação. Só teríamos $A_S(\mathbf{X}) = A'_S(\mathbf{X})$ quando $\varepsilon = 0,5$ e $P = 0,5$. Por outro lado, se as definições (2.4) e (2.5) fossem alteradas para

$$P = \frac{Y}{T},\ (1-P) = \frac{X}{T} \qquad \text{e} \tag{2.4$'$}$$

$$P_j = \frac{Y_j}{T_j},\ (1-P_j) = \frac{X_j}{T_j}, \tag{2.5$'$}$$

a família de Atkinson para segregação como definida por James e Taeuber (1985) será igual à dada pela eq. (3.11).

Pela eq. (3.10), é fácil verificar que se todos os estratos são ocupados exclusivamente por pessoas de uma categoria, e $\varepsilon \in (0,1)$, então o valor máximo da medida será um. Por outro lado, seu valor mínimo é zero (com $x_j = y_j$, $\forall j$). James e Taeuber (1985, p. 9) apresenta uma pequena discussão para utilização da medida de Atkinson para segregação com $\varepsilon \in (0,1)$.

Considerando $\varepsilon > 0$ e $\varepsilon \neq 1$, de (3.9), verificamos que

$$\frac{\partial A_S(\mathbf{X})}{\partial I_S(\mathbf{X})} = \varepsilon\left[\sum_{j=1}^{k} y_j \left(\frac{x_j}{y_j}\right)^{1-\varepsilon}\right]^{\frac{\varepsilon}{1-\varepsilon}} > 0. \tag{3.12}$$

Portanto, a família de medidas de Atkinson para segregação é uma transformação monotonicamente crescente de $I_S(\mathbf{X})$.

Com a última medida apresentada, a família de Atkinson para segregação, foi demonstrado que é possível encontrar medidas de segregação análogas às medidas de desigualdade geral, Gini, Dissimilaridade, T e L de Theil, Atkinson e coeficiente de variação. A tabela a seguir sintetiza essas medidas nos dois contextos, assim como seus respectivos intervalos de variação (ou imagens).

Tabela 3.1. Tabela de comparação das medidas de desigualdade e segregação

Medida	Desigualdade	Segregação
Gini (com ordenação)	$G(\mathbf{z}) = 1 - \sum_{i=1}^{n}(\Phi_i + \Phi_{i-1})/n$ $[0, 1)$	$G_S(\mathbf{X}) = 1 - \sum_{j=1}^{k}(\Psi_j + \Psi_{j-1})y_j$ $[0, 1]$
Gini (sem ordenação)	$G(\mathbf{z}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\lvert z_i - z_j\rvert/(2n\mu)$ $[0, 1)$	$G_S(\mathbf{X}) = \frac{\sum_{i=1}^{k}\sum_{j=1}^{k} T_i T_j \lvert P_i - P_j\rvert}{2T^2 P(1-P)}$ $[0, 1]$
Dissimilaridade	$D_{max} = \sum_{i=1}^{h}(\mu - z_i)/(n\mu)$ $[0, 1)$	$D(\mathbf{X}) = \sum_{j=1}^{k}\Big\lvert x_j - y_j\Big\rvert/2$ $[0, 1]$
Geral	$I(\mathbf{z}) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \frac{1}{n}\sum_{i=1}^{n}\left(z_i/\mu\right)^{1-\varepsilon}\right]$ $[0, (1 - n^{-\varepsilon})/[\varepsilon(1-\varepsilon)]]$ se $\varepsilon < 1$ $[0, \infty)$ se $\varepsilon \geq 1$	$I_S(\mathbf{X}) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{j=1}^{k} y_j\left(x_j/y_j\right)^{1-\varepsilon}\right]$ $[0, 1/[\varepsilon(1-\varepsilon)]]$ se $\varepsilon \in (0, 1)$ $[0, \infty)$ se $\varepsilon \in (-\infty, 0] \cup [1, \infty)$
T de Theil ($\varepsilon = 0$)	$T(\mathbf{z}) = \sum_{i=1}^{n}(z_i/n\mu)\ln(z_i/\mu)$ $[0, \ln(n)]$	$T_S(\mathbf{X}) = \sum_{j=1}^{k} x_j \ln(x_j/y_j)$ $[0, \infty)$
Raiz Quadrada ($\varepsilon = 0,5$)	$I(\mathbf{z}; \varepsilon = 0,5) = 4\left[1 - \frac{1}{n}\sum_{i=1}^{n}\sqrt{z_i/\mu}\right]$ $[0, 4(1 - 1/\sqrt{n})]$	$I_S(\mathbf{X}; \varepsilon = 0,5) = 4\left(1 - \sum_{j=1}^{k}\sqrt{x_j y_j}\right)$ $[0, 4]$
L de Theil ($\varepsilon = 1$)	$L(\mathbf{z}) = -\frac{1}{n}\sum_{i=1}^{n}\ln(z_i/\mu)$ $[0, \infty)$	$L_S(\mathbf{X}) = \sum_{j=1}^{k} y_j \ln(y_j/x_j)$ $[0, \infty)$
Atkinson ($\varepsilon > 0$)	$A(\mathbf{z}) = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}(z_i/\mu)^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}$ $[0, 1)$ se $0 < \varepsilon < 1$ $[0, 1]$ se $\varepsilon \geq 1$	$A_S(\mathbf{X}) = 1 - \frac{1-P}{P}\left[\frac{\sum_{j=1}^{k} P_j^{1-\varepsilon}(1-P_j)^{\varepsilon} T_j}{(1-P)T}\right]^{\frac{1}{1-\varepsilon}}$ $[0, 1]$ se $\varepsilon > 0$, $\varepsilon \neq 1$
Coeficiente de variação ($\varepsilon = -1$)	$CV(\mathbf{z}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(z_i/\mu)^2 - 1}$ $[0, \sqrt{n^2 - 1}]$	$CV_S(\mathbf{X}) = \sqrt{\sum_{j=1}^{k} y_j (x_j/y_j)^2 - 1}$ $[0, \infty)$

Fonte: elaboração própria.
Nota: os intervalos abaixo das medidas são suas respectivas imagens.

3.4. Um exemplo para decomposição

A Tabela 3.2 apresenta um exemplo numérico para ilustrar a decomposição da medida geral de segregação. As duas categorias de análise (X e Y) são distribuídas em dois grupamentos, com dois estratos cada. O grupamento 1 é constituído pelo primeiro e quarto estrato do exemplo da Tabela 2.1. Por sua vez, o segundo grupamento é constituído pelos demais estratos (dois e três). A Tabela 3.2 apresenta os valores da medida geral, das parcelas entre e dentro dos grupamentos, e os pesos referentes a cada grupamento para o cálculo da parcela dentro dos grupamentos. São utilizados os índices T e L de Theil e o índice Raiz Quadrada.

Tabela 3.2. Decomposição das medidas de segregação para o exemplo da Tabela 2.1

	$T_S(\mathbf{X})$		Raiz Quadrada		$L_S(\mathbf{X})$	
Grupamento	peso	medida	peso	medida	peso	medida
1	0,5455	1,0730	0,5017	1,0186	0,4615	1,0730
2	0,4545	0,0592	0,4947	0,0593	0,5385	0,0596
Dentro dos grup.	0,6122		0,5404		0,5273	
Entre os grup.	-	0,0141	-	0,0141	-	0,0141
Total	-	0,6263	-	0,5545	-	0,5414

Fonte: elaboração própria.

Os resultados indicam que 97,75%, 97,46% e 97,39% da segregação entre todos os estratos é explicada pela segregação dentro dos grupamentos para os índices T de Theil, Raiz Quadrada e L de Theil, respectivamente.

Em todos os casos era esperado que a medida de segregação dentro do primeiro grupamento fosse maior do que a do segundo, uma vez que o primeiro foi criado a partir dos dois estratos com maiores diferenças entre suas participações relativas. Isso se evidencia no fato de a medida de segregação dentro do primeiro grupamento ser aproximadamente 17 vezes o valor da referente ao segundo, utilizando o índice Raiz Quadrada.

Pode-se verificar que multiplicando a quantidade de pessoas em cada estrato por uma constante não se altera a segregação dentro de cada grupamento, os pesos e a medida de segregação entre grupamentos. Em consequência, o valor da segregação entre estratos também não mudará. Esta é uma propriedade que alguns autores acreditam que deve ser necessariamente obedecida por uma medida de segregação. A próxima seção buscará aprofundar a discussão sobre as propriedades desejáveis de uma medida de segregação.

3.5. Propriedades desejáveis das medidas de segregação

Diversos autores discutem as propriedades desejáveis das medida de segregação. São exemplos os trabalhos de James e Taeuber (1985), Hutchens (1991, 2001, 2004, 2015), Watts (1998a, 1998b), Blackburn, Siltanen e Jarman (1995), Charles e Grusky (1995), Chakravarty e Silber (1994, 2007) e Alonso-Villar e Río (2007). Nesta seção serão apresentadas algumas dessas propriedades e discutida suas relevâncias. Quando necessário, será utilizada a literatura sobre desigualdade para exemplificar as semelhanças entre as medidas nos dois contextos. Os trabalhos de referência sobre as propriedades das medidas de desigualdade são Shorrocks (1980, 1984), Cowell e Kuga (1981a, 1981b) e Foster (1983, 1985). Esta seção é desenvolvida com base nas apresentações de Hutchens (2004) e Chakravarty e Silber (2007).

- **Propriedade 1 (Invariância de escala)** *O valor de uma medida de segregação não deve sofrer alteração se a quantidade de pessoas de uma mesma categoria aumentar ou diminuir na mesma proporção em todos os estratos.*

Essa propriedade indica que, caso exista uma matriz diagonal de ordem 2 com os elementos da diagonal principal positivos, ou seja,

$$\mathbf{\Gamma} = \begin{bmatrix} \alpha^* & 0 \\ 0 & \beta^* \end{bmatrix}, \; \alpha^*, \beta^* > 0,$$

temos que $\Omega(\mathbf{X}) = \Omega(\,\mathbf{\Gamma X})$. Essa propriedade também pode ser chamada de homogeneidade (HUTCHENS, 1991) ou consistência de escala (CHAKRAVARTY; SILBER, 2007).

É mais intuitiva a situação em que $\alpha^* = \beta^*$. Esse caso particular é conhecido como *invariância de tamanho.* Nessa situação, se $\alpha^* = 2$, por exemplo, temos que a quantidade de pessoas em cada categoria, em cada estrato, dobra. Como consequência, tem-se que a quantidade de pessoas de cada estrato dobra e a população também. Temos, portanto, que $\Omega(\mathbf{X}) = \Omega(\alpha^*\mathbf{X})$, pois as participações de cada estrato nas duas categorias não são alteradas.

Essa propriedade, para medidas de desigualdade, é equivalente a dizer que a medida é homogênea de grau zero nas rendas e que uma n-plicação da população não altera a medida, isto é, uma duplicação, triplicação, etc., da população também não afeta as medidas de desigualdade. Caso todas as pessoas tenham sua renda multiplicada por uma constante, a desigualdade não diminuirá e tampouco aumentará. Essas são propriedades atendidas pelas medidas de dispersão relativa. Muitas vezes, como ocorre por exemplo com o índice de Gini, medidas de dispersão relativa e de desigualdade se confundem.

- **Propriedade 2 (Simetria nos estratos)** *O valor da medida de segregação não deve ser alterado com permutações na ordenação dos estratos.*

Seja $\mathbf{P}_k$ uma matriz de permutação de ordem k. Deste modo, essa propriedade implica $\Omega(\mathbf{X}) = \Omega(\mathbf{X}\mathbf{P}_k)$. A consequência desta propriedade é que a ordenação dos estratos é cardinal. Assim, por exemplo, não há distinção entre setores em uma economia, quando se trata da análise da segregação de gênero no mercado de trabalho. Esta propriedade foi proposta por Hutchens (1991). Para Chakravarty e Silber (2007), como a análise é sobre medidas de segregação ocupacional, essa propriedade chama-se simetria nas ocupações.

Para medidas de desigualdade, essa propriedade é equivalente a dizer que apenas o valor das rendas é relevante, pois a correspondência de cada pessoa com cada renda não altera o valor da medida. Considerando que a medida é insensível a permutações entre indivíduos, essa propriedade é conhecida como *simetria* nas rendas. Desse modo, o índice de Gini, tanto para desigualdade quanto para segregação, atende a essa propriedade. De modo similar, é fácil verificar que o índice de Dissimilaridade também não é afetado por permutações nos estratos.

- **Propriedade 3 (Mudança entre estratos)** *Uma transferência de pessoas entre estratos, que resulte em maior diferença das participações relativas, provocará aumento da medida de segregação.*

Essa propriedade é equivalente à condição de Pigou-Dalton para desigualdade de renda. Considere os estratos i e h de modo que

$$\frac{x_i}{y_i} \leq \frac{x_h}{y_h}. \tag{3.13}$$

Assim, se uma matriz de distribuição $\mathbf{X}^*$ puder ser obtida pela transferência de membros da categoria X, do i-ésimo para o h-ésimo estrato, de modo que

$$\frac{x_i^*}{y_i^*} = \frac{x_i - \theta}{y_i} < \frac{x_h + \theta}{y_h} = \frac{x_h^*}{y_h^*}, \tag{3.14}$$

com $0 < \theta \leq x_i$ e esta propriedade for válida, temos que $\Omega(\mathbf{X}) < \Omega(\mathbf{X}^*)$. Esse tipo de transferência recebe o nome de *mudança regressiva entre estratos*. Ressalta-se que não são alteradas as participações de cada estrato na categoria Y, nem as participações dos demais estratos na categoria X, isto é, tem-se: a) $y_i^* = y_i$ e $y_h^* = y_h$, e; b) $x_j^* = x_j$ e $y_j^* = y_j$, $\forall j \neq i, h$. Os mesmos conceitos e resultados se aplicam a mudanças regressivas na categoria Y, considerando as devidas adaptações.

Essa propriedade, conforme apresentada neste trabalho, difere ligeiramente da encontrada na literatura, pois consideramos a diferença nas participações relativas entre dois estratos. Desse modo, não é necessária a hipótese $y_i = y_h$ apresentada em Hutchens (2004, p. 559) e Chakravarty e Silber (2007, p. 188). Por sua vez, a exposição de Alonso-Villar e Río (2007, p. 8) difere substancialmente das demais apresentações, uma vez que a participação relativa do autor é definida como x_j/t_j.

A condição de Pigou-Dalton foi enunciada na p. 21, a partir da definição de transferências regressivas de renda. Entretanto, pode-se definir a condição de Pigou-Dalton a partir de transferências progressivas de renda. Nesse caso, se um indivíduo mais rico transferir uma quantia de dinheiro para um mais pobre, reduzindo a diferença entre suas rendas, necessariamente a medida de desigualdade diminuirá. De modo semelhante, podemos definir uma *mudança progressiva* entre estratos. Assim, a mudança de pessoas entre estratos, de modo que a diferença entre as participações relativas dos estratos envolvidos na mudança diminua, necessariamente implicará diminuição da medida de segregação.

- **Propriedade 4 (Insensibilidade a divisões proporcionais nos estratos)** *Uma medida de segregação não altera seu valor se diversos estratos forem criados a partir da divisão proporcional, em ambas as categorias, de um único estrato original.*

Considere que um determinado estrato, digamos o último (k), seja subdividido de forma igual em m estratos. Desse modo, sendo $\mathbf{X}^*$ a nova matriz de distribuição, teremos que as participações relativas de cada categoria nos estratos, com exceção do k-ésimo, permanecem inalteradas ($x_j^* = x_j$ e $y_j^* = y_j$, $\forall j \neq k$). Porém, as participações das categorias nos novos estratos são definidas como $x_h^* = x_k/m$, $y_h^* = y_k/m$, $\forall h = 1, ..., m$. É claro que a ordem da nova matriz de distribuição $\mathbf{X}^*$ é $2 \times (k + m - 1)$.

A definição dos estratos pode ser feita de forma arbitrária. Imagine um bairro, que por determinação da prefeitura, seja subdividido em dois ou mais. Nesse caso, se as proporções de pessoas em cada categoria (brancos e negros) se mantiverem nos novos bairros, a medida de segregação calculada a partir dos novos estratos deverá ser igual à anterior. Segundo Hutchens (2004, p. 559), "quando vinho novo é despejado em garrafas velhas, continua sendo vinho novo".

Essa propriedade não é aplicável para desigualdade de renda. É difícil imaginar como "criar" indivíduos a partir de um único.

- **Propriedade 5 (Decomposição aditiva)** *Uma medida de segregação deve ser aditivamente decomponível.*

Essa propriedade para medidas de desigualdade é utilizada desde a década de 60 com os índices de Theil. As ideias sobre medidas de segregação aditivamente decomponíveis foram introduzidas por Theil e Finizza (1971) e Theil (1972), mas foi só na década dos anos 2000, com os trabalhos de Mora e Ruiz-Castillo (2003) e Hutchens (2004), que a discussão dessa propriedade foi retomada.

A classe de medidas aditivamente decomponíveis é mais restrita que a classe de decomponíveis. O índice de Gini, por exemplo, é uma medida que pode ser decomposta, porém, não é possível decompô-la de forma aditiva. Nesse caso, além das parcelas dentro e entre os grupos, aparece uma terceira parcela referente à sobreposição delas.

- **Propriedade 6 (Simetria nas categorias)** *O valor de uma medida de segregação deve ser o mesmo considerando qualquer uma das categorias como a de interesse.*

Essa propriedade foi proposta por Chakravarty e Silber (1994). Considerando uma matriz de permutação de ordem 2 ($\mathbf{P}_2$), essa propriedade implica $\Omega(\mathbf{X}) = \Omega(\mathbf{P}_2\mathbf{X})$. Assim, o valor da medida será igual se considerarmos, por exemplo, as mulheres como a categoria X ou Y. Essa propriedade não é aplicável para a distribuição de renda.

- **Propriedade 7 (Imagem)** *Uma medida de segregação deve assumir valores não negativos, com valor mínimo (zero) se a proporção de pessoas de cada categoria em cada estrato for a mesma.*

A propriedade equivalente para medidas de desigualdade é que ela seja não negativa e que seu valor mínimo ocorra quando todas as rendas são iguais. É curioso notar que essa propriedade é enunciada em diversos trabalhos sobre medidas de desigualdade (por exemplo Shorrocks (1980, 1984)), porém não é frequentemente usada para o caso da segregação. O trabalho de Hutchens (2004, p. 565), por exemplo, prefere utilizar uma propriedade também relacionada à imagem da medida, porém diferente da enunciada aqui.

Segundo o autor, a sétima propriedade de seu trabalho propõe que uma medida de segregação deve assumir valores de zero a um, com um representando o caso de completa segregação, enquanto que o valor zero é a situação inversa. Imagine uma medida que assuma valores não negativos e máximo finito. Se dividirmos essa medida por seu valor máximo, necessariamente ela assumirá valores no intervalo de zero a um. Conforme os resultados derivados no início deste capítulo, se $\varepsilon \in (0,1)$, então o valor máximo da medida geral de segregação será $1/[\varepsilon(1-\varepsilon)]$. Se dividirmos a medida geral por esse valor, obtemos a seguinte expressão $\varepsilon(1-\varepsilon)I_S(\mathbf{X}) = O_c(\mathbf{X})$, com $\varepsilon \in (0,1)$. Essa é a medida geral de segregação de Hutchens (eq. (3.2)). Talvez isso explique o fato de o autor preferir usar essa transformação ao longo de todo seu artigo, pois, para qualquer valor de $\varepsilon \in (0,1)$, a medida sempre assumirá valores de zero a um. Em consequência, toda a classe de medidas em (3.2) atenderá a essa propriedade proposta pelo autor.

Embora o argumento do autor seja coerente, não é plausível utilizá-lo como propriedade que deve ser atendida por uma medida de segregação. Limitar a imagem de uma medida a $[0,1]$ só é útil quando existe uma interpretação clara a respeito de seu valor. O índice de Dissimilaridade, por exemplo, indica a proporção de pessoas da categoria de interesse que deve mudar de estratos para que a segregação passe a ser zero (veja demonstração desse resultado na página 58). Nesse contexto, é plausível limitar a imagem ao intervalo fechado $[0,1]$. No caso de desigualdade, por exemplo, nenhum dos índice T e L de Theil estão limitados à unidade.

Desse modo, seguindo os trabalhos de Shorrocks (1980, 1984), preferimos adotar a propriedade de que as medidas de segregação devem assumir valores não-negativos. Entretanto, é importante destacar que o enunciado dessa propriedade é, de certo modo, "redundante". Caso a condição de normalização ($x_j = y_j$, $\forall j \Rightarrow \Omega(\mathbf{X}) = 0$) e a propriedade de movimentos entre estratos forem válidas, a consequência é que a medida assumirá, somente, valores não-negativos, isto é, teremos $\Omega(\mathbf{X}) \geq 0$.

Conforme demonstrado no Apêndice A.1, a medida geral de segregação atende a todas essas propriedades, com exceção da Propriedade 6 (Simetria nas categorias). Entretanto, para o índice Raiz Quadrada, que corresponde à medida geral com $\varepsilon = 0,5$, todas as sete propriedades listadas anteriormente são atendidas.

O próximo capítulo apresenta a análise de sensibilidade das medidas de segregação a movimentos regressivos entre estratos.

4. SENSIBILIDADE DAS MEDIDAS DE SEGREGAÇÃO

A literatura trata da sensibilidade das diversas medidas de desigualdade a transferências regressivas de renda (KAKWANI, 1980; SHORROCKS; FOSTER, 1987; HOFFMANN, 1992; COWELL; FLACHAIRE, 2007). Entretanto, com exceção de uma breve menção à sensibilidade conforme valores de ε para a família de medidas de Atkinson em James e Taeuber (1985, p. 9), não foram encontradas referências sobre sensibilidade a mudanças regressivas nas medidas de segregação. Essa é a principal contribuição teórica deste capítulo.

Para ilustrar o conceito de sensibilidade no contexto da segregação, vamos considerar o exemplo numérico apresentado por Hoffmann (1998, p. 254), porém, com algumas mudanças entre estratos na categoria X. A Tabela 4.1 apresenta os dados hipotéticos para quatro situações distintas.

Tabela 4.1. Distribuição das categorias X e Y em situações hipotéticas

Estrato	Y	$X(A)$	$X(B)$	$X(C)$	$X(D)$
1	200	12	6	12	12
2	100	12	18	6	12
3	100	36	36	42	30
4	100	180	180	180	186

Fonte: adaptado de Hoffmann (1998, p. 254).

Nas quatro situações que consideraremos, de A a D, a distribuição da categoria Y nos estratos permanecerá inalterada. A situação A é a situação inicial. Na segunda, B, observe que foi feita uma mudança regressiva de seis pessoas do primeiro estrato para o segundo. No caso C, comparado ao primeiro, foi realizada uma mudança regressiva de seis pessoas do segundo para o terceiro estrato. Por fim, na situação D, por mudança regressiva, seis pessoas do terceiro estrato do caso A passaram a ocupar o quarto estrato. A Figura 4.1 mostra como a medida geral de segregação se comporta para esses quatro casos, conforme valores de ε de menos um a um.

