Considerações metodológicas sobre as medidas de desigualdades

RAMONAVAL AUGUSTO COSTA

I — INTRODUÇÃO

O objetivo deste estudo é apresentar alguns exemplos de variações em indicadores, em função do tamanho da amostra e da metodologia empregada. Para o cálculo deles os indicadores utilizados como exemplo são medidas de desigualdade de renda.[1] Estas medidas estão sendo usadas devido à disponibilidade da informação e à grande importância que se da às oscilações nos valores destes indicadores. Quando se constatam aumentos na desigualdade das rendas sugere-se que a posição relativa dos indivíduos não melhorou. Tal indicação gera muitos problemas de ordem política e econômica. Nestas circunstâncias seria necessário evitar que se considerasse como aumento efetivo simples variações dos indicadores devido ao tamanho da amostra e ao método de cálculo utilizado. Nosso interesse é chamar atenção para esta possibilidade, a fim de incutir nos técnicos maior preocupação com a metodologia quando da utilização de indicadores de desigualdade de renda (*Costa,* 1975).

Os indicadores utilizados como exemplo são: o Coeficiente de Pareto, o Coeficiente ou Índice de Gini, a Razão de Concentração de Gini e o Índice de Theil. Aproveitar-se-á a oportunidade para oferecer alguns detalhes sobre o verdadeiro significado destes indicadores e suas propriedades como medidas de desigualdade das rendas.

---

1 Estas medidas não são exclusivamente as da desigualdade de rendas, são também as de desigualdade de qualquer outro atributo que se julgue interessante investigar. Historicamente estas medidas estão ligadas aos estudos da situação de desigualdade das rendas pessoais, daí o experimento ter sido montado com base numa distribuição de renda. Existem vários pontos em comum nos trabalhos do autor divulgados na *Rev. Bras. de Estatística* n.º 144 e no *Boletim Geográfico* n.º 238.

I.1 — Tamanho da amostra

No caso o Censo Demográfico de 1970 a Fundação IBGE tem disponível duas amostras. A primeira é constituída de aproximadamente 1,3% da população e dos domicílios. A segunda se refere à amostra de 25% da população e dos domicílios com a qual foram feitas as perguntas adicionais sobre a situação sócio-econômica das pessoas e dos domicílios. A maior amostra serviu de base para divulgação definitiva dos dados do Censo Demográfico de 1970, ao passo que a amostra de 1,3% se constituiu na fonte de informação para a elaboração e divulgação das tabulações avançadas do Censo Demográfico de 1970, em circulação em meados de 1971. Neste trabalho as medidas de desigualdade de rendas foram calculadas para as duas amostras, a fim de evidenciar as possíveis diferenças entre os valores obtidos para cada índice.

Sem dúvida, ambas as amostras foram obtidas por processos de amostragem probabilística da população e dos domicílios. Aqui se admite que a Fundação IBGE tenha tomado todos os cuidados necessários para garantir a representatividade das amostras utilizadas; sendo assim, o trabalho que se segue não contém nenhum julgamento sobre a qualidade das amostras.

I.2 — Elementos representativos

Quando do cálculo das medidas da desigualdade das rendas a escolha do elemento representativo de cada classe de renda é muito importante. Em geral, os elementos representativos são os pontos médios. Quando não se tem a possibilidade de obter a verdadeira média da classe e não se tem razão suficiente para admitir que o ponto médio não seja um bom elemento representativo, então ele se destaca e é usado em quase todas as situações. No entanto, se a disponibilidade de dados permite calcular a média de cada classe, ela se torna o elemento representativo mais interessante e mais adequado. Devido às características de determinados índices, como os de Pareto e Gini, às vezes encontramos o limite inferior, ou mesmo o limite superior da classe, sendo utilizado como elemento representativo da mesma (*Tório*, 1973).

Nestas circunstâncias, as medidas de desigualdades serão calculadas de várias maneiras. Tal procedimento metodológico é, em princípio, muito simples, mas será extremamente útil para evidenciar as possíveis diferenças. Um índice de Pareto, calculado utilizando-se o ponto médio como elemento representativo e o mesmo índice, considerando-se a verdadeira média como elemento representativo, apresenta diferenças relativamente grandes. Tais diferenças podem induzir o pesquisador a acreditar que de fato houve uma variação no grau de desigualdade. Por isto mesmo é que neste trabalho, além de considerarmos as diferenças introduzidas pelo tamanho da amostra, também consideramos a escolha do elemento representativo da distribuição como de suma importância, se pretendemos medir a desigualdade da renda de uma maneira honesta e imparcial.

Os elementos representativos aqui considerados são o ponto médio, a média da renda, o limite inferior e o limite superior de cada classe de renda. Dos quatro elementos representativos, somente a média apresenta maiores variações entre as amostras, pois os demais são fixos, já que as classes de renda são mantidas.

I.3 — Dados utilizados

Já nos referimos aos dados como sendo do Censo Demográfico de 1970, com duas amostras que serviram de base para as divulgações das tabulações avançadas e dos dados definitivos do Censo Demográfico de 1970. Para efeito de comparação e facilidade de computação escolheu-se uma distribuição de renda que não apresentasse problema de última classe aberta, pelo menos no caso da Amostra de 1,3%.[2] Feita a escolha desta maneira e tendo disponível as verdadeiras médias de cada classe, esta distribuição ofereceria condições mais favoráveis à computação das medidas de desigualdade. A distribuição que encontramos com tais características foi a distribuição das rendas dos Homens em Ocupações da Agropecuária, da Produção Extrativa Vegetal, do Estado de São Paulo, em 1970.

Sem dúvida, seria interessante obter várias distribuições com estas características, a fim de tornar o trabalho mais definitivo e menos provocativo. No entanto, é muito raro obter uma distribuição de rendas, como esta, com a renda média de cada classe disponível, inclusive para a última classe. O mais interessante da distribuição utilizada é que ela permite comparar diversas maneiras de computar as medidas de desigualdade, livre da arbitrariedade introduzida pela abertura da última classe de renda (*Costa,* 1975).

II — ÍNDICES QUE DEPENDEM DE AJUSTAMENTO

Consideramos, neste estudo, uma classificação das medidas de desigualdade em função da necessidade ou não de ajustamento dos dados à alguma função específica. Dos índices estudados somente o Índice de Pareto e o Índice de Gini se enquadram como medidas que necessitam do ajustamento de uma determinada função aos dados. Os índices são parâmetros da função ajustada. Portanto, além das diferenças de ajustamento, existem diferenças amostrais e de escolha dos elementos representativos.

Apesar do Índice de Pareto (ou Coeficiente de Pareto) ser uma medida bem conhecida, entre as medidas de desigualdades, antes de apresentar os resultados obtidos tentaremos ressaltar algumas de suas características, bem como mostrar certas relações entre o Índice de Pareto e o Índice de Gini que tem sido pouco divulgadas. Este último, apesar de ser contemporâneo do Índice de Pareto e de ter sido apresentado por *Corrado Gini* como uma medida de desigualdade com melhores propriedades do que o Índice de Pareto, é praticamente desconhecido.

II.1 — Índice de Pareto

O Índice de Pareto tem sido bem divulgado, mesmo no Brasil. Basta observarmos o trabalho de *Kingston* em 1962, *Hoffman* em 1969, *Duarte* em 1971, *Langoni* em 1973 e mais recentemente *Costa* em 1975. Por este motivo não entraremos em maiores detalhes sobre a descrição desta medida.

Considerando que os indivíduos recebem rendas iguais a X_1, X_2 ... X_n e dado que N_x é o número de pessoas com renda X ou mais (às

2 No caso da amostra de 25%, a última classe é totalmente fechada, porém a renda média está subestimada em função dos problemas de codificação do Censo de 1970, onde as rendas acima de Cr$ 9.997,00 foram codificadas como sendo Cr$ 9.998,00. Tal procedimento certamente subestima a renda média.

vezes alguns autores consideram as pessoas com renda maior que X). *Pareto* admitia que N_x e X eram relacionados através da seguinte função:

$$N_x = A_x^{-\alpha} \quad \text{ou} \quad \boxed{N_x = \frac{A}{x^{\alpha}}}$$

Onde A e α são parâmetros. A medida de desigualdade de *Pareto* seria dada pelo expoente α. A interpretação de α pode ser feita como se fosse uma elasticidade.[3] Admitindo-se que a função densidade da distribuição das rendas seja f (x), então a função de distribuição será:

$$\int_{-\infty}^{x} f(x)\, dx = F(x)$$

Nestes termos podemos interpretar N_x como sendo igual a:

$$N_x = 1 - F(x)$$

Observamos que a função de *Pareto* é uma curva obtida a partir de F (x), ou seja, a Ogiva. Daí a explicação para N_x significar o número de pessoas que possuem a renda x ou mais (ou que possua renda maior que x).

Na informação original de *Pareto* o ajustamento da curva acima mencionada seria muito bom para valores de $x > x_0$, onde x_0 seria considerada uma renda mínima a partir da qual a função se ajustaria aos dados. Tal fato coloca a função de *Pareto* como um instrumento de descrição parcial da distribuição de rendas. Considerando-se a existência de uma renda mínima x_0, podemos sugerir uma outra interpretação para α. Considerando que:

$$N = Ax_0^{-\alpha} \quad \text{e} \quad N_x = Ax^{-\alpha}$$

obtemos:

$$\frac{Nx}{N} = \frac{Ax^{-\alpha}}{Ax_0^{-\alpha}} = \frac{x_0^{\alpha}}{x^{\alpha}} = \left(\frac{x_0}{x}\right)^{\alpha}$$

Sendo assim α poderia ser interpretado como o expoente ao qual se elevaria a fração que x_0 (renda mínima) representa da renda x, a fim de igualá-la à fração de contribuintes que recebem renda igual e maior que x (*Kingston*, 1967).

Uma outra maneira de vislumbrar a medida de desigualdade introduzida por *Pareto* é aquela derivada do processo de linearização a que submetemos a fração, a fim de poder estimar os seus parâmetros, ou seja:

$$N_x = Ax^{-\alpha} e^{\mu}$$

onde A e α são os parmetros a serem estimados. Particularmente, α é a própria medida de desigualdade; μ é um termo de erro na equação que

3 Para esta interpretação vide *Ramonaval Augusto Costa, "Medidas de Desigualdade de Renda" — Boletim Geográfico* n.º 238, jan./fev. de 1974, Ano 33, pp. 45/72.

possui média zero e variância constante.[4] Após o processo de linearização por logaritmos obtemos:

$$\log N_x = \log A - \alpha \log x + \mu$$

Então podemos interpretar α como sendo a inclinação da reta cuja ordenada é $\log N_x$ e abcissa $\log x$, respectivamente o logaritmo das pessoas com renda maior ou igual a x (ou maior que x) e o logaritmo de x. Sendo assim, também podemos lembrar que a representação gráfica desta reta seria uma maneira alternativa de utilização da medida de *Pareto.*

Todas as medidas de desigualdade apresentam suas vantagens e desvantagens e o Índice de Pareto tem sido muito discutido através dos tempos, tendo acumulado, portanto, muitas informações sobre o seu desempenho (*Costa,* 1975).

II.2 — Índice de Gini

O Índice de Gini é uma outra medida de desigualdade da renda pouco conhecida, apesar de ter sido apresentada por *Corrado Gini* como uma alternativa melhor do que a medida de *Pareto,* na compreensão da situação da distribuição da renda pessoal dos países.[5] Daremos um tratamento mais completo no caso deste índice, em função de sua falta de divulgação.

Primeiramente, vejamos a sua forma geral. O índice de Gini, como coeficiente de Pareto, é também o parâmetro de uma função cuja forma é:

$$N_x = KA_x^{\delta}$$

Onde N_x é o número total de pessoas com renda maior que x, A_x é a renda total das N_x pessoas, K é uma constante e δ é o que chamamos de Índice de Gini.

Uma interpretação semelhante ao Índice de Pareto é obtido através da observação de que δ também é uma elasticidade, mas neste caso o seu valor é positivo, ou seja $\frac{dNx}{dA_x} > 0$ ao passo que para *Parreto* $\frac{dNx}{dx} <$ < 0.

Para obter uma interpretação[6] bem clara do que significa o Índice de Gini consideremos que n seja o total de pessoas com as rendas x_1, x_2 ... x_n e m seja o total de pessoas com as maiores rendas, ou seja, $x_{n-m+1}, x_{n-m+2} \ldots x_n$. Consideremos que $x_1 < x_2 < \ldots < x_n$, nestas circunstâncias, podemos dizer que:

$$\bar{x}(n) = \frac{x_1 + x_2 \ldots x_n}{n}$$

$$\bar{x}(m) = \frac{x_{n-m+1} + x_{n-m+2} + \ldots x_n}{m}$$

4 Estas são hipóteses padrões nos métodos modernos de ajustamento através de regressão simples, usando o método de estimação dos mínimos quadrados.

