

Estimação da desigualdade dentro de estratos no cálculo do índice de Gini e da redundância *

RODOLFO HOFFMANN **

1 — Sumário

Quando calculamos medidas do grau de concentração ou desigualdade para uma distribuição da renda ou da riqueza e dispomos apenas dos dados referentes ao número de elementos em cada estrato e a correspondente renda ou riqueza, surge o problema de estimar a desigualdade dentro dos estratos. O cálculo de medidas de concentração desprezando a desigualdade dentro dos estratos pode levar a uma séria subestimação do grau de desigualdade real. Neste trabalho apresentamos um método geral de estimar a desigualdade dentro dos estratos quando se calcula o índice de Gini ou a redundância. Para os estratos com limites finitos, pressupomos que a distribuição dentro do estrato tem função de densidade linear e dentro do estrato de rendas mais altas, quando ele é aberto à direita, pressupomos que a distribuição é a de Pareto com dois parâmetros. Também é analisado um outro método, baseado na obtenção de um limite inferior (admitindo perfeita igualdade dentro dos estratos) e de um limite superior (supondo um máximo de desigualdade dentro dos estratos) para o valor do índice de Gini ou da redundância.

* Agradeço os comentários críticos que o corpo editorial de *Pesquisa e Planejamento Económico* fez a uma versão anterior deste trabalho.

** Professor do Departamento de Economia e Sociologia Rural da ESALQ-USP.

2 — Introdução

Os dados disponíveis para o cálculo de medidas de concentração da distribuição da renda ou da riqueza em uma população consistem, freqüentemente, no número de pessoas e nos totais de renda ou riqueza referentes a um certo número de estratos, e não dispomos (ou são de difícil acesso) das informações referentes à renda ou riqueza de cada indivíduo. É comum, nesses casos, calcular-se a medida de concentração referente à desigualdade entre estratos, desprezando-se a desigualdade dentro destes. Esse procedimento pode levar a uma séria subestimação do grau de desigualdade real.

Consideremos a distribuição da renda em uma população dividida em k estratos. Seja n_h ($h = 1, \dots, k$) o número de elementos no h -ésimo estrato e seja x_{hi} ($h = 1, \dots, k; i = 1, \dots, n_h$) a renda recebida pelo i -ésimo elemento do h -ésimo estrato. O número total de elementos na população é:

$$N = \sum_{h=1}^k n_h$$

Se a renda média da população é μ , a fração da renda total apropriada pelo i -ésimo elemento do h -ésimo estrato é:

$$y_{hi} = \frac{x_{hi}}{N\mu}$$

A proporção da população que se situa no h -ésimo estrato é:

$$\pi_h = \frac{n_h}{N}$$

E a correspondente proporção da renda total é:

$$Y_h = \sum_{i=1}^{n_h} y_{hi} = \frac{1}{N\mu} \sum_{i=1}^{n_h} x_{hi}$$

A renda média do h -ésimo estrato é:

$$\mu_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} = \frac{Y_h}{\pi_h} \mu$$

Segue-se que:

$$\frac{\mu_h}{\mu} = \frac{Y_h}{\pi_h} \quad (1)$$

isto é, a renda (média) relativa em um estrato é igual à razão entre a proporção da renda total e a proporção da população correspondente a esse estrato.

No ponto correspondente ao limite superior do h -ésimo estrato, as coordenadas da curva de Lorenz são:

$$p_h = \frac{1}{N} \sum_{j=1}^h n_j = \sum_{j=1}^h \pi_j \quad (2)$$

e:

$$\Phi_h = \sum_{j=1}^h Y_j = \frac{1}{\mu} \sum_{j=1}^h \mu_j \pi_j = \frac{1}{N\mu} \sum_{j=1}^h n_j \mu_j \quad (3)$$

Pode-se demonstrar¹ que o valor do índice de Gini (G) para a população é dado por:

$$G = G_e + \sum_{h=1}^k \pi_h Y_h G_h \quad (4)$$

onde G_h é o índice de Gini referente à desigualdade dentro do h -ésimo estrato e G_e é o índice de Gini referente à desigualdade entre os estratos, isto é, G_e seria o índice de Gini para a distribuição da

¹ Algumas das demonstrações omitidas neste trabalho podem ser encontradas em R. Hoffmann, *Medidas de Concentração de uma Distribuição e a Desigualdade Econômica em uma Sociedade*, Série Estudos (Departamento de Economia e Sociologia Rural da ESALQ-USP, 1976), n.º 20, ou no Capítulo 16 de R. Hoffmann, *Estatística para Economistas*, a ser publicado pela Editora Pioneira.

renda nessa população se dentro dos estratos a renda fosse equitativamente distribuída ($x_{hi} = \mu_h$ para $i = 1, \dots, n_h$). O valor de G_e é dado por:

$$G_e = 1 - \sum_{h=1}^k (\Phi_{h-1} + \Phi_h) \pi_h \quad (5)$$

com $\Phi_0 = 0$. Note-se que para calcular G_e precisamos apenas conhecer a proporção da população (π_h) e da renda (Y_h) em cada estrato. Entretanto, para calcular os índices de Gini referentes às desigualdades dentro dos estratos (G_h , $h = 1, \dots, k$) precisamos conhecer a renda recebida por indivíduo.