Como as situações B, C e D foram obtidas de A por mudanças regressivas, como era esperado, a curva para a medida geral para a situação A, independentemente dos valores de ε, está sempre abaixo das demais curvas. Isso ocorre porque a medida geral de segregação obedece à propriedade de mudanças entre os estratos.

Para ε baixo (negativo), o aumento na medida de segregação cresce com a participação relativa (x_j/y_j) dos estratos nos quais ocorreu a mudança. Para valores elevados de ε ocorre o contrário: o acréscimo em $I_S(\mathbf{X})$ é tanto menor quanto maior for a participação relativa dos estratos afetados pela mudança.

Um dos objetivos desse capítulo é entender como mudanças regressivas (ou progressivas) entre estratos afetam a medida geral de segregação, dado um valor do parâmetro ε e o nível inicial da participação relativa x_i/y_i do estrato do qual mudam as pessoas. Também serão analisadas as sensibilidades do índice de Gini para segregação e do índice de Dissimilaridade.

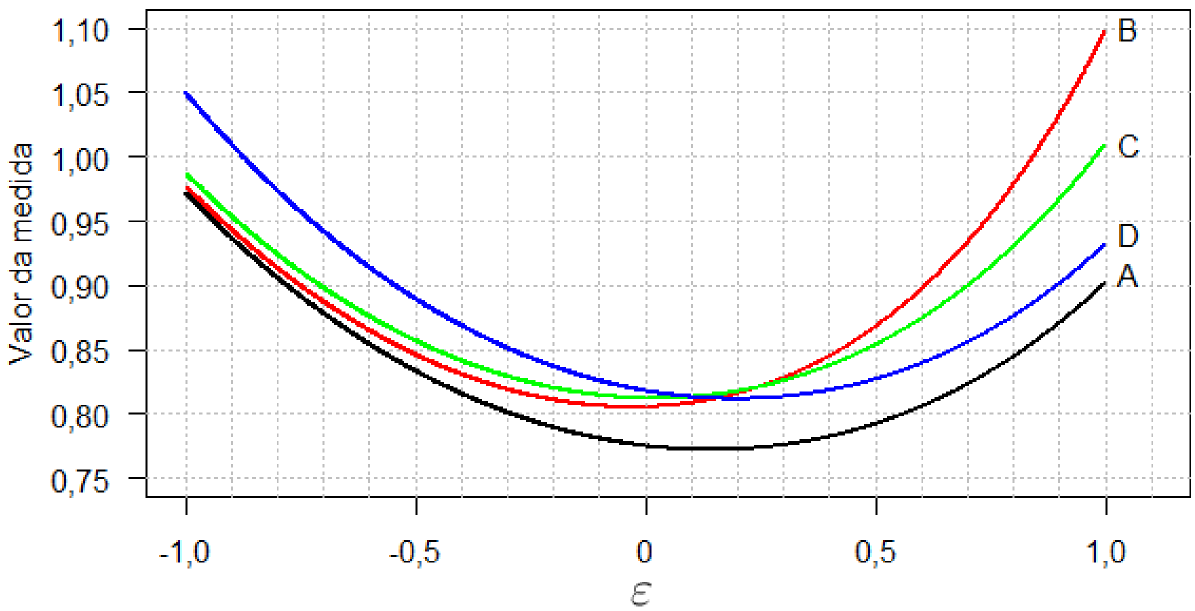

Figura 4.1. Medida geral de segregação para quatro situações hipotéticas, conforme valor de ε

Conforme visto no capítulo anterior, uma mudança regressiva na categoria X é definida pela transferência de θX pessoas do i-ésimo para o h-ésimo estrato, com $x_i/y_i < x_h/y_h$. O *efeito de uma mudança regressiva entre estratos* é dado pela razão entre a variação da medida e o número de pessoas que mudou de estrato em relação ao total da categoria. Matematicamente, temos

$$\frac{\Omega(\mathbf{X}^*) - \Omega(\mathbf{X})}{\theta} = \frac{\Delta\Omega(\mathbf{X}^*)}{\theta}.$$

Muitas vezes as expressões que medem esse efeito são complicadas. O uso do cálculo diferencial, na grande maioria dos casos, facilita a álgebra e o entendimento. Entretanto, é importante ressaltar que as conclusões sobre os efeitos independem do uso do cálculo infinitesimal da mudança entre estratos. Assim, definiremos o efeito, na medida de segregação, de uma mudança infinitesimal em alguma das categorias por

$$\lim_{\theta\to 0} \frac{\Delta\Omega(\mathbf{X}^*)}{\theta} = \lim_{\theta\to 0} \frac{\partial\Omega(\mathbf{X}^*)}{\partial\theta}.$$

Tratando-se de pessoas, o valor mínimo de θ é $1/X$ ou $1/Y$. Desse modo, usar a derivada em relação a θ é uma aproximação. Entretanto, poder-se-ia imaginar uma sequência de mudança (regressivas) infinitesimais positivas, cuja soma seja ao menos igual ao valor mínimo de θ. Assim, as conclusões para essas mudanças infinitesimais são generalizadas para o caso de uma mudança com valor mínimo θ igual a $1/X$ ou $1/Y$.

Considere o Brasil, por exemplo, cuja população economicamente ativa é da ordem de 100 milhões. Considerando que a proporção de pessoas de cada categoria (gênero) é a mesma na população, o valor mínimo para θ é 2×10^{-8}. Desse modo, os resultados para os casos infinitesimal ou discreto são aproximadamente iguais. Como as derivadas são contínuas e, conforme demonstrado no apêndice, não mudam de sinal no intervalo relevante de ε, a demonstração usando derivada é válida.

Nas seções posteriores são desenvolvidas e analisadas as sensibilidades da medida geral e seus casos particulares, do índice de Gini e do índice de Dissimilaridade.

4.1. Medida geral e seus casos particulares

Consideraremos inicialmente mudanças regressivas na categoria X. Uma mudança entre os estratos i e h, conforme a propriedade 3 (ver p. 46), resultará no efeito

$$\frac{\partial I_S(\mathbf{X}^*)}{\partial \theta} = \frac{1}{\varepsilon}\left[\left(\frac{x_i - \theta}{y_i}\right)^{-\varepsilon} - \left(\frac{x_h + \theta}{y_h}\right)^{-\varepsilon}\right], \varepsilon \neq 0 \text{ e } \varepsilon \neq 1.$$

O desenvolvimento pormenorizado das equações e seus resultados podem ser encontrados no apêndice. Levando em consideração uma mudança infinitesimal, temos

$$\lim_{\theta \to 0} \frac{\partial I_S(\mathbf{X}^*)}{\partial \theta} = \frac{1}{\varepsilon}\left[\left(\frac{x_i}{y_i}\right)^{-\varepsilon} - \left(\frac{x_h}{y_h}\right)^{-\varepsilon}\right]. \tag{4.1}$$

De modo semelhante, o efeito de uma mudança infinitesimal da categoria X no índice T de Theil para segregação será

$$\lim_{\theta \to 0} \frac{\partial T_S(\mathbf{X}^*)}{\partial \theta} = \ln\left(\frac{x_h/y_h}{x_i/y_i}\right). \tag{4.2}$$

Observe que a equação acima pode ser obtida, a partir da eq. (4.1), utilizando o limite quando ε tende a zero.

Para a medida $L_S(\mathbf{X})$, teremos

$$\lim_{\theta \to 0} \frac{\partial L_S(\mathbf{X}^*)}{\partial \theta} = \frac{y_i}{x_i} - \frac{y_h}{x_h}. \tag{4.3}$$

Esse resultado é igual à eq. (4.1), considerando $\varepsilon = 1$.

Vimos que o coeficiente de variação e a família de medidas de Atkinson são transformações monotônicas da medida geral de segregação. Seja $F(.)$ uma transformação monotonicamente crescente em seu argumento. Desse modo, considerando a medida geral, temos que, se

$$\Omega(\mathbf{X}) = F(I_S(\mathbf{X})), \text{ com } \frac{dF}{dI_S(\mathbf{X})} > 0,$$

então

$$\lim_{\theta \to 0} \frac{\partial \Omega(\mathbf{X}^*)}{\partial \theta} = \frac{dF}{dI_S(\mathbf{X})}\left[\lim_{\theta \to 0} \frac{\partial I_S(\mathbf{X}^*)}{\partial \theta}\right].$$

É interessante observar que a primeira derivada do lado direito depende somente do valor prévio da medida de segregação[1]. Entretanto, as derivadas parciais dependerão do novo valor da medida, obtido após a mudança regressiva.

Outra consideração deve ser feita a respeito da desigualdade estrita na expressão anterior. Desde que exista pelo menos um estrato tal que as proporções de cada categoria no total do estrato sejam diferentes, a medida geral de segregação necessariamente assumirá valores positivos. Assim, a desigualdade estrita na equação anterior sempre será válida. Essa é uma consequência das transformações e da Propriedade 7 (Imagem.)

Lembrando que o coeficiente de variação para segregação é igual a $\sqrt{2I_S(\mathbf{X}; \varepsilon = -1)}$, temos

$$\lim_{\theta \to 0} \frac{\partial CV_S(\mathbf{X}^*)}{\partial \theta} = \frac{1}{\sqrt{2I_S(\mathbf{X}; \varepsilon = -1)}} \left[\lim_{\theta \to 0} \frac{\partial I_S(\mathbf{X}^*; \varepsilon = -1)}{\partial \theta} \right].$$

Utilizando $\varepsilon = -1$ na eq. (4.1) e substituindo acima, temos

$$\lim_{\theta \to 0} \frac{\partial CV_S(\mathbf{X}^*)}{\partial \theta} = \frac{1}{\sqrt{2I_S(\mathbf{X}; \varepsilon = -1)}} \left[\frac{x_h}{y_h} - \frac{x_i}{y_i} \right].$$

De modo semelhante, para a família de medidas de Atkinson para segregação, pela eq. (3.12), temos

$$\lim_{\theta \to 0} \frac{\partial A_S(\mathbf{X}^*)}{\partial \theta} = \varepsilon[1 - \varepsilon(1-\varepsilon) I_S(\mathbf{X})]^{\frac{\varepsilon}{1-\varepsilon}} \left[\lim_{\theta \to 0} \frac{\partial I_S(\mathbf{X}^*)}{\partial \theta} \right].$$

Observe que o efeito de uma mudança regressiva na categoria X, para o coeficiente de variação e a família de medidas de Atkinson, são iguais a uma constante (fixados ε e $I_S(\mathbf{X})$) multiplicado pelo efeito na medida geral. Isso significa que a maneira como o efeito de uma mudança infinitesimal varia em função de x_i/y_i e x_h/y_h para as medidas de Atkinson ou o coeficiente de variação é a mesma que para a medida geral correspondente.

As equações (4.1), (4.2) e (4.3) mostram os efeitos de uma mudança regressiva na categoria X na medida geral e no T e L de Theil para segregação, respectivamente. Veremos para cada uma das medidas, fixando o valor de ε, como esse efeito varia com o nível das participações relativas dos estratos envolvidos.

Na literatura sobre distribuição de renda, os autores, predominantemente, analisam o efeito de uma transferência regressiva quando a distância entre as rendas dos dois indivíduos é fixada. Entretanto, ao fazer tal análise, não se leva em consideração que, em termos de bem estar, a importância de uma transferência de 100 reais entre dois indivíduos pobres é muito maior que quando o mesmo montante é transferido entre dois indivíduos milionários. Como alternativa, Hoffmann (1992) propõe analisar o efeito de

[1] Isso é uma consequência direta da mudança infinitesimal utilizada. Caso surja alguma dúvida sobre esse resultado, recomenda-se fazer uma mudança regressiva na medida em sua forma original (sem a transformação monotônica), derivar em relação a θ e aplicar o limite quando θ tender a zero. Esse resultado aparecerá diretamente.

uma transferência regressiva fixando, percentualmente, quanto um indivíduo é mais rico que outro. As situações em que a diferença entre as rendas é fixa e em que a razão entre as rendas é dada percentualmente, adaptadas para segregação, matematicamente, são

1. $x_h/y_h = x_i/y_i + \zeta$, com $\zeta \geq 0$;

2. $x_h/y_h = \gamma x_i/y_i$, com $\gamma \geq 1$.

Vamos supor que as participações dos estratos nos totais das categorias sejam estritamente positivas para evitar indefinições matemáticas e, claramente, para que sejam possíveis as mudanças regressivas. Embora a quantidade ζ seja a diferença entre as rendas no contexto de desigualdade, essa interpretação obviamente não é válida para ζ no caso da segregação. Entretanto, o valor de γ representa, percentualmente, o quanto o estrato h é predominante na categoria X, comparado ao estrato i. Por exemplo, caso $y_i = y_h$ e $\gamma = 2$, temos que a proporção de pessoas na categoria X é 100% maior no estrato h, comparado ao i. Desse modo, ao analisar a sensibilidade das medidas de segregação a mudanças regressivas, consideraremos a diferença percentual entre as participações relativas envolvidas. Assim, considerando $x_h/y_h = \gamma x_i/y_i$, com $\gamma > 1$, e a eq. (4.1), temos, para os valores indicados de ε, que o efeito de uma mudança regressiva na categoria X será igual a

$$\varepsilon = -2 \quad \Rightarrow \quad \frac{1}{2}\left(\frac{x_i}{y_i}\right)^2 \left[\gamma^2 - 1\right] \tag{4.4}$$

$$\varepsilon = -1 \quad \Rightarrow \quad \left(\frac{x_i}{y_i}\right) \left[\gamma - 1\right] \tag{4.5}$$

$$\varepsilon \to 0 \quad \Rightarrow \quad \ln(\gamma) \tag{4.6}$$

$$\varepsilon \to 1 \quad \Rightarrow \quad \frac{y_i}{x_i}\left[1 - \frac{1}{\gamma}\right] \tag{4.7}$$

$$\varepsilon = 2 \quad \Rightarrow \quad \frac{1}{2}\left(\frac{y_i}{x_i}\right)^2 \left[1 - \frac{1}{\gamma^2}\right]. \tag{4.8}$$

A eq. (4.6) indica que o efeito de uma mudança infinitesimal na categoria X, para a medida T de Theil para segregação, não depende da participação relativa inicial x_i/y_i. Por outro lado, as eqs. de (4.4) a (4.8), com exceção da eq. (4.6), mostram que esse efeito depende de x_i/y_i, com o sentido de variação dependendo do sinal de ε. Para $\varepsilon = -1$ (transformação do coeficiente de variação), o efeito da mudança regressiva em X é proporcional a x_i/y_i, enquanto que o efeito é inversamente proporcional a x_i/y_i para $\varepsilon \to 1$ (L de Theil para segregação).

Para valores negativos de ε, conforme eqs. (4.4) e (4.5), fixado esse valor, o efeito de uma mudança regressiva será tanto menor quanto menor for o valor da participação relativa x_i/y_i. Para valores da participação relativa próximos de zero, o efeito da mudança regressiva em X se aproxima de zero, considerando ε negativo. Por outro lado, para valores positivos de ε (eqs. (4.7) e (4.8)), o efeito da mudança regressiva varia inversamente com o valor da participação relativa. Nesse caso, quanto menor for o valor da participação relativa, maior será o efeito dessa mudança regressiva.

Desse modo, para mudanças regressivas na categoria X, concluímos que para ε igual a zero, a medida geral de segregação é insensível aos valores de x_i/y_i . Por outro lado, com ε negativo, o efeito dessa mudança aumenta quando a participação relativa cresce. Para ε positivo, o efeito da mudança regressiva varia em sentido contrário ao da participação relativa.

A Fig. 4.2 apresenta as *curvas de sensibilidade relativa*, para valores de ε das eqs. de (4.4) a (4.8), conforme a participação relativa do i-ésimo estrato. Para esses cálculos, foi considerada a proporção fixa $\gamma = 1,01$.

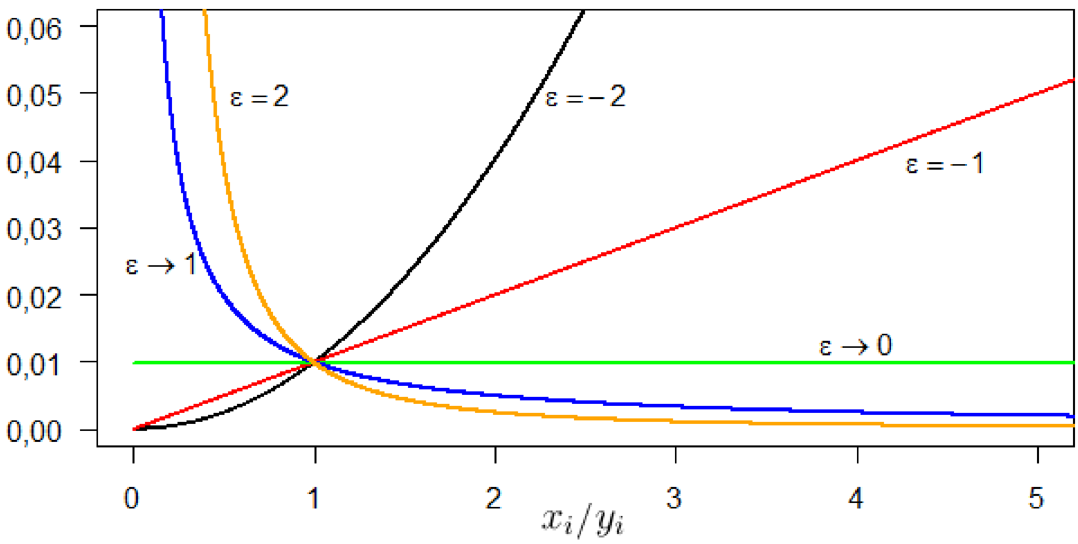

Figura 4.2. Curvas de sensibilidade relativa da medida geral para mudanças regressivas na categoria X, para vários valores de ε e com $\gamma = 1,01$

Agora vamos considerar mudanças regressivas na categoria Y. Para evitar confusão de notação, denominaremos de $\mathbf{Y}^*$ a matriz de distribuição resultante de uma mudança regressiva nessa categoria. Observe que essa mudança corresponde em transferir θY pessoas do h-ésimo para o i-ésimo estrato[2]. Desse modo, essa mudança regressiva infinitesimal, fixada a proporção entre as participações relativas (γ), terá o seguinte efeito sobre a medida geral

$$\lim_{\theta \to 0} \frac{\partial I_S(\mathbf{Y}^*)}{\partial \theta} = \frac{1}{1-\varepsilon} \left(\frac{x_i}{y_i} \right)^{1-\varepsilon} \left[\gamma^{1-\varepsilon} - 1 \right].$$

Podemos, de modo semelhante às eqs. de (4.4) a (4.8), calcular o efeito de uma mudança regressiva infinitesimal na categoria Y, para diversos valores de ε, conforme o

[2] Para maiores detalhes, veja a página 90, no apêndice.

nível de x_i/y_i. As curvas de sensibilidade relativa resultantes dessas equações estão no Fig. 4.3, para valores inteiros de ε entre -2 e 2, inclusive os extremos. Utiliza-se $\gamma = 1,01$.

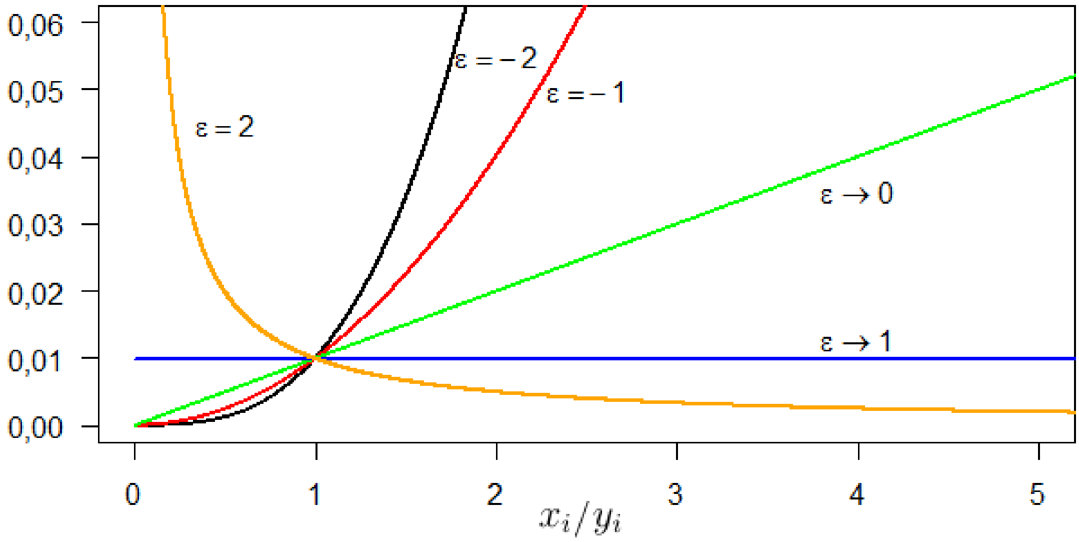

Figura 4.3. Curvas de sensibilidade relativa da medida geral para mudanças regressivas na categoria Y, para vários valores de ε e com $\gamma = 1,01$

Para $\varepsilon \to 1$, ou seja, para o L de Theil para segregação, o efeito de uma mudança regressiva infinitesimal na categoria Y não depende do nível inicial de x_i/y_i. Para valores de ε maiores que um, quanto menor for a participação relativa x_i/y_i, maior será o efeito dessa mudança sobre a medida de segregação.

Por outro lado, valores de ε igual a zero ou negativos mostram que a medida de segregação é mais sensível para valores elevados da participação relativa x_i/y_i. Nota-se que o efeito de uma mudança regressiva infinitesimal em X, considerando ε, é igual ao efeito de uma mudança regressiva infinitesimal em Y, considerando $\varepsilon + 1$, fixando, nos dois casos, os valores x_i/y_i e γ.

As Figuras 4.2 e 4.3 dão a impressão que o efeito de uma mudança regressiva será sempre igual a $\ln(\gamma)$ quando $x_i = y_i$. Esse resultado é aproximado conforme valores de γ. Para $\gamma > 1$, porém próximo da unidade, utilizando a aproximação de Taylor, temos $\ln(\gamma^n) \approx \gamma^n - 1$, com $n = ... - 1, -2, 1, 2, ...$. Utilizando esses resultados nas equações de (4.4) a (4.8), considerando $x_i = y_i$, teremos que esses efeitos serão aproximadamente iguais a $\ln(\gamma)$. Para γ maior do que dois, por exemplo, é nítido que essas curvas de sensibilidade não convergiriam para o mesmo ponto quando a participação do estrato i é a mesma nas duas categorias.

Em resumo, como exemplo ilustrativo desse resultado, vamos considerar a categoria de interesse (X) como as mulheres e a categoria Y como os homens. Serão considerados os setores de atividade como os estratos. Os resultados indicam que, independentemente da categoria na qual sejam feitas as mudanças regressivas, para valores negativos de ε, o efeito dessas mudanças será maior para estratos cuja participação relativa seja maior.

Por outro lado, medidas que consideram valores de ε altos são mais sensíveis para estratos cuja participação relativa seja pequena.