5 Este índice foi apresentado originalmente no estudo de Corrado Gini *"Índice di Concentrazione e di Dipendenza"*, Biblioteca Dell'Economista, Roma Editice, Torinese Milano, 1922, páginas 5/137.

6 Esta interpretação encontra-se em Corrado Gini *"Índice di Concentrazione e di Dipendenza"*, op. cit. 1922.

e que $\bar{x}(n) < \bar{x}(m)$. Portanto, temos que:

$$\frac{x_{n-m+1} + x_{n-m+2} + x_n}{m} > \frac{x_1 + x_2 + \ldots + x_n}{n}$$

e

$$\frac{x_{n-m+1} + x_{n-m+2} \ldots + x_n}{x_1 + x_2 + \ldots + x_n} > \frac{m}{n}$$

Ou seja, o percentual da renda recebida pelas m pessoas é maior que o percentual de pessoas, e será tanto maior quanto maior for a concentração da renda nas mãos destas m pessoas. Nestes termos, para *Gini,* deveríamos elevar a fração da renda e um determinado expoente δ que a igualaria à fração de pessoas que recebem esta renda. Ou seja:

$$\left(\frac{x_{n-m+1} + x_{n-m+2} \ldots x_n}{x_1 + x_2 + \ldots x_n}\right)\delta = \frac{m}{n}$$

Então, quanto maior for o valor de δ maior será a concentração.

Para podermos mostrar que na função $N_x = KA^\delta$ δ possui o mesmo significado do expoente δ que iguala a fração de rendas à fração de pessoas acima, basta fazermos:

$$A = x_1 + x_2 + \ldots x_n \quad e \quad N = n$$

$$A_x = x_{n-m+1} + x_{n-m+2} + \ldots + x_n \quad e \quad N_x = m$$

Assumindo que $N = KA^\delta$ e $N_x = KA_x^\delta$, temos $\log N_x = \log K + \delta \log A_x$ e $\log N = \log K + \delta \log A$; subtraindo-se uma expressão da outra obtemos.

$$\log m = \log K + \delta \log (x_{n-m+1} + x_{n-m+2} \ldots + x_n)$$

$$\log n = \log K + \delta \log (x_1 + x_2 + \ldots + x_n)$$

$$\log m - \log n = \delta\,[\log(x_{n-m+1} + \ldots + x_n) - \log(x_1 + x_2 + \ldots + x_n)]$$

$$\log \frac{m}{n} = \delta \log \frac{x_{n-m+1} + \ldots + x_n}{x_1 + a_2 + \ldots x_n}$$

$$\frac{m}{n} = \left(\frac{x_{n-m+1} + \ldots + x_n}{x_1 + x_2 + \ldots + x_n}\right)^\delta$$

Não há dúvida de que o significado de δ se torna mais claro e fácil de se compreender quando é interpretado como o expoente ao qual devemos elevar a fração de renda que um determinado número de pessoas (m) recebe a fim de que ela se iguale à fração que estas pessoas representam no total.[7]

7 Observamos aqui o mesmo critério de igualdade que encontraremos por ocasião da definição da curva de Lorenz, onde o percentual acumulado de pessoas deve ser igual ao correspondente percentual acumulado de renda recebida por elas, a fim de se obter perfeita igualdade.

II.3 — A Relação entre o Índice de Gini e o Índice de Pareto

O Índice de Gini foi idealizado a fim de contrapor as idéias de *Pareto* de que a distribuição das rendas apresentava pouca variabilidade entre as nações. Como a opinião de *Pareto* se baseou nos resultados empíricos obtidos com o seu índice calculado para vários países europeus, *Corrado Gini* tenta mostrar que a pouca variabilidade era mais uma deficiência do instrumental que *Pareto* utilizara do que uma característica da distribuição das rendas dos países. Assim, as origens do Índice de Gini justificam porque as duas medidas da situação da distribuição da renda estão intimamente relacionadas.

O objetivo principal desta parte é divulgar o Índice de Gini que, apesar de ter surgido logo depois da medida de *Pareto*, vem sendo ofuscado por outra medida também proposta por *Gini*, a razão de concentração tão criticada e tão usada na literatura de distribuição da renda pessoal.

A fim de demonstrar a relação que existe entre α e δ necessitamos antes explicitar algumas hipóteses e simplificações envolvidas:

1. $N = Ax^{-\alpha}$ e $N_1 = A(x+1)^{-\alpha}$
2. N pessoas com renda maior que x
3. N_1 pessoas com renda maior que $x + 1$
4. $n = N - N_1$ número de pessoas com renda entre x e $x + 1$
5. A_x = renda total das N pessoas com renda superior a x
6. r = a renda total de n pessoas com renda compreendida entre x e $x + 1$

Além das hipóteses acima é também preciso admitir as seguintes aproximações:

$$1 - \left(\frac{x}{x+1}\right)^{\alpha} = \frac{\alpha}{x} \quad (7)$$

$$r = nx \quad (8)$$

Em seguida, multiplicamos relação (7) por $\frac{A}{x^{\alpha}}$ e obtemos:

$$\frac{A}{x^{\alpha}} - \frac{A}{x^{\alpha}}\left(\frac{x}{x+1}\right)^{\alpha} = \frac{A\alpha}{x^{\alpha}x} = \frac{A\alpha}{x^{\alpha+1}} \quad (9)$$

A partir de (4) temos, $n = N - N_1 = \frac{A}{x^{\alpha}} = \frac{A}{(x+1)^{\alpha}}$.

Considerando-se o resultado (9), $n = \frac{A}{x^{\alpha}} - \frac{A}{(x+1)^{\alpha}} = \frac{A\alpha}{x^{\alpha+1}}$ (10)

Linearizando a relação $r = nx$, temos: $\log r = \log n + \log x$. Dado

$$\log n = \log A\alpha - (\alpha + 1) \log x$$

e

$$\log n = \log r - \log x$$

obtemos:

$$\log A\alpha - (\alpha + 1) \log x = \log r - \log x$$

$$\log A\alpha - \alpha \log x - \log x = \log r - \log x$$

$$\log r = \log A\alpha - \alpha \log x$$

$$r = A\alpha\, x^{-\alpha} \quad (11)$$

Integrando (11) temos:

$$A_x = \int r\, dx = A\alpha \int x^{-\alpha}\, dx$$

$$= \frac{A\alpha}{\alpha - 1} \cdot x^{-(\alpha-1)}$$

Fazendo $\frac{A\alpha}{\alpha - 1} = T$ temos $A_x = Tx^{-(\alpha-1)}$ Linearizando:

$A_x = Tx^{-(\alpha-1)}$ obtemos $\log A_x = \log T - (\alpha - 1) \log x$ (12). Considerando a relação (12) e $\log N = \log A - \alpha \log x$, eliminando-se $\log x$ obtemos:

$$\log N = \frac{\alpha}{\alpha - 1} \log A_x + \log A - \frac{\alpha}{\alpha - 1} \log T.$$

Chamando de

$$\log K = \log A - \frac{\alpha}{\alpha - 1} \log T$$

temos

$$\log N = \log K + \frac{\alpha}{\alpha - 1} \log A_x$$

então

$$N = KA_x^{\frac{\alpha}{\alpha - 1}}$$

onde obtemos

$$\frac{\alpha}{\alpha - 1} = \delta$$

$$N = KA_x^{\delta}$$

Não há dúvida das íntimas relações entre a medida sugerida por *Pareto* e a medida desenvolvida por *Gini*. É preciso, no entanto, observar que a relação $\frac{\alpha}{\alpha - 1} = \delta$ é uma relação matemática, obtida a partir de algumas hipóteses e simplificações apontadas anteriormente (*Gini*, 1922).

Antes de descrevermos os resultados empíricos obtidos para os dois índices seria interessante que destacássemos algumas vantagens que δ apresenta sobre α. A principal vantagem reside na maior sensibilidade de δ, como foi registrado por *Corrado Gini*,[8] tendo sido ressaltado posteriormente por *Mary Jean Bowman* no seu trabalho clássico sobre a distribuição pessoal da renda nos Estados Unidos.[9] Outra vantagem diz respeito ao significado das medidas de desigualdade, ao contrário de α, δ tem um significado muito preciso. Ele é o expoente ao qual se deve elevar a fração da renda recebida para igualá-la à fração de pessoas que a recebe. Quando δ cresce, aumenta o grau de concentração, não havendo nenhuma ambigüidade na sua interpretação.[10]

II.4 — Resultados Empíricos

II.4.1 — Índice de Pareto

Nesta parte tentaremos comentar rapidamente os resultados obtidos para o Índice de Pareto no que se refere às diferenças do tamanho da amostra e do elemento representativo de cada classe no cálculo da medida de desigualdade de renda.

A primeira informação que se depreende da tabela 1 é de que ao passarmos de uma amostra de 1,3% para uma amostra de 25% o Índice de Pareto aumenta, qualquer que seja o elemento representativo da

TABELA 1

Valores do índice de Pareto segundo o tamanho da amostra e os elementos representativos das classes de renda

AMOSTRA / SUBAMOSTRA — ELEMENTOS REPRESENTATIVOS	AMOSTRA 25%	SUBAMOSTRA 1,3%	VARIAÇÃO	
			Absoluta	%
Ponto Médio (PM)	1,61065	1,59998	0,01067	0,6
Renda Média (RM)	1,84852	1,82204	0,02648	1,4
Limite Inferior (LI)	1,81118	1,79800	0,01318	0,7
Limite Superior (LS)	1.63947	1.62908	0,01039	0,6

FONTE: Dados do Censo Demográfico de 1970 — FIBGE.

8 Vide *Corrado Gini, Índice di Concentrazione e di Dipendenza*, 1922, op. cit.

9 Vide *Mary Jean Bowman, A Graphical Analysis of Regional Income Distribution*. AER. XXXV, setembro 1945, pp. 607/628.

10 No caso do coeficiente de *Pareto*, alguns autores e o próprio *Pareto* admitiam que a medida que o índice crescesse a concentração aumentaria, mas a relação $G = \frac{1}{2\alpha - 1}$ sugere justamente o contrário, quanto maior o valor de α menor a concentração. No caso do Índice de Gini não existe nenhuma discordância sobre o que ele representa.

classe usado. O aumento do índice de Pareto significa uma diminuição da renda,[11] ou seja, pequenas amostras tendem a exagerar o grau de desigualdade. Apesar de ser flagrante o fato de que quanto maior a amostra menor a desigualdade, é preciso observar também que as diferenças são muito pequenas, no máximo atingindo a segunda casa decimal. Elas não vão além de 2%, quando passamos de uma amostra para outra. No caso em que elemento representativo é a renda média, a diferença é bem maior. Esta maior diferença pode ser explicada pelo fato de que ao se mudar de amostra muda-se também as médias, assim como o total de pessoas em cada classe. Ao passo que no caso dos outros elementso representativos (ponto médio, limite inferior e limite superior), somente varia o número de pessoas. Esta observação é importante, já que ela deve ser válida no caso de qualquer medida de desigualdade. *A priori,* espera-se que no caso do elemento representativo ser a renda média, as diferenças amostrais introduzam diferenças mais intensas no valor da medida de desigualdade.

Outras informações sobre as mudanças no valor da medida de desigualdade da renda são oferecidas pelo exame da tabela 2, onde procuramos apontar as diferenças que surgem em função da utilização de diferentes elementos representativos no cálculo do coeficiente de *Pareto.* A maior diferença surge quando comparamos os valores de α obtidos a partir do ponto médio e da renda média, como elementos representativos das classes de renda. A diferença, neste caso, atinge valores relativamente grandes, chegando a casa dos 14% de variação tanto para a amostra de 25% como para a amostra de 1,3%. Com base nas informações da tabela 1 podemos verificar que o índice de Pareto difere até a primeira casa decimal, isto é, quando α foi calculado utilizando o ponto médio da classe como elemento representativo, obtivemos 1,61065 para a amostra de 25% e 1,59998 para a subamostra de 1,3%. Quando a renda média foi usada como elemento representativo os valores de α foram 1,84852 e 1,82204, respectivamente, para a amostra e subamostra. Observa-se também que o ponto médio e a renda média são responsáveis pelos valores extremos obtidos para α. O ponto médio, quando usado como elemento representativo da classe de renda, dá origem aos menores valores de α, tanto na amostra como na subamostra, ao passo que a renda média, aos maiores valores de α. Os demais elementos representativos oferecem valores de α entre estes extremos, sendo que o limite inferior se aproxima de α obtido com a renda média e o limite superior de α obtido com o ponto médio.