A redundância da distribuição da renda descrita é dada por:²

$$R = R_e + \sum_{h=1}^k Y_h R_h \quad (6)$$

onde R_h é a redundância dentro do h -ésimo estrato e R_e é a redundância entre estratos, dada por:

$$R_e = \sum_{h=1}^k Y_h \log \frac{Y_h}{\pi_h} \quad (7)$$

Para calcular os valores de R_h ($h = 1, \dots, k$) precisamos conhecer a fração da renda total recebida por indivíduo.

Apenas quando há perfeita igualdade na distribuição da renda dentro dos estratos é que $G = G_e$ e $R = R_e$. Em qualquer outro caso os valores de G_e e R_e constituem subestimativas do verdadeiro grau de desigualdade.

Neste trabalho apresentaremos um método de estimar os valores de G_h e R_h . Para os estratos com limites finitos vamos pressupor que a distribuição dentro do estrato tem função de densidade linear e dentro do estrato de rendas mais altas, quando ele é aberto à direita, vamos pressupor que a distribuição é a de Pareto com dois parâmetros.

² Ver H. Theil, *Economics and Information Theory* (Chicago: Rand McNally, 1967), Capítulo 4.

3 — O índice de Gini e a redundância para uma distribuição da renda com função de densidade linear e para uma distribuição de Pareto com dois parâmetros

Seja x uma variável aleatória contínua cuja distribuição tem função de densidade linear, isto é:

$$f(x) = \alpha + \beta x \text{ para } a \leq x \leq b$$

e:

$$f(x) = 0 \text{ para } x < a \text{ e } x > b$$

De:

$$\int_a^b f(x) dx = 1$$

obtemos:

$$\alpha = \frac{1}{\theta} - \beta \left(a + \frac{\theta}{2} \right) \quad (8)$$

onde $\theta = b - a$.

Pode-se verificar que a média da distribuição é:

$$m = \int_a^b xf(x) dx = a + \frac{\theta}{2} + \beta \frac{\theta^3}{12} \quad (9)$$

Tendo em vista obter o índice de Gini, interessa-nos a diferença absoluta média, definida por:

$$\Delta = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y| dF(y) dF(x)$$

onde x e y representam dois valores quaisquer da variável e $F(x)$ é a função de distribuição, isto é:

$$F(x) = \int_{-\infty}^x f(y) dy$$

Para uma distribuição com média finita, pode-se demonstrar ³ que:

$$\begin{aligned}\Delta &= 4 \int_{-\infty}^{\infty} xF(x) dF(x) - 2\mu = \\ &= 4 \int_{-\infty}^{\infty} x \left[F(x) - \frac{1}{2} \right] dF(x) = \\ &= 2 \int_{-\infty}^{\infty} F(x) [1 - F(x)] dx\end{aligned}$$

Se x é uma variável não-negativa com média finita, temos: ⁴

$$\Delta = 2\mu - 2 \int_0^{\infty} [1 - F(x)]^2 dx$$

Com base em qualquer uma dessas expressões pode-se verificar que para uma distribuição com função de densidade linear a diferença absoluta média é:

$$\Delta = \frac{\theta}{3} \left(1 - \frac{\beta^2 \theta^4}{20} \right)$$

Como $G = \Delta / (2m)$, obtemos: ⁵

$$G = 2 \left(1 - \frac{a}{m} \right) \left[1 - \frac{(1 + 3\lambda)^2}{15\lambda} \right] \quad (10)$$

onde:

$$\lambda = (m - a) / \theta \quad (11)$$

Uma vez que $f(a) \geq 0$ e $f(b) \geq 0$, pode-se deduzir que:

$$\frac{1}{3} \leq \lambda \leq \frac{2}{3} \quad (12)$$

³ Ver R. Hoffmann, *Medidas de Concentração...*, *op. cit.*

⁴ Ver R. Dorfman, "A Formula for the Gini Coefficient", in *Review of Economic and Statistics*, vol. 61, n.º 1 (fevereiro de 1979), pp. 146-149.

⁵ A fórmula (10) é apresentada em S. Goldsmith, G. Jaszi, H. Kaitz e M. Liebenberg, "Size Distribution of Income Since the Mid-thirties", in *Review of Economics and Statistics*, vol. 36, n.º 1 (1954), pp. 1-32, onde é utilizada para estimar o índice de Gini referente à desigualdade dentro dos estratos.