Setores de atividade como construção civil e agrícola são casos em que, se definidos como estratos no cálculo da medida geral de segregação, as medidas serão mais sensíveis para ε altos, pois os valores de x_i/y_i são pequenos. Por exemplo, o efeito de uma mudança regressiva de uma mulher do setor agrícola para o da administração pública (cuja participação de homens e mulheres é mais próxima), terá um efeito maior se considerarmos $\varepsilon = 2$, por exemplo, comparado à medida geral com $\varepsilon = -2$. No caso extremo em que houvesse apenas uma ou duas mulheres ocupadas no setor agrícola, considerando $\varepsilon = 2$, o efeito dessa mudança regressiva seria extremamente elevado. Para $\varepsilon = -2$, esse efeito seria próximo de zero.

Para ilustrar o caso em que a participação relativa é alta, consideraremos o setor de serviços domésticos. Esse é um exemplo de atividade em que existem poucos homens ocupados em relação às mulheres. Atribuindo um valor muito baixo para ε, como -2, por exemplo, o efeito da mudança de um homem ocupado no setor de serviços domésticos para a administração pública, será tanto maior quanto menor for a participação y_i (ou quanto maior for a participação relativa x_i/y_i). No caso extremo, em que só houvesse 2 ou 3 homens ocupados no setor de serviços domésticos, o efeito dessa mudança regressiva seria extremamente elevado. Caso fosse considerado $\varepsilon = 2$, esse efeito seria próximo de zero.

As próximas seções são dedicadas à análise de sensibilidade do índice de Gini e do índice de Dissimilaridade, respectivamente.

4.2. Índice de Gini

A análise de sensibilidade do índice de Gini para segregação será feita em duas etapas. A primeira considerará uma mudança regressiva infinitesimal em dois estratos consecutivos, obedecendo à ordenação dos estratos pelas participações relativas (ou pela razão X_j/Y_j). A segunda etapa irá generalizar o resultado para o caso de estratos não consecutivos.

Considere a ordenação

$$\frac{x_1}{y_1} \leq \ldots \leq \frac{x_i}{y_i} \leq \frac{x_{i+1}}{y_{i+1}} \leq \ldots \leq \frac{x_k}{y_k}. \tag{4.9}$$

A mudança regressiva entre os estratos i e $i+1$, na categoria X, resultará em

$$\frac{x_i - \theta}{y_i} < \frac{x_{i+1} + \theta}{y_{i+1}}.$$

Assim, as coordenadas das abscissas da curva de segregação para a nova distribuição entre os estratos serão

$$\Psi_j^* = \Psi_j, \text{ se } j < i \text{ ou } j \geq i+1, \text{ e}$$
$$\Psi_j^* = \Psi_j - \theta, \text{ se } j = i.$$

Logo, de acordo com a eq. (2.31), o novo valor para o índice de Gini será

$$G_S(\mathbf{X}^*) = 1 - \sum_{j \notin \{i,i+1\}} (\Psi_j^* + \Psi_{j-1}^*)y_j - (\Psi_i^* + \Psi_{i-1}^*)y_i - (\Psi_{i+1}^* + \Psi_i^*)y_{i+1}$$

ou

$$G_S(\mathbf{X}^*) = 1 - \sum_{j=1}^{k} (\Psi_j + \Psi_{j-1})y_j + \theta y_i + \theta y_{i+1}.$$

Logo, o efeito de uma mudança na categoria X, entre os estratos i e $i+1$, é

$$\frac{\partial G_S(\mathbf{X}^*)}{\partial \theta} = y_i + y_{i+1}.$$

Assim, o efeito de uma mudança regressiva na categoria X depende, apenas, das participações dos estratos (consecutivos) envolvidos na mudança no total da categoria Y.

De modo semelhante, considerando que uma mudança regressiva na categoria Y, para os mesmos estratos, altera as participações y_i e y_{i+1} para $y_i + \theta$ e $y_{i+1} - \theta$, temos que esse efeito será

$$\frac{\partial G_S(\mathbf{Y}^*)}{\partial \theta} = x_i + x_{i+1}.$$

Observe que os resultados encontrados anteriormente são válidos para mudanças entre estratos consecutivos, ordenados por (4.9). Generalizando essa situação, vamos considerar a mesma ordenação, porém a quantidade θ é transferida do i-ésimo para o $(i+h)$-ésimo estrato. Desse modo, teremos

$$\Psi_j^* = \Psi_j, \text{ se } j < i \text{ ou } j \geq i+h, \text{ e}$$
$$\Psi_j^* = \Psi_j - \theta, \text{ se } j \geq i \text{ e } j < i+h.$$

Logo, o efeito de uma mudança regressiva na categoria X entre os estratos i e $i+h$ será

$$\frac{\partial G_S(\mathbf{X}^*)}{\partial \theta} = y_i + 2y_{i+1} + 2y_{i+2} + \ldots + 2y_{i+h-2} + 2y_{i+h-1} + y_{i+h}.$$

É curioso notar que o efeito de uma mudança regressiva na categoria X entre os estratos i e $i+h$ depende de todas as participações da categoria Y nos estratos intermediários. Assim, como as quantidades y_j são não-negativas, o efeito de uma mudança regressiva será tanto maior quanto maior for a distância h entre os dois estratos. Como o índice de Gini para segregação atende à propriedade de Simetria nas categorias (Propriedade 6), podemos generalizar esse resultado para mudanças regressivas na categoria Y.

4.3. Dissimilaridade

Vamos considerar mudanças regressivas entre os estratos i e h para analisar o efeito dessa mudança no valor do índice de Dissimilaridade. Pela equação da medida (2.38), considerando a mudança regressiva (3.14), teremos

$$D(\mathbf{X}^*) = \frac{1}{2}\sum_{j\neq i,h} |x_j - y_j| + \frac{1}{2}|x_i - \theta - y_i| + \frac{1}{2}|x_h + \theta - y_h|.$$

Devemos, obviamente, considerar a definição da função modular para analisar apropriadamente as mudanças regressivas. Lembrando essa definição, ou seja,

$$|v| = \begin{cases} v & \text{se } v \geq 0; \\ -v & \text{se } v < 0, \end{cases}$$

temos que se x_i/y_i e x_h/y_h são maiores (ou menores) que um, notamos que não haverá efeito sobre o valor da medida ($D(\mathbf{X}^*) - D(\mathbf{X}) = 0$). Por outro lado, se $x_i/y_i \leq 1$ e $x_h/y_h \geq 1$, teremos que

$$\frac{\partial D(\mathbf{X}^*)}{\partial \theta} = 1. \tag{4.10}$$

Desse modo, provamos o que já foi anunciado em capítulos anteriores. O índice de Dissimilaridade não atende à propriedade de Mudança entre estratos (Propriedade 3). Assim, somente algumas mudanças regressivas terão efeito sobre o valor da medida. Observe, também, que esse resultado é válido para mudanças regressivas na categoria Y, uma vez que a medida atende à propriedade de Simetria nas categorias (Propriedade 6).

A eq. (4.10) representa uma importante interpretação do índice de dissimilaridade. Ao invés de uma mudança regressiva, consideraremos uma mudança progressiva. Desse modo, para o caso não-infinitesimal, temos

$$D(\mathbf{X}) - D(\mathbf{X}^*) = \theta.$$

Vamos admitir que após essa mudança progressiva o valor final da medida seja zero (ausência de segregação). Nesse caso, teríamos que o nível inicial de segregação era $D(\mathbf{X}) = \theta$. Portanto, o valor de θ é interpretado como a proporção de pessoas de uma das categorias que deve se mudar de um estrato para outro de modo que não haja segregação. Essa interpretação do índice de Dissimilaridade é bastante conhecida e utilizada na literatura sobre segregação (JAMES; TAEUBER, 1985). Como o índice de Dissimilaridade não atende à propriedade de Mudanças entre estratos (Propriedade 3), não há sentido em analisar a sensibilidade a mudanças regressivas (ou progressivas) entre os estratos.

No próximo capítulo é feita a análise da segregação por gênero em setores de atividade no Brasil utilizando o índice de Gini para segregação, o índice de Dissimilaridade (Capítulo 2) e a medida geral de segregação (Capítulo 3). Também serão utilizados os resultados do presente capítulo (Capítulo 5) para explicar a variação na tendência da evolução da medida geral de segregação considerando diferentes valores para ε.

5. SEGREGAÇÃO POR GÊNERO EM SETORES DE ATIVIDADE NO BRASIL

O conceito de segregação não deve ser confundido com discriminação ou segmentação. Segregação, no mercado de trabalho, indica que uma das categorias - homens, por exemplo - tendem a se concentrar em alguns ramos de atividade, enquanto que a outra categoria (mulheres), se concentra em outras. Na literatura especializada, discriminação e segmentação estão relacionados ao rendimento. Por outro lado, para o cálculo da segregação, não é levado em consideração o rendimento, pois só é relevante em qual ramo de atividade o indivíduo está ocupado. Uma discussão abrangente sobre discriminação e segmentação para o Brasil pode ser encontrado em Barros, Franco e Mendonça (2007).

Uma nota deve ser feita sobre as expressões referentes à segregação no mercado de trabalho. A expressão "segregação no mercado de trabalho" é frequentemente utilizada em lugar de "segregação ocupacional" (*occupational segregation* (WATTS, 1998a) ou *employment segregation* (SILBER, 1989)). Neste trabalho, essas expressões serão usadas como sinônimos. Dito isso, é apropriado diferenciar os conceitos de segregação ocupacional por gênero e por cor.

Além do estudo da segregação ocupacional por gênero, tornou-se comum, após o artigo de Albelda (1986), a análise da segregação por cor no mercado de trabalho. Alguns autores como King (2009) e Salardi (2016) analisam essas duas situações para o Brasil. Ribeiro e Araújo (2016), por exemplo, preferem analisar somente a distinção por cor e ampliam a discussão para o nível de escolaridade. Embora o desenvolvimento das medidas de segregação permita fazer ambas análises, os fenômenos sociológicos nas duas situações são substancialmente diferentes.

Mesmo que o objetivo de Fryer Jr. (2011) não seja a análise de segregação entre negros e brancos no mercado de trabalho, o autor destaca que a segregação por cor, em diversas situações, está relacionada às diferenças geográficas entre as categorias, como já indicavam os trabalhos sobre a distribuição espacial entre brancos e negros nas cidades americanas (JAHN; SCHMID; SCHRAG, 1947; TAEUBER; TAEUBER, 1965). Uma excelente discussão sobre o assunto pode ser encontrada em Duncan e Duncan (1955b). Por outro lado, Anker (1997) destaca que a segregação ocupacional por gênero está relacionada aos aspectos de demanda e oferta de trabalho por parte dos homens e mulheres. Para evitar comparações espúrias nas situações de cor e de gênero, este trabalho tem por objetivo mensurar somente a segregação por gênero em setores de atividade econômica no Brasil.

Sobre a segregação no mercado de trabalho por gênero no Brasil, para a década de 80, destacam-se os trabalhos de Oliveira (1997) e Ometto, Hoffmann e Alves (1997). Ambos utilizam os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) e fazem suas análises com base no índice de Dissimilaridade. Oliveira (1997) utiliza os dados de 1981 e 1990, considerando os indivíduos ocupados em atividades e ocupações não agrícolas para todo o Brasil. Ometto, Hoffmann e Alves (1997), por sua vez, utilizam os dados de 1981 a 1990, exclusive 1982 e 1986, para os Estados de São Paulo e Pernambuco. Os autores consideram na análise pessoas ocupadas com dez anos ou mais de idade, cuja condição no domicílio seja chefe, cônjuge, filho ou outro parente e cuja posição na ocupação seja assalariado, empregador ou autônomo com rendimentos positivos. Os trabalhos levam às conclusões de que a segregação diminuiu durante a década de 80 para o Brasil, embora pouco (OLIVEIRA, 1997), e para o Estado de São Paulo (OMETTO; HOFFMANN; ALVES, 1997), porém se elevou substancialmente em Pernambuco.

Para a década de 90 em diante, destacamos os trabalhos de Salas e Leite (2007), King (2009) e Vaz e Hoffmann (2011). Os dois primeiros artigos utilizam os dados da PNAD. Salas e Leite (2007), com os dados de 1995 e 2004, e King (2009), com dados de 1989 e 2001, chegam a conclusões similares a Oliveira (1997).

Por sua vez, Vaz e Hoffmann (2011) analisam a segregação ocupacional por sexo no setor público brasileiro de 1995 a 2008, utilizando os dados da Relação Anual de Informações Sociais (RAIS). Os índices utilizados pelos autores são o de Gini e o de Dissimilaridade. Ao se analisar o setor público, onde a grande maioria das ocupações é emprego formal, a base de dados da RAIS é uma excelente fonte de informações, pois, além dessa característica sobre o emprego formal, também é uma fonte de informações censitária, e não de caráter amostral como é a PNAD. As conclusões dos autores indicam que a segregação diminuiu de modo geral nos três âmbitos de governo, em especial nas esferas municipal e federal. Dos trabalhos encontrados sobre segregação ocupacional por gênero no Brasil, esse é o único que não utiliza os dados da PNAD.

Outros trabalhos que analisam essa questão para o Brasil são Fresneda (2007) (PNAD de 2004) e Madalozzo, Martins e Lico (2015) (PNAD de 2013). Como esses dois trabalhos utilizam dados de um corte transversal, não é possível afirmar sobre a evolução da segregação no tempo. Salardi (2016) utiliza os dados da PNAD para alguns anos de 1987 a 2006 e conclui que, embora a segregação tenha diminuído, esse decréscimo não foi elevado. As próximas seções descrevem, respectivamente, a base de dados e os resultados encontrados.

5.1. Base de dados

Para ilustrar a utilização da medida geral de segregação, apresentada e discutida no Capítulo 3, utilizaremos os dados da PNAD. Também serão consideradas, para efeito de comparação, as duas medidas já consagradas na literatura: o índice de Gini para segregação e o índice de Dissimilaridade.

Os dados utilizados referem-se às PNADs de 1992 a 2014. Nesse período, não houve coleta de dados nos anos de 1994, 2000 e 2010. Até 2003, a PNAD não abrangia a área rural da antiga região Norte (atual região Norte exclusive Tocantins). Deste modo, as observações referentes a esta região de 2004 em diante foram excluídas. São consideradas todas as pessoas ocupadas, de 10 anos ou mais, com rendimentos positivos e com declaração de cor, sexo e escolaridade[1].

Até 2001, as atividades na PNAD eram classificadas de acordo com a Classificação Nacional de Atividades Econômicas (CNAE) e a Classificação Brasileira de Ocupações (CBO). A partir de 2002, a PNAD passou a adotar a CNAE-Domiciliar e a CBO-Domiciliar. A CNAE-Domiciliar apresenta diferenças significativas sobre a classificação das atividades, quando comparada à CNAE. Desse modo, a variável que indica o setor de atividade na PNAD (V9907) de 2002 em diante não deve ser confundida com a variável de mesmo nome nas PNADs anteriores.

Embora em nível mais desagregado não seja viável a compatibilização da variável de atividade, é possível a harmonização para as variáveis de grupamento da atividade. Na PNAD, até 2001, essa variável correspondia à V4709, enquanto que de 2002 em diante corresponde à V4809. Essa harmonização é realizada para que possam ser comparados os períodos antes e depois de 2002.

Será analisada a evolução e decomposição da segregação por gênero no mercado de trabalho no Brasil, para as pessoas ocupadas. Serão realizadas duas etapas. Na primeira, compara-se a evolução de várias medidas de segregação de gênero e, na segunda, é apresentada a aplicação da decomposição da medida geral de segregação.

Para a análise da evolução das medidas, são considerados sete estratos de ocupação. São eles: 1) agrícola; 2) indústria (exclusive construção civil); 3) construção civil; 4) comércio e reparação; 5) alojamento, transporte e similares; 6) administração pública e; 7) demais serviços. Foi utilizada a variável de grupamento da atividade principal (V4809) para essa estratificação. As outras atividades ou atividades mal definidas foram excluídas da amostra. Os dados para essa análise são das PNADs de 1992 a 2014. Lembra-se, todavia, que foi realizada a compatibilização da variável de grupamento da atividades, conforme procedimento sugerido pelo Instituto Brasileiro de Geografia e Estatística (IBGE), para os anos de 1992 a 2001.

Para a decomposição da segregação serão considerados seis grupamentos. A única diferença para os grupamentos da seção sobre a evolução das medidas é que o grupamento construção civil será agregado com indústria. Assim, são utilizados 53 atividades divididas em seis grupamentos. A Tabela B.1, disponível no apêndice, descreve pormenorizadamente os estratos considerados dentro de cada grupamento. Como dito anteriormente, não é possível compatibilizar a variável de atividade da PNADs de 2002 a 2014 com as de

[1] Após os recortes da amostra, para o ano de 2011, há uma observação com declaração do sexo masculino, ramo de atividade intermediação financeira (V9907=65000) e grupamento da atividade principal transporte, armazenagem e comunicação (V4809=7). Como a atividade intermediação financeira compõe o grupamento de outras atividades (V4809=12), claramente, ou a atividade ou o grupamento está com valor declarado equivocado. Desse modo, optou-se por excluí-la da amostra.

1992 a 2001. Desse modo, a decomposição da segregação será feita com os dados de 2002 a 2014. Para todas as informações estatísticas apresentadas, como proporção de pessoas nas categorias, nas ocupações e medidas de segregação, por exemplo, são considerados os fatores de ponderação amostral de cada observação.

5.2. Resultados

A Tabela 5.1 apresenta o número de homens e mulheres ocupados na população e na amostra por ano. Embora tanto a população economicamente ativa ocupada de homens e mulher cresça, é mais acentuado o crescimento do número de mulheres. Em termos relativos, esse aumento é de 86,3%, que é equivalente a aproximadamente 14,6 milhões de mulheres. Para a população de homens, houve aumento de aproximadamente 12,9 milhões de ocupados, ou seja, crescimento de 40,2%. A população economicamente ativa ocupada aumentou de 49 para 76,4 milhões de pessoas (56,1%).

A última coluna da Tabela 5.1 apresenta a evolução da participação das mulheres no mercado de trabalho, em valores percentuais, representada, graficamente, na Figura 5.1. A figura mostra, claramente, o aumento da participação feminina que, de 1992 a 2014, passou de 34,5% para 41,2%. Esse aumento é menor que 7 pontos percentuais (p.p.). Em termos relativos, seria um aumento de 19,3%. Uma regressão da participação das mulheres contra o ano correspondente permite concluir que há evidências estatísticas para afirmar que a participação das mulheres no mercado de trabalho aumentou, ao nível de significância de 0,1%. Com exceção de uma leve diminuição de quase 0,3 p.p. na participação das mulheres em 1997, para todos os demais anos houve aumento dessa participação. Uma discussão abrangente e detalhada sobre a participação das mulheres no mercado de trabalho no Brasil pode ser encontrada em Soares e Izaki (2002).

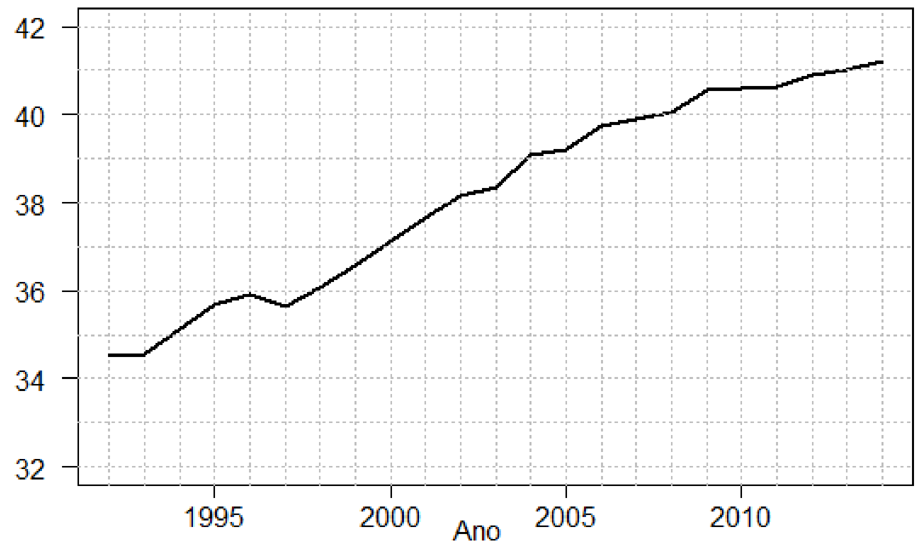

Figura 5.1. Participação percentual das mulheres no mercado de trabalho. Brasil, 1992 a 2014

Tabela 5.1. Número de homens e mulheres ocupados por ano. Brasil, 1992 - 2014

Ano	Homens	Mulheres	% de mulheres
1992	32.053.355	16.911.359	34,54
	(69.269)	(37.675)	
1993	32.731.140	17.300.868	34,58
	(70.221)	(38.321)	
1995	33.925.118	18.815.243	35,68
	(73.552)	(42.285)	
1996	33.291.070	18.671.261	35,93
	(70.189)	(41.001)	
1997	34.149.749	18.923.784	35,66
	(74.571)	(42.866)	
1998	34.028.104	19.211.563	36,09
	(72.878)	(42.764)	
1999	35.238.765	20.316.372	36,57
	(74.298)	(44.502)	
2001	36.958.132	22.314.326	37,65
	(80.014)	(49.399)	
2002	37.897.641	23.362.819	38,14
	(82.188)	(51.681)	
2003	38.126.906	23.728.586	38,36
	(81.533)	(51.603)	
2004	39.294.882	25.212.893	39,09
	(84.078)	(54.740)	
2005	40.179.964	25.918.889	39,21
	(86.521)	(56.937)	
2006	40.724.062	26.853.808	39,74
	(87.315)	(58.382)	
2007	41.392.434	27.491.726	39,91
	(85.526)	(57.714)	
2008	42.797.517	28.598.730	40,06
	(85.813)	(58.326)	
2009	42.843.556	29.233.297	40,56
	(86.623)	(60.233)	
2011	42.614.652	29.180.657	40,64
	(75.937)	(52.526)	
2012	43.703.222	30.226.762	40,89
	(77.986)	(54.866)	
2013	43.631.034	30.334.559	41,01
	(76.909)	(54.108)	
2014	44.949.271	31.508.013	41,21
	(79.142)	(56.226)	

Fonte: elaboração própria com base nos microdados da PNAD de 1992 a 2014.
Nota: valores entre parênteses indicam o tamanho da amostra por gênero e ano.

5.2.1. Evolução da segregação de 1992 a 2014

A Figura 5.2 apresenta a curva de segregação para os anos de 1992 (0,5828), 2003 (0,5717) e 2014 (0,5413). Os valores entre parênteses são os respectivos índices de Gini. Mesmo que pela visualização gráfica das curvas de segregação não seja claro que a segregação tenha diminuído, os valores para o índice de Gini apontam decréscimo de mais de 7% (4,1 p.p.) entre 1992 e 2014. Embora os dados apresentados nas tabelas desta seção sejam referentes apenas aos anos de 1992, 2003 e 2014, os dados para os demais anos podem ser encontrados nas tabelas disponíveis no Apêndice B.