No caso do coeficiente de Pareto, o ponto médio e o limite superior superestimam as informações da desigualdade de renda, oferecendo valores menores de α. Já a renda média e o limite inferior dão origem à valores de α que mostram um menor grau de desigualdade. A tabela 2 apresenta todas as combinações possíveis 2 a 2 das diferenças entre os valores de α obtidos conforme o indicador utilizado como elemento representativo de cada classe.

Parece que a partir das informações existentes nas tabelas 1 e 2 podemos derivar pelo menos duas conclusões importantes de caráter metodológico para quem precisa utilizar o coeficiente de Pareto como medida de desigualdade. A primeira delas é que as diferenças amostrais são pequenas. Não obstante, o índice de Pareto apresenta uma direção bem definida, independente do elemento representativo das classes considerado. Em outras palavras, quando passamos de uma amostra me-

11 Para uma explicação sobre a controvérsia a respeito do significado de α vide o trabalho de *Corrado Gini, Indice di Concentrazione e di Dipendenza,* 1922, op. cit.

TABELA 2

Variações do Índice de Pareto segundo o tamanho da amostra e os elementos representativos utilizados na representação das classes de renda

VARIAÇÕES / AMOSTRA E SUBAMOSTRA	[PM, RM]		[PM, LI]		[PM, LS]	
	Absoluta	%	Absoluta	%	Absoluta	%
Amostra (25%)	0,23787	19,77	0,20053	12,45	0,02882	1,78
Subamostra (1,3%)	0,22206	13,87	0,19802	12,37	0,02910	1,8

VARIAÇÕES / AMOSTRA E SUBAMOSTRA	[RM, LI]		[RM, LS]		[LI, LS]	
	Absoluta	%	Absoluta	%	Absoluta	%
Amostra (25%)	0,03734	2,0	0,20905	11,30	0,17171	9,48
Subamostra (1,3%)	0,02404	1,3	0,10296	5,65	0,16892	9,39

FONTE: Censo Demográfico de 1970 — FIBGE (só valores absolutos).

nor para uma maior o coeficiente de Pareto aumenta, isto é, a desigualdade diminui. A outra conclusão é de que o uso de diferentes elementos representativos introduzem diferenças significantes nos valores do coeficiente de Pareto. A renda média e o limite inferior tendem a aumentar o valor de α, ao passo que o ponto médio e o limite superior das classes tendem a diminuir o valor de α, por conseguinte, diminuir e aumentar a desigualdade.

II.4.2 — Índice de Gini

Os resultados para o caso do Índice de Gini podem ser observados nas tabelas 3 e 4. A primeira enfatiza as diferenças entre as amostras. A segunda ressalta as diferenças entre os vários métodos de computação.

A tabela 3 mostra como os efeitos da mudança de tamanho de uma amostra não são os mesmos para todos os métodos de computação do Índice de Gini. Ao passarmos de uma pequena amostra para uma grande amostra, o ponto médio e a renda média sendo escolhidos como os elementos representativos das classes de renda, apresentam uma diminuição na desigualdade da renda. Esta diminuição é mais intensa no caso do ponto médio chegando a casa dos 4%. No caso da renda média esta redução fica em torno dos 2%. Quando os elementos representativos da classe de renda são os limites inferior e o limite superior observa-se um aumento na desigualdade ao passarmos de uma amostra pequena para uma maior. No entanto, este aumento é quase que desprezível, não ultrapassando a casa dos 0,3%. Neste particular o Índice de Gini difere um pouco do coeficiente de Pareto, já que os efeitos da mudança amostral não apresentam a mesma regularidade observada para o coeficiente de Pareto.

Outra diferença entre os resultados obtidos para o Índice de Gini e o coeficiente de Pareto refere-se aos valores extremos do grau de concentração da renda. O coeficiente de Pareto apresentou a renda média e o limite inferior como os elementos representativos que proporcionaram os maiores valores do coeficiente de Pareto. Agora, com o Índice de Gini ocorre justamente o contrário. Os mesmos elementos representativos são responsáveis pelos menores valores do Índice de Gini. Isto é, a renda média e o limite inferior tendem a subestimar o valor do grau de concentração, ao passo que o limite superior e o ponto médio a superestimar, justamente o contrário do que se verificou anteriormente para os resultados do coeficiente de Pareto.

TABELA 3

Valores do Índice de Gini segundo o tamanho da amostra e os elementos representativos das classes de rendas

AMOSTRA SUBAMOSTRA / ELEMENTOS REPRESENTATIVOS DAS CLASSES DE RENDA	AMOSTRA 25%	SUBAMOSTRA 1,3%	VARIAÇÃO	
			Absoluta	%
Ponto Médio (PM)	1,65306	1,73527	—0,082210	4,97
Renda Média (RM)	1,53360	1,56511	—0,031510	2,05
Limite Inferior (LI)	1,43835	1,43662	0,001730	0,12
Limite Superior (LS)	1,90017	1,89768	0,002490	0,13

FONTE: Censo Demográfico de 1970 — FIBGE.

A influência das diferentes maneiras de se calcular o Índice de Gini pode ser observada através da tabela 4. O par de elementos representativos que apresentam maior discrepância no cálculo do Índice de Gini é aquele formado pelo limite inferior e o limite superior. Tal discrepância chega a atingir o nível de 32% de diferença, demonstrando a importância das diferenças metodológicas no cálculo das medidas de desigualdade. O par que apresentou menor discrepância foi aquele formado pela renda média e o limite inferior, a discrepância neste caso atingindo os níveis de 6% a 8%, que é bem superior ao máximo observado para as diferenças em função do tamanho da amostra.

Novamente, para o Índice de Gini, observamos que a diferença em relação ao tamanho da amostra não constitui um elemento que introduz grandes discrepâncias. Mas as diferenças da metodologia no cálculo das medidas de desigualdade parecem de suma importância, especialmente no caso do Índice de Gini para o qual os valores das discrepâncias atingiram percentuais muito grandes e bem maiores do que os observados no caso do coeficiente de Pareto. Não há dúvida que a alegada insensibilidade do coeficiente de Pareto deve contribuir muito para a reduzida variação nos valores deste índice tanto em função do tamanho da amostra como das diversas maneiras de calcular. [12]

Os dois exemplos anteriormente descritos devem também ser qualificados no que se refere ao método geral de obtenção das duas medidas

12 Sobre a insensibilidade do coeficiente de *Pareto* vide *Mary Jean Bowman — A Graphical Analysis of Personal Income Distribution*, op. cit. e *Corrado Gini, Índice di Concentrazione*, op. cit.

de concentração. Ambas estão sujeitas ao ajustamento das respectivas curvas, o que também se constitui numa outra fonte de diferença, apesar das duas funções proporcionarem graus de ajustamento sempre em torno dos 90%.[13]

Apesar da limitação do experimento, estes dois exemplos deixam bem claro que as diferentes maneiras de calcular os índices são elementos que não devem ser esquecidos quando houver necessidade de comparação no tempo, entre nações ou entre regiões geográficas. As diferenças são enormes, principalmente no caso do Índice de Gini, que se mostrou extremamente sensível ao uso de diferentes elementos representativos das classes de renda. As diferenças em função do tamanho da amostra não foram de grande monta, não ultrapassando níveis de 5% a 6%, perfeitamente aceitáveis do ponto de vista prático.

Em seguida verificaremos como se comportam algumas medidas de desigualdade que independem do ajustamento a *priori* de curvas, isto é, da forma da distribuição em questão como, por exemplo, a Razão de Concentração de *Gini* e o Índice de Theil.

III — ÍNDICES QUE NÃO DEPENDEM DE AJUSTAMENTO

Até o momento verificamos a influência do tamanho da amostra e das diversas maneiras de se calcular medidas de desigualdade obtidas a partir do ajustamento de uma função à distribuição da renda em ques-

TABELA 4

Variações do Índice de Gini segundo o tamanho da amostra e os elementos representativos utilizados na representação das classes de renda

VARIAÇÕES / AMOTRA E SUBAMOSTRA	[PM, RM]		[PM, LI]		[PM, LS]	
	Absoluto	%	Absoluto	%	Absoluto	%
Amostra (25%)	0,11946	7,2	0,21471	12,99	0,24711	14,95
Subamostra (13%)	0,17016	9,8	0,29865	17,21	0,16241	9,36

VARIAÇÕES / AMOSTRA E SUBAMOSTRA	[RM, LI]		[RM, LS]		[LI, LS]	
	Absoluto	%	Absoluto	%	Absoluto	%
Amostra (25%)	0,09525	6,2	0,36657	23,90	0,46182	32,10
Subamostra (1,3%)	0,12849	8,20	0,33257	21,24	0,46106	32,09

FONTE: Censo Demográfico de 1970 — FIBGE (só valores absolutos).

13 Para o coeficiente de *Pareto,* no caso da subamostra e renda média, o resultado do agrupamento foi $\log N_x$ com o coeficiente de determinação $R^2 = 0{,}963$. Para o índice de Gini, nas mesmas condições, o ajustamento foi $\log N_x = -15{,}67887 + 1{,}56511 \log A_x$ com o coeficiente de determinação $R^2 = 0{,}97587$. Os valores de $\alpha = -1{,}82204$ e $\delta = 1{,}56511$ podem ser observados respectivamente nas tabelas 1 e 3.

tão. O fato deles dependerem de um bom ajustamento das respectivas funções aos dados constitui uma nova fonte de variação que se deve levar em conta quando da utilização destes índices.

Nesta parte tentaremos observar o comportamento de duas medidas de desigualdade de renda que independem, *a priori,* de ajustamento de qualquer função aos dados. Verificaremos o caso da razão de Concentração de Gini e o Índice de Theil. O primeiro devido a sua grande popularidade e o uso muito difundido como indicador do grau de desigualdade da renda das pessoas, numa série de tempo ou *cross-section.* O segundo em virtude de sua recente introdução nos estudos da distribuição de renda do Brasil.[14] Estas duas medidas independem da forma da distribuição da renda e foram experimentados nas mesmas condições dos Índices de Pareto e Gini. Antes da apresentação dos resultados faremos uma descrição do significado da Razão de Concentração de Gini e do Índice de Theil.

III.1 — Razão de Concentração de Gini

A Razão de Concentração de Gini é, sem dúvida, a medida de desigualdade mais conhecida, usada e criticada nos estudos da distribuição da renda pessoal, tendo sido calculada para a maioria dos países existentes (*Paukert,* 1967). A pesar de sua divulgação ter sido apresentada praticamente como uma medida vinculada à Curva de Lorenz, a Razão de Concentração de Gini tem raízes bem estabelecidas do ponto de vista estatístico.

Estatisticamente, a Razão de Concentração de Gini é definida como uma medida de dispersão relativa, ou seja, a razão entre a média das diferenças e o dobro da média aritmética:

$$G = \frac{\triangle}{2\mu}$$

Considerando-se $\triangle$ como sendo a média das diferenças, isto é:

$$\Delta = \frac{\sum\limits_{i=i} \sum\limits_{j=1} \left| X_i - X_j \right| f_i \; f_j}{N^2}$$

Não há dúvida, portanto, que a Razão de Concentração é uma medida cujas bases estatísticas são relativamente sólidas, podendo ser comparada ao do Coeficiente de Variação. Trata-se, portanto, de um índice adimensional cujo limite inferior é zero e o limite superior é igual a um.

Recentemente, o reconhecimento destas características da Razão de Concentração levou *Joseph L. Gastwirth* a superar algumas de suas deficiências mais comumente apontadas. Primeiro foi a eliminação da crítica de subestimação da Razão de Concentração, apresentando um limite superior para tal medida (*Gastwirth,* 1972). Em seguida permitindo a decomposição da desigualdade entre classes e intraclasses (*Soltow,* 1960).