Interessa-nos o caso particular em que $f(b) = 0$. Então $\lambda = 1/3$ e:

$$G = \frac{2}{5} \left(1 - \frac{a}{m} \right) \quad (13)$$

Também nos interessa o caso particular em que $f(a) = 0$. Então $\lambda = 2/3$ e:

$$G = \frac{2}{5} \left(\frac{b}{m} - 1 \right) \quad (14)$$

De acordo com Theil,⁶ se x é uma variável não-negativa com distribuição contínua, a redundância (como medida de desigualdade) da distribuição é dada por:

$$R = \int_0^{\infty} \left(\frac{x}{m} \log \frac{x}{m} \right) f(x) dx$$

Com $f(x) = \alpha + \beta x$ para $a \leq x \leq b = a + \theta$, e lembrando (8), (9) e (11), obtemos:

$$R = \frac{a^2}{2\theta m} \left[1 - \left(\frac{2a}{\theta} + 3 \right) (2\lambda - 1) \right] \log \frac{b}{a} + \log \frac{b}{m} + \frac{1}{\theta m} (2\lambda - 1) \left(a^2 + a\theta + \frac{\theta^2}{12} \right) - \frac{1}{2m} \left(a + \frac{\theta}{2} \right) \quad (15)$$

Para o caso particular em que $f(b) = 0$, temos $\lambda = 1/3$:

$$\theta = 3 (m - a) \quad (16)$$

e:

$$b = 3m - 2a \quad (17)$$

Para o caso particular em que $f(a) = 0$, temos $\lambda = 2/3$:

$$\theta = 3 (b - m) \quad (18)$$

⁶ H. Theil, *op. cit.*, Capítulo 4.

e:

$$a = 3m - 2b \quad (19)$$

Consideremos agora uma variável aleatória com distribuição de Pareto com dois parâmetros, isto é, a função de distribuição de x é:

$$F(x) = 0 \text{ para } x < a \text{ e}$$

$$F(x) = 1 - \left(\frac{a}{x}\right)^\alpha, \text{ com } \alpha > 1 \text{ e } a > 0, \text{ para } x \geq a$$

Pode-se verificar ⁷ que a média da distribuição é:

$$m = \frac{\alpha a}{\alpha - 1} \quad (20)$$

que o índice de Gini é dado por:

$$G = \frac{1}{2\alpha - 1} \quad (21)$$

e que a redundância é dada por:

$$R = \frac{1}{\alpha - 1} - \log \frac{\alpha}{\alpha - 1} \quad (22)$$

De (20) e (21) obtemos:

$$G = \frac{m - a}{m + a} \quad (23)$$

De (20) e (22) obtemos:

$$R = \frac{m}{a} - 1 - \log \frac{m}{a} \quad (24)$$

⁷ Ver H. Theil, *op. cit.*, e R. Hoffmann, *A Distribuição de Pareto*, Série Estudos (Departamento de Economia e Sociologia Rural da ESALQ-USP, 1974), n.º 18.

4 — Estimação da desigualdade dentro de estratos no cálculo do índice de Gini e da redundância de uma distribuição da renda

Consideremos, novamente, a população descrita na introdução. Admitimos que são conhecidos os valores de π_h , Y_h e μ_h para os k estratos, mas que não são dadas as rendas de cada indivíduo. Vamos admitir, também, que são conhecidos os limites dos estratos. Sejam E_{h-1} e E_h os limites inferior e superior, respectivamente, do h -ésimo estrato.

De acordo com (4) e (6), para obter o índice de Gini e a redundância da população, precisamos estimar os índices de Gini (G_h) e as redundâncias (R_h) referentes às desigualdades dentro dos estratos.

Vamos considerar, inicialmente, um estrato com limites finitos. Neste caso o índice de Gini (G_h) e a redundância (R_h) referentes à desigualdade dentro do estrato serão estimados admitindo-se que a distribuição dentro do estrato tem função de densidade linear. É importante assinalar que não pretendemos afirmar que a distribuição real tenha função de densidade linear, a qual é utilizada apenas para obter uma estimativa da desigualdade dentro do estrato,⁸ e está claro que o método pode levar à subestimação ou superestimação dessa desigualdade. De qualquer maneira, este método é melhor do que simplesmente ignorar as desigualdades dentro dos estratos ou admitir que a distribuição dentro de um estrato com limites finitos é necessariamente uma distribuição uniforme. Além disso, admitir que a função de densidade dentro do estrato é linear (ou alguma outra função com *dois* parâmetros) é, num certo sentido, o melhor que se pode fazer com os dados disponíveis, que são os limites e a renda média do estrato. Se tentássemos utilizar uma distribuição cuja função de densidade tivesse três ou mais parâmetros, não seria possível determinar os valores desses parâmetros. Uma outra solução para o problema, no caso do índice de Gini, foi

⁸ No programa para computador que elaboramos, a distribuição com função de densidade linear e (no caso do estrato de rendas mais altas, quando não é delimitado à direita) a distribuição de Pareto são utilizadas também para a determinação de percentis da distribuição global.

elaborada por Gastwirth,⁹ e consiste em abandonar a tentativa de obter *uma* estimativa para a medida da desigualdade da distribuição e considerar um intervalo que contenha o valor real da medida de desigualdade. Devemos assinalar que o método descrito aqui conduz a uma estimativa do índice de Gini que está necessariamente dentro do intervalo obtido pelo método de Gastwirth.