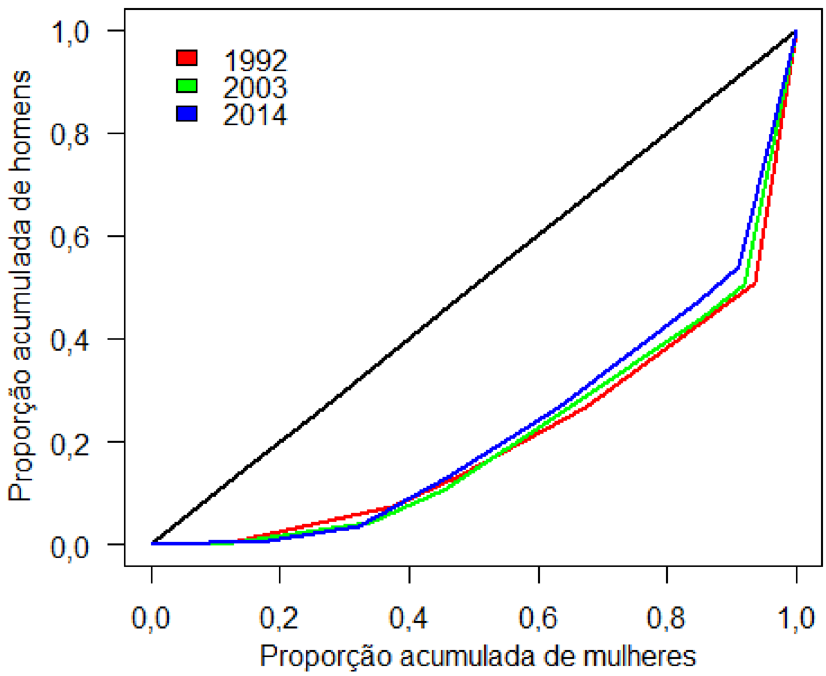

Figura 5.2. Curvas de segregação para sete grupamentos de atividade principal. Brasil, 1992, 2003 e 2014

É importante fazer uma observação sobre o número de algarismos significativos apresentados para as medidas de segregação. Em geral, apenas dois ou três algarismos após a vírgula são reportados em trabalhos sobre segregação ou desigualdade. Neste trabalho, são utilizados quatro algarismos para fins de verificação. A Tabela 5.2 apresenta as participações de cada grupamento de atividade nos totais de homens e mulheres em cada ano, que permitem a construção da curva de segregação para 1992, 2003 e 2014.

Consideraremos, para efeito de ilustração, o ano de 2014. Análise similar pode ser feita para os demais anos. Ordenando os grupamentos por suas participações relativas (razão entre a participação do grupamento no total das mulheres sobre a respectiva participação para homens), temos a seguinte ordenação: 1) construção civil; 2) agrícola; 3) alojamento, transporte e similares; 4) indústria; 5) comércio e reparação; 6) administração pública; e 7) demais serviços. Desses sete grupamentos, somente para os dois últimos

Tabela 5.2. Participação de cada grupamento no total das mulheres (x_j) e dos homens (y_j) para sete grupamentos de atividade (em %). Brasil, 1992, 2003 e 2014

	1992		2003		2014	
Grupamento	x_j	y_j	x_j	y_j	x_j	y_j
Agrícola	6,86	25,81	3,84	21,02	2,80	13,87
Indústria	17,96	20,08	15,29	18,56	13,51	17,17
Const. civil	0,48	11,87	0,38	12,75	0,79	18,32
Comércio e rep.	13,53	19,10	18,31	21,62	20,65	21,72
Aloj., transp, e similares	5,96	10,60	6,51	11,66	9,51	13,78
Adm. pública	5,89	6,32	6,05	6,46	6,56	6,31
Demais serviços	49,33	6,23	49,62	7,92	46,18	8,83
Total	100	100	100	100	100	100

Fonte: com base nos microdados da PNAD.

a participação das mulheres é maior que a dos homens. Destaca-se, além disso, que a participação das mulheres nos demais serviços é mais de cinco vezes a participação dos homens.

Os dados mostram que, em 2014, menos de 3% dos ocupados na construção civil eram mulheres. Considerando todos os anos de 1992 a 2014, essa proporção ficou entre 1,6% (1997) a 3,2% (1999). O segundo grupamento com a menor participação feminina é o agrícola. A Tabela 5.3 apresenta esses dados para os anos selecionados.

Tabela 5.3. Proporção de mulheres em cada um dos sete grupamentos de atividade (em %). Brasil, 1992, 2003 e 2014

	Ano		
Grupamento	1992	2003	2014
Agrícola	12,30	10,20	12,41
Indústria	32,06	33,89	35,56
Const. civil	2,08	1,84	2,93
Comércio e rep.	27,20	34,51	39,99
Aloj., transp, e similares	22,86	25,79	32,60
Adm. pública	32,99	36,82	42,16
Demais serviços	80,69	79,59	78,57

Fonte: com base nos microdados da PNAD.

Embora não haja evidências estatísticas para afirmar que a participação das mulheres nos grupamentos da construção civil e agrícola tenha aumentado, podemos afirmar, ao nível de significância de 5%, que essa participação no grupamento demais serviços diminuiu (valor-p igual a 0,021). Por outro lado, para os demais grupamentos (indústria, comércio e reparação, alojamento, transportes e similares e administração pública), há evidências estatísticas para afirmar que a participação das mulheres no grupamento aumentou (valor-p menor que 0,001 em todos os casos).

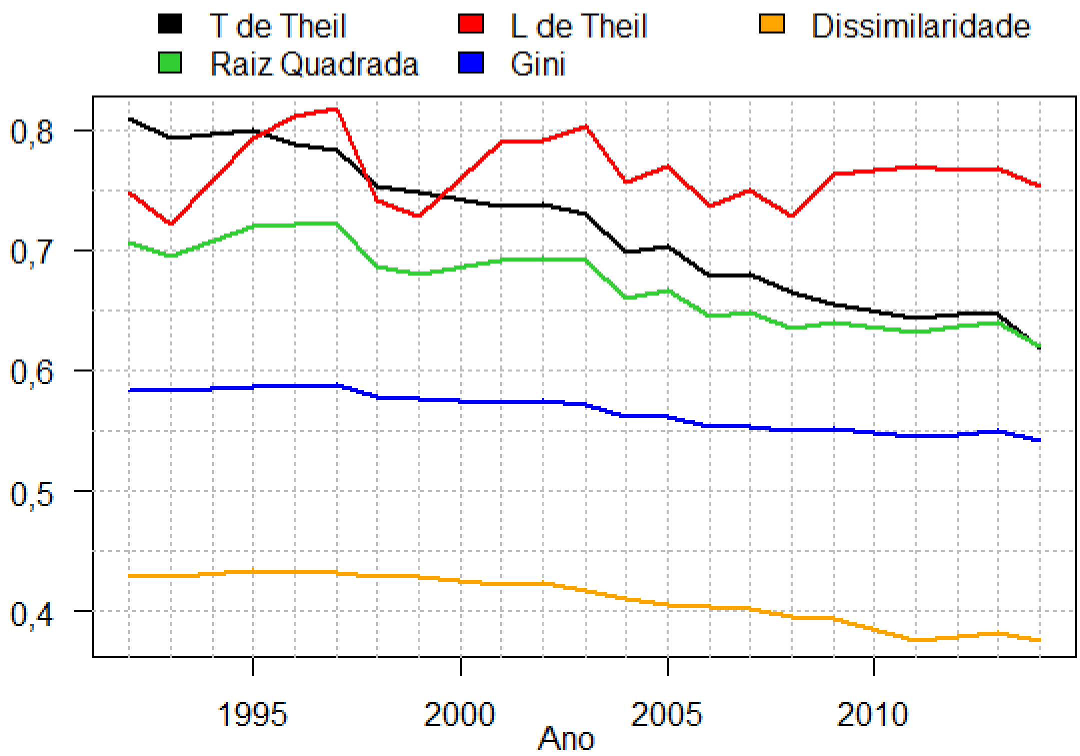

Figura 5.3. Evolução da segregação de gênero no mercado de trabalho para sete grupamentos de atividade. Brasil, 1992 - 2014

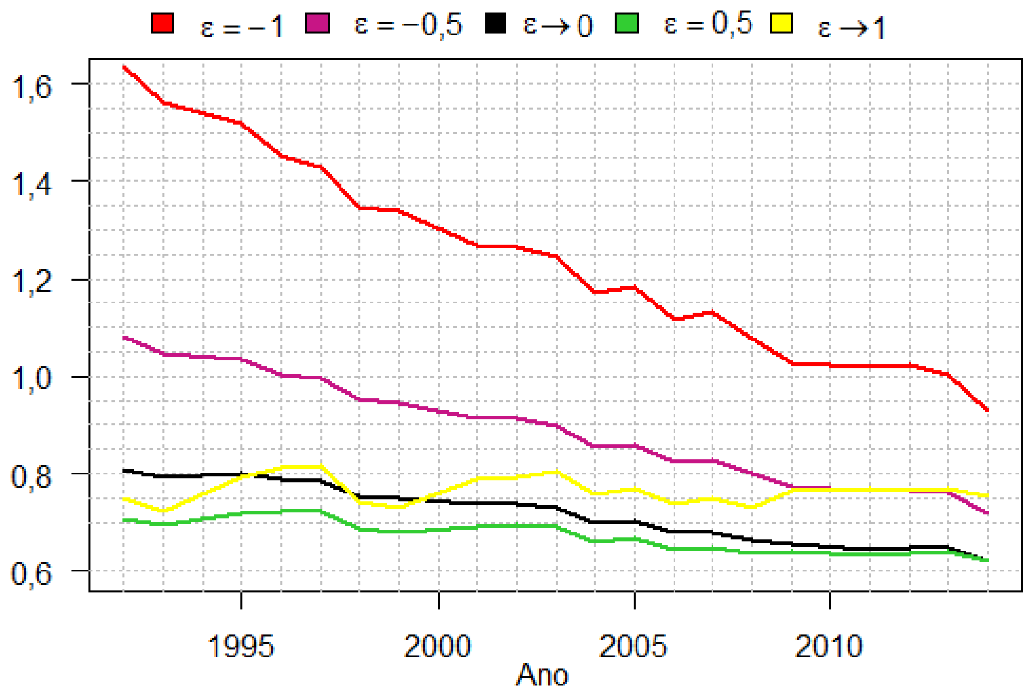

Figura 5.4. Medida geral de segregação para valores de ε, de -1 a 1, inclusive os extremos. Brasil, 1992 - 2014

A Figura 5.3 apresenta a evolução das medidas de segregação Gini, Dissimilaridade, T e L de Theil e Raiz Quadrada. Com exceção da medida L de Theil, todas as demais apresentam tendência de queda. A 0,1% de significância, podemos afirmar, para essas medidas, que a segregação diminuiu de 1992 a 2014. Para o índice T de Theil houve queda de 23,5%, seguido pelo índice de Dissimilaridade (12,8%), Raiz Quadrada (12,4%) e Gini (7,1%).

O índice de Dissimilaridade mostra que, em 2014, 37,6% das mulheres deveriam mudar dos grupamentos predominantemente femininos para os predominantemente masculinos, para que a segregação fosse nula. Esse valor era 43,1% em 1992. Esse tipo de interpretação para o índice de Dissimilaridade foi demonstrado na seção 4.3.

Nota-se que há evidências para afirmar que a segregação diminuiu, utilizando a medida geral, para valores de ε tendendo a zero ou igual a 0,5, porém, não podemos afirmar o mesmo para ε tendendo a um. Desse modo, é interessante analisar como a medida geral de segregação evolui temporalmente e, ao mesmo tempo, como essa evolução varia para diversos valores do parâmetro ε. A Figura 5.4 apresenta os resultados para ε assumindo valores de -1 a 1, inclusive os extremos, para intervalos de 0,5.

Pela figura, é claro que, conforme os valores de ε aumentam, a inclinação da linha que mostra a evolução da medida de segregação tende a ser menos negativa. Para $\varepsilon = -1$, houve queda de 43,2% na segregação entre 1992 e 2014, seguido de queda de 33,6% para $\varepsilon = -0,5$. Conforme os valores de ε aumentam de -1 a 1, a queda na segregação de 1992 a 2014 diminui, progressivamente, de 43,2% adotando $\varepsilon = -1$ para zero por cento considerando $\varepsilon \to 1$.

As Figuras 5.5 e 5.6 apresentam a evolução das medida de segregação para $0 < \varepsilon \leq 0,5$ e para $0,5 \leq \varepsilon < 1$, variando de 0,1 em 0,1. Conforme demonstrado no Capítulo 3, para esses valores de ε as medidas de segregação são limitadas superiormente (exclusive os extremos para ε igual a 0 ou 1). Para valores de ε de 0 a 0,5 (Figura 5.5), as tendências de queda são similares. Por outro lado, para ε de 0,5 a 1 (Figura 5.6), o comportamento das medidas são similares, porém, conforme o valor de ε aumenta, não fica claro que há tendência de queda ao longo do tempo. As regressões dos valores das medidas contra a tendência temporal indicam que há evidências estatísticas, a 0,1%, para afirmar que a segregação diminuiu utilizando valores de ε menores que 0,8. Para esse valor, rejeitamos a hipótese nula de que não há tendência linear de queda ao nível de significância de 1%. Para ε igual a 0,9 ou um, não há evidências estatísticas, a 10%, de que a segregação tenha diminuído.

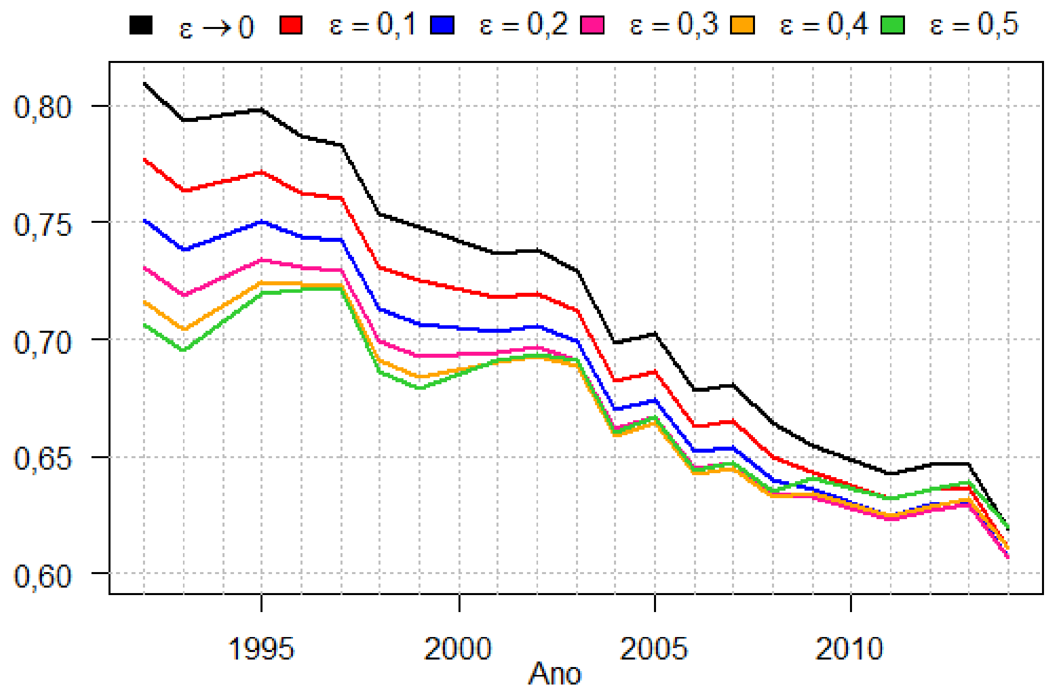

Figura 5.5. Medida geral de segregação para valores $0 < \varepsilon \leq 0,5$, para sete grupamentos de atividade. Brasil, 1992 - 2014

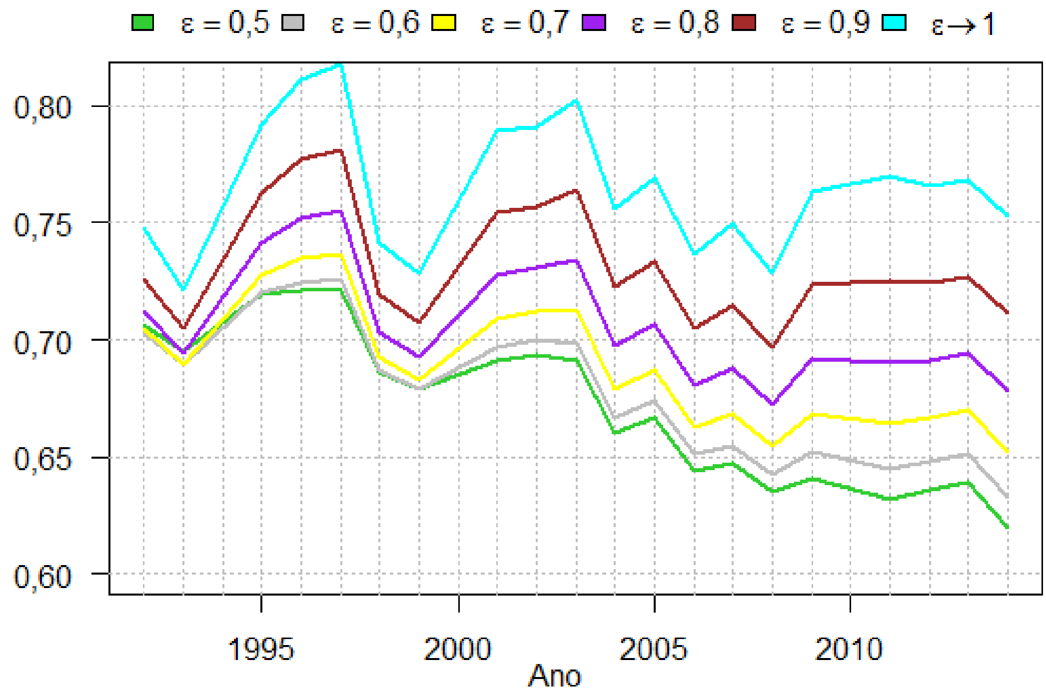

Figura 5.6. Medida geral de segregação para valores $0,5 \leq \varepsilon < 1$, para sete grupamentos de atividade. Brasil, 1992 - 2014

Curiosamente, considerando valores de ε maiores que um, como 1,5 e 2, por exemplo, podemos afirmar, a 5% de significância, que houve aumento na segregação. Buscaremos explicar a razão pela qual, ao utilizar valores para ε elevados, a medida geral indica essa tendência de aumento. Para isso, a análise é baseada na evolução das participações relativas, representadas na Figura 5.7. Nota-se que não aparece nessa figura a evolução da participação relativa do setor de demais serviços, que é o único dos sete grupamentos considerados com participação relativa de mulheres acima de 1, com tendência claramente decrescente no período analisado.

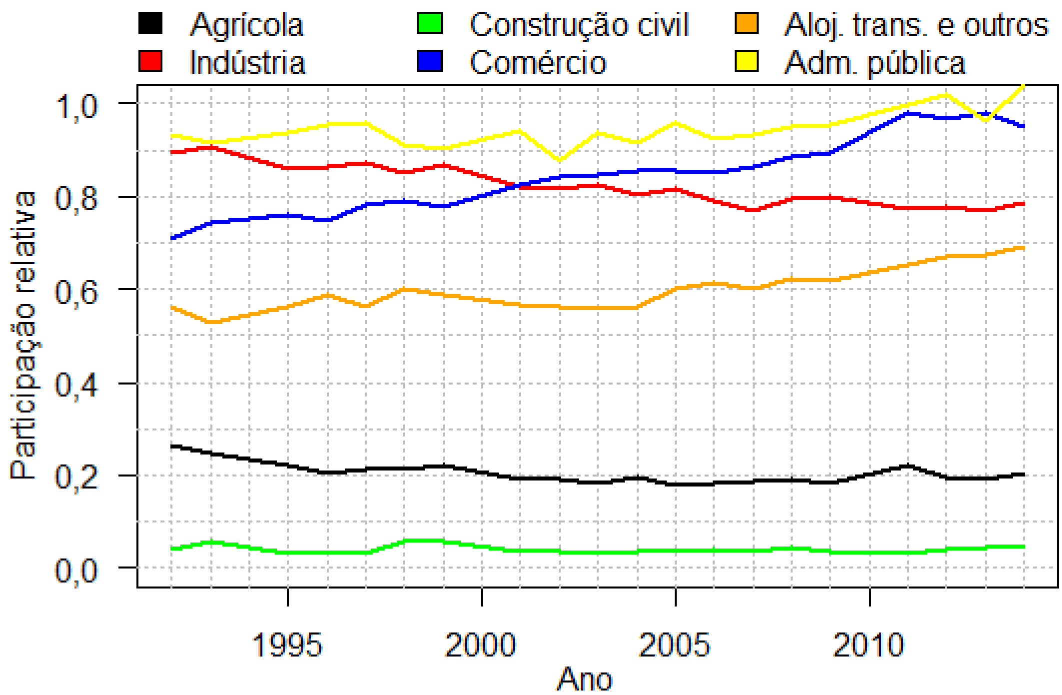

Figura 5.7. Evolução da participação relativa por grupamento (exclusive demais serviços). Brasil, 1992 - 2014

Estatisticamente, podemos afirmar, aos níveis de significância entre parênteses, que as participações relativas referentes aos grupamentos alojamento, transporte e similares (0,1%), comércio (0,1%) e administração pública (1%) apresentaram tendência de crescimento. Como a participação dos homens nesses grupamentos, em geral, é maior que a das mulheres, esses grupamentos contribuíram para a diminuição da segregação. A participação relativa do setor de demais serviços (0,1%) apresentou forte tendência de queda, porém, como a participação das mulheres predomina nesse grupamento, esse também contribuiu para a diminuição da segregação.

A participação relativa do grupamento da construção civil, estatisticamente, não apresentou tendência alguma. Já para a indústria (0,1%) e o agrícola (1%), houve tendência de queda na participação relativa. Como esses dois grupamentos são predominantemente masculinos, eles contribuíram para o aumento da segregação. É importante visualizar que a queda da participação relativa da indústria pode ser compensada pelo aumento da participação relativa do comércio. Assim, considerando valores elevados para ε, o único grupamento que parece contribuir fortemente para o aumento da segregação é o agrícola. Há uma explicação metodológica para esse resultado.

Conforme resultados demonstrados nos Capítulos 3 e 4, se considerarmos valores de ε elevados (maiores que 1, por exemplo) as medidas de segregação são mais sensíveis para estratos cuja participação relativa de mulheres seja muito pequena. Como a delimitação dos nossos grupamentos inclui dois grupamentos cuja participação relativa é muito baixa, temos que, para valores de ε altos, os grupamentos agrícola e construção civil merecem especial atenção. Como vimos, para a participação relativa do grupamento da construção civil, não foi verificada tendência de queda ou aumento. Desse modo, resta-nos analisar o comportamento da participação relativa do grupamento agrícola.