Em verdade, parece fora de dúvida de que a interpretação da Razão de Concentração como uma medida de dispersão não foi muito divulgada na literatura econômica anglo-saxônica. Mas se observarmos as contribuições dos estatísticos italianos, nas primeiras três décadas do

14 Vide *C. G. Langoni,* Distribuição da Renda e Desenvolvimento Econômico do Brasil, Rio de Janeiro, Expressão e Cultura e *Albert Fishlow,* Brazilian Size Distribution of Income AER, LXII, maio, 1972 pp. 391-402.

século XX, principalmente em publicações na *Revista Metron*, internacionalmente conhecida, observaremos que o cálculo da média das diferenças (ou diferença média) foi motivo de muita discussão. Estas discussões giravam em torno do objetivo de simplificar sua obtenção, já que se tratava de uma medida de dispersão que, ao invés de aferir as diferenças em relação a uma medida de tendência central (como a variância), pretendia aferir todas as diferenças possíveis entre os atributos, o que dificultava, em demasia, o seu cálculo (*Finetti*, 1931). A Razão de Concentração de Gini não está necessariamente vinculada à Curva de Lorenz, ela tem um significado estatístico próprio, ou seja, uma medida de dispersão relativa representada pela relação:

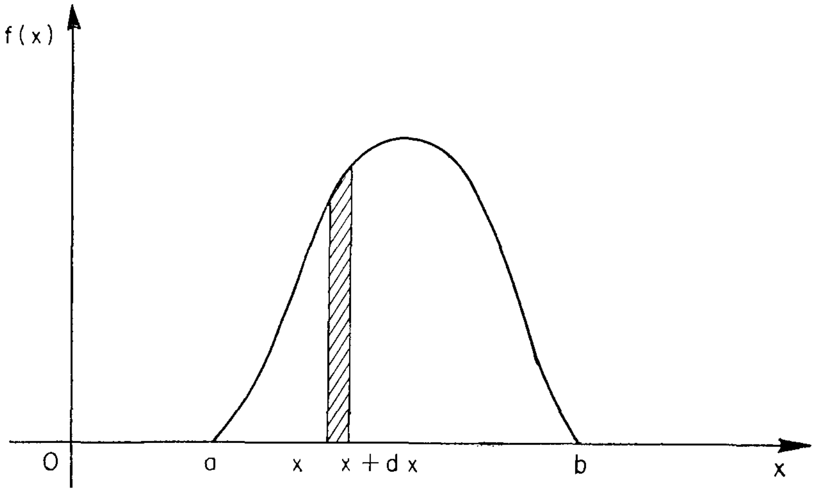
$$\frac{\Delta}{2\mu} = G$$

III.1.1 — Curva de Lorenz

A Razão de Concentração de Gini tem sido muito divulgada como a razão entre a área de concentração real e a área de concentração máxima, em termos da curva de Lorenz. Tal fato obscurece um pouco o seu significado estatístico, apesar de facilitar enormemente a compreensão e, por conseguinte, a sua ampla divulgação e utilização como medida de desigualdade.

Historicamente, o advento da curva de Lorenz como um instrumento adequado para observações de variações no grau de concentração de um atributo (renda), foi introduzido através dos trabalhos de *Lorenz* (*Lorenz*, 1905), *Chatelain* (*Chatelain*, 1907), *Seailles* (*Seailles*, 1910) e *Corrado Gini* (*Gini* 1914). Todos estes autores foram contemporâneos, no entanto *M. O. Lorenz* tornou-se mais conhecido, tendo a curva de concentração em questão recebido o seu nome.[15]

Antes de apresentarmos a curva de Lorenz na sua forma mais simples, seria interessante dar-lhe uma apresentação mais técnica. Para isso assumimos um atributo quantitativo e variável no intervalo [a, b] tal que $a < b$. Seja $f(x)$ a função densidade de freqüência e $f(x)\,dx$ a freqüência com que o atributo assume valores no intervalo $x + dx$. Graficamente temos:



15 Vide *M. O. Lorenz*, *Methods of Measuring the Concentration of Wealth*, ASA Nova Série, junho 1965, 209.

Considerando x o atributo e p e r como coordenadas cartesianas de um ponto qualquer P da curva de Lorenz, temos, portanto, as respectivas equações:

$$p = \frac{\int_a^x f(x)\,dx}{\int_a^b f(x)\,dx} \quad (1) \quad \text{e}$$

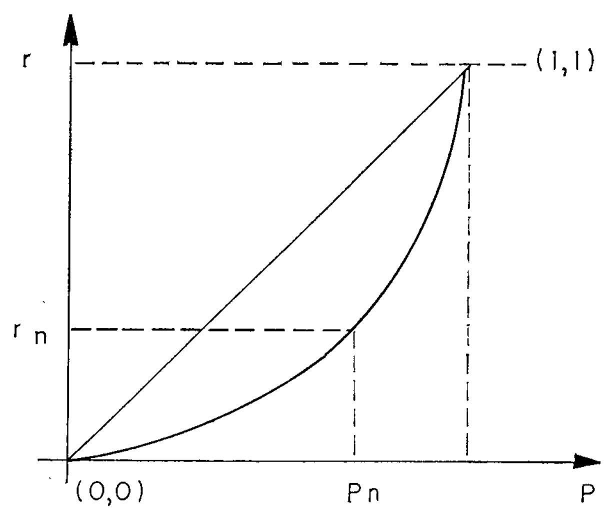
$$r = \frac{\int_a^x \times f(x)\,dx}{\int_a^b \times f(x)\,dx} \quad (2)$$

A equação (1) é a razão entre o momento incompleto e completo de ordem 0 da função densidade f(x). Ao passo que a equação (2) representa a razão entre momento incompleto e completo de ordem 1 da mesma função. Portanto, podemos escrever p e r como segue:

$$p = \frac{m_0^{(x)}}{m_0} = \frac{\int_a^x f(x)\,dx}{N} \quad \text{e}$$

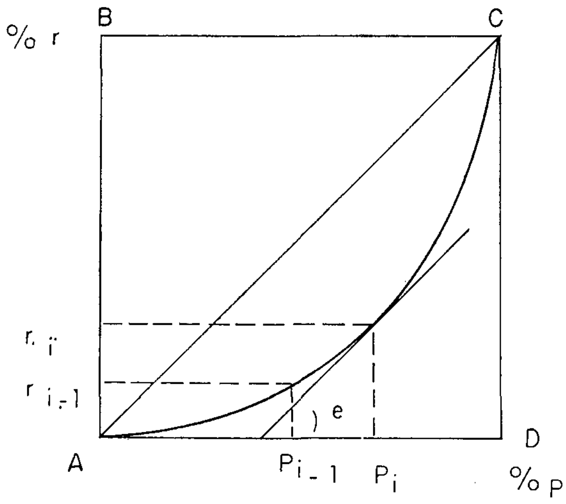
$$r = \frac{m_1\,(x)}{m_1} = \frac{\int_a^x \times f(x)\,dx}{N\overline{X}}$$

Onde N e X são respectivamente a freqüência total e a média aritmética do atributo x; evidentemente NX representa o total do atributo da distribuição em questão. A curva de Lorenz, que é representada num diagrama cartesiano tendo como abcissa p e como ordenada r, é uma curva convexa em relação a p, cujos valores correspondentes aos limites a e b são respectivamente (0,0) e (1,1). Fazendo-se o gráfico correspondente temos:



Pela simples observação do gráfico concluímos que $P_i > r_i$ para todos os pontos com exceção dos extremos onde $P_i = r_i$ (*Galvani,* 1932). Nestas condições, fica bem claro que a curva de Lorenz existe para a distribuição de qualquer atributo, não havendo necessidade de que x seja renda, apesar de *Lorenz* ter sugerido tal curva como a forma mais adequada para observar mudanças no grau de concentração da renda. (*Lorenz,* 1905)

Com o intuito de simplificar o entendimento da curva de Lorenz através da eliminação das formalidades matemáticas e estatísticas, como foi feito no trabalho pioneiro de Lorenz, tentaremos expor rapidamente o significado da curva de Lorenz em termos mais simples. Consideremos uma distribuição de freqüências para o atributo x com n classes. As freqüências são f_1, f_2 ... f_n e os respectivos elementos representativos de cada classe são x_1, x_2 ... x_n. A curva de Lorenz relaciona os percentuais acumulados das rendas com os percentuais acumulados das pessoas que recebem tais rendas, a partir dos níveis de renda mais baixos. Em geral, utiliza-se um quadrado de lado igual a unidade para representar a curva de Lorenz.



Nesta figura, AD representa o eixo das abcissas no qual registramos

$p_i = \dfrac{\sum_{i<n} f_i}{N}$ o percentual acumulado das pessoas em ordem crescente

da renda recebida. O lado AB do quadro representa o eixo das ordenadas

onde $r_i = \dfrac{\sum_{i<n} f_i \, x_i}{\sum_{i=1}^{n} f_i \, x_i}$ o percentual das recebidas pelas pessoas

registradas na abcissa. O conjunto dos pares ordenados (r_i, p_i) representa o que chamamos de curva de Lorenz. No seu trabalho original, *M. O. Lorenz* colocou r_i nas abcissas e p_i nas ordenadas, mas isto não altera em nada as propriedades da curva. Apenas a curva de Lorenz ficará acima da diagonal AC (*Lorenz,* 1905).

III.1.2 — Algumas Propriedades da Curva de Lorenz

Esta curva tão utilizada quanto criticada desde o advento de sua divulgação por *M. O. Lorenz,* em 1905, apresenta muitas propriedades que a tornam quase indispensável para aqueles que estudam as características de um dado atributo, mormente a renda. Enunciaremos algumas das principais características analíticas e práticas relacionadas com a curva de Lorenz:

1. É uma curva não decrescente e convexa em relação aos percentuais acumulados de pessoas que recebem renda;

2. Com exceção dos dois primeiros e últimos pares ordenados, r_i será sempre menor que p_i, no caso do atributo x apresentar uma dispersão diferente de zero;

3. A inclinação da Curva de Lorenz em qualquer ponto é igual ao limite da razão $\frac{r_i - r_{i-1}}{p_i - p_{i-1}}$ (*Ames,* 1941).

4. Observando-se que a curva de Lorenz é convexa, percebe-se que existe um ponto em que a sua inclinação é paralela à diagonal AC. Justamente neste ponto a inclinação é igual à unidade, ou seja, o percentual de renda é igual ao percentual de pessoas na classe em questão. Esta propriedade permite identificar, aproximadamente, a classe que contém a renda média (x) (*Ames,* 1941).

5. A curva de Lorenz possibilita a visualização das modificações na distribuição da renda através do tempo ou as comparações, no mesmo período, entre diferentes regiões.

6. Quando a curva de Lorenz se reduz à diagonal AC, temos que $p_i = r_i$ para todo i. Estamos diante de um caso extremo de perfeita igualdade. Quando a curva de Lorenz se reduz ao triângulo ACD, onde $p_i = r_i = 0$, $p_j = r_j = 1$ para $j \neq i$, então estamos em outro extremo, o de máxima desigualdade, uma pessoa (ou família) com toda a renda. Estes dois extremos são apenas pontos de referência, pois, na realidade, o que existe mesmo é uma curva convexa como AEC. Quanto mais próxima AEC estiver de AC menor a concentração, quanto mais afastada de AC maior a concentração.

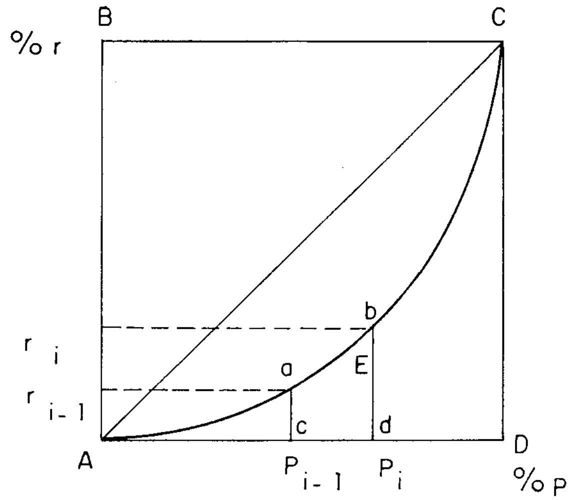
7. Apesar de suas propriedades e indicações sobre o que se passa com a distribuição da renda, a curva de Lorenz apresenta situações de ambigüidade quando existe o cruzamento de duas curvas. A distinção é feita pela área entre a diagonal AC e a curva de Lorenz AEC, isto é, a área de concentração. Quando as áreas de concentração são iguais temos a situação de duas distribuições com a mesma dispersão relativa (*Costa,* 1974).

Não podemos nos alongar na descrição da curva de Lorenz, pois não é o objetivo do trabalho em questão. A curva de Lorenz é um instrumento cujas características e propriedades ainda não foram completamente discutidas num só trabalho. O material aqui apresentado é o suficiente para compreendermos melhor o significado da Razão de Concentração de Gini quando derivada diretamente da curva de Lorenz. Porém, a curva de Lorenz, pela sua utilidade, mereceria um trabalho a parte se quiséssemos esgotar todas as informações a seu respeito.