Consideremos o estrato com limites finitos E_{h-1} e E_h e cuja renda média é μ_h . Se tivermos:

$$\frac{1}{3} \leq \frac{\mu_h - E_{h-1}}{E_h - E_{h-1}} \leq \frac{2}{3}$$

fazemos $E_{h-1} = a$, $E_h = b$ e $\mu_h = m$, e obtemos as estimativas de G_h e R_h através de (10) e (15), respectivamente.

Se tivermos:

$$\frac{\mu_h - E_{h-1}}{E_h - E_{h-1}} < \frac{1}{3}$$

fazemos $E_{h-1} = a$, $\lambda = 1/3$ e $\mu_h = m$, e obtemos a estimativa de G_h através de (13). A estimativa de R_h é obtida através de (15) com os valores de θ e b dados por (16) e (17). Estamos admitindo, nesse caso, que a função de densidade é igual a zero no intervalo de b até E_h .

Finalmente, se tivermos:

$$\frac{\mu_h - E_{h-1}}{E_h - E_{h-1}} > \frac{2}{3}$$

fazemos $E_h = b$, $\lambda = 2/3$ e $\mu_h = m$, e obtemos a estimativa de G_h através de (14). A estimativa de R_h é obtida através de (15) com os valores de θ e a dados por (18) e (19). Estamos admitindo, nesse caso, que a função de densidade é igual a zero no intervalo de E_{h-1} até a .

Quando o estrato de rendas mais altas, cujo limite inferior é E_{k-1} , não é delimitado à direita, admitimos que a distribuição dentro desse

⁹ Ver J. L. Gastwirth, "The Estimation of the Lorenz Curve and Gini Index", in *Review of Economics and Statistics*, vol. 54, n.º 3 (agosto de 1972), pp. 306-316.

estrato é a de Pareto com dois parâmetros. Valem aqui comentários semelhantes àqueles que fizemos em relação à pressuposição de que a distribuição dentro de um estrato com limites finitos tem função de densidade linear. Os valores do índice de Gini (G_k) e da redundância (P_k) dentro do último estrato serão, então, obtidos através de (23) e (24), respectivamente, fazendo $E_{k-1} = a$ e $\mu_k = m$.

Para ilustrar, vamos apresentar aqui dois dos exemplos numéricos artificiais utilizados para testar o programa para computador que elaboramos, tendo em vista o cálculo de medidas de concentração e a interpolação de percentis. O primeiro exemplo é apresentado na Tabela 1.

TABELA 1

Distribuição da renda em uma população dividida em três estratos

Estrato (h)	Limite Inferior (E_{h-1})	Limite Superior (E_h)	Número de Indi- viduos (n_h)	Proporção da População (π_h)	Renda Total	Renda Média (μ_h)
1	0	50	30	0,25	1.200	40,0
2	50	120	60	0,50	4.200	70,0
3	120	135	30	0,25	3.825	127,5
			120	1,00	9.225	

A renda média da população é $\mu = 76,875$.

De acordo com (3) e (5), obtemos $G_c = 35/164 = 0,213415$.

De acordo com (7), obtemos $R_e = 0,082142$ nits.¹⁰ O correspondente valor do índice de Theil¹¹ é $T_e = 1 - \exp \{-R_e\} = 0,078859$.

¹⁰ O termo "nits" é formado pelas letras inicial e finais da expressão *natural units*, indicando que no cálculo da redundância foram utilizados logaritmos neperianos ou naturais.

¹¹ Souza mostrou que o índice de concentração de Theil é um caso particular do conceito de *dual* de uma medida de concentração. Ver J. de Souza, "Dualidade e Concentração", trabalho apresentado no II Encontro Anual da ANPEC (CEDEPLAR-UFMG, 1974). Alguns autores chamam de índice de Theil ao que foi denominado aqui de redundância; ver, por exemplo, C. G. Langoni, *Distribuição da Renda e Desenvolvimento Económico do Brasil* (Rio de Janeiro: Expressão e Cultura, 1973).