Podemos imaginar transferências regressivas desse grupamento para outros, de 1992 a 2014, de modo que diminua a participação relativa do grupamento agrícola (princípio das mudanças regressivas). Segundo os resultados do Capítulo 4, sintetizados nos últimos parágrafos da seção 4.1 (p. 55), para ε relativamente alto, os efeitos dessas transferências serão tanto maiores quanto menores forem os valores das participações relativas. Para o grupamento agrícola, de 1992 a 2014, a participação relativa diminuiu de 0,2658 para 0,2021. Conforme dito anteriormente, podemos afirmar que há tendência de queda para essa participação relativa. Isso explica por que há tendência de aumento da segregação no Brasil, ao considerar valores elevados de ε, para esses sete grupamentos. Assim, será verificado aumento na segregação por gênero, porém, essa evolução é explicada somente pelo grupamento agrícola, pois são atribuídos pesos relativamente baixos aos demais grupamentos.

A análise de correlação entre a medida geral de segregação, considerando diversos valores de ε, corrobora esses resultados. A correlação entre as medidas para os 20 anos disponíveis, considerando dois valores de ε de -1 a 0,8 são, sempre, correlacionados positivamente e significativos a 5%. Considerando a correlação entre os valores de duas medidas, a primeira com $\varepsilon = 2$ e, a segunda, com ε menor ou igual a 0,6, essas correlações são negativas, porém não estatisticamente significativas.

Em resumo, com as medidas mais utilizadas na literatura sobre segregação, o índice de Gini e o Dissimilaridade, pode-se comprovar que houve diminuição da segregação por gênero, entre pessoas ocupadas, considerando os grupamentos de atividades no Brasil de 1992 a 2014. Por outro lado, não se pode fazer tal afirmação utilizando a medida geral de segregação para qualquer valor de ε.

Para valores de ε menores que 0,8, a 5% de significância, há evidências de que a segregação por gênero diminuiu ao longo do tempo. Considerando valores do parâmetro de 0,8 a 1, por exemplo, não há indício de tendência temporal. Por outro lado, com ε igual a 1,5 ou 2, há evidências estatísticas, a 5%, para afirmar que a segregação aumentou. Como explicado, ao se utilizar valores elevados para ε nesse contexto, é dado peso muito elevado para o grupamento agrícola. Como a participação relativa desse grupamento diminuiu com o tempo, constata-se aumento da segregação. Ao utilizarmos, por exemplo, o índice Raiz Quadrada, que é a medida que atende a todas as propriedades enunciadas no Capítulo 3, podemos afirmar que a segregação diminuiu. Se desejarmos evitar uma medida de segregação que possa ser demasiadamente sensível ao que acontece em um dos estratos, e muito pouco sensível ao que acontece nos demais, cabe limitar os valores de ε na medida geral ao intervalo de -1 a 1.

A seção seguinte descreve os resultados para a decomposição da segregação de 2002 a 2014.

5.2.2. Decomposição da segregação de 2002 a 2014

A Figura 5.8 apresenta a evolução da segregação, considerando 53 atividades, para valores de ε de 0 a 1. A evolução temporal para o T e L de Theil, o índice Raiz Quadrada e considerando ε igual a 0,1 e 0,9 são similares aos apresentados na Figura 5.3, na qual eram considerados sete grupamentos. Como verificado anteriormente, para o T de Theil, para $\varepsilon = 0,1$ e o índice Raiz Quadrada, há evidências, a 0,1%, para afirmar que a segregação diminuiu. Para esses índices, essa queda, considerando as 53 atividades, foi de 17,3%, 15,8% e 10,5%, respectivamente. Não há evidências para afirmar o mesmo considerando $\varepsilon = 0,9$ ou o índice L de Theil, para o mesmo nível de significância. Também havíamos encontrado esse mesmo resultado, considerando sete grupamentos, na seção anterior.

Para a segregação entre os seis grupamentos, a Figura 5.9 apresenta essa evolução temporal. Curiosamente, embora não se possa afirmar que a segregação diminuiu considerando as 53 atividades para ε igual a 0,9 ou 1, podemos afirmar, a 0,1%, que a segregação entre os grupamentos diminuiu entre 2002 e 2014, para qualquer valor de ε de 0 a 1. Considerando essas medidas, temos que a queda na segregação entre os seis grupamentos foi de 18% (T de Theil), 17,3% ($\varepsilon = 0,1$), 14,9% (Raiz Quadrada), 13,2% ($\varepsilon = 0,9$) e 12,9% (L de Theil).

Considerando as medidas com $\varepsilon = 0,9$ e igual a 1 (L de Theil), as medidas entre os grupamentos são praticamente idênticas. Por outro lado, considerando as medidas com ε tendendo a 0 e igual a 0,1, a diferença entre essas duas medidas é maior, porém diminui ao longo do tempo de 2,7 p.p. em 2002, para 1,7 p.p. em 2014.

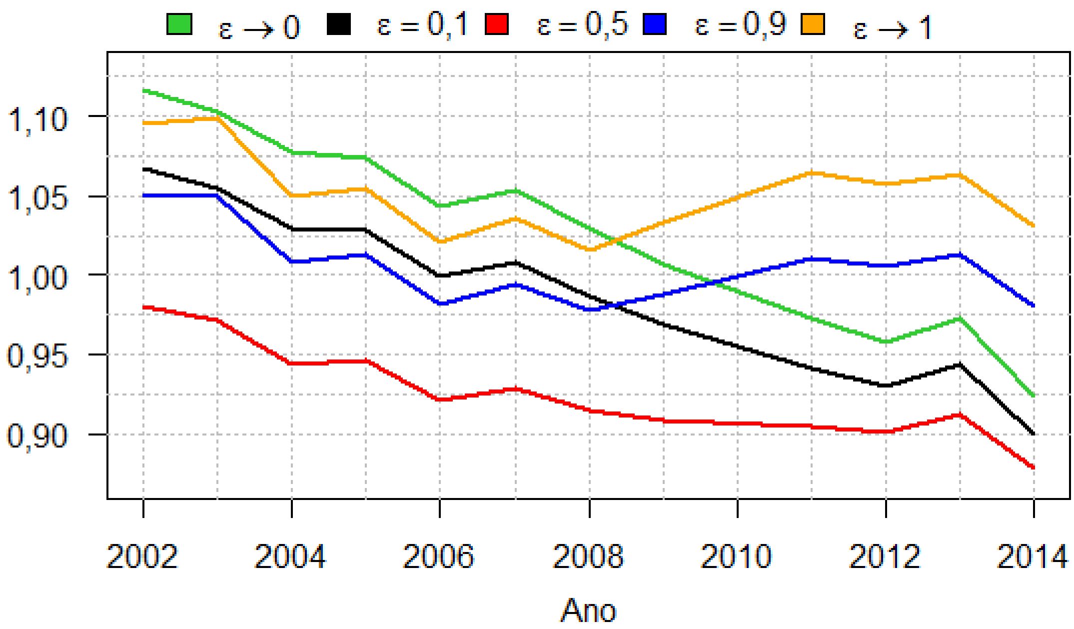

Figura 5.8. Evolução da segregação para 53 atividades. Brasil, 2002 - 2014

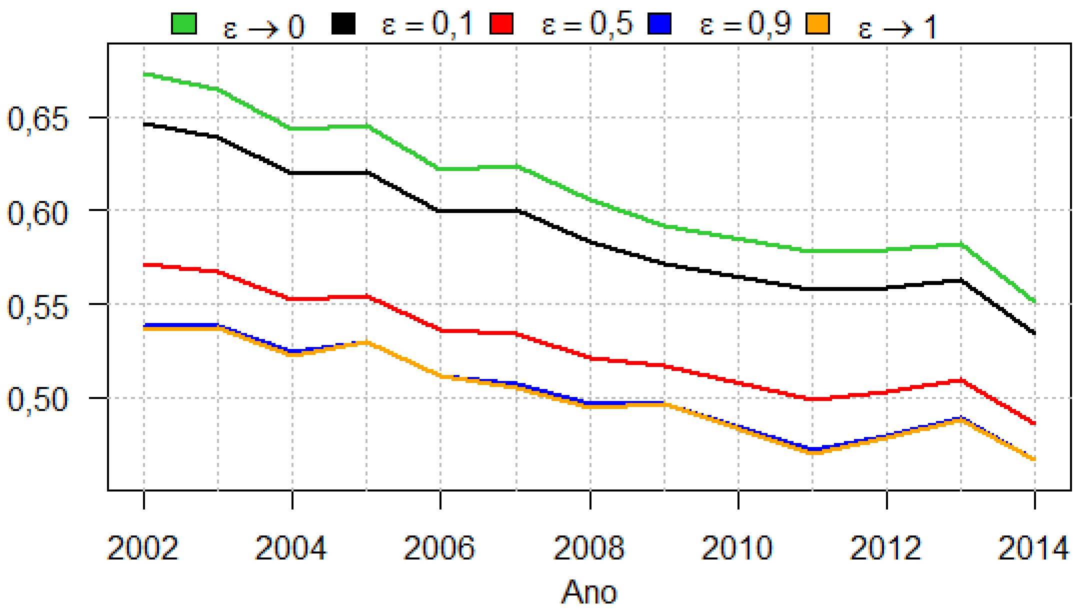

Figura 5.9. Evolução da segregação para seis grupamentos. Brasil, 2002 - 2014

A Figura 5.10 apresenta a evolução da parcela entre os grupamentos na segregação total. Considerando o índice Raiz Quadrada, o L de Theil ou $\varepsilon = 0,9$, existem evidências, a 1%, de que há tendência de queda, ao longo do tempo, da parcela entre os grupamentos na segregação total. Porém, para o índice T de Theil e para $\varepsilon = 0,1$, não rejeitamos essa hipótese nula ao mesmo nível de significância. Para a medida L e para $\varepsilon = 0,9$, entre 2002 e 2014, a queda dessa participação na segregação total foi de 7% (ou 3,6 p.p. para os dois casos). Para 2014, a participação da parcela entre grupamentos na segregação total era de 59,74% para o índice T de Theil, enquanto que para a medida L essa valor era de 45,3%.

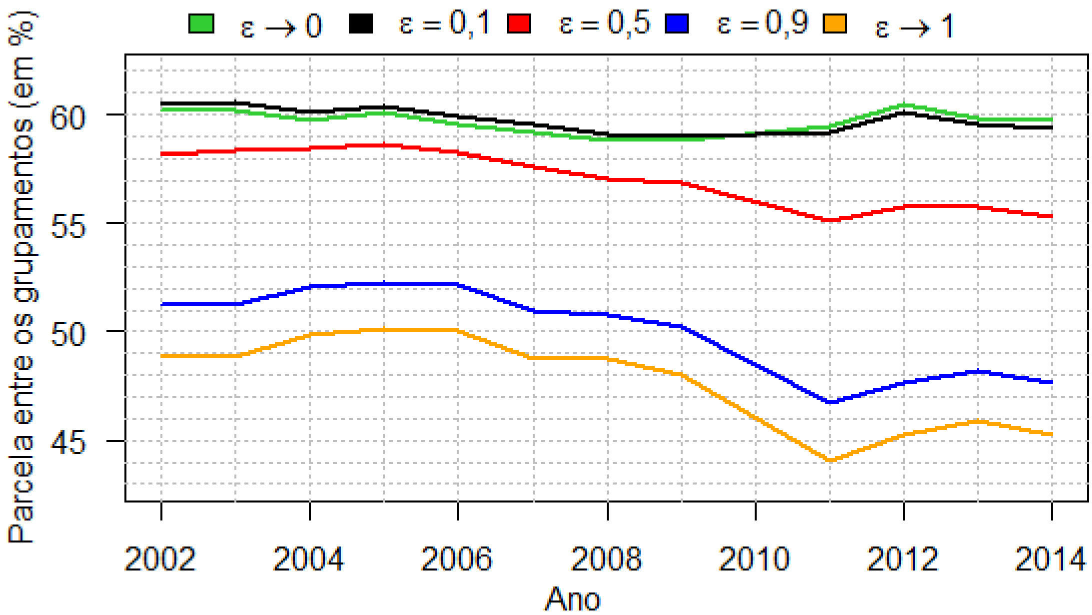

Figura 5.10. Participação da parcela “entre grupamentos” na segregação total. Brasil, 2002 - 2014

Em resumo, para a decomposição da segregação, consideraremos o índice Raiz Quadrada. Para as 53 atividades, de 2002 a 2014, a segregação diminuiu em 10,5%. A segregação entre os grupamentos, por sua vez, diminuiu 14,9%. Estatisticamente, podemos afirmar que a participação da parcela entre os grupamentos, em relação a segregação total, também diminuiu. Desse modo, embora à segregação entre atividades por gênero tenha diminuído, esse decréscimo é puxado pela segregação entre os grupamentos.

6. OUTROS RESULTADOS

6.1. Discussão sobre o limite superior para os índices de Gini

Nesta seção procuramos deixar claro por que o limite superior do índice de Gini para segregação ($G_S(\mathbf{X})$) é 1, embora o limite superior do índice de Gini para distribuição de rendas não negativas ($G(\mathbf{z})$), em uma população de n pessoas, seja $1 - 1/n$.

Vamos pressupor que um indivíduo detém toda a renda (concentração máxima), ou seja, $z_i = 0$, $i \in \{1, 2, ..., n-1\}$ e $0 < z_n < \infty$. Da eq. (2.17), temos que a inclinação do último segmento da poligonal de Lorenz será

$$\begin{aligned} d_n &= \frac{z_n}{\mu} = \frac{z_n}{z_n \backslash n} \qquad \text{ou} \\ d_n &= n. \end{aligned}$$

Como pressupomos população finita ($n < \infty$), a inclinação do último segmento dessa poligonal não poderá ser vertical. Por outro lado, a declividade da curva de segregação referente ao j-ésimo estrato é

$$d_j^S = \frac{X_j}{Y_j}\frac{Y}{X}.$$

Temos que $X, Y > 0$, porém a restrição não é estrita para X_j e Y_j ($X_j, Y_j \geq 0$). Pela ordenação dada em (2.24), se somente um estrato for ocupado exclusivamente por pessoas da categoria X, esse estrato necessariamente será o k-ésimo. Sob essas condições, se $Y_k = 0$, então $d_k^S \to \infty$. Isso implica que a inclinação do k-ésimo segmento da poligonal será vertical, precisamente, em algum trecho sobre o eixo AB na Figura 2.2. Consideremos, a seguir, uma situação mais geral, pressupondo que até o h-ésimo estrato haja apenas pessoas da categoria Y e do (h+1)-ésimo em diante todos incluam apenas pessoas da categoria X. Nesse caso temos

$$\begin{aligned} &X_j = 0,\, Y_j > 0, \qquad j \in \{1, 2, ..., h\} \quad \text{e} \\ &X_j > 0,\, Y_j = 0, \qquad j \in \{h+1, h+2, ..., k\}. \end{aligned}$$

Para $j \in \{1, 2, ..., h\}$, os pontos da curva de segregação estarão sobre o eixo OA, pois serão da forma $(\rho_j, 0)$ (inclinação de cada segmento da poligonal igual a zero). O último ponto deste segmento, (ρ_h, Ψ_h), coincidirá com $A = (1, 0)$. Para $j \in \{h+1, h+2, ..., k\}$, a inclinação de cada segmento da poligonal será vertical. Desse modo, todos os pontos estarão sobre o eixo AB, ou seja, serão da forma $(1, \Psi_j)$. Temos, portanto, que na situação de máxima segregação α_S será a área do triângulo OAB, ou seja, igual a 1/2. Nessa situação, teremos $G_S(\mathbf{X}) = 1$.

Assim, na análise da distribuição de renda, quando a população tende a infinito e um indivíduo concentra toda a renda temos, no limite, $G(\mathbf{z}) = 1$. No caso da segregação, isso acontece quando todos os extratos são ocupados exclusivamente por uma das duas categorias. Não é necessária a condição $k \to \infty$ para que $G_S(\mathbf{X}) = 1$. Se todos os estratos são ocupados exclusivamente por pessoas de uma categoria, é condição suficiente $2 \leq k < \infty$ para que o índice de Gini para segregação assuma valor igual a 1.

6.2. Relação entre o índice de Centralização Relativa e a razão de concentração

Uma medida proposta para medir a segregação geográfica é o índice de Centralização Relativa ($RCE(\mathbf{X})$). Essa medida parte da pressuposição de que existem diferenças nas proporções de pessoas de cada categoria em diferentes bairros. Além disso, também se pressupõe que existe uma tendência de aumento da proporção de uma das categorias à medida que os bairros estão mais distantes do centro. Utilizando a notação (2.25), porém adotando a ordenação dada pela distância do bairro até o centro, o $RCE(\mathbf{X})$ é definido por Duncan e Duncan (1955b) como

$$RCE(\mathbf{X}) = \sum_{j=1}^{k} \Psi_{j-1}\rho_j - \sum_{j=1}^{k} \Psi_j \rho_{j-1}. \tag{6.1}$$

É importante destacar que essa medida varia de -1 a 1, inclusive os extremos. Valores positivos indicam que a categoria de interesse (X) está mais concentrada no centro, enquanto que valores negativos indicam que a categoria Y está mais concentrada em relação ao centro. As duas situações extremas ocorrem quando, para todos os bairros até a distância h do centro, só residam pessoas de uma categoria e para bairros cuja distância seja maior que h só residam pessoas da outra.

Podemos adicionar e subtrair a expressão $(\sum_{j=1}^{k} \Psi_j \rho_j - \sum_{j=1}^{k} \Psi_{j-1}\rho_{j-1})$ ao lado direito da eq. (6.1). Fazendo as operações algébricas necessárias[1] e observando que $\sum_{j=1}^{k} (\Psi_j \rho_j - \Psi_{j-1}\rho_{j-1}) = \Psi_k \rho_k = 1$, tem-se

$$RCE(\mathbf{X}) = 1 - \sum_{j=1}^{k} (\Psi_j - \Psi_{j-1})(\rho_j + \rho_{j-1}). \tag{6.2}$$

De modo semelhante às operações anteriores, também se pode chegar à expressão

$$RCE(\mathbf{X}) = \sum_{j=1}^{k} (\Psi_j + \Psi_{j-1})(\rho_j - \rho_{j-1}) - 1. \tag{6.3}$$

Observe que o primeiro termo do lado direito da equação acima corresponde a duas vezes a área β_S no gráfico da curva de segregação menos um (ver eq. (2.28)). Assim, de (6.3) temos

$$RCE(\mathbf{X}) = 2\beta_S - 1. \tag{6.4}$$

[1] Ver seção A.3, p. 94, no apêndice.

Não podemos dizer que o $RCE(\mathbf{X})$ é o negativo do índice de Gini, comparando as eqs. (6.4) com (2.31). Para a construção do índice de Gini, os indivíduos são ordenados conforme a renda e o eixo das ordenadas é a proporção acumulada da renda desses indivíduos. Quando a ordenação é feita por uma variável, mas a proporção acumulada no eixo das ordenadas é outra, a formulação é equivalente ao do índice de Gini, porém essa medida recebe o nome de *razão de concentração* (PYATT; CHEN; FEI, 1980).

Assim, a expressão do $RCE(\mathbf{X})$, multiplicada por menos um, é uma forma equivalente às expressões da razão de concentração encontradas na literatura. Uma vez que a população está ordenada por valores não decrescentes de uma variável e o eixo das ordenadas é a proporção acumulada dessa mesma variável, a expressão do $-RCE(\mathbf{X})$ é a mesma que do índice de Gini. Os limites inferior e superior da razão de concentração são, respectivamente, segundo Hoffmann (2014), $-1 + 1/n$ e $1 - 1/n$, quando a ordenação é realizada pela renda em população finita. A razão pela qual temos que $RCE(\mathbf{X}) \in [-1, 1]$ é a mesma que a discutida na seção 6.1.

No Brasil, só foram encontrados dois trabalhos que fazem referência ao índice de Centralização Relativa. São eles Carvalho et al. (2013) e Fernandes (2012), porém nenhum dos dois trabalhos aplicam essa medida.

O $RCE(\mathbf{X})$ apareceu na literatura sobre medidas de segregação após o trabalho de Massey e Denton (1988), porém acreditamos que é inapropriado tratá-lo como medida de segregação. Um exemplo numérico ilustra esse fato. Imagine uma cidade com três bairros, ordenados por ordem decrescente de distância ao centro, com a distribuição apresentada na Tabela 6.1.

Tabela 6.1. Distribuição das categorias X e Y nos bairros de uma cidade hipotética

j	X_j	Y_j	x_j	y_j	Ψ_j	ρ_j	$\Psi_{j-1}\rho_j$	$\Psi_j\rho_{j-1}$
1	1000	0	0,5	0	0,5	0	0	0
2	0	50	0	1	0,5	1	0,5	0
3	1000	0	0,5	0	1	1	0,5	1

Fonte: elaboração própria.

O cálculo do $RCE(\mathbf{X})$ indicará valor 0, enquanto o índice de Dissimilaridade é 1. De fato, a segregação é máxima nessa situação, pois em todos os bairros residem exclusivamente pessoas de uma categoria. Portanto, o índice de Centralização Relativa medirá a *concentração* dos grupos em relação ao centro, mas não a *segregação* existente entre grupos.

Hoffmann (2014) destaca uma importante propriedade da razão de concentração. Se as proporções de pessoas de ambos os grupos forem iguais em todos os bairros, a razão de concentração será 0, porém o inverso não é válido. Se o RCE é igual a 0, não se pode afirmar que as proporções de pessoas de cada grupo em cada bairro serão iguais. Assim, temos que $x_j = y_j$, para todo j, é condição suficiente para $RCE(\mathbf{X}) = 0$, mas $x_j = y_j$

não é condição necessária para $RCE(\mathbf{X}) = 0$. O exemplo numérico da Tabela 6.1 ilustra esse fato.

Portanto, considerando que a distribuição espacial dos dois grupos é a mesma quando as proporções são iguais em todos os bairros, deve-se tomar cuidado com a afirmação de que quando o valor do índice é zero, os dois grupos têm a mesma distribuição espacial, como foi afirmado em Massey e Denton (1988, p. 292). Por fim, reafirmamos que o $RCE(\mathbf{X})$ não deve ser utilizado como medida de segregação, pois ele medirá a concentração dos grupos em relação ao centro, e não a segregação como desejado.

6.3. O índice proposto por Karmel e MacLachlan (1988)

O trabalho de Karmel e MacLachlan (1988) propõe uma "nova medida" de segregação por gênero. Adaptando a notação para a utilizada neste trabalho e admitindo que o grupo de interesse é o X, temos que o $I_p(\mathbf{X})$ proposto pelos autores é:

$$I_p(\mathbf{X}) = \frac{1}{T} \sum_{j=1}^{k} |PY_j - (1-P)X_j|. \tag{6.5}$$

Desenvolvendo os termos dentro da função modular e lembrando que $T_j = X_j + Y_j$, podemos encontrar a equação

$$I_p(\mathbf{X}) = \frac{\sum_{j=1}^{k} |PT_j - X_j|}{T}. \tag{6.6}$$

Observe que $PT_j = (X/T)T_j$ é a quantidade esperada de pessoas da categoria X, no j-ésimo estrato, caso não houvesse segregação. Deste modo, o numerador da expressão (6.6) é a soma das diferenças entre o valor esperado e o valor observado do número de pessoas da categoria X em cada estrato, desconsiderando o sinal da diferença. Comparando a equação (6.6) com o índice $I_8(\mathbf{X})$ apresentado em Jahn, Schmid e Schrag (1947, p. 147)[2] temos que $I_p(\mathbf{X}) = I_8(\mathbf{X})$. Além do trabalho de Karmel e MacLachlan (1988), os autores Walker, Stinchcombe e McDill (1967, p. 6) também haviam proposto esse mesmo índice, que denominaram de *replacement index* (conforme eq. (6.7) adiante).