III.1.3 — A Razão de Concentração de Gini e a Curva de Lorenz

Juntando as informações sobre a Razão de Concentração de Gini e sobre a curva de Lorenz podemos mostrar rapidamente como se deriva uma medida prática da Razão de Concentração de Gini. Esta derivação é a forma mais simples de mensuração da Razão de Concentração de Gini. O método aqui apresentado se deve a *James Morgan,* um estudioso da distribuição de renda (*Morgan,* 1960).

Consideremos, como antes, um quadrado ABCD com lados iguais à unidade, conforme figura abaixo:



Onde AC representa uma situação de perfeita igualdade. A área ACD = 1/2 representa uma situação de máxima desigualdade. Ao passo que a área ACE, entre a diagonal AC e a curva de Lorenz AEC, representa a área de concentração real, a qual pretendemos medir para construir um índice que representa a razão entre esta área de concentração real ACE e a área máxima de concentração ACD, ou seja:

$$G = \frac{\text{área ACE}}{\text{área ACD}} = \frac{\text{concentração real}}{\text{concentração máxima}}$$

Calculando a área AECD e subtraindo-a da área ACD obteremos a área ACE, ou seja:

$$G = \frac{\text{área ACD} - \text{área AECD}}{\text{área ACD}} =$$

$$= \frac{1/2 - \text{área AECD}}{1/2} = 1 - 2 \text{ área AECD}$$

Então a fórmula prática para a Razão de Concentração é obtida após o cálculo da área AECD. Mas esta área pode ser avaliada dividindo-a em pequenos trapézios do tipo abcd. A soma total das áreas de todos estes pequenos trapézios resultará na área AECD, um pouco superestimada porque o segmento ab, em verdade, não é uma reta, mas sim uma curva [16] (*Hoffman*, 1971). A conseqüência disto é a obtenção de um valor para G menor do que realmente se verifica. Quando o número de classes é muito grande, esta diferença não chega a constituir um grande obstáculo para aqueles que buscam apenas uma medida aproximada, sem maiores ambições.

16 Vide *R. Hoffman, Contribuição à Análise da Distribuição da Renda e da Posse da Terra no Brasil.* Dissertação de Doutorado E.S.A.L. *Queiroz.* 1971.

Aplicando a definição da área de um trapézio temos que:

$$\text{área abcd} = \frac{bd + ac}{2} \,.\, cd$$

Mas,

$$bd = r_i,\ ac = r_{i-1} \quad e \quad cd = p_i - p_{i-1},$$

ou seja:

$$\text{área abcd} = \frac{r_i + r_{i-1}}{2} \cdot (p_i - p_{i-1})$$

Como temos k trapézios deste tipo, a área

$$AECD = \sum_{i=1}^{k} \frac{(r_i + r_{i-1})}{2} \cdot (p_i - p_{i-1})$$

em termos das coordenadas da curva de Lorenz. Por conseguinte

$$G = 1 - 2 \frac{1}{2} \sum_{i=1}^{k} (r_i + r_{i-1}) \ (p_i - p_{i-1})$$

$$G = 1 - \sum_{i=1}^{k} (r_i + r_{i-1}) \ (p_i - p_{i-1})$$

Esta é a fórmula que se utiliza, na prática, quando se quer obter o valor da Razão de Concentração de Gini. Os limites da Razão de Concentração de Gini ficam bem evidentes. $G = 0$ quando a área de concentração ACE for igual a zero, ou seja, a igualdade perfeita; $G = 1$ quando a área de concentração ACE for igual a ACD, ou a máxima desigualdade.

Existem outras fórmulas práticas para a avaliação numérica de G, todas elas fornecem valores aproximados de G mais ou menos equivalentes. A nossa preocupação nesta descrição foi mostrar como a Razão de Concentração de Gini tem raízes estatísticas bem fundamentadas, não sendo simplesmente um número que se obtém dividindo a área de concentração real pela área de concentração máxima condicionada pela curva de Lorenz. É uma medida de dispersão relativa, como o Coeficiente de Variação e outras medidas de desigualdade, não se podendo confundir o método aproximado de cálculo com o verdadeiro significado do índice.

Esta medida também foi utilizada no exercício que estamos realizando para verificação das variações nos valores das medidas de desigualdade de renda em função do tamanho da amostra e do método do cálculo, que utiliza diferentes elementos representativos das classes. Após este longo esclarecimento ainda resta observar as peculiaridades que a Razão de Concentração de Gini apresenta nas circunstâncias em que o experimento vem sendo realizado. Antes descreveremos rapidamente o Índice de Theil, também incluído no experimento como uma das medidas que independem da forma da distribuição.

III.2 — Índice de Theil

O Índice de Theil é uma medida de desigualdade que independe da forma da distribuição estudada. Tem sua origem na Teoria da Informação.[17] Tem sido usado em Economia como uma medida do grau de desigualdade de renda.[18] Não há dúvida de que qualquer informação sobre o comportamento deste índice, nas mesmas condições em que foram experimentados as demais medidas de desigualdade de renda, será importante quando a preocupação for uma utilização mais adequada dos indicadores.

Veremos que a idéia de desigualdade contida no Índice de Theil é relativamente simples, no entanto o seu entendimento fica condicionado à compreensão de alguns conceitos básicos da Teoria de Informação. Como resultado desta dependência da Teoria de Informação, o Índice de Theil também exige certa familiaridade na manipulação de logaritmos. Portanto, antecipadamente, percebe-se que não se trata de uma medida de fácil penetração popular; mesmo entre os profissionais de economia, já que Teoria de Informação não faz parte do *curriculum* básico e logaritmo, infelizmente, é mal aprendido e, às vezes, subestimada a sua utilização pelo economista. Nestas circunstâncias, fica claro que o Índice de Theil será pouco entendido pelo público acadêmico ou não acadêmico que se interessa e acompanha as discussões sobre os problemas da distribuição da renda.

Os conceitos básicos [19] da Teoria de Informação que se fazem necessários para chegarmos ao Índice de Theil são os seguintes: Conteúdo de Informação, Entropia ou Informação Esperada e Informação Esperada de uma Mensagem Indireta (*Costa*, 1975). Nos restringiremos aqui às idéias emitidas por *Henry Theil* no seu livro *Economics and Information Theory*. Qualquer divergência de nomenclatura sobre os conceitos básicos da Teoria de Informação deve ser encarada com certa tolerância por se tratar de área relativamente recente de conhecimento (*Shannon*, 1949). Para ilustrar tal situação reproduziremos a observação feita por *Elnyn Edwards* em 1964: *"Infelizmente a momenclatura de Teoria de Informação não está padronizada. Como observaram McGill* e *Quastler* (1965): *'Constitui uma ironia o fato de a Teoria de Informação colocar problemas de comunicação para os que a utilizam! O que chamamos de 'Incerteza' foi denominado 'Entropia' por Shannon* (1948), ...".

Tentaremos descrever o que se entende por Conteúdo de Informação. Dado um evento E, cuja probabilidade de acontecer é igual a p, chamamos de Conteúdo de Informação da mensagem, afirmando que E aconteceu a uma função inversamente proporcional a sua probabilidade. No caso da Teoria de Informação utiliza-se uma função log na base 2 e, em alguns casos, log na base e. Logo, o conteúdo de informação de um evento E ter acontecido é dado por:

$$h(p) = \log \frac{1}{p} = -\log p, \text{ onde } 0 < p < 1$$

17 Vide *Henry Theil* — Economics and Information Theory, Amsterdam, North, Holland Publishing co., 1967, página 123.

18 Vide *Carlos Geraldo Langoni*, Distribuição da Renda e Desenvolvimento Econômico do Brasil, Rio de Janeiro, Expressão e Cultura, 1973.

19 Estes conceitos já foram apresentados em outro trabalho de *Ramonaval A. Costa*, "Medidas de Desigualdade de Renda" — *Boletim Geográfico* n.º 238, jan./fev. de 1974, Ano 33 pp. 45/72.

Quando a base dos logaritmos é 2 as unidades de conteúdo de informação denominam-se *bits,* quando a base é de logaritmos neperianos, as unidades são chamadas de *nits.* Na área de comunicação, onde a Teoria de Informação é muito aplicada, as unidades mais comumente encontradas são *bits.* Quando da aplicação dos conceitos da Teoria de Informação em outras áreas, como Economia, têm-se utilizado os logaritmos neperianos com mais freqüência (*Theil,* 1967).

O outro conceito básico da Teoria de Informação relacionado com o Índice de Theil é o que se chama de Entropia ou Informação Esperada de várias mensagens. Neste caso teríamos um conjunto de eventos $E_1, E_2 \ldots E_n$ com as seguintes probabilidades $p_1, p_2 \ldots p_n$. A Entropia não é nada mais nada menos que o valor esperado dos respectivos conteúdos de informação, ou seja:

$$H\,(p) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} = - \sum_{i=1}^{n} p_i \log p_i$$

No caso da Entropia observamos uma distribuição, cujos valores do atributo são os eventos E_i e as respectivas probabilidades são os valores de p_i. Juntamente no caso deste conceito é que se tem algumas discordâncias de nomenclatura. *Edwards* (1964) chamou-o de grau de Incerteza, *Shannon* (1948) deu o nome de Entropia, ao passo que *Wiener* (1948) preferiu chamá-lo de *"Entropia negativa"* (Edwards, 1964). O que interessa no caso em questão é que se compreenda H(p) como um valor médio do conteúdo de informação de vários eventos.

Finalmente o último conceito que se faz necessário para uma melhor compreensão do Índice de Theil chama-se Informação Esperada de uma Mensagem Indireta. Consideremos $E_1, E_2 \ldots E_n$ um conjunto de eventos e seja $p_1, p_2 \ldots p_n$ suas probabilidades *a priori,* ou a probabilidade deles acontecerem sem nenhuma outra informação. Seja $t_1, t_2 \ldots t_n$ as respectivas probabilidades *a posteriori,* ou as probabilidades corrigidas depois dos eventos terem acontecido (*Theil,* 1967). Podemos calcular facilmente o conteúdo de informação de uma mensagem indireta para cada evento E_i levando-se em consideração as duas informações sobre as probabilidades p_i e t_i:

$$h\,(p_i) - h\,(t_i) = \log \frac{1}{p_i} - \log \frac{1}{t_i} = \log \frac{t_i}{p_i}$$

Agora podemos obter o conceito Informação Esperada de uma mensagem Indireta. Basta calcular, como anteriormente, o valor esperado de $\log \frac{t_i}{p_i}$, ou seja, do conteúdo de informação de uma mensagem indireta. A única diferença que surge é que as ponderações tanto podem ser as possibilidades *à priori* como as probabilidades *à posteriori,* ou seja:

$$I\,(t_i : p) = \sum_{i-1}^{n} t_i \log \frac{t_i}{p_i}$$

ou

$$I\,(p, t) = \sum_{i=1}^{n} p_i \log \frac{t_i}{p_i}$$

Apontados estes três conceitos poderemos descrever com mais clareza o que chamamos de Índice de Theil no nosso experimento.

III.2.1 — Entropia, Índice de Theil ou Redundância

Os conceitos anteriormente descritos — quando aplicados ao estudo do grau de desigualdade de uma distribuição de rendas — nos fornecem uma medida de desigualdade de renda para informações individuais de rendas e outra medida para as situações em que os dados de renda estão grupados em classes.

Consideremos o primeiro caso. Seja $r_1, r_2 \ldots r_n$ as frações da renda individual entre N indivíduos. Tomando-se o conceito de Entropria e r_1 a fração da renda total em poder do indivíduo i, obtemos facilmente uma medida do grau de igualdade entre as diversas frações de renda, ou seja:

$$H(r) = \sum_{i=1}^{N} r_1 \log \frac{1}{r_i}$$

Para mostrarmos como a Entropia nos fornece uma medida de igualdade, basta examinarmos os dois casos extremos relacionados com a distribuição de rendas. O primeiro caso é o de perfeita igualdade, onde r_1 seria igual para todos os indivíduos $i = 1, 2 \ldots N$. Nesta situação, o valor de r_i seria igual a $\frac{1}{N}$ e o valor da Entropia seria igual a:

$$H(r) = \sum_{j=1}^{N} \frac{1}{N} \log \frac{1}{1/N} = N \cdot \frac{1}{N} \log N = \log N$$

A máxima igualdade ou máxima Entropia seria dada por log N. É necessário observarmos que a máxima igualdade depende do número de indivíduos envolvidos na distribuição (*Young,* 1935).