Para o primeiro estrato temos $\frac{\mu_1 - E_0}{E_1 - E_0} = \frac{40}{50} > \frac{2}{3}$. Então, de acordo com (14) obtemos $G_1 = 1/10$ e através de (18), (19) e (15), com $\lambda = 2/3$, obtemos $R_1 = 0,016381$ nits.

Para o segundo estrato temos $\frac{\mu_2 - E_1}{E_2 - E_1} = \frac{20}{70} < \frac{1}{3}$. Então, de acordo com (13) obtemos $G_2 = 4/35$ e através de (16), (17) e (15), com $\lambda = 1/3$, obtemos $R_2 = 0,019926$ nits.

Para o terceiro estrato temos $\frac{\mu_3 - E_2}{E_3 - E_2} = \frac{7,5}{15} = \frac{1}{2}$. Através de (10) e (15), com $\lambda = 1/2$, obtemos $G_3 = 1/51$ e $R_3 = 0,000577$ nits.

Finalmente, substituindo esses resultados em (4) e (6), obtemos $G = 301/1230 = 0,244715$ e $R = 0,093584$ nits. O índice de Theil correspondente a essa redundância é $T = 0,089339$.

Como segundo exemplo vamos considerar uma população dividida em apenas dois estratos, como mostra a Tabela 2. Note-se que o segundo estrato não é limitado à direita.

TABELA 2

Distribuição da renda em uma população dividida em dois estratos

Estrato (h)	Limite Inferior (E_{h-1})	Limite Superior (E_h)	Número de Indi- víduos (n_h)	Proporção da População (π_h)	Renda Total	Renda Média (μ_h)
1	58	72	28	0,28	1.820	65
2	72	∞	72	0,72	10.368	144
			100	1,00	12.188	

A renda média da população é $\mu = 121,88$.

Obtemos $G_e = 0,130673$, $R_e = 0,047998$ nits e $T_e = 0,046864$.

Para o primeiro estrato temos $G_1 = 7/195$ e $R_1 = 0,001935$ nits.

Para o segundo estrato, considerando uma distribuição de Pareto com dois parâmetros, obtemos, através de (23) e (24), $G_2 = 1/3$ e $R_2 = 0,306853$ nits.

Substituindo esses resultados em (4) e (6), obtemos $G = 0,336335$ e $R = 0,309318$ nits. O índice de Theil correspondente a essa redundância é $T = 0,266053$.

5 — Outro método

Nesta seção veremos um outro método para obter os valores do índice de Gini e da redundância de uma distribuição da renda levando em consideração a provável desigualdade dentro dos estratos de renda. Este método se baseia na determinação de valores mínimos e máximos para os índices de Gini (G_h) e redundâncias (R_h) dentro dos estratos.

Consideremos o h -ésimo estrato, com renda média μ_h e cujos limites inferior e superior são, respectivamente, E_{h-1} e E_h . Temos, obviamente, $E_{h-1} \leq \mu_h \leq E_h$. Teremos um máximo de desigualdade dentro desse estrato quando uma parte dos indivíduos tiverem renda igual a E_{h-1} e os demais tiverem renda igual a E_h . É óbvio que isso maximiza o valor da diferença média (Δ_h) dentro desse estrato. Seja n_h o número de indivíduos do estrato e seja ϕ a fração deles que tem renda igual a E_{h-1} . Então temos ϕn_h indivíduos com renda igual a E_{h-1} e $(1-\phi) n_h$ indivíduos com renda igual a E_h . Para que a renda média seja μ_h devemos ter:

$$\mu_h = \phi E_{h-1} + (1-\phi) E_h$$

Segue-se que:

$$\phi = \frac{E_h - \mu_h}{E_h - E_{h-1}} \quad (25)$$

Não é difícil verificar que a diferença média dentro do estrato é dada por:

$$\Delta_h = 2\phi(1-\phi)(E_h - E_{h-1}) \quad (26)$$

De (25) e (26), e lembrando que $G_h = \Delta_h / (2\mu_h)$, concluímos que o valor máximo do índice de Gini dentro do estrato é dado por:

$$G_h = \frac{(E_h - \mu_h)(\mu_h - E_{h-1})}{\mu_h(E_h - E_{h-1})} \quad (27)$$

Quando o estrato de rendas mais altas, cujo limite inferior é E_{k-1} , não é limitado à direita, o valor máximo para o índice de Gini dentro desse estrato é dado por:

$$G_k = \frac{\mu_k - E_{k-1}}{\mu_k} \quad (28)$$

Esta expressão é o limite de (27), com $h = k$, quando E_k tende a infinito.