Portanto, o novo índice proposto pelos autores em 1988 já havia sido descrito anteriormente na literatura sobre medidas de segregação. O $I_p(\mathbf{X})$ de Karmel e MacLachlan (1988) não é comum na literatura recente sobre segregação, porém os trabalhos de Watts (1998a, 1998b) recomendam seu uso em detrimento ao tradicional índice de Dissimilaridade, argumentando que o índice possui as propriedades desejáveis de equivalência organizacional, invariância de escala e simetria de gênero. Nota-se, ainda, que Karmel e MacLachlan (1988) e Watts (1998a, 1998b) não fazem referência ao trabalho de Jahn, Schmid e Schrag (1947) ou Walker, Stinchcombe e McDill (1967).

[2] $I_8 = \frac{\sum \textit{(observed number of Negroes- expected number of Negroes)}}{\sum \textit{total population of each census tract}} (\textit{disregard signs})$.

Esse índice apresenta, todavia, uma propriedade não desejável. Desenvolvendo a eq. (6.5), chega-se a

$$I_p(\mathbf{X}) = P(1-P)\sum_{j=1}^{k}\left|\frac{Y_j}{(1-P)T} - \frac{X_j}{PT}\right|.$$

Utilizando a definição (2.4) e a expressão para o índice de Dissimilaridade (eq. (2.38)) temos

$$I_p(\mathbf{X}) = 2P(1-P)D(\mathbf{X}). \tag{6.7}$$

A expressão $P(1-P)$ é uma função crescente para $P \in (0, 1/2)$ e decrescente para $P \in (1/2, 1)$. Portanto, em $P = 0,5$ temos um ponto de máximo (global). Assim, o valor do índice assumirá valor máximo, sob a hipótese de máxima segregação, quando $P = 0,5$ e este valor será $2P(1-P)$. A Figura 6.1 mostra como varia o valor da medida $I_p(\mathbf{X})$ de acordo com as proporções da categoria X na população total, considerando o valor 0,5 para o índice de Dissimilaridade.

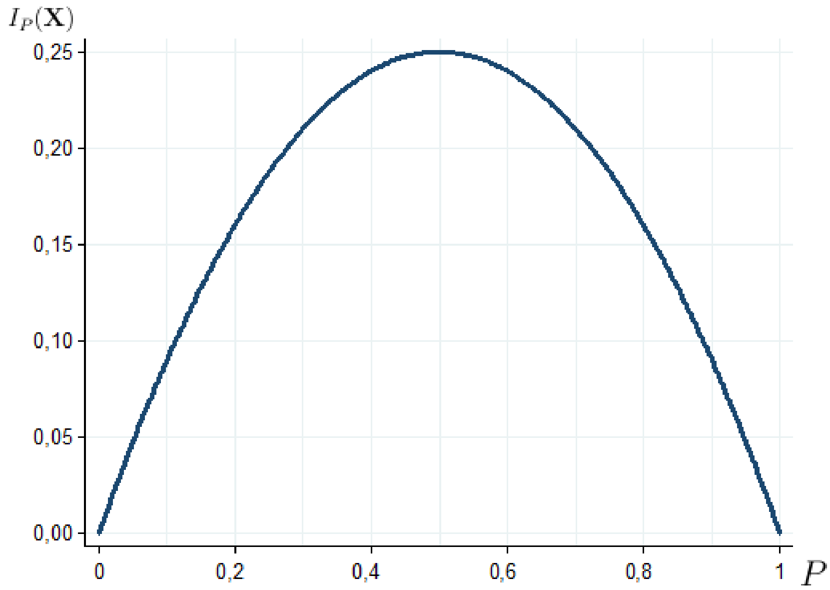

Figura 6.1. Variação do $I_p(\mathbf{X})$ conforme proporção das categorias para $D(\mathbf{X}) = 0,5$

A propriedade indesejável do índice é que o valor do $I_p(\mathbf{X})$ é sensível à proporção das pessoas de cada categoria. Por exemplo, no Brasil e na maior parte dos países, a participação das mulheres no mercado de trabalho vem crescendo ano após ano, embora permaneça abaixo de 50%. Deste modo, o valor do $I_p(\mathbf{X})$ crescerá, mesmo que seja constante o grau de segregação de acordo com o índice de Dissimilaridade.

Para ilustrar o problema com um exemplo real, a Tabela 6.2 apresenta a proporção de homens e mulheres no mercado de trabalho no Brasil, de acordo com os dados da PNAD, para os anos de 2002 a 2014. Os recortes da amostra seguem os apresentados no Capítulo 5.

Os resultados mostram que a participação das mulheres no mercado de trabalho cresceu 8,05%, ou 3,07 p.p. no período. Por outro lado, a segregação diminuiu ano após ano de 2002 a 2014, com exceção dos aumentos em 2012 e 2013, segundo o índice de Dissimilaridade. Note que no ano de 2006 a segregação era menor quando comparada a 2005, em 0,16 p.p. - ou $-0,39\%$ - utilizando o índice de Dissimilaridade, porém, devido ao aumento da participação das mulheres no mercado de trabalho, o índice $I_p(\mathbf{X})$ indica aumento da segregação em 2006, quando comparado a 2005, em 0,01 p.p. (ou $0,05\%$).

Tabela 6.2. Segregação de gênero no mercado de trabalho. Brasil, 2002 - 2014

Ano	Prop. de homens	Prop. de Mulheres	Dissimilaridade	$I_p(\mathbf{X})$
2002	0,6186	0,3814	0,4242	0,2001
2003	0,6164	0,3836	0,4170	0,1972
2004	0,6091	0,3909	0,4107	0,1956
2005	0,6079	0,3921	0,4056	0,1934
2006	0,6026	0,3974	0,4040	0,1935
2007	0,6009	0,3991	0,4022	0,1929
2008	0,5994	0,4006	0,3958	0,1901
2009	0,5944	0,4056	0,3933	0,1896
2011	0,5936	0,4064	0,3759	0,1814
2012	0,5911	0,4089	0,3779	0,1827
2013	0,5899	0,4101	0,3813	0,1845
2014	0,5879	0,4121	0,3760	0,1822

Fonte: elaboração própria de acordo com os microdados da PNAD.

Nota-se, ainda, que os próprios autores Jahn, Schmid e Schrag (1947, p. 294) enfatizam que uma das propriedades desejáveis de um índice de segregação é que essa medida não seja afetada pelo tamanho da população e, tampouco, pela proporção das categorias. Esta propriedade ao qual o índice não atende é a propriedade 1 (invariância de escala) apresentada no Capítulo 3, para o caso em que $\alpha^* \neq \beta^*$. Assim, concluímos que o $I_p(\mathbf{X})$ não deve ser utilizado como medida de segregação, em especial quando é analisada a segregação de gênero no mercado de trabalho.

O crescimento da participação das mulheres no mercado de trabalho é um fenômeno importante, mas é melhor diferenciá-lo de segregação. Não é conveniente misturar o crescimento dessa participação e mudanças no grau de segregação em um mesmo índice.

7. CONSIDERAÇÕES FINAIS

As relações entre medidas de desigualdade e segregação são conhecidas na literatura há algumas décadas. Esta dissertação buscou aprofundar a análise dessas relações. Ao abordar as medidas consagradas na literatura sobre segregação e desigualdade, no Capítulo 2, sempre que possível, foi feito o exercício de relacioná-las, para os índices de Gini e de Dissimilaridade. Além disso, a exposição da medida geral de desigualdade, ao final do capítulo, serviu como base para os principais resultados do capítulo posterior.

No Capítulo 3 foi abordado o desenvolvimento de uma medida geral de segregação aditivamente decomponível. Foi demonstrado que, para os valores 0 e 1 do parâmetro ε, é possível deduzir medidas de segregação análogas às medidas T e L de Theil de desigualdade. Uma delas, o T de Theil para segregação, já havia sido descrita por Hutchens (1991). Além disso, podem-se definir duas transformações monotônicas dessa medida geral, que resultam na família de medidas de Atkinson para segregação (JAMES; TAEUBER, 1985) e no coeficiente de variação para segregação de Hutchens (1991).

Outra contribuição desse capítulo é a análise da decomposição conforme valores de ε. Para uma medida de segregação aditivamente decomponível, temos que a segregação total (por estratos) pode ser decomposta em duas parcelas; a primeira referente à segregação entre os grupamentos e a segunda sendo uma soma ponderada das medidas de segregação dentro dos grupamentos. Foi demonstrado que ε igual a zero ou um é condição suficiente para afirmar que a parcela dentro dos grupamentos é uma média ponderada, ou seja, a soma dos fatores de ponderação é igual a um. Entretanto, caso a segregação entre os grupamentos seja igual a zero, a segregação total será igual a uma média ponderada das medidas de segregação dentro dos grupamentos (os pesos também somarão um). Para ε entre zero e um, a soma desses pesos será menor que um, enquanto que para ε maior que um ou menor que zero, essa soma excederá a unidade.

A terceira e última contribuição desse capítulo refere-se às propriedades das medidas de segregação. Foram enunciadas sete propriedades desejáveis e foi demonstrado que a medida geral de segregação atende a seis delas. Se considerarmos o caso especial do índice Raiz Quadrada, em que $\varepsilon = 0,5$, então todas as sete propriedades são atendidas. Embora a propriedade Imagem seja recorrentemente utilizada na literatura sobre desigualdade, este trabalho a incorporou no contexto da mensuração da segregação. Outro diferencial é que a propriedade de Mudanças entre os estratos foi enunciada de forma diferente da encontrada na literatura, o que permitiu a flexibilização de uma hipótese.

Embora a análise de sensibilidade das medidas de desigualdade seja conhecida há algumas décadas, não foram encontradas referências sobre análises similares no caso da segregação. O Capítulo 4 buscou preencher essa lacuna. Para a medida geral de segregação, foi constatado que a sensibilidade da medida a mudanças regressivas é função do nível da participação relativa dos estratos e essa sensibilidade se altera substancialmente conforme valores do parâmetro ε.

A medida geral de segregação é mais sensível a mudanças regressivas para valores altos da participação relativa x_i/y_i quanto menor for o valor do parâmetro ε. Por outro lado, para valores baixos da participação relativa, medidas que considerem valores de ε elevados serão mais sensíveis às mudanças regressivas, independentemente da categoria.

Para a análise de sensibilidade do índice de Gini para segregação, os resultados indicaram que o efeito de uma mudança regressiva em uma categoria só depende da participação nos estratos no total da outra categoria. Curiosamente, esse efeito será tanto maior quanto maior for a diferença entre as participações relativas, pois será dependente de todas as participações nos estratos intermediários da outra categoria aos estratos envolvidos na mudança. Para o índice de Dissimilaridade, como essa medida não atende à condição de mudanças entre estratos, os resultados indicam que somente algumas mudanças regressivas terão efeito sobre o valor da medida.

O Capítulo 5 teve como principal objetivo ilustrar a utilização das medidas de segregação apresentadas nos capítulos anteriores. Para isso, foram utilizados os dados da PNAD de 1992 a 2014, considerando, para o Brasil, a segregação por gênero em setores de atividade. Considerando os índices de Gini, de Dissimilaridade e o Raiz Quadrada, há indícios de que a segregação diminuiu no período. Ao analisar a decomposição da segregação, podemos concluir, pelo índice Raiz Quadrada, que essa queda na segregação é especialmente puxada pela segregação entre os grupamentos de atividade.

São três os resultados do Capítulo 6. O primeiro conclui que o limite superior para o índice de Gini para segregação pode ser um, pois é possível que para algum estrato a inclinação da curva de segregação seja vertical. Para o caso da desigualdade, considerando população finita, o limite superior para o índice de Gini é sempre menor que a unidade. O segundo resultado refere-se ao índice de Centralização Relativa. As conclusões indicam que essa medida é inapropriada para medir segregação. Esse índice somente deve ser utilizado como medida de concentração, na qual o objetivo é analisar a distribuição de duas categorias conforme o bairro esteja mais ou menos distante do centro de uma cidade, por exemplo. Por fim, o terceiro resultado indica que o índice proposto por Karmel e MacLachlan (1988) já havia sido proposto na literatura, pelo menos, duas vezes. Além disso, argumenta-se que essa não é uma boa medida para segregação para o mercado de trabalho, pois é afetada pelo crescimento da participação de mulheres nesse mercado.

REFERÊNCIAS

ALBELDA, R. P. Occupational segregation by race and gender, 1958-1981. **Industrial and Labor Relations Review**, v. 39, n. 3, p. 404-411, 1986.

ALLEN, R.; VIGNOLES, A. What should an index of school segregation measure? **Oxford Review of Education**, v. 33, n. 5, p. 643-668, Nov. 2007.

ALONSO-VILLAR, O.; RÍO, C. del. **An alternative proposal for measuring occupational segregation**. Palma de Mallorca, ECINEQ Working paper, Society for the Study of Economic Inequality, n. 82, p. 29, 2007.

_____. Local versus overall segregation measures. **Mathematical Social Sciences**, v. 60, n. 1, p. 30-38, July 2010.

ANKER, R. Theories of occupational segregation by sex: an overview. **International Labor Review**, v. 136, p. 315-339, 1997.

ATKINSON, A. B. On the measurement of inequality. **Journal of Economic Theory**, v. 2, n. 3, p. 244-263, Sept. 1970.

BARROS, R. P. de; FRANCO, S.; MENDONÇA, R. Discriminação e Segmentação no Mercado de Trabalho e Desigualdade de Renda no Brasil. In: BARROS, R. P. de; FOGUEL, M. N.; ULYSSEA, G. (Org.). **Desigualdade de Renda no Brasil**: uma análise da queda recente. Brasília: IPEA, 2007. p. 371-399, v. 2.

BLACKBURN, R. M.; SILTANEN, J.; JARMAN, J. The measurement of occupational gender segregation: current problems and a new approach. **Journal of the Royal Statistical Society. Series A (Statistics in Society)**, v. 158, n. 2, p. 319-331, 1995.

BOJANOWSKI, M.; CORTEN, R. Measuring segregation in social networks. **Social Networks**, v. 39, p. 14-32, 2014.

BOURGUIGNON, F. Decomposable income inequality measures. **Econometrica**, v. 47, n. 4, p. 901-920, July 1979.

BUTLER, R. J. New indices of segregation. **Economics Letters**, v. 24, n. 4, p. 359-362, 1987.

CARD, D.; MAS, A.; ROTHSTEIN, J. Tipping and the dynamics of segregation. **The Quarterly Journal of Economics**, v. 123, n. 1, p. 177-218, Feb. 2008.

CARVALHO, A. X. Y. de; LAURETO, C. R.; PENA, M. G.; ALBUQUERQUE, P. H. M.; JUNIOR, W. R. Um estudo das metodologias e funcionalidades dos índices de segregação. **Revista Brasileira de Estudos de População**, v. 30, n. 3, p. 567-594, jul./dez. 2013.

CHAKRAVARTY, S. R.; SILBER, J. Employment segregation indices: an axiomatic characterization. In: EICHHORN, W. (Ed.). **Models and Measurement of Welfare and Inequality**. Berlin, Heidelberg: Springer, 1994. p. 912-920.

_____. A generalized index of employment segregation. **Mathematical Social Sciences**, v. 53, n. 2, p. 185-195, Mar. 2007.

CHARLES, M.; GRUSKY, D. B. Models for describing the underlying structure of sex segregation. **American Journal of Sociology**, v. 100, n. 4, p. 931-971, Jan. 1995.

CORTESE, C. F.; FALK, R. F.; COHEN, J. K. Further considerations on the methodological analysis of segregation indices. **American Sociological Review**, v. 41, n. 4, p. 630-637, 1976.

COWELL, F. A. On the structure of additive inequality measures. **The Review of Economic Studies**, v. 47, n. 3, p. 521-531, Apr. 1980.

COWELL, F. A.; FLACHAIRE, E. Income distribution and inequality measurement: The problem of extreme values. **Journal of Econometrics**, v. 141, n. 2, p. 1044-1072, 2007.

COWELL, F. A.; KUGA, K. Additivity and the entropy concept: an axiomatic approach to inequality measurement. **Journal of Economic Theory**, v. 25, n. 1, p. 131-143, Aug. 1981a.

_____. Inequality measurement: an axiomatic approach. **European Economic Review**, v. 15, n. 3, p. 287-305, Mar. 1981b.

CURRARINI, S.; JACKSON, M. O.; PIN, P. An economic model of friendship: homophily, minorities, and segregation. **Econometrica**, v. 77, n. 4, p. 1003-1045, July 2009.

DUNCAN, O. D.; DUNCAN, B. A methodological analysis of segregation indexes. **American Sociological Review**, v. 20, n. 2, p. 210-217, Apr. 1955a.

_____. Residential distribution and occupational stratification. **American Journal of Sociology**, v. 60, n. 5, p. 493-503, Mar. 1955b.

ECHENIQUE, F.; FRYER JR., R. G. A measure of segregation based on social interactions. **The Quarterly Journal of Economics**, v. 122, n. 2, p. 441-485, May 2007.

ERRATA: Measures of Segregation. **Sociological Methodology**, v. 16, p. xi-xi, 1986.

FERNANDES, G. A. de A. L. **Os aspectos econômicos da discriminação racial no Brasil**. 2012. Tese (Doutorado em Teoria Econômica) — Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, São Paulo, 2012.

FOSTER, J. E. An axiomatic characterization of the Theil measure of income inequality. **Journal of Economic Theory**, v. 31, n. 1, p. 105-121, Oct. 1983.

_____. Inequality measurement. **Proceedings of Symposia in Applied Mathematics. Fair allocation**, v. 33, p. 31-68, 1985.

FRANKEL, D. M.; VOLIJ, O. Measuring school segregation. **Journal of Economic Theory**, v. 146, n. 1, p. 1-38, Jan. 2011.

FRESNEDA, B. Segregação ocupacional versus discriminação salarial por gênero no mercado de trabalho brasileiro - 2004. In: **XIII Congresso Brasileiro de Sociologia**. Recife: [s.n.], 2007.

FRYER JR., R. G. The importance of segregation, discrimination, peer dynamics, and identity in explaining trends in the racial achievement gap. In: BENHABIB, J.; BISIN, A.; JACKSON, M. O. (Ed.). **Handbook of Social Economics**. Amsterdam: North-Holland, 2011. v. 1, p. 1165-1191.

GASTWIRTH, J. L. A general definition of the Lorenz curve. **Econometrica**, v. 39, n. 6, p. 1037-1039, Nov. 1971.

GROSS, E. Plus ca change...? The sexual structure of occupations over time. **Social Problems**, v. 16, n. 2, p. 198-208, 1968.

HARRISON, R. Segregation indices. In: SMELSER, N. J.; BALTES, P. B. (Ed.). **International Encyclopedia of the Social & Behavioral Sciences**. Oxford: Pergamon, 2001. p. 13791-13795.

HOFFMANN, R. Sensibilidade das medidas de desigualdade a transferências regressivas. **Pesquisa e Planejamento Econonômico**, v. 22, n. 2, p. 289-304, ago. 1992.

_____. **Desigualdade de renda**: medidas de desigualdade e pobreza. São Paulo: Editora da Universidade de São Paulo, 1998.

_____. Uso e interpretação das razões de concentração e sua aplicação à análise da insegurança alimentar no Brasil. **Segurança Alimentar e Nutricional**, v. 21, n. 2, p. 481-498, 2014.

HUTCHENS, R. Segregation curves, Lorenz curves, and inequality in the distribution of people across occupations. **Mathematical Social Sciences**, v. 21, n. 1, p. 31-51, Feb. 1991.

_____. Numerical measures of segregation: desirable properties and their implications. **Mathematical Social Sciences**, v. 42, n. 1, p. 13-29, July 2001.

_____. One measure of segregation. **International Economic Review**, v. 45, n. 2, p. 555-578, May 2004.

_____. Symmetric measures of segregation, segregation curves, and Blackwell's criterion. **Mathematical Social Sciences**, v. 73, p. 63-68, Jan. 2015.

JAHN, J.; SCHMID, C. F.; SCHRAG, C. The measurement of ecological segregation. **American Sociological Review**, v. 12, n. 3, p. 293-303, June 1947.

JAMES, D. R.; TAEUBER, K. E. Measures of segregation. **Sociological Methodology**, v. 15, p. 1-32, 1985.

KAKWANI, N. C. **Income inequality and poverty: Methods of Estimation and Policy Applications**. Washington D.C.: Oxford University Press, 1980.

KARMEL, T.; MACLACHLAN, M. Occupational sex segregation - increasing or decreasing? **Economic Record**, v. 64, n. 3, p. 187-195, 1988.

KING, M. C. Occupational segregation by race and sex in Brazil, 1989-2001. **The Review of Black Political Economy**, v. 36, n. 2, p. 113-125, 2009.

MADALOZZO, R.; MARTINS, S.; LICO, M. R. **Segregação ocupacional e hiato salarial entre os gêneros**. São Paulo, Insper Working Paper, WPE: 357/2015, 24 p., 2015.

MASSEY, D. S.; DENTON, N. A. The dimensions of residential segregation. **Social Forces**, v. 67, n. 2, p. 281-315, 1988.

MORA, R.; RUIZ-CASTILLO, J. Additively decomposable segregation indexes. the case of gender segregation by occupations and human capital levels in spain. **The Journal of Economic Inequality**, v. 1, n. 2, p. 147-179, Aug. 2003.

MORRILL, R. L. On the measure of geographic segregation. **Geography Research Forum**, v. 11, p. 25-36, July 1991.

OLIVEIRA, A. M. H. C. de. **A segregação ocupacional por sexo no Brasil**. 1997. Dissertação (Mestrado em Demografia) — Faculdade de Ciências Econômicas, Universidade Federal de Minas Gerais, Belo Horizonte, 1997.

OMETTO, A. M. H.; HOFFMANN, R.; ALVES, M. C. A segregação por gênero no mercado de trabalho nos estados de São Paulo e Pernambuco. **Economia Aplicada**, v. 1, n. 3, p. 393-423, Jul. 1997.

OPPENHEIMER, V. K. **The Female Labor Force in The United States**: Demographic and Economic Factors Governing its Growth and Changing Composition. Berkeley: University of California Press, 1970.

PYATT, G.; CHEN, C.; FEI, J. The distribution of income by factor components. **The Quarterly Journal of Economics**, v. 95, n. 3, p. 451-473, 1980.

REARDON, S. F.; FIREBAUGH, G. Measures of multigroup segregation. **Sociological Methodology**, v. 32, n. 1, p. 33-67, 2002.

REARDON, S. F.; O'SULLIVAN, D. Measures of spatial segregation. **Sociological Methodology**, v. 34, n. 1, p. 121-162, 2004.