O segundo caso é o de máxima desigualdade, quando um único indivíduo estaria de posse de toda a renda, ou seja $r_i = 0$ para todo $i \neq j$, ao passo que $r_j = 1$. Nestas circunstâncias a Entropia se reduz a um único termo:

$$H(r) = r_j \log \frac{1}{r_j} = 1 \cdot \log \frac{1}{1} = \log 1 = 0$$

A máxima desigualdade significa Entropia zero. Na realidade, nunca lidamos com nenhum dos extremos, portanto o valor da Entropia, no caso da informação individual de renda, está sempre compreendido entre log N e 0:

$$\log N < H(r) < 0$$

Uma medida de desigualdade de renda pode ser obtida, imediatamente, a partir da medida de Entropia de uma distribuição. Basta subtraírmos da Entropia máxima, log N, o valor da verdadeira Entropia e obteremos uma medida do grau de desigualdade das rendas, isto é:

$$T(r) = \log N - H(r) = \log N - \sum_{i=1}^{N} r_i \log \frac{1}{r_i} =$$

$$= \log N + \sum_{i=1}^{N} r_i \log r_i = \sum_{i=1}^{N} r_i \log N r_i$$

Tal medida é chamada de Redundância. Esta é a medida de desigualdade de renda que obtemos utilizando as noções da Teoria de Informação. A redundância é o que chamamos de Índice de Theil, já que *Henry Theil* tem sido o economista que mais tem divulgado a aplicação da Teoria de Informação em Economia (*Theil,* 1967).

Não há dificuldade para se perceber que o Índice de Theil (ou Redundância) possui o valor zero quando temos uma situação de máxima igualdade e um valor igual a log N numa situação de máxima desigualdade (*Costa,* 1975). Sendo assim, é uma medida que depende do valor de N. Apesar do logaritmo reduzir substancialmente este aspecto indesejável do Índice de Theil, ainda assim tal dependência de N constitui uma deficiência, mormente quando se fazem necessárias comparações envolvendo distribuições com valores de N substancialmente diferentes. Às vezes, o uso do Índice de Theil (Redundância ou de Entropia), sem se dar conta da influência de N, pode levar a evidências obscuras, imprecisas e ambíguas (*Meirelles,* 1976).

Para evitar os problemas introduzidos por N basta dividirmos tanto a Entropia como o Índice de Theil por log N e diminuiremos este aspecto indesejável das medidas em questão:

$$h(r) = \frac{\sum_{i=1}^{N} r_i \log \frac{1}{r_i}}{\log N} \quad \text{(Entropia relativa)}$$

e

$$t(r) = \frac{\sum_{i=1}^{N} r_i \log_2 N r_i}{\log N} \quad \text{(Redundância relativa)}$$

O Índice de Theil será então entendido como a medida de Redundância relativa que é igual a 1 (um) menos a Entropia relativa:

$$t(r) = \frac{T(r)}{\log N} = 1 - \frac{H(r)}{\log N}$$

Com esta modificação os valores para o Índice de Theil ficariam compreendidos entre 0 e 1, como no caso da Razão de Concentração de Gini.

$$0 < t(r) < 1$$

A máxima igualdade corresponderia a zero, como antes, e a máxima desigualdade corresponderia ao valor 1, eliminando os problemas de medidas, ou seja, não haveria necessidade de se preocupar com os *bits* ou *nits*. A única dificuldade seria a diminuição da sensibilidade do índice de Theil, cuja discussão está fora do alcance deste trabalho.

III.2.2 — Informação de uma Mensagem Indireta

Até agora apresentamos o índice de Theil com a hipótese de que os dados individuais estariam disponíveis. Como o nosso experimento foi realizado com base numa distribuição de renda com dados grupados é necessário, então, mostrar como o índice de Theil assim obtido é interpretado. Na ausência de dados individuais o conceito da Teoria de Informação utilizado é o valor esperado do conteúdo de informação de uma mensagem indireta.

Numa distribuição de freqüência, cujo atributo é a renda R, observamos dois tipos de informações. O primeiro são as freqüências relativas da população correspondentes a cada classe de renda, isto é, n_i. O segundo são as frações de renda, r_i, correspondentes a cada classe de renda, ou seja, a fração da renda r_i recebida pela fração n_i da população. Estes dois tipos de informações podem ser interpretados como as probabilidades *"a priori"* e *"a posteriori"*. Conforme consideremos as frações de renda ou as frações da população, como a probabilidade *"a priori"*, obteremos medidas de desigualdade de renda que diferem simplesmente pelos pesos utilizados. Para este experimento escolhemos a fração de renda em cada classe como o peso:

$$\bar{I}(R, N) = \sum_{i=1}^{n} r_i \log \frac{r_i}{n_i}$$

O índice de Theil empregado no caso em que a renda é divulgada em classes ou estratos possui como limite inferior o valor $I(R, N) = 0$ e como limite superior $I(R, N) = \log N$. O primeiro corresponde a uma situação de perfeita igualdade, entendida como $r_i = n_i$, implicando, por conseguinte, que:

$$I\ (R, N) = \sum_{i=1}^{n} r_i \log \frac{r_i}{n_i} = \sum_{i=1}^{N} r_i \log 1 = 0$$

O segundo representa uma situação de máxima desigualdade, ou seja, $r_i = 0$, $r_j = 1$ para $i \neq j$. Aqui novamente $I(R, N)$ se reduz a um único termo cujo valor é o seguinte:

$$I\ (R, N) = r_j \log \frac{r_j}{1/N} = 1 \cdot \log \frac{1}{1/N}$$

III.2.3 — Interpretação do Índice de Theil para Dados Grupados

Esta medida empregada no experimento, além da interpretação, como a informação esperada de uma mensagem indireta, na qual r_1, $r_2 \ldots r_n$ seria as probabilidades *a posteriori* e n_1, $n_2 \ldots n_n$ as probabilidades *a priori*, possui uma outra interpretação relacionada como o conceito de média geométrica. Esta segunda interpretação é mais simples do que a anterior. Lembremos a fórmula de I(R, N):

$$I\ (R,\ N) = \sum_{i=1}^{n} r_i \log \frac{r_i}{n_i}\ ,\ i = 1,\ 2,\ \ldots\ n$$

Considerando que r_i/n_i é a renda *per capita* da classe de renda i, deflacionada pela renda *per capita* total, é possível interpretar I(R, N) como o logaritmo da média geométrica ponderada das rendas *per capita* de cada classe deflacionada pela renda *per capita* total, isto é:

$$M_G = {}_eI\ (R,\ N) = \prod_{i=1}^{n} \left(\frac{r_i}{r_i}\right)^{r_i}\ \ i = 1,\ 2,\ \ldots\ n$$

Tal interpretação mostra como esta medida do Índice de Theil diminui a influência das diferenças absolutas de renda, ora pelo uso do logaritmo ora porque a média geométrica é menos sensível aos valores extremos de uma distribuição.

Parece fora de dúvida de que com as informações oferecidas a respeito da Razão de Concentração de Gini e do Índice de Theil estamos aptos a descrever como estas medidas, independentes de qualquer tipo de ajustamento estatístico ou da forma da distribuição, se comportaram no experimento como a variação do tamanho da amostra e a dos elementos representativos da distribuição.

III.3 — Resultados Empíricos

III.3.1 — Resultados para a Razão de Concentração de Gini

Agora tentaremos ressaltar os resultados obtidos para a Razão de Concentração de Gini quando examinamos os efeitos das diferenças de tamanho da amostra e dos elementos representativos de cada classe de renda, por ocasião do cálculo desta medida de desigualdade.

Primeiramente observamos na tabela 5 que os valores da Razão de Concentração de Gini são menores para a amostra maior, ou seja, ao passarmos de uma amostra de 1,3% da população para uma amostra de 25%, o grau de desigualdade diminui qualquer que seja o elemento representativo. A diminuição do valor da Razão de Concentração de Gini é mais intensa quando o elemento representativo é a Renda Média, atingindo um valor máximo de 2% aproximadamente. Amostras pequenas tendem a exagerar o grau de desigualdade medido pela Razão de Concentração de Gini. No entanto, as diferenças, apesar de existirem, só se manifestam depois da segunda casa decimal, o que não representa uma grande mudança do ponto de vista prático, já que a Razão de Concentração é praticamente divulgada com duas casas decimais. [20]

Novamente, para a razão de concentração de Gini se repetem os fatos observados para os dois primeiros índices (Pareto e Gini), ou seja, quando o elemento representativo é a renda média obtemos uma variação maior. Esta variação maior é explicada pelo fato de que os outros elementos representativos não se modificam quando passamos de uma amostra para outra, é o caso do ponto médio, limite superior e limite

---

20 O uso do computador na mensuração da Razão de Concentração permite-nos obter maior aproximação, fazendo com que se vá além das duas casas decimais.

inferior. Portanto, apesar da renda média das classes ser uma informação mais adequada, pois permite uma melhor aproximação na estimativa da média geral e, por conseguinte, da Renda Total, ela é justamente o elemento representativo que apresenta as maiores diferenças quando passarmos de uma amostra para outra. Isto parece muito coerente, pois, nesta situação, além de mudar a freqüência relativa de cada classe também muda a média de cada classe.

TABELA 5

Valores da razão de concentração de Gini segundo o tamanho da amostra e os elementos representativos das classes de renda

AMOSTRA / SUBAMOSTRA — ELEMENTOS REPRESENTATIVOS DAS CLASSES DE RENDA	AMOSTRA 25%	SUBAMOSTRA 1,3%	VARIAÇÃO	
			Absoluta	%
Ponto Médio (PM)	0,358070	0,361354	—0,003284	0,92
Renda Média (RM)	0,328930	0,335041	—0,006111	1,86
Limite Inferior (LI)	0,371520	0,373858	—0,002338	0,63
Limite Superior (LS)	0,339270	0,343185	—0,003915	1,15

FONTE: Censo Demográfico de 1970 — FIBGE.

Como anteriormente fizemos para o Coeficiente de Pareto e o Índice de Gini, na tabela 6 observamos as diferenças entre os valores da Razão de Concentração em função dos elementos representativos. As maiores diferenças para a Razão de Concentração verificam-se em três situações. A primeira maior diferença verificamos entre a Renda Média e o limite Inferior quando observamos uma diferença absoluta de aproximadamente 0,04259, o que corresponderia a uma variação em torno de 12,9% para a amostra de 25% e de 11,6% para a de 1,3%. Não há dúvida de que se trata de uma variação muito significativa. A segunda maior diferença observamos quando comparamos os valores da Razão de Concentração em termos Limite Inferior e Limite Superior. Neste caso, a variação fica em torno de 8,68% para a amostra de 25%, e cerca de 8,20% para a de 1,3% da população. A terceira maior diferença aparece quando comparamos os valores da Região de Concentração em função do Ponto Médio e Renda Média. As diferenças ficam em volta de 8,14 para a amostra de 25% e 7,28% para a de 1,3%.

Estas informações são facilmente entendidas quando voltamos à tabela 5 e observamos que a Renda Média, como elemento representativo, fornece o menor valor para a Razão de Concentração, ou seja, 0,328930, já que o Limite Inferior propicia o valor máximo, isto é, ... 0,371520. Os menores valores para a variação da Razão de Concentração de Gini também são observados na tabela 6 quando comparamos os valores obtidos através da Renda Média com os valores obtidos a partir do limite superior. Apesar desta comparação apresentar as menores variações, ainda assim são maiores do que aquelas observadas em função das diferenças de tamanho de amostras, já que o menor valor observado no Quadro 6 é 0,008144, referente à diferença entre o valor da Razão de Concentração quando se utiliza a Renda Média e o Limite Superior como elementos representativos, no caso da amostra de 1,3%.

TABELA 6

Variações da razão de concentração de Gini segundo o tamanho da amostra e os tipos de elementos representativos utilizados na representação das classes de renda

VARIAÇÕES / AMOSTRA E SUBAMOSTRA	[PM, RM]		[PM, LI]		[PM, RS]	
	Absoluto	%	Absoluto	%	Absoluto	%
Amostra (25%)	0,02914	8,19	0,01345	3,76	0,01880	5,25
Subamostra (1,3%)	0,026313	7,28	0,012504	3,46	0,018169	5,03

VARIAÇÕES / AMOSTRA E SUBAMOSTRA	[RM, LI]		[RM, LS]		[LI, LS]	
	Absoluto	%	Absoluto	%	Absoluto	%
Amostra (25%)	0,04259	12,95	0,01034	3,14	0,03225	8,68
Subamostra (1,3%)	0,038817	11,59	0,008144	2,43	0,030673	8,20

FONTE: Censo Demográfico de 1970 — FIBGE (só valores absolutos).