Substituindo os valores obtidos de (27) e (28) em (4), obtemos um limite superior (G_s) para o valor do índice de Gini da distribuição da renda na população. Quaisquer que sejam as distribuições dentro dos estratos, o índice de Gini (G) para a população é ao menos igual a G_e e no máximo igual a G_s , isto é, $G_e \leq G \leq G_s$.¹² É razoável, então, tomar a média aritmética de G_e e G_s como uma estimativa do índice de Gini da população:

$$G = \frac{1}{2} (G_e + G_s) \quad (29)$$

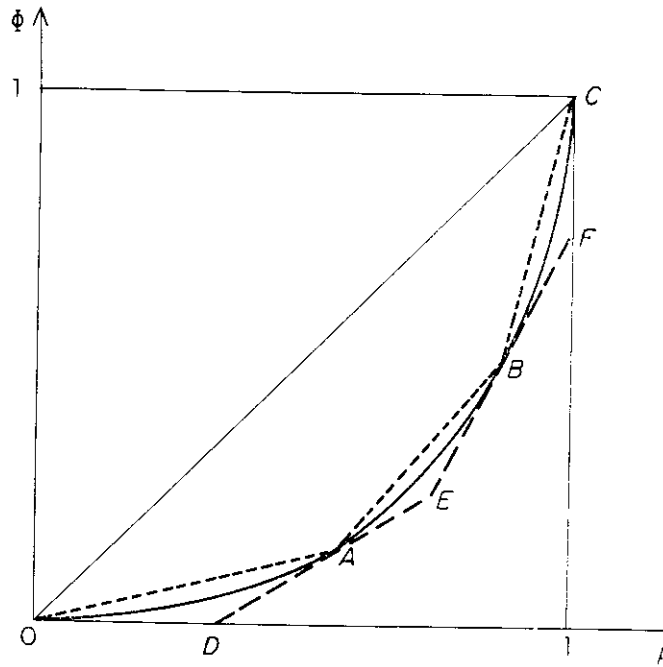
É interessante examinar a interpretação geométrica para os valores de G_e e G_s . Vamos admitir que a população está dividida em apenas três estratos, que o limite inferior do primeiro estrato é $E_0 = 0$ e que o terceiro estrato não tem limite superior finito. No gráfico a seguir traçamos uma curva de Lorenz hipotética, e os pontos A , B e C correspondem aos limites superiores dos três estratos.

A área do polígono $OABC$ é a área de concentração correspondente ao valor do índice de Gini da população quando se admite que

¹² Além de considerar o intervalo delimitado por G_e e G_s , Gastwirth, fazendo algumas pressuposições sobre as distribuições dentro dos estratos, determinou um intervalo mais estreito dentro do qual deve estar o valor do índice de Gini da população. Ver J. L. Gastwirth, *op. cit.*

dentro de cada estrato todos os indivíduos têm a mesma renda, isto é, $G_e = 2$ (área $OABC$). Traçando retas tangentes à curva de Lorenz nos pontos A e B obtemos o polígono $ODEFC$, cuja área é a área de concentração quando se supõe um máximo de desigualdade dentro de cada estrato, isto é, $G_s = 2$ (área $ODEFC$). Pode-se demonstrar que a declividade da reta tangente à curva de Lorenz no ponto que corresponde ao limite superior (E_h) de um estrato é igual a E_h/μ .

CURVAS DE LORENZ (HIPOTÉTICAS)



Vejamos, agora, a determinação do valor da redundância dentro do h -ésimo estrato, com renda média μ_h e limites E_{h-1} e E_h , quando se admite que os indivíduos têm renda igual a E_{h-1} ou igual a E_h . Os ϕn_h indivíduos com renda igual a E_{h-1} se apropriam da fração $(\phi E_{h-1})/\mu_h$ da renda total do estrato e os $(1-\phi) n_h$ indivíduos com

renda igual a E_h se apropriam da fração $[(1-\phi)E_h]/\mu_h$ da renda total do estrato. Então a redundância dentro do estrato é:

$$R_h = \frac{\phi E_{h-1}}{\mu_h} \log \frac{E_{h-1}}{\mu_h} + \frac{(1-\phi)E_h}{\mu_h} \log \frac{E_h}{\mu_h}$$

Lembrando (25), obtemos:

$$R_h = \frac{l}{(E_h - E_{h-1})\mu_h} \left[(E_h - \mu_h) E_{h-1} \log \frac{E_{h-1}}{\mu_h} + (\mu_h - E_{h-1}) E_h \log \frac{E_h}{\mu_h} \right] \quad (30)$$

De acordo com essa expressão, quando o estrato de rendas mais altas, cujo limite inferior é E_{k-1} , não é limitado à direita, o valor máximo da redundância dentro desse estrato é infinito, pois de (30), com $h = k$, obtemos:

$$\lim_{E_k \rightarrow \infty} R_k = \infty$$

Isso levaria a um limite superior infinito para a redundância da população. Theil¹³ sugere, então, que a redundância dentro do k -ésimo estrato, quando ele não é limitado à direita, seja estimada admitindo que a distribuição dentro do estrato é a de Pareto com dois parâmetros. De acordo com (24) temos:

$$R_k = \frac{\mu_k}{E_{k-1}} - l - \log \frac{\mu_k}{E_{k-1}} \quad (31)$$

Um limite superior (R_s) para o valor da redundância da distribuição da renda na população é obtido substituindo em (6) os valores de R_h dados por (30) e (31).