RIBEIRO, R.; ARAÚJO, G. S. Segregação ocupacional no mercado de trabalho segundo cor e nível de escolaridade no Brasil contemporâneo. **Nova Economia**, v. 26, n. 1, p. 147-177, 2016.

SALARDI, P. The evolution of gender and racial occupational segregation across formal and non-formal labor markets in brazil, 1987 to 2006. **Review of Income and Wealth**, v. 62, n. S1, p. S68-S89, Aug. 2016.

SALAS, C.; LEITE, M. Segregación sectorial por género: una comparación Brasil-México. **Cadernos PROLAM/USP**, v. 6, n. 11, p. 241-259, jul. 2007.

SHANNON, C. E. A mathematical theory of communication. **Bell System Techinical Journal**, v. 27, p. 379-423, July-Oct. 1948.

SHORROCKS, A. F. The class of additively decomposable inequality measures. **Econometrica**, v. 48, n. 3, p. 613-625, Apr. 1980.

_____. Inequality decomposition by population subgroups. **Econometrica**, v. 52, n. 6, p. 1369-1385, Nov. 1984.

SHORROCKS, A. F.; FOSTER, J. E. Transfer sensitive inequality measures. **The Review of Economic Studies**, v. 54, n. 3, p. 485-497, 1987.

SILBER, J. G. On the measurement of employment segregation. **Economics Letters**, v. 30, n. 3, p. 237-243, Sept. 1989.

SOARES, S.; IZAKI, R. S. **A participação feminina no mercado de trabalho**. Rio de Janeiro, Texto para discussão, IPEA, Instituto de Pesquisa Econômica Aplicada, n. 923, 2002.

TAEUBER, K. E.; TAEUBER, A. F. **Negroes in Cities**: Residential Segregation and Neighborhood Change. Chicago: Aldine Publishing, 1965.

THEIL, H. **Economics and Information Theory**. Chicago, Amsterdam: North-Holland, Rand McNally, 1967.

_____. **Statistical Decomposition Analysis**. Amsterdam: North Holland, 1972.

THEIL, H.; FINIZZA, A. J. A note on the measurement of racial integration of schools by means of informational concepts. **The Journal of Mathematical Sociology**, v. 1, n. 2, p. 187-193, July 1971.

VAZ, D. V.; HOFFMANN, R. Segregação ocupacional por sexo no setor público brasileiro no período 1995 e 2008. **Revista ABET**, v. 10, n. 1, p. 120-141, jan./jun. 2011.

WALKER, D.; STINCHCOMBE, A. L.; MCDILL, M. S. **School desegregation in Baltimore**. Baltimore, The Johns Hopkins University Report, ED013168, p. 52, 1967.

WATTS, M. Occupational gender segregation: index measurement and econometric modeling. **Demography**, v. 35, n. 4, p. 489-496, Nov. 1998a.

_____. The analysis of sex segregation: when is index measurement not index measurement? **Demography**, v. 35, n. 4, p. 505-508, Nov. 1998b.

WHITE, M. J. The measurement of spatial segregation. **American Journal of Sociology**, v. 88, n. 5, p. 1008-1018, 1983.

APÊNDICES

Apêndice A. Demonstrações e deduções

A.1. Demonstrações referentes à medida geral de segregação

Atendimento à propriedade 1 (Invariância de escala)

Temos que $\mathbf{\Gamma X} \Rightarrow \alpha^* X_j, \forall j$. Por sua vez, $\alpha^* X_j, \forall j \Rightarrow \alpha^* X$. Assim, $(\alpha^* X_j)/(\alpha^* X) = X_j/X = x_j$. De modo semelhante, temos $\mathbf{\Gamma X} \Rightarrow (\beta^* Y_j)/(\beta^* Y) = Y_j/Y = y_j$. Logo, as participações de cada estrato nas duas categorias não serão alteradas a partir da multiplicação de escala. Desse modo, a medida geral de segregação não alterará seu valor, ou seja, $I_S(\mathbf{\Gamma X}) = I_S(\mathbf{X})$. $\square$

Atendimento à propriedade 2 (Simetria nos estratos)

Da equação da medida geral de segregação, temos que alterações na ordenação do somatório não alteram o valor da medida. $\square$

Atendimento à propriedade 3 (Mudança entre estratos)

A demonstração ao atendimento dessa propriedade pela medida geral de segregação se dará em duas partes. A primeira, considerará mudanças regressivas na categoria X, pressupondo inalterada a distribuição da Y nos estratos. Na segunda, será considerada a situação inversa.

Considerando a primeira mudança regressiva, definida em (3.14), temos

$$I_S(\mathbf{X}^*) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{j \neq i,h} y_j \left(\frac{x_j}{y_j}\right)^{1-\varepsilon} - y_i \left(\frac{x_i - \theta}{y_i}\right)^{1-\varepsilon} - y_h \left(\frac{x_h + \theta}{y_h}\right)^{1-\varepsilon}\right].$$

Derivando essa expressão em relação a θ, considerando $\varepsilon \neq 1$, obtêm-se

$$\frac{\partial I_S(\mathbf{X}^*)}{\partial \theta} = \frac{1}{\varepsilon}\left[\left(\frac{x_i - \theta}{y_i}\right)^{-\varepsilon} - \left(\frac{x_h + \theta}{y_h}\right)^{-\varepsilon}\right].$$

Aplicando o limite quando esta mudança é infinitesimal, tem-se

$$\lim_{\theta \to 0} \frac{\partial I_S(\mathbf{X}^*)}{\partial \theta} = \frac{1}{\varepsilon}\left[\left(\frac{x_i}{y_i}\right)^{-\varepsilon} - \left(\frac{x_h}{y_h}\right)^{-\varepsilon}\right]. \tag{A.1}$$

Para uma mudança infinitesimal, com $x_i/y_i \leq x_h/y_h$, temos

$$\lim_{\theta \to 0} \frac{\partial I_S(\mathbf{X}^*)}{\partial \theta} \geq 0,$$

com igualdade válida apenas se $x_i/y_i = x_h/y_h$.

Entretanto, na prática o que ocorre são mudanças não infinitesimais. Matematicamente, uma mudança não infinitesimal pode ser pensada como uma sucessão de mudanças infinitesimais. Após a primeira dessas mudanças já teríamos $x_i/y_i < x_h/y_h$ e o efeito das mudanças seguintes seria estritamente positivo. Conclui-se que, para mudanças não infinitesimais o efeito na medida de segregação é positivo, isto é,

$$\frac{\Delta I_S(\mathbf{X}^*)}{\theta} > 0, \text{ se } \frac{x_i}{y_i} \leq \frac{x_h}{y_h}. \tag{A.2}$$

Para $\varepsilon = 1$, devemos utilizar o índice $L_S(\mathbf{X})$. Assim, temos que a mudança regressiva na categoria X resultará em

$$L_S(\mathbf{X}^*) = \sum_{j \neq i,h} y_j \ln\left(\frac{y_j}{x_j}\right) + y_i \ln\left(\frac{y_i}{x_i - \theta}\right) + y_h \ln\left(\frac{y_h}{x_h + \theta}\right).$$

Derivando essa expressão em relação a θ e considerando uma mudança infinitesimal, temos

$$\lim_{\theta \to 0} \frac{\partial L_S(\mathbf{X}^*)}{\partial \theta} = \frac{y_i}{x_i} - \frac{y_h}{x_h} \geq 0.$$

Repetindo o raciocínio feito a partir de (A.1), verifica-se que o resultado (A.2) é válido para o efeito de mudanças regressivas na categoria X sobre o $L_S(\mathbf{X})$.

Para concluir esta parte da demonstração, é necessário analisar a situação em que $\varepsilon = 0$. Assim, utilizando o índice T de Theil para segregação, temos

$$T_S(\mathbf{X}^*) = \sum_{j \neq i,h} x_j \ln\left(\frac{x_j}{y_j}\right) + (x_i - \theta) \ln\left(\frac{x_i - \theta}{y_i}\right) + (x_h + \theta) \ln\left(\frac{x_h + \theta}{y_h}\right).$$

Derivando esta expressão em relação a θ e aplicando o limite quando $\theta \to 0$, chega-se a

$$\lim_{\theta \to 0} \frac{\partial T_S(\mathbf{X}^*)}{\partial \theta} = -\ln\left(\frac{x_i}{y_i}\right) + \ln\left(\frac{x_h}{y_h}\right)$$

ou

$$\lim_{\theta \to 0} \frac{\partial T_S(\mathbf{X}^*)}{\partial \theta} = \ln\left(\frac{x_h/y_h}{x_i/y_i}\right) \geq 0.$$

Logo, para $\varepsilon = 0$ e mudanças regressivas não infinitesimais entre estratos na categoria X, chega-se, novamente, à eq. (A.2).

Para mudanças regressivas na categoria Y, consideraremos a mesma condição inicial (3.13). Assim, para que essa mudança resulte em maior diferença entre as participações relativas entre os dois estratos, a mudança deve consistir na transferência de θY pessoas do h-ésimo para o i-ésimo estrato, obtendo-se

$$\frac{x_i}{y_i + \theta} < \frac{x_h}{y_h - \theta}.$$

Desse modo, da equação da medida geral temos

$$I_S(\mathbf{Y}^*) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{j\neq i,h} y_j\left(\frac{x_j}{y_j}\right)^{1-\varepsilon} - (y_i+\theta)\left(\frac{x_i}{y_i+\theta}\right)^{1-\varepsilon} - (y_h-\theta)\left(\frac{x_h}{y_h-\theta}\right)^{1-\varepsilon}\right],$$

ou

$$I_S(\mathbf{Y}^*) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{j\neq i,h} y_j\left(\frac{x_j}{y_j}\right)^{1-\varepsilon} - (y_i+\theta)^{\varepsilon}(x_i)^{1-\varepsilon} - (y_h-\theta)^{\varepsilon}(x_h)^{1-\varepsilon}\right],$$

considerando $\varepsilon \neq 1$. Se $\varepsilon \neq 0$, derivando essa expressão em relação a θ, e utilizando uma mudança infinitesimal, temos

$$\lim_{\theta\to 0}\frac{\partial I_S(\mathbf{Y}^*)}{\partial\theta} = \frac{-1}{1-\varepsilon}\left[\left(\frac{x_i}{y_i}\right)^{1-\varepsilon} - \left(\frac{x_h}{y_h}\right)^{1-\varepsilon}\right].$$

Se $\varepsilon \in (-\infty, 1)$, então $(1-\varepsilon) > 0$. Logo, da condição (3.13), considerando $\theta \to 0$, teremos

$$\left(\frac{x_i}{y_i}\right)^{1-\varepsilon} \leq \left(\frac{x_h}{y_h}\right)^{1-\varepsilon}.$$

Para uma mudança θ positiva e não infinitesimal, chega-se a (A.2). Se $\varepsilon > 1$, temos que $(1-\varepsilon) < 0$, que implica $\big(x_i/y_i\big)^{1-\varepsilon} \geq \big(x_h/y_h\big)^{1-\varepsilon}$. Assim, para uma mudança regressiva não infinitesimal, mostra-se, novamente, a validade de (A.2).

Resta analisar os casos de $\varepsilon = 1$ ($L_S(\mathbf{Y}^*)$) e $\varepsilon = 0$ ($T_S(\mathbf{Y}^*)$). Dada a simetria entre esses dois índices, verifica-se que, para mudanças regressivas na categoria Y, conclui-se que para ambos é válido o resultado (A.2). Isso encerra a demonstração. $\square$

Atendimento à propriedade 4 (Insensibilidade a divisões proporcionais nos estratos)

Vamos considerar as suposições do enunciado

$$\begin{aligned} x_j^* = x_j,\, y_j^* = y_j, &\qquad \forall j \neq k \qquad \text{e} \\ x_h^* = x_k/m,\, y_h^* = y_k/m, &\qquad h = 1, ..., m. \end{aligned}$$

Assim, a medida geral, considerando os $k+m-1$ estratos, será

$$I_S(\mathbf{X}^*) = \frac{1}{\varepsilon(1-\varepsilon)}\left[1 - \sum_{j\neq k} y_j\left(\frac{x_j}{y_j}\right)^{1-\varepsilon} - \frac{1}{m}\sum_{h=1}^{m} y_k\left(\frac{x_k}{y_k}\right)^{1-\varepsilon}\right]. \tag{A.3}$$

Como os valores y_h são iguais para qualquer estrato nesse conjunto (divisão proporcional), o mesmo ocorrendo com os x_h, concluímos que

$$\frac{1}{m}\sum_{h=1}^{m} y_k\left(\frac{x_k}{y_k}\right)^{1-\varepsilon} = y_k\left(\frac{x_k}{y_k}\right)^{1-\varepsilon}.$$

Utilizando esse resultado em (A.3), concluímos que $I_S(\mathbf{X}^*) = I_S(\mathbf{X})$.

Observe que o mesmo resultado é válido para as medidas T e L de Theil para segregação, pois as participações relativas não são alteradas com a divisão ($x_h^*/y_h^* = x_k/y_k$), que implica $\ln(x_h^*/y_h^*) = \ln(x_k/y_k)$. Assim, para os m novos estratos teremos uma média (aritmética) das quantidades $\ln(x_h^*/y_h^*)$, que são todas iguais. Logo, chegamos ao mesmo valor da medida de segregação que antes da divisão proporcional. □

Atendimento à propriedade 5 (Decomposição aditiva)

Alguns autores preferem o termo agregativa ao invés de decomponível (HUTCHENS, 2004, p. 560; BOURGUIGNON, 1979, p. 903). Segundo Hutchens (2004) e Foster (1985), a condição de ser aditivamente decomponível é mais restrita que ser agregativa. Deste modo, toda medida aditivamente decomponível é agregativa (ou decomponível), mas o inverso não é válido. Foi demonstrado na seção 3.2 que a medida geral de segregação é aditivamente decomponível. Logo, a medida geral de segregação atende tanto à versão mais geral desta propriedade, quanto a sua versão mais restrita. □

Entretanto, a propriedade mais fraca de decomposição é atendida por uma classe maior de medidas que inclui, por exemplo, o índice de Gini. A definição de medida decomponível (fraca) é a seguinte.

Divida os k estratos em dois grupamentos mutuamente exclusivos e coletivamente exaustivos, com n_1 e n_2 estratos, cujas matrizes de distribuição são, respectivamente, $\mathbf{X}^1$ e $\mathbf{X}^2$. Segundo Shorrocks (1984, p. 1.371), utilizando uma adaptação similar à de Hutchens (2004, p. 562) para segregação, $\Omega(\mathbf{X})$ é agregativa se, e somente se, existe uma função (agregadora) $A(.)$, tal que

$$\Omega(\mathbf{X}^1, \mathbf{X}^2) = A[\Omega(\mathbf{X}^1), \boldsymbol{\xi}(\mathbf{X}^1), \Omega(\mathbf{X}^2), \boldsymbol{\xi}(\mathbf{X}^2)]$$

onde $A(.)$ é contínua e estritamente crescente em $\Omega(\mathbf{X}^1)$ e $\Omega(\mathbf{X}^2)$, e $\boldsymbol{\xi}(\mathbf{X}^g) = (\mu_S^g, 1/y_1, ..., 1/y_{ng})$ é um vetor de parâmetros referente ao g-ésimo grupamento. Foster (1985) demonstra, por exemplo, que o índice se Gini para desigualdade atende a essa forma de agregação. É claro que essa forma de decomposição pode ser expandida para mais de dois grupamentos sem perda de generalidade.

Atendimento à propriedade 6 (Simetria nas categorias)

Da seção 3.1 temos, para o j-ésimo estrato do índice T de Theil para segregação, que no $\lim_{x_j \to 0} x_j \ln(x_j/y_j) = 0$ e $y_j \to 0 \Rightarrow T_S(\mathbf{X}) \to \infty$. De modo semelhante, para o índice L de Theil para segregação, $\lim_{y_j \to 0} y_j \ln(y_j/x_j) = 0$ e $x_j \to 0 \Rightarrow L_S(\mathbf{X}) \to \infty$. Assim, utilizando como exemplo os casos em que $\varepsilon \in \{0, 1\}$, mostra-se que a medida geral de segregação não atende à essa propriedade. Por outro lado, vamos demonstrar que somente para o índice Raiz Quadrada, como caso particular da medida geral de segregação, essa propriedade é atendida.

Notemos que, sendo $\mathbf{P}_2$ uma matriz de permutação de ordem dois, se $I_S(\mathbf{X}) = I_S(\mathbf{P}_2\mathbf{X})$, considerando que existe pelo menos um estrato tal que $x_j \neq y_j$, teremos para cada estrato que

$$y_j^{\varepsilon} x_j^{1-\varepsilon} = x_j^{\varepsilon} y_j^{1-\varepsilon}, \forall j \Leftrightarrow \varepsilon = 0,5.$$

Assim, somente para o índice Raiz Quadrada esta propriedade é atendida. Outra maneira de abordar a questão é notar, inicialmente, com base na expressão da medida geral de segregação (eq. (3.1)), que $I_S(\mathbf{X};\varepsilon) = I_S(\mathbf{P}_2\mathbf{X}; 1-\varepsilon)$. É facil verificar que $\varepsilon = 0,5$ (índice Raiz Quadrada) é solução para a equação $\varepsilon = 1 - \varepsilon$ e, por isso, nesse caso considerar a categoria X ou a Y como a de interesse resulta no mesmo valor da medida (quando $\varepsilon = 0,5$ é indiferente utilizar a matriz de permutação). $\square$

Atendimento à propriedade 7 (Imagem)

Da equação da medida geral de segregação temos que, se $x_j = y_j \, \forall j$, então $I_S(\mathbf{X}) = 0$. Nessa situação, todas as participações relativas assumem o mesmo valor (um). Qualquer mudança de pessoas entre dois estratos acarretará em maior diferença nas participações relativas dos estratos envolvidos na mudança e, em consequência, teremos $I_S(\mathbf{X}^*) > 0$. Nessa situação, qualquer mudança corresponde a uma mudança regressiva pela Propriedade 3 (Mudança entre estratos). Logo, desses dois resultados, concluímos que $I_S(\mathbf{X}) \geq 0$. $\square$

A.2. Dedução de fórmulas especiais para o índice de Gini e a discrepância máxima com $n = 2$

Se $n = 2$, de acordo com a eq. (2.14), o índice de Gini será

$$G(\mathbf{z}) = \frac{|z_1 - z_1| + |z_1 - z_2| + |z_2 - z_1| + |z_2 - z_2|}{2^2 2\mu} \quad \text{ou}$$
$$G(\mathbf{z}) = \frac{|z_1 - z_2|}{4\mu}. \tag{A.4}$$

De (2.23), supondo dois indivíduos teremos

$$D_{max} = \frac{\frac{1}{2}[|z_1 - \mu| + |z_2 - \mu|]}{2\mu}. \tag{A.5}$$

Como $\mu = (z_1 + z_2)/2$, substituindo esse resultado nos valores absolutos do numerador de (A.5) temos

$$D_{max} = \frac{|z_1 - z_2|}{4\mu}. \tag{A.6}$$

Comparando (A.6) com (A.4), conclui-se que o índice de Gini coincidirá com a discrepância máxima quando o número de indivíduos for igual a dois.

A.3. Dedução de outra expressão para o RCE

Na definição do índice de Centralização Relativa (eq. 6.1), adicionando e subtraindo o termo $[\sum_{j=1}^{k}(\Psi_j\rho_j - \Psi_{j-1}\rho_{j-1})]$ do lado direito da expressão temos

$$RCE(\mathbf{X}) = \sum_{j=1}^{k}(\Psi_{j-1}\rho_j - \Psi_j\rho_{j-1}) + \sum_{j=1}^{k}(\Psi_j\rho_j - \Psi_{j-1}\rho_{j-1}) - \sum_{j=1}^{k}(\Psi_j\rho_j - \Psi_{j-1}\rho_{j-1}) \quad \text{(A.7)}$$

Como Ψ_j e ρ_j são proporções acumuladas, $\Psi_k = \rho_k = 1$, de modo que $\Psi_k\rho_k = 1$. Assim, $[\sum_{j=1}^{k}(\Psi_j\rho_j - \Psi_{j-1}\rho_{j-1})]$ é igual a $\Psi_k\rho_k$, ou 1. Então,

$$\begin{aligned}
RCE(\mathbf{X}) &= \sum_{j=1}^{k}(\Psi_{j-1}\rho_j - \Psi_j\rho_{j-1} + \Psi_j\rho_j - \Psi_{j-1}\rho_{j-1}) - 1 \\
&= \sum_{j=1}^{k}[\Psi_{j-1}(\rho_j - \rho_{j-1}) + \Psi_j(\rho_j - \rho_{j-1})] - 1 \\
RCE(\mathbf{X}) &= \sum_{j=1}^{k}(\Psi_{j-1} + \Psi_j)(\rho_j - \rho_{j-1}) - 1. \qquad \text{(A.8)}
\end{aligned}$$

A expressão (A.8) é a mesma que (6.3). De modo semelhante, tomando a eq. (A.7) e substituindo $[\sum_{j=1}^{k}(\Psi_j\rho_j - \Psi_{j-1}\rho_{j-1})]$ por 1,

$$\begin{aligned}
RCE(\mathbf{X}) &= \sum_{j=1}^{k}(\Psi_{j-1}\rho_j - \Psi_j\rho_{j-1}) + 1 - \sum_{j=1}^{k}(\Psi_j\rho_j - \Psi_{j-1}\rho_{j-1}) \\
&= 1 + \sum_{j=1}^{k}(\Psi_{j-1}\rho_j - \Psi_j\rho_{j-1} - \Psi_j\rho_j + \Psi_{j-1}\rho_{j-1}) \\
&= 1 + \sum_{j=1}^{k}[\Psi_{j-1}(\rho_j + \rho_{j-1}) - \Psi_j(\rho_j + \rho_{j-1})] \\
&= 1 + \sum_{j=1}^{k}(\Psi_{j-1} - \Psi_j)(\rho_j + \rho_{j-1}) \\
RCE(\mathbf{X}) &= 1 - \sum_{j=1}^{k}(\Psi_j - \Psi_{j-1})(\rho_j + \rho_{j-1}). \qquad \text{(A.9)}
\end{aligned}$$

A eq. (A.9) é a mesma que a eq. (6.2).