No caso da Razão de Concentração de Gini parece claro que as variações da medida de desigualdade, em função do tamanho da amostra, são pequenas. Tudo indica que será sempre na mesma direção, isto é, quanto maior o tamanho da amostra menor a desigualdade. É um resultado à primeira vista coerente, já que a Razão de Concentração de Gini é uma medida de dispersão relativa. Além disso, este resultado era esperado em função do que se verificou para o Índice de Pareto — o completamente oposto — já que existe uma relação matemática entre ambos os índices:

$$G = \frac{1}{2\alpha - 1}$$

O resultado empírico mais importante se refere à situação em que os diversos elementos representativos são utilizados para a obtenção da Razão de Concentração de Gini. Neste caso as variações, ainda que pequenas, chegam a ser significativas, pois são da ordem de 12,9%, justificando uma certa cautela na escolha do elemento representativo, quando se pretende comparar esta medida de desigualdade no tempo e/ou no espaço.

III.3.2 — Resultados para o Índice de Theil

O Índice de Theil como a Razão de Concentração de Gini é também um índice que independe da forma da distribuição, não necessitando, portanto, de qualquer tipo de ajustamento estatístico a fim de atingir algum significado como medida de desigualdade. Este índice quando submetido ao mesmo experimento que as demais medidas de desigualdade apresentadas também mostrou certas diferenças em função do tamanho da amostra, bem como algumas referentes à utilização dos vários elementos representativos.

Também neste caso os valores do índice de Theil que apresentaram maior variação ao passarmos de uma amostra de 25% para uma amostra de 1,3% da população foram aqueles obtidos quando a média é o elemento representativo da classe de renda. As razões para tais variações são as mesmas apresentadas anteriormente na discussão das outras medidas, isto é, a média muda de amostra para amostra, ao passo que os demais elementos representativos de cada classe são fixos. No entanto, para o índice de Theil esta variação não é tão desprezível assim, pois atingiu valores ao redor de 4,5% que já faz uma certa diferença em comparações em dois momentos de tempo. Os demais valores do índice de Theil não ultrapassam muito o limite de 1%. Esta maior sensibilidade do índice de Theil poderia ser explicada quando o elemento representativo é a média, lembrando que o índice depende do tamanho da população.

TABELA 7

Valores do índice de Theil segundo o tamanho da amostra e os elementos representativos das classes de renda

AMOSTRA / SUBAMOSTRA — ELEMENTOS REPRESENTATIVOS DAS CLASSES DE RENDA	AMOSTRA 25%	SUBAMOSTRA 1,3%	VARIAÇÃO	
			Absoluta	%
Ponto Médio (PM)	0,133270	0,134560	—0,00129	0,97
Renda Média (RM)	0,106115	0,110984	—0,00487	4,59
Limite Inferior (LI)	0,116956	0,118236	—0,00128	1,09
Limite Superior (LS)	0,142903	0,143776	—0,00087	0,61

FONTE. Censo Demográfico de 1970 — FIBGE.

Na tabela 7 observamos que os valores máximos obtidos para o índice de Theil surgem quando o elemento representativo é o limite superior para os dois tamanhos da amostra. Já os valores mínimos são obtidos quando a Renda Média é o elemento representativo. O ponto médio e o limite inferior proporcionam valores do índice de Theil que se aproximam, respectivamente, do limite superior e da Renda Média. Sendo assim, o uso costumeiro do ponto médio como elemento representativo induziria a uma estimativa superestimada da desigualdade, em oposição a uma estimativa subestimada proporcionada pela Renda Média.

Quando observamos a tabela 8, onde comparamos os valores do índice de Theil de acordo com o elemento representativo utilizado, novamente observamos o mesmo fenômeno constatado para os demais índices, isto é, as discrepâncias são consideráveis. A mudança de elemento representativo introduz diferenças no valor do índice de Theil da ordem de 34,6%, como limite superior, e de 17,6%, como limite inferior. Nestas circunstâncias é preciso tomar o máximo cuidado nas comparações com o índice de Theil.

Os valores mais díspares estão obviamente ligados às comparações entre os valores extremos, ou seja, o ponto médio e a Renda Média, o Limite Superior e o Limite Inferior, tanto na amostra como na subamostra. Ao compararmos o limite superior e a renda média obtemos a diferença máxima que atinge o valor de 0,036788 correspondendo aos 34,6% de diferença. Apesar deste elemento representativo não ser muito utilizado, constitui um exemplo bem claro sobre a importância de uma metodologia uniforme quando se pretende comparar medidas de desigualdade. No caso da amostra de 25%, a segunda maior discrepância é quando se mede o Índice de Theil através da utilização do ponto médio e da renda média como elementos representativos. Tal discrepância atinge cerca de 20,38% de diferesça. Este último caso é o mais relevante, já que a troca entre o ponto médio e a renda média torna-se naturalmente mais freqüente à medida que se obtém melhores informações sobre a renda.

Observamos a mesma característica no caso da amostra de 1,3%. Porém para esta amostra menor, a discrepância que aparece em segundo lugar não é a devida à comparação entre o ponto médio e a renda média, mas sim entre os limites inferior e superior. Os níveis de discrepâncias também não são nada modestos, variando de 21,60 a 29,54%. As informações da tabela 8 tornam imprescindível a necessidade de se tomar o máximo cuidado na manipulação do Índice de Theil, procurando sempre explicitar qual o elemento representativo utilizado na sua computação.

O Índice de Theil, como um representante das medidas de desigualdade de renda que não dependem da forma da distribuição, mostrou-se mais sensível às mudanças amostrais do que a Razão de Concentração de Gini, mas as diferenças obtidas estão aquém dos 6% — nível de

TABELA 8

Variações do Índice de Theil segundo o tamanho da amostra e os tipos de elementos representativos utilizados na representação das classes de renda

VARIAÇÕES / AMOSTRA E SUBAMOSTRA	[PM, RM]		[PM, LI]		[PM, LS]	
	Absoluto	%	Absoluto	%	Absoluto	%
Amostra (25%)	0,027155	20,38	0,016314	12,24	0,009633	7,23
Subamostra (1,3%)	0,023576	17,52	0,016324	12,13	0,009216	8,30

VARIAÇÕES / AMOSTRA E SUBAMOSTRA	[RM, LI]		[RM, LS]		[LI, LS]	
	Absoluto	%	Absoluto	%	Absoluto	%
Amostra (25%)	0,010841	10,22	0,036788	34,67	0,025947	22,19
Subamostra (1,3%)	0,007252	6,13	0,032792	29,55	0,025540	21,60

FONTE: Censo Demográfico de 1970 — IBGE (só valores absolutos).

erro tradicionalmente aceito. No entanto, quando se trata das diferenças em função do elemento representativo, o Índice de Theil não difere substancialmente das evidências obtidas tanto para a Razão de Concentração de Gini como para os dois outros índices estudados.

Enfim, no caso do índice de Theil, a magnitude das diferenças observadas na tabela 8 compromete as evidências, exigindo, portanto, uma maior preocupação e cuidado a fim de se obter uma metodologia uniforme no cômputo do índice em questão, quando se fizer necessário o uso de comparações.

IV — CONCLUSÕES

Em função das informações anteriores, não resta a menor dúvida de que o tamanho da amostra e os diversos elementos representativos das classes de renda de uma distribuição de freqüência introduzem variações nos valores das medidas de desigualdade de renda mais comumente utilizados nos estudos da situação da renda pessoal. A seguir tentaremos ressaltar os pontos mais importantes que observamos no decorrer do trabalho.

IV.1 — Variações em Função do Tamanho da Amostra

As variações em função do tamanho da amostra não apresentaram valores muito sigunificantes do ponto de vista prático, já que o máximo de variação não ultrapassou a casa dos 5%. Tal cifra não representa um valor muito significante, já que na prática erros de até 5% são toleráveis, mormente no caso especial de indicadores de desigualdade, que são medidas aproximadas da dispersão relativa.

Observando os índices que dependem de ajustamento, como é o caso do Coeficiente de Pareto e o Índice de Gini, constatou-se para o Coeficiente de Pareto um aumento do valor de α ao passarmos de uma amostra menor para uma maior (ou seja uma diminuição da desigualdade de renda). Para o Índice de Gini não se observou a mesma regularidade na direção da variação amostral, como no caso do Coeficiente de Pareto. Houve uma diminuição de δ quando os elementos representativos eram o Ponto Médio e a Renda Média e um aumento de δ, ainda que pequeno, quando os elementos eram os Limites Inferior e Superior. Tal fato evidencia, para o Índice de Gini, uma maior sensibilidade às variações amostrais do que para o Coeficiente de Pareto.

Para os índices que independem de ajustamento, como é o caso da Razão de Concentração de Gini e do Índice de Theil, verificou-se o mesmo fenômeno: pequenas variações nos valores dos Índices quando se muda de uma amostra para outra. Neste caso, porém, a regularidade foi para os dois índices. A Razão de Concentração de Gini diminuiu quando passamos de uma amostra menor para uma maior (ou seja, diminuiu a desigualdade coerentemente com o valor de α independentemente obtido). O Índice de Theil apresentou a mesma característica que a Razão de Concentração de Gini, seus valores diminuíram, isto é, indicaram uma subestimação da desigualdade ao passarmos para uma amostra maior, ou seja, para pequenas amostras a desigualdade é maior. Estes fatos são perfeitamente coerentes com o significado dos dois índices, isto é, medidas de dispersão relativa (*Costa,* 1974).

No experimento que fizemos o tamanho da amostra introduz variações, mas elas são pequenas do ponto de vista prático. Não podemos generalizar que sempre acontecerá assim, por dois motivos. O primeiro é que trabalhamos com apenas uma distribuição e seriam necessários muitas distribuições para estabelecermos uma posição definitiva sobre o efeito da variação do tamanho da amostra. O segundo é que a amostra, no caso específico, foi derivada do próprio Censo, ou seja, a amostra de 1,3% é, em verdade, uma subamostra, com base na amostra de 25%. Este último fato contribui muito para diminuir as discrepâncias entre as duas amostras. O mais interessante é que a distribuição estudada não representa uma parcela muito grande do universo, o que nos leva a concluir que as duas amostras do Censo Demográfico são de boa qualidade, pois indicadores secundários foram obtidos com variações muito pequenas.

IV.2 — Escolha dos Elementos Representativos das Classes de Renda

O trabalho lidou com quatro possíveis elementos representativos, o Ponto Médio (PM), a Renda Média (RM), o Limite Superior (LS) e o Limite Inferior (LI). Tradicionalmente utiliza-se o Ponto Médio (PM) e, às vezes, a Renda Média (RM). Quando os dados são de melhor qualidade, os outros dois são raramente usados, por esta razão nos deteremos apenas na qualificação dos efeitos dos dois primeiros elementos representativos nas variações das medidas de desigualdade. Tentaremos contrapor o Ponto Médio e a Renda Média em termos de medidas que dependem de ajustamento de função e as que independem de ajustamento.

Independente de qualquer condição, a mudança do elemento representação no cálculo das medidas introduz variações significativas, atingindo percentuais acima de 20% de diferença. Para o Coeficiente de Pareto o Ponto Médio é o Limite Inferior e a Renda Média o Limite Superior, ou seja, constituem os valores extremos. O Ponto Médio superestima a desigualdade, ao passo que a Renda Média subestima (quanto maior α, menor a desigualdade), conseqüentemente a maior diferença aparece quando comparamos α calculado em função do Ponto Médio e α calculado através da Renda Média. A diferença, neste caso, atinge a casa dos 14% (para a distribuição em questão) uma diferença relativamente elevada. No caso do Índice de Gini a situação é inversa, o Ponto Médio dá origem aos maiores valores de δ, a Renda Média aos menores valores, portanto, novamente o Ponto Médio superestima e a Renda Média subestima a desigualdade de renda (os valores extremos são dados pelo Limite Superior e Limite Inferior).

O Índice de Gini apresentou algumas discrepâncias muito elevadas, em torno de 32%, quando comparamos δ obtido em função do Limite Superior com o δ calculado usando o Limite Inferior. Este é um caso gritante de variação nos valores das medidas única e exclusivamente em função da troca dos elementos representativos.

Quando observamos os índices que não dependem de ajustamento de função, percebemos que, para a Razão de Concentração de Gini, a Renda Média proporciona o menor valor da desigualdade e o Ponto Médio o segundo maior valor.[21] Conseqüentemente, a diferença entre estas duas maneiras de computar a Razão de Concentração proporciona diferenças em torno de 8,14%. Tal diferença está acima dos 5%, tolerados na prática, constituindo-se, portanto, numa diferença significativa. Em relação ao Índice de Theil, o comportamento é o mesmo verificado para a Razão de Concentração. A Renda Média apresenta o menor valor do

21 O primeiro maior valor é introduzido pelo Limite Inferior, tanto na amostra de 25 como na de 1,3%.

Índice de Theil, ao passo que o Ponto Médio proporciona o segundo maior valor. Comparando-se os respectivos valores do Índice de Theil obtém-se a segunda maior variação dos valores deste índice.