Theil comete o engano de indicar a média aritmética dos valores de R_e e R_s como uma estimativa para o valor da redundância na população. Isso só é razoável quando todos os estratos tiverem limites finitos e os valores de R_h ($h = 1, \dots, k$) utilizados no cálculo de R_s tiverem todos sido obtidos através de (30). O valor de R_k dado por (31) não é um limite superior para redundância dentro

¹³ Ver H. Theil, *op. cit.*, pp. 128-134.

do k -ésimo estrato, mas sim uma boa estimativa do valor dessa redundância, já que, como tem sido observado em vários estudos, a distribuição de Pareto se ajusta bastante bem para rendas relativamente altas. Então, quando o estrato de rendas mais altas não é limitado à direita e o valor de R_k é obtido através de (31), uma estimativa razoável para a redundância da distribuição da renda na população é dada por:

$$R = R_e + \sum_{h=1}^{k-1} Y_h \frac{R_h}{2} + Y_k R_k \quad (32)$$

onde os valores de R_h para $h = 1, \dots, k-1$ são obtidos através de (30). Note-se que apenas os valores de R_h obtidos através de (30) são considerados como valores máximos, adotando-se, então, $(1/2)R_h$ como estimativa da redundância dentro do estrato.

Devemos ressaltar que para os exemplos apresentados em Theil a diferença entre a média aritmética de R_e e R_s e o valor dado por (32) é muito pequena, porque o peso do último estrato, dado por sua participação na renda total, é pequeno.

Para o exemplo numérico apresentado na Tabela 1, os valores máximos do índice de Gini dentro dos estratos, obtidos através de (27), são $G_1 = 1/5$, $G_2 = 10/49$ e $G_3 = 1/34$. Já vimos que $G_e = 35/164 = 0,213415$. Substituindo esses resultados em (4) obtemos o limite superior para o índice de Gini da população, que é $G_s = 0,269425$. A média aritmética dos limites inferior e superior para o valor do índice de Gini é $(1/2)(G_e + G_s) = 0,241420$, semelhante ao valor do índice de Gini obtido admitindo que a distribuição dentro dos estratos tem função de densidade linear, que é $G = 0,244715$.

Para o exemplo da Tabela 1 os valores máximos da redundância dentro dos estratos, obtidos através de (30), são $R_1 = 0,223144$ nits (no cálculo desse valor, uma vez que $\lim_{a \rightarrow 0} a \log a = 0$, toma-se $0 \log 0 = 0$), $R_2 = 0,092329$ nits e $R_3 = 0,001731$ nits. Já vimos que $R_e = 0,082142$ nits. Substituindo esses resultados em (6) obtemos $R_s = 0,153923$. A média aritmética de R_e e R_s é $0,118033$ nits e o correspondente valor do índice de Theil é $T = 0,111333$.

Para o exemplo numérico apresentado na Tabela 2, os valores máximos do índice de Gini dentro dos estratos, obtidos através de (27)

e (28), são $G_1 = 7/130$ e $G_2 = 1/2$. Já vimos que $G_c = 0,130673$. Substituindo esses valores em (4) obtemos $G_s = 0,439166$. A média aritmética dos limites inferior e superior para o valor do índice de Gini é $(1/2)(G_r + G_s) = 0,284920$. Este valor difere bastante da estimativa do índice de Gini obtida, admitindo que a distribuição dentro do primeiro estrato tem função de densidade linear e que no segundo estrato a distribuição é a de Pareto com dois parâmetros, que, como vimos, é $G = 0,336335$. A diferença, neste caso, se explica pelo fato de termos apenas dois estratos.

Para o exemplo da Tabela 2, o valor máximo da redundância no primeiro estrato é $R_1 = 0,005810$ nits. Já vimos que $R_2 = 0,047998$ nits e que dentro do segundo estrato, admitindo uma distribuição de Pareto com dois parâmetros, temos $R_2 = 0,306853$ nits. Substituindo esses resultados em (6) obtemos $R_s = 0,309897$ nits e substituindo em (32) obtemos $R = 0,309463$ nits. A essa última redundância corresponde o índice de Theil $T = 0,266159$. Note-se que a média aritmética de R_c e R_s é igual a $0,178947$ nits, o que certamente subestima o grau de desigualdade na população se a distribuição dentro do segundo estrato é ou se assemelha a uma distribuição de Pareto.