Apêndice B. Tabelas

Tabela B.1. Classificação das atividades em grupamentos e estratos

Grupamento	Cód. V9907	Estrato
Agrícola	1101 a 1500	Agricultura e pecuária
	2001 e 2002	Silvicultura e exploração florestal
	5001 e 5002	Pesca e aquicultura
Indústria	15010 a 15055	Produtos alimentícios e bebidas
	16000	Produtos do fumo
	17001 e 17002	Produtos têxteis
	18001 e 18002	Vestuário e acessórios
	19011 a 19020	Artefatos de couro e similares
	20000	Produtos de madeira
	21001 e 21002	Celulose, papel e produtos de papel
	22000	Edição, impressão e gravações
	23010 a 23400	Petróleo, álcool e outros
	24010 a 24090	Produtos químicos
	25010 e 25020	Artigos de borracha e plástico
	26010 a 26092	Produtos de minerais não-metálicos
	27001 a 27003	Metalurgia básica
	28001 e 28002	Metal - exclusive máquinas e equipamentos
	29001 e 29002	Máquinas e equipamentos
	30000	Máquinas para escritório e de informática
	31001 e 31002	Máquinas, aparelhos e materiais elétricos
	32000	Material eletrônico e similares
	33001 a 33005	Instrumentação diversa
	34001 a 34003	Montagem de veículos automotores
	35010 a 35090	Outros equipamentos de transporte
	36010 e 36090	Móveis e indústrias diversas
	37000	Reciclagem
	10000 e 11000	Extr. de petróleo, gás natural e similares
	12000 a 13002	Extr. de minerais metálicos
	14001 a 14004	Extr. de minerais não-metálicos
	40010 e 40020	Eletricidade, gás e água quente
	41000	Captação, tratamento e distribuição de água
	45005 e 45999	Construção
Administração pública	75011 e 75015	Administração pública federal
	75012 e 75016	Administração pública estadual
	75013 e 75017	Administração pública municipal
	75014	Defesa
	75020	Seguridade social
Comércio e reparação	50010 a 50050	Veículos e combustíveis
	53010 a 53113	Atacado e varejo
Alojamento, transporte e similares	55010	Alojamento
	55020 e 55030	Alimentação
	60010 a 60092	Transporte terrestre
	61000	Transporte aquaviário
	62000	Transporte aéreo
	63010 a 63030	Auxiliares do transporte
	64010 e 64020	Correios e telecomunicações
Demais serviços	80011 a 80090	Educação
	85011 a 85030	Saúde e serviços sociais
	90000	Limpeza urbana e esgoto
	91010 a 91092	Atividades associativas
	92011 a 92040	Atividades recreativas
	93010 a 93092	Serviços pessoais
	95000	Serviços domésticos

Fonte: elaboração própria com base na variável V9907 da PNAD.

Tabela B.2. População total e porcentagem de mulheres em sete grupamentos de atividade principal. Brasil, 1992 - 2014

	Agrícola		Indústria		Construção civil		Comércio e rep.		Aloj., transp. etc.		Adm. pública		Demais serviços	
Ano	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
1992	9.432.366	12,30	9.473.467	32,06	3.884.203	2,08	8.408.421	27,20	4.406.662	22,86	3.021.786	32,99	10.337.809	80,69
1993	9.289.632	11,51	9.638.722	32,39	4.172.409	3,00	8.822.250	28,21	4.391.022	21,86	3.050.287	32,61	10.667.686	80,07
1995	9.380.732	10,98	9.664.268	32,29	4.103.650	1,78	9.575.185	29,63	4.999.874	23,78	3.214.182	34,19	11.802.470	80,21
1996	8.696.353	10,24	9.437.199	32,69	4.171.567	1,70	9.743.658	29,57	4.936.740	24,83	3.177.047	34,86	11.799.767	79,75
1997	9.010.789	10,55	9.433.589	32,63	4.452.764	1,65	9.702.021	30,21	5.289.401	23,78	3.162.941	34,65	12.022.028	79,33
1998	8.632.718	10,75	9.176.314	32,47	4.822.128	3,08	9.840.994	30,92	5.293.866	25,29	3.183.554	33,94	12.290.093	78,87
1999	9.273.395	11,40	9.351.765	33,32	4.817.986	3,18	10.345.437	30,94	5.460.958	25,34	3.332.481	34,26	12.973.115	79,11
2001	8.661.728	10,42	10.282.659	33,19	4.931.058	2,04	11.509.576	33,27	6.137.376	25,55	3.640.115	36,24	14.109.946	79,25
2002	8.738.798	10,64	10.478.056	33,54	5.308.079	2,24	12.070.158	34,24	6.044.270	25,68	3.782.767	35,04	14.838.332	79,45
2003	8.926.680	10,20	10.706.866	33,89	4.951.996	1,84	12.586.824	34,51	5.991.671	25,79	3.898.269	36,82	14.793.186	79,59
2004	9.063.255	11,14	11.485.340	33,98	5.027.329	2,18	12.982.372	35,48	6.252.321	26,58	4.055.166	37,04	15.641.992	79,41
2005	8.998.743	10,25	11.733.796	34,47	5.288.883	2,25	13.747.491	35,51	6.471.395	27,99	4.119.724	38,27	15.738.821	79,82
2006	8.772.323	10,80	12.058.363	34,32	5.434.709	2,62	13.835.235	35,98	6.709.144	28,87	4.276.270	37,87	16.491.826	79,38
2007	8.323.286	11,02	12.664.800	33,81	5.619.262	2,29	14.378.084	36,46	6.954.522	28,58	4.274.081	38,28	16.670.125	79,77
2008	8.367.161	11,23	13.031.918	34,66	6.485.983	2,99	14.413.146	37,20	7.468.106	29,34	4.362.500	38,80	17.267.433	79,35
2009	8.323.290	11,04	12.709.771	35,33	6.507.202	2,43	14.810.279	37,95	7.338.632	29,70	4.589.358	39,49	17.798.321	78,96
2011	7.607.855	13,21	11.611.586	34,57	7.251.031	2,24	15.015.148	40,12	8.786.183	30,92	4.806.143	40,62	16.718.540	79,59
2012	7.272.935	11,97	12.267.022	35,03	7.808.938	2,68	15.302.292	40,06	8.987.700	31,72	4.963.771	41,34	17.327.326	79,73
2013	7.016.731	11,75	11.756.340	34,83	8.177.547	2,94	15.324.937	40,53	8.868.577	31,87	5.020.424	40,12	17.801.037	79,34
2014	7.118.046	12,41	11.973.808	35,56	8.484.655	2,93	16.270.957	39,99	9.187.482	32,60	4.905.428	42,16	18.516.908	78,57

Fonte: elaboração própria de acordo com os microdados da PNAD.

Tabela B.3. Participações (x_j para mulheres e y_j para homens) de grupamentos de atividade nos totais de mulheres e homens ocupados em cada ano (em %). Brasil, 1992 - 2014

	Agrícola		Indústria		Const. civil		Comércio e rep.		Aloj., transp. etc.		Adm. pública		Demais serviços	
Ano	x_j	y_j	x_j	y_j	x_j	y_j	x_j	y_j	x_j	y_j	x_j	y_j	x_j	y_j
1992	6,86	25,81	17,96	20,08	0,48	11,87	13,53	19,10	5,96	10,60	5,89	6,32	49,33	6,23
1993	6,18	25,12	18,05	19,91	0,72	12,37	14,39	19,35	5,55	10,48	5,75	6,28	49,37	6,50
1995	5,47	24,62	16,58	19,29	0,39	11,88	15,08	19,86	6,32	11,23	5,84	6,24	50,32	6,88
1996	4,77	23,45	16,52	19,08	0,38	12,32	15,43	20,61	6,56	11,15	5,93	6,22	50,40	7,18
1997	5,02	23,60	16,27	18,61	0,39	12,82	15,49	19,83	6,65	11,81	5,79	6,05	50,40	7,28
1998	4,83	22,64	15,51	18,21	0,77	13,73	15,84	19,98	6,97	11,62	5,62	6,18	50,45	7,63
1999	5,20	23,32	15,34	17,70	0,75	13,24	15,76	20,27	6,81	11,57	5,62	6,22	50,52	7,69
2001	4,04	20,99	15,29	18,59	0,45	13,07	17,16	20,78	7,03	12,36	5,91	6,28	50,11	7,92
2002	3,98	20,61	15,04	18,37	0,51	13,69	17,69	20,95	6,64	11,85	5,67	6,48	50,46	8,05
2003	3,84	21,02	15,29	18,56	0,38	12,75	18,31	21,62	6,51	11,66	6,05	6,46	49,62	7,92
2004	4,00	20,50	15,48	19,30	0,43	12,52	18,27	21,32	6,59	11,68	5,96	6,50	49,27	8,19
2005	3,56	20,10	15,61	19,14	0,46	12,87	18,84	22,06	6,99	11,60	6,08	6,33	48,47	7,91
2006	3,53	19,21	15,41	19,45	0,53	13,00	18,53	21,75	7,21	11,72	6,03	6,52	48,75	8,35
2007	3,34	17,89	15,58	20,25	0,47	13,26	19,07	22,07	7,23	12,00	5,95	6,37	48,37	8,15
2008	3,29	17,36	15,79	19,90	0,68	14,70	18,75	21,15	7,66	12,33	5,92	6,24	47,91	8,33
2009	3,14	17,28	15,36	19,18	0,54	14,82	19,23	21,45	7,46	12,04	6,20	6,48	48,07	8,74
2011	3,44	15,49	13,76	17,83	0,56	16,63	20,64	21,10	9,31	14,24	6,69	6,70	45,60	8,01
2012	2,88	14,65	14,22	18,24	0,69	17,39	20,28	20,99	9,43	14,04	6,79	6,66	45,70	8,04
2013	2,72	14,19	13,50	17,56	0,79	18,19	20,48	20,89	9,32	13,85	6,64	6,89	46,56	8,43
2014	2,80	13,87	13,51	17,17	0,79	18,32	20,65	21,72	9,51	13,78	6,56	6,31	46,18	8,83

Fonte: elaboração própria de acordo com os microdados da PNAD.

Tabela B.4. Proporção de homens e mulheres e evolução da segregação de gênero no mercado de trabalho para sete grupamentos de atividade. Brasil, 1992 - 2014

Ano	Prop. homens	Prop. mulheres	$G_S(\mathbf{X})$	$D(\mathbf{X})$	$T_S(\mathbf{X})$	$I_S(\mathbf{X}; \varepsilon = 0,5)$	$L_S(\mathbf{X})$
1992	0,6546	0,3454	0,5828	0,4310	0,8093	0,7069	0,7478
1993	0,6542	0,3458	0,5830	0,4287	0,7933	0,6948	0,7208
1995	0,6432	0,3568	0,5870	0,4343	0,7985	0,7193	0,7922
1996	0,6407	0,3593	0,5872	0,4322	0,7871	0,7210	0,8113
1997	0,6434	0,3566	0,5885	0,4312	0,7831	0,7212	0,8175
1998	0,6391	0,3609	0,5778	0,4282	0,7533	0,6866	0,7414
1999	0,6343	0,3657	0,5768	0,4283	0,7478	0,6791	0,7281
2001	0,6235	0,3765	0,5731	0,4219	0,7366	0,6909	0,7892
2002	0,6186	0,3814	0,5744	0,4242	0,7383	0,6936	0,7913
2003	0,6164	0,3836	0,5717	0,4170	0,7297	0,6908	0,8027
2004	0,6091	0,3909	0,5614	0,4107	0,6986	0,6602	0,7563
2005	0,6079	0,3921	0,5608	0,4056	0,7025	0,6664	0,7687
2006	0,6026	0,3974	0,5531	0,4040	0,6782	0,6443	0,7362
2007	0,6009	0,3991	0,5526	0,4022	0,6804	0,6473	0,7495
2008	0,5994	0,4006	0,5504	0,3958	0,6640	0,6350	0,7285
2009	0,5944	0,4056	0,5508	0,3933	0,6544	0,6405	0,7631
2011	0,5936	0,4064	0,5456	0,3759	0,6427	0,6322	0,7700
2012	0,5911	0,4089	0,5472	0,3779	0,6467	0,6360	0,7659
2013	0,5899	0,4101	0,5503	0,3813	0,6468	0,6391	0,7680
2014	0,5879	0,4121	0,5413	0,3760	0,6189	0,6194	0,7531

Fonte: elaboração própria de acordo com os microdados da PNAD.

Tabela B.5. Evolução da medida geral de segregação para valores de ε de -1 a 2, considerando sete grupamentos de atividade. Brasil, 1992 - 2014

	Valor de ε														
Ano	-1	$-0,5$	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	1,5	2
1992	1,6349	1,0796	0,8093	0,7774	0,7514	0,7311	0,7163	0,7069	0,7029	0,7045	0,7122	0,7264	0,7478	1,0128	1,8354
1993	1,5597	1,0460	0,7933	0,7631	0,7385	0,7190	0,7045	0,6948	0,6899	0,6898	0,6947	0,7049	0,7208	0,9129	1,4459
1995	1,5187	1,0345	0,7985	0,7715	0,7502	0,7344	0,7241	0,7193	0,7202	0,7273	0,7411	0,7623	0,7922	1,1399	2,2527
1996	1,4510	1,0043	0,7871	0,7629	0,7442	0,7310	0,7232	0,7210	0,7247	0,7348	0,7519	0,7770	0,8113	1,2016	2,4570
1997	1,4287	0,9942	0,7831	0,7598	0,7419	0,7295	0,7225	0,7212	0,7258	0,7368	0,7551	0,7815	0,8175	1,2255	2,5484
1998	1,3480	0,9507	0,7533	0,7306	0,7127	0,6995	0,6908	0,6866	0,6871	0,6925	0,7030	0,7191	0,7414	0,9798	1,6184
1999	1,3383	0,9446	0,7478	0,7249	0,7067	0,6931	0,6839	0,6791	0,6789	0,6833	0,6927	0,7074	0,7281	0,9522	1,5542
2001	1,2705	0,9119	0,7366	0,7177	0,7036	0,6944	0,6901	0,6909	0,6971	0,7093	0,7280	0,7543	0,7892	1,1673	2,3246
2002	1,2660	0,9116	0,7383	0,7196	0,7058	0,6968	0,6927	0,6936	0,6999	0,7121	0,7308	0,7568	0,7913	1,1596	2,2603
2003	1,2447	0,8979	0,7297	0,7120	0,6992	0,6913	0,6884	0,6908	0,6988	0,7131	0,7344	0,7638	0,8027	1,2266	2,5737
2004	1,1713	0,8545	0,6986	0,6819	0,6697	0,6619	0,6587	0,6602	0,6668	0,6789	0,6972	0,7227	0,7563	1,1176	2,2188
2005	1,1831	0,8596	0,7025	0,6860	0,6741	0,6668	0,6642	0,6664	0,6739	0,6871	0,7066	0,7335	0,7687	1,1408	2,2539
2006	1,1165	0,8226	0,6782	0,6629	0,6519	0,6450	0,6424	0,6443	0,6510	0,6629	0,6805	0,7046	0,7362	1,0642	2,0065
2007	1,1308	0,8284	0,6804	0,6649	0,6537	0,6470	0,6447	0,6473	0,6549	0,6681	0,6876	0,7143	0,7495	1,1234	2,2574
2008	1,0786	0,8000	0,6640	0,6499	0,6399	0,6340	0,6323	0,6350	0,6423	0,6547	0,6727	0,6970	0,7285	1,0480	1,9336
2009	1,0253	0,7742	0,6544	0,6431	0,6359	0,6330	0,6344	0,6405	0,6518	0,6688	0,6922	0,7232	0,7631	1,1719	2,3858
2011	1,0203	0,7635	0,6427	0,6317	0,6249	0,6226	0,6249	0,6322	0,6450	0,6642	0,6905	0,7253	0,7700	1,2404	2,7035
2012	1,0223	0,7665	0,6467	0,6358	0,6291	0,6268	0,6290	0,6360	0,6483	0,6665	0,6915	0,7242	0,7659	1,1894	2,4221
2013	1,0038	0,7604	0,6468	0,6367	0,6308	0,6290	0,6317	0,6391	0,6517	0,6700	0,6948	0,7270	0,7680	1,1749	2,3216
2014	0,9289	0,7166	0,6189	0,6109	0,6068	0,6066	0,6107	0,6194	0,6330	0,6524	0,6781	0,7112	0,7531	1,1666	2,3363

Fonte: elaboração própria de acordo com os microdados da PNAD.

Tabela B.6. Correlação entre valores da medida geral para diversos valores de ε, considerando sete grupamentos de atividade. Brasil, 1992 - 2014

	Valor de ε														
	-1	$-0,5$	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1	1,5	2
-1	1														
$-0,5$	0,9979*	1													
0	0,9873*	0,9953*	1												
0,1	0,9824*	0,9919*	0,9995*	1											
0,2	0,9752*	0,9866*	0,9977*	0,9993*	1										
0,3	0,9642*	0,9778*	0,9931*	0,9962*	0,9988*	1									
0,4	0,9466*	0,9628*	0,9834*	0,9885*	0,9935*	0,9979*	1								
0,5	0,9169*	0,9362*	0,9637*	0,9712*	0,9795*	0,9882*	0,9960*	1							
0,6	0,8647*	0,8877*	0,9241*	0,9349*	0,9474*	0,9619*	0,9776*	0,9924*	1						
0,7	0,7716*	0,7988*	0,8460*	0,8611*	0,8791*	0,9011*	0,9272*	0,9568*	0,9852*	1					
0,8	0,6156*	0,6467*	0,7053*	0,7252*	0,7496*	0,7804*	0,8189*	0,8664*	0,9211*	0,9741*	1				
0,9	0,3910	0,4240	0,4917*	0,5158*	0,5460*	0,5851*	0,6356*	0,7010*	0,7829*	0,8777*	0,9633*	1			
1	0,1402	0,1724	0,2436	0,2700	0,3035	0,3478	0,4064	0,4848*	0,5878*	0,7172*	0,8560*	0,9633*	1		
1,5	$-0,4268$	$-0,4086$	$-0,3539$	$-0,3309$	$-0,3005$	$-0,2584$	$-0,1998$	$-0,1162$	0,0026	0,1701	0,3855	0,6169*	0,8040*	1	
2	$-0,4333$	$-0,4191$	$-0,3710$	$-0,3499$	$-0,3218$	$-0,2823$	$-0,2268$	$-0,1469$	$-0,0322$	0,1312	0,3439	0,5760*	0,7677*	0,9935*	1

Fonte: elaboração própria de acordo com os microdados da PNAD.

Notas: número de observações igual a 20; asterisco indica que a correlação é significativa a 5%.

Tabela B.7. Participações relativas (x_j/y_j) para sete grupamentos de atividade[1]. Brasil, 1992 - 2014

Ano	Agrícola	Indústria	Const. civil	Comércio e rep.	Aloj., trans. etc.	Adm. pública	Demais serviços
1992	0,2658	0,8943	0,0403	0,7083	0,5618	0,9329	7,9199
1993	0,2460	0,9064	0,0585	0,7434	0,5294	0,9157	7,5985
1995	0,2224	0,8597	0,0327	0,7591	0,5627	0,9365	7,3093
1996	0,2034	0,8661	0,0308	0,7487	0,5889	0,9542	7,0208
1997	0,2128	0,8741	0,0303	0,7811	0,5630	0,9569	6,9260
1998	0,2134	0,8516	0,0564	0,7929	0,5995	0,9101	6,6096
1999	0,2232	0,8667	0,0570	0,7772	0,5888	0,9037	6,5687
2001	0,1927	0,8228	0,0345	0,8256	0,5685	0,9412	6,3256
2002	0,1932	0,8187	0,0372	0,8445	0,5606	0,8752	6,2716
2003	0,1825	0,8237	0,0302	0,8468	0,5584	0,9365	6,2645
2004	0,1954	0,8020	0,0347	0,8569	0,5642	0,9168	6,0124
2005	0,1771	0,8156	0,0357	0,8537	0,6027	0,9610	6,1301
2006	0,1837	0,7925	0,0407	0,8521	0,6156	0,9244	5,8380
2007	0,1864	0,7691	0,0353	0,8641	0,6024	0,9340	5,9364
2008	0,1893	0,7939	0,0462	0,8866	0,6212	0,9489	5,7509
2009	0,1818	0,8008	0,0365	0,8964	0,6193	0,9564	5,4998
2011	0,2223	0,7716	0,0335	0,9784	0,6536	0,9990	5,6952
2012	0,1966	0,7797	0,0399	0,9665	0,6718	1,0188	5,6872
2013	0,1915	0,7687	0,0436	0,9803	0,6728	0,9638	5,5222
2014	0,2021	0,7872	0,0430	0,9506	0,6900	1,0399	5,2308

Fonte: elaboração própria de acordo com os microdados da PNAD.
[1] x_j é a participação do j-ésimo grupamento no total de mulheres e y_j é a participação do j-ésimo grupamento no total de homens.

Tabela B.8. Valor da medida geral de segregação de gênero para 53 atividades, conforme valores de ε. Brasil, 2002 - 2014

	Valor de ε				
Ano	0	0,1	0,5	0,9	1
2002	1,1168	1,0676	0,9809	1,0506	1,0951
2003	1,1033	1,0545	0,9713	1,0498	1,0984
2004	1,0778	1,0298	0,9440	1,0076	1,0492
2005	1,0743	1,0277	0,9462	1,0130	1,0553
2006	1,0438	0,9996	0,9213	0,9821	1,0209
2007	1,0541	1,0081	0,9280	0,9943	1,0363
2008	1,0294	0,9872	0,9147	0,9772	1,0158
2009	1,0072	0,9688	0,9088	0,9881	1,0332
2011	0,9729	0,9412	0,9039	1,0104	1,0652
2012	0,9580	0,9301	0,9011	1,0052	1,0568
2013	0,9732	0,9443	0,9124	1,0130	1,0632
2014	0,9235	0,8993	0,8778	0,9805	1,0304

Fonte: com base nos microdados da PNAD.

Tabela B.9. Valor da medida geral de segregação de gênero entre os seis grupamentos, conforme valores de ε (parcela entre os grupamentos). Brasil, 2002 - 2014

	Valor de ε				
Ano	0	0,1	0,5	0,9	1
2002	0,6726	0,6458	0,5710	0,5383	0,5358
2003	0,6643	0,6384	0,5671	0,5384	0,5371
2004	0,6436	0,6192	0,5518	0,5242	0,5228
2005	0,6453	0,6209	0,5545	0,5298	0,5295
2006	0,6219	0,5991	0,5366	0,5121	0,5113
2007	0,6236	0,6000	0,5344	0,5069	0,5053
2008	0,6055	0,5833	0,5218	0,4963	0,4949
2009	0,5921	0,5719	0,5168	0,4964	0,4962
2011	0,5783	0,5574	0,4985	0,4719	0,4695
2012	0,5794	0,5591	0,5026	0,4793	0,4780
2013	0,5821	0,5626	0,5092	0,4887	0,4882
2014	0,5517	0,5342	0,4858	0,4670	0,4664

Fonte: com base nos microdados da PNAD.

Tabela B.10. Participação da parcela "entre os grupamentos" na segregação de gênero para 53 atividades divididos em seis grupamentos, conforme valores de ε. Brasil, 2002 - 2014

	Valor de ε				
Ano	0	0,1	0,5	0,9	1
2002	60,22	60,49	58,21	51,24	48,92
2003	60,21	60,54	58,39	51,28	48,89
2004	59,72	60,13	58,45	52,03	49,83
2005	60,07	60,42	58,60	52,30	50,18
2006	59,57	59,94	58,24	52,15	50,08
2007	59,16	59,52	57,59	50,98	48,76
2008	58,82	59,09	57,05	50,79	48,73
2009	58,78	59,03	56,86	50,24	48,03
2011	59,43	59,22	55,15	46,70	44,08
2012	60,48	60,10	55,77	47,68	45,23
2013	59,81	59,58	55,80	48,24	45,92
2014	59,74	59,40	55,34	47,63	45,27

Fonte: com base nos microdados da PNAD.