Ao contrário da variação no tamanho da amostra a amostra do elemento representativo introduz diferenças muito significativas na avaliação da desigualdade. Por isto é necessário, para efeito de comparação, calcular as medidas de desigualdade utilizando-se sempre o mesmo elemento representativo a fim de evitar variações destas medidas introduzidas única e exclusivamente pela troca dos elementos representativos.

Outro fato interessante observado em todos os quatro casos: a Renda Média foi o elemento representativo que provocou maior variação nos valores dos índices nas comparações entre as amostras. Parece que a explicação para este fato é de que a Renda Média muda muito mais de amostra para amostra do que os outros elementos representativos, pois permanecem fixos, uma vez que as classes de renda são as mesmas.

IV.3 — Considerações Metodológicas

O trabalho em questão levanta dois pontos de caráter metodológico quando do cálculo da medida de desigualdade de renda para comparações de qualquer natureza. O primeiro ponto diz respeito à necess dade de se tomar o devido cuidado com as diferenças introduzidas pelo tamanho da amostra que, no caso particular deste experimento, não acusou diferenças acima de 5%, no entanto não devem ser esquecidas. O segundo ponto diz respeito à escolha do elemento representativo de cada classe de renda. Em função das diferenças constatadas, chegando, às vezes, ao redor de 32%, é preciso que se tenha uma metodologia uniforme para o cálculo dos índices, a fim de evitar evidências devido a diferenças de metodologia. Se existe a necessidade de comparação de duas medidas no tempo ou no espaço é necessário que estas medidas sejam calculadas utilizando-se o mesmo elemento representativo. Ou seja, deve-se escolher a Renda Média, ou o Ponto Médio de classe, ou o Limite Superior, ou o Limite Inferior que não se deve fazer é misturar os vários métodos.

Apesar das limitações do experimento, ele serve de alerta para aqueles que não se importam com a metodologia de cálculo das medidas de desigualdade. Se mesmo com um experimento feito em condições excepcionais, as diferenças são de magnitudes relativamente grandes, podendo desviar o sentido das evidências, é imprescindível um mínimo de preocupação com a metodologia de cálculo destas medidas. Mormente em comparações da Razão de Concentração de Gini entre estado, entre países ou entre períodos de tempo, é necessário estar atento para o método de cálculo desta medida, sob pena de estar fazendo afirmações que não são necessariamente verdadeiras.

As medidas da desigualdade tentam resumir em um só índice informações sobre toda a distribuição, logo, devem merecer um certo cuidado no seu cálculo, do contrário estaremos apenas em busca de um número, cujo significado se torna vazio pela arbitrariedade com que foi obtido.

IV.4 — A Mensagem que Fica

Pode parecer um exagero a preocupação demasiada com a metodologia do cálculo de uma modesta medida de desigualdade. No entanto, é um problema muito sério para aqueles que querem saber o que de fato se passa com a dispersão relativa da distribuição de renda. Podemos encontrar em *Corrado Gini* — um dos cientistas que mais contribuíram para o estudo da desigualdade da renda pessoal — uma preocupação semelhante com a que tentamos transmitir neste modesto experimento, pois, em 1922, quando comparava o seu índice de concentração δ e o índice α de Pareto, fazia a seguinte afirmativa:[22]

"É arbitrario seguire per il calcolo dei δ l'uno o l'atro dei metodi suesposti. Quando pero si voglia procedere a confronti, é opportuno constantemente un dato metodo, per non conceudere a differenze essenziale nella distribuizione dei redditi da differenze nei valori di δ che dipendono unicamente dall'aver usato metodi differenti".

Um exemplo onde o pesquisador teve uma grande preocupação metodológica no cálculo das medidas de desigualdade é o nosso trabalho sobre a distribuição da renda pessoal no Brasil, em 1970, pois a pesquisa envolvia uma seqüência de comparações destas medidas através das unidades geográficas estudadas. Não há dúvida que os trabalhos de *Corrado Gini* sempre levaram em consideração esta necessidade de uma certa uniformidade metodológica quando qualquer comparação das medidas de desigualdade se fazia necessária.

No entanto, a literatura atual sobre a distribuição da renda pessoal é rica em trabalhos, onde a comparação de medidas de desigualdade de renda de países diferentes é feita sem a menor preocupação com a metodologia envolvida no cálculo dessas medidas. Alguns autores compararam informações sobre a Razão de Concentração de Gini de diversos países simplesmente coletando os valores a partir dos trabalhos que foram feitos para as respectivas unidades geográficas sem, pelo menos, indagar como foi calculada cada Razão de Concentração, qual o tamanho da amostra, se as classes de renda eram fechadas ou se os elementos representativos utilizados no cálculo para cada país eram os mesmos ou não.

Para ilustrar, podemos citar o trabalho de *Félix Paukert*[23], onde compara as razões de concentração para sessenta e um países com os respectivos PIB, procurando estabelecer uma relação entre as duas medidas a fim de justificar certas heresias sobre desigualdade de renda e crescimento econômico. O outro exemplo é o trabalho de *Montek S. Ahluwalia*[24] para o Banco Mundial que, apesar de reconhecer as limitações dos dados utilizados, usou razões de concentração calculadas para os vários países sem a menor preocupação com a metodologia envolvida.

Apesar das intenções dos autores serem dignas de respeito, apesar da ausência de informação justificar a utilização do melhor dado disponível no momento, ainda assim seria necessário um cuidado especial com as medidas utilizadas, pois, como foi observado, elas estão sujeitas a variações significantes quando simplesmente usamos o Ponto Médio no lugar da Renda Média. Sem esta preocupação de uniformidade no cálculo das medidas a serem comparadas, corremos um grande risco de apresentar evidências que são apenas uma conseqüência das diferentes

---

22 Vide *Corrado Gini* — *Indice di Concentrazione e di Dipendenza*, Roma, Biblioteca Dell'Economista, Editice Torinese, Milano, 1922, pp. 5/137.

23 Vide *Felix Paukert*, "Income Distribution at Different Levels of Development: a Survey of Evidence" — *International Labour Review*, 108, agosto/setembro de 1973, pp. 97/123.

24 Vide *Montek S. Ahluwalia* e outros — *Redistribution and Growth*, publicado pelo Banco Mundial, Oxford University Press, 1974.

maneiras como foram calculadas as medidas utilizadas. Além das críticas mais sofisticadas sobre a utilização de dados de *cross-section* para obter referências sobre a evolução de cada unidade geográfica, as diferenças existentes em função da diversidade de cálculo podem marcar evidências interessantes, bem como apresentar evidências as mais absurdas possíveis, que podem se transformar em regra para a maioria dos economistas.

Para ilustrar como este problema de metodologia é muito importante nas comparações no tempo ou entre unidades geográficas, podemos utilizar um exemplo extraído do trabalho de *Cline*,[25] onde, para 1960, foi obtido um valor para a Razão de Concentração de Gini ao redor de 0,61. Se utilizarmos pura e simplesmente esta informação sem averiguar em que condições foi obtida e a compararmos com o valor da Razão de Concentração de Gini calculada por *Langoni,* para 1970, cujo valor foi 0,5684 chegaríamos a conclusão de que não houve nenhum aumento de concentração de renda no Brasil, mas sim uma diminuição. Nestas condições, porque tanta celeuma a respeito do aumento da desigualdade ocorrido entre 1960 e 1970? O absurdo está na comparação das duas medidas, as quais foram obtidas por métodos completamente diferentes e a partir de dados de qualidade muito diversa.

Anteriormente, afirmamos que no caso da Razão de Concentração o Ponto Médio fornece um valor mais elevado do que aquele que se obtém usando a Renda Média. Os dados de 1960 só permitem uma aferição da Razão de Concentração de Gini, via Ponto Médio da classe, ao passo que, em 1970, a medida foi calculada utilizando-se a Renda Média. No nosso experimento a Renda Média proporcionou sempre o valor mínimo para a Razão de Concentração de Gini. Nestas circunstâncias, ao compararmos a Razão de Concentração de Gini, para 1960, com a medida, para 1970, consideramos que o acréscimo de desigualdade está sendo subestimado por duas razões. A primeira porque o Ponto Médio, em 1960, fornece um valor maior para a Razão de Concentração de Gini. A segunda porque a Renda Média, em 1970, subestima o valor da desigualdade. Em outras palavras, para comparar 1960 com 1970 seria necessário calcular a desigualdade para 1970 usando o Ponto Médio, ou calculá-la para 1960 usando a Renda Média, a fim de se obter uma estimativa mais realista do aumento de desigualdade ocorrido entre 1960 e 1970.

Finalmente este trabalho procura mostrar que precisamos ser mais cautelosos com os números que usamos, calculamos e comparamos, a fim de evitar absurdos e injustiças. Nunca devemos nos esquecer que os números foram também inventados pelo homem.

25 Vide *William R. Cline — Patential Effects of Income Redistribution on Economic Growth* — Latin American Cases, Praeger Publishers, 1972.

BIBLIOGRAFIA

(1) AMES, Edward — *"A Method for Estimating the Size Distribution of a Given Aggregate Income"* — The Review of Economics and Statistics, julho de 1942.

(2) BOWMAN, Mary J. — *"A Graphical Analysis of Personal Income Distribution"* — A. E. R. XXXV, setembro de 1945, pp. 607/628.

(3) CHENERY, Hollis; AHLUWALIA, Montek S.; BELL, G. L. G.; DULOAY, John H. e JOLLY, Richard — Redistribution and Growth, publicado pelo Banco Mundial, Oxford University Press, 1974.

(4) CLINE, William R. — Potential Effects of Income Redistribution on Economics Growth — Latin American Cases, Praeger Publishers, 1972.

(5) COSTA, Ramonaval A. — Size Income Distribution of Brazil in 1970 — A Cross-section Analysis of Income Distribution by Occupations — Tese de Doutorado apresentada na Universidade de Vanderbilt, agosto de 1975.

(6) COSTA, Ramonaval A. — *"Bem Estar e Indicadores de Desiqualdade"* — Revista Brasileira de Estatística — Rio de Janeiro, Fundação IBGE, julho de 1974.

(7) COSTA, RAMONAVAL A. — *Medidas de Desigualdade de Renda* Boletim Geográfico — n.º 238, jan./fev. de 1974, ano 33, pp. 45/72.

(8) DE FINETTI, B. Paciello, U. — *"Calcolo della differenza Media"* — metron, Vol. VIII, n.º 3, fevereiro de 1960.

(9) DUARTE, J. C. — *"Aspectos da Distribuição de Renda no Brasil em 1970"* — Piracicaba, Tese de Mestrado da E. S. A. L. Queiroz, 1971.

(10) EDWARDS, Elwyn — *Introdução à Teoria da Informação* — Editora Cultura, Editora da USP, 1971.

(11) GALVANI, L. — *Sulle Curve di Concentrazione Relative a Caratteri non Limitati e Limitati* — metron, Vol. X, n.º 3, Roma, outubro de 1932.

(12) GASTWIRTH, Joseph L. — *"The Estimation of the Lorenz Curve and Gini Index"* — The Review of Economics and Statistics, LIV, agosto de 1972.

(13) GINI, Corrado — *Indici di Concentrazione e di Dipendenza,* Roma, Biblioteca Dell'Economista, Editice Torinese Milano, 1922, pp. 5/137.

(14) HOFFMAN, R. — *"Contribuição à Análise da Distribuição da Renda e da Posse da Terra no Brasil"* — Piracicaba, Tese não publicada, E. S. A. L. Queiroz, 1971.

(15) KINGSTON, Jorge — *Desigualdade na Distribuição das Rendas* — Revista Brasileira de Economia, Rio de Janeiro, Fundação Getúlio Vargas, março de 1955, p. 5.

(16) LANGONI, C. G. — Distribuição da Renda e Desenvolvimento Econômico do Brasil, Rio de Janeiro, Expressão e Cultura, 1973.

(17) LORENZ, M. O. — *"Methods of Measuring the Concentration of Wealth"* — A. S. A., New Series, junho de 1905, p. 209.

(18) MEIRELLES, Antonio C. — *"A Evolução da Estrutura do Sistema Bancário Brasileiro no Período 1966-1974"* — Sindicato dos Bancos do Estado da Guanabara, 1975, ano III, n.º 8.