6 — Índice de Gini e redundância para a distribuição da renda entre empregados na indústria e no comércio e serviços, no Brasil, de 1969 a 1974

Como ilustração final dos diferentes métodos de cálculo do índice de Gini e da redundância discutidos neste trabalho, apresentamos, nas Tabelas 3, 4 e 5, o valor dessas medidas de concentração para a distribuição da renda recebida como salário pelos empregados na indústria e no comércio e serviços, no Brasil, de 1969 a 1974. Os dados básicos consistem no total de empregados e o respectivo montante de salários para 17 estratos de salário e foram publicados pelo Centro de Documentação e Informática (CDI) do Ministério do

TABELA 3

Índice de Gini e redundância para a distribuição da renda entre empregados da indústria, no Brasil, de 1969 a 1974 — dados básicos relativos à “Lei dos 2/3”, referentes a abril de cada ano

Anos	G_e	G_s	Média	G^a	R_e	R_s	R^b	R^a
1969	0,407	0,414	0,410	0,411	0,340	0,374	0,370	0,363
1970	0,422	0,429	0,426	0,427	0,376	0,401	0,394	0,389
1971	0,422	0,429	0,426	0,427	0,378	0,404	0,399	0,397
1972	0,438	0,448	0,443	0,444	0,403	0,442	0,438	0,436
1973	0,460	0,468	0,464	0,465	0,435	0,507	0,504	0,503
1974	0,504	0,519	0,512	0,515	0,524	0,734	0,732	0,731

^aAdmitindo distribuição com função de densidade linear nos estratos com limites finitos e distribuição de Pareto com dois parâmetros no estrato de rendas mais altas, que não é limitado à direita.

^bCalculada de acordo com (32).

TABELA 4

Índice de Gini e redundância para a distribuição da renda entre empregados do setor de comércio e serviços, no Brasil, de 1969 a 1974 — dados básicos relativos à “Lei dos 2/3”, referentes a abril de cada ano

Anos	G_e	G_s	Média	G^a	R_e	R_s	R^b	R^a
1969	0,449	0,456	0,453	0,454	0,384	0,421	0,418	0,417
1970	0,470	0,475	0,472	0,473	0,437	0,461	0,456	0,453
1971	0,475	0,482	0,478	0,479	0,445	0,474	0,470	0,468
1972	0,487	0,497	0,492	0,493	0,466	0,515	0,512	0,511
1973	0,498	0,510	0,504	0,506	0,480	0,558	0,556	0,555
1974	0,517	0,537	0,527	0,530	0,521	0,679	0,677	0,676

^aVer Tabela 3.

^bIbid.

TABELA 5

Índice de Gini e redundância para a distribuição da renda entre empregados na indústria, no comércio e no setor de serviços, no Brasil, de 1969 a 1974 — dados básicos relativos à “Lei dos 2/3”, referentes a abril de cada ano

Anos	G_e	G_s	Média	G^a	R_e	R_s	R^b	R^a
1969	0,430	0,436	0,433	0,434	0,367	0,401	0,398	0,397
1970	0,448	0,454	0,451	0,452	0,411	0,436	0,430	0,426
1971	0,451	0,458	0,454	0,455	0,418	0,445	0,441	0,439
1972	0,466	0,475	0,470	0,471	0,441	0,485	0,481	0,480
1973	0,479	0,489	0,484	0,486	0,460	0,534	0,531	0,530
1974	0,519	0,527	0,519	0,522	0,522	0,706	0,703	0,703

^aVer Tabela 3.

^b*Ibid.*

Trabalho,¹⁴ com base em informações coletadas em abril de cada ano por força da “Lei dos 2/3”.

O estrato de rendas mais altas não é limitado à direita, e o CDI manteve seu limite inferior fixado, a partir de 1970, em Cr\$ 2.400,00 em moeda corrente. Então, devido à inflação, com o passar dos anos uma proporção cada vez maior da população é classificada nesse estrato. Para a indústria, em 1970 esse estrato incluía 0,8% dos empregados, aos quais correspondia 8,6% da renda total, e em 1974 esse estrato incluía 6,0% dos empregados, aos quais correspondia 36,2% da renda total. Isso faz com que aumente a diferença $G_s - G_e$, mostrando que é cada vez mais importante, no cálculo das medidas de concentração, levar em consideração a provável desigualdade dentro dos estratos. Note-se também que, especialmente nos últimos anos, quando uma proporção relativamente alta da renda total corresponde ao estrato de rendas mais altas, a média aritmética entre R_e e R_s é significativamente menor que o valor da redundância calculado de acordo com (32).

¹⁴ Ver *Boletim Técnico* do CDI, do Centro de Documentação e Informática do Ministério do Trabalho. Até 1971, essa publicação era denominada *Boletim Técnico do SEPT* (Serviço de Estatística da Previdência e Trabalho, do Ministério do Trabalho e Previdência Social). Os dados utilizados estão nos *Boletins* de n.os 21, 22, 26, 32, 41 e 45.