Nº 34

MEDIDAS DE CONCENTRAÇÃO

FERNANDO DE HOLANDA BARBOSA

MEDIDAS DE CONCENTRAÇÃO

Fernando de Holanda Barbosa*

I - INTRODUÇÃO

O objetivo deste trabalho consiste em apresentar, de um ponto de vista puramente estatístico, as medidas de concentração mais frequentemente encontradas na literatura econômica. Mais especificamente trataremos, pela ordem de apresentação, da razão de concentração, do coeficiente de Gini, do índice de Hirschman-Herfindahl, da entropia e da redundância e da variância logarítmica.

Cabe mencionar que algumas destas medidas de concentração têm sido aplicadas com maior frequência em determinadas áreas. Com efeito, o coeficiente de Gini tem sido quase que exclusivamente aplicado no estudo da distribuição de renda, o índice de Hirschman-Herfindahl e a razão de concentração são mais familiares entre aqueles que se dedicam a estudos de organização industrial. Todavia, vale salientar que do ponto de vista da teoria econômica raramente existe uma indicação clara de qual a medida mais apropriada a ser empregada em uma área específica em virtude do próprio estágio de desenvolvimento da teoria. Em geral, a seleção da medida de concentração a ser usada é baseada em considerações que dizem respeito às propriedades da própria medida, tanto no que toca

*Professor da Escola de Pós-Graduação em Economia da Fundação Getúlio Vargas.

a sua interpretação como no que diz respeito a sua tratabilidade algébrica.

A principal justificativa para a realização deste trabalho é de carater didático e prende-se ao fato de que dificilmente encontra-se em uma única referência bibliográfica as medidas de concentração aqui apresentadas. Acreditamos, também, que a apresentação de diferentes medidas dentro de um arcabouço estatístico, além de propiciar um tratamento unificado, permite uma avaliação das diferenças existentes entre as várias medidas de concentração.

II - ALGUNS CONCEITOS BÁSICOS

Imagine que se deseje estudar a concentração da distribuição de uma variável aleatória X. Antes de estudar as medidas que podem descrever a concentração de tal distribuição introduziremos alguns conceitos básicos com o objetivo de tornar este trabalho auto-suficiente.

A variável aleatória X é do tipo contínuo quando a sua função de densidade de probabilidade for dada por $f(x) \geq 0$ e a função de distribuição igual a:

$$F(t) = P\ (X \leq t) = \int_0^t f(x)\ dx$$

onde $P\ (X \leq t)$ indica a probabilidade da variável aleatória X ser menor ou igual ao valor t*. Quando a variável aleató-

*Estamos considerando que a variável aleatória assume apenas valores não negativos, pois as variáveis econômicas são, em geral, deste tipo.

ria X for do tipo discreto a função de densidade $f(x_i) = p_i$ é igual a probabilidade da variável aleatória X assumir o valor x_i. A função de distribuição neste caso é igual a:

$$F(t) = P(X \leq t) = \sum_{i=1}^{k} f(x_i)$$

onde x_k é o maior valor de x_i que é menor ou igual a t.

A esperança matemática da variável aleatória X, $EX = \mu$ é definida por:

$$EX = \int_0^{\infty} xf(x)\, dx = \mu$$

quando X for do tipo contínua, e igual a:

$$EX = \sum_{i=1}^{n} x_i\, f(x_i) = \sum_{i=1}^{n} x_i p_i = \mu$$

quando X for do tipo discreta.

A função de distribuição do primeiro momento da variável aleatória X é definida através de

$$F_1(t) = \frac{\int_0^t xf(x)\ dx}{\int_0^\infty xf(x)\ dx} = \int_0^t \frac{x}{\mu}\ f(x)\ dx$$

quando X for uma variável aleatória contínua e por

$$F_1(t) = \frac{\sum_{i=1}^{k} x_i f(x_i)}{\sum_{i=1}^{n} x_i f(x_i)}\ ,\ x_k \leqslant t$$

quando X for uma variável discreta.

No que se segue trataremos apenas do caso em que a variável X é discreta e cuja distribuição de probabilidade seja a seguinte:

TABELA 1

Distribuição de Variável Aleatória X

X	x_1	x_2	...	x_n
P (X = x)	p_1	p_2	...	p_n

Admitiremos, também, que os valores de x_i estejam ordenados de tal modo que $x_1 < x_2 < \ldots < x_n$.

III - A RAZÃO DE CONCENTRAÇÃO

Com a finalidade de tornar mais fácil a apresentação da razão de concentração, admitiremos que a variável X represente a distribuição do tamanho das empresas de um determinado setor industrial, tamanho este medido, por exemplo, pelo número de empregados em cada empresa. Assim p_i representaria a proporção de empresas, em relação ao número total de empresas do setor, cujo número de empregados é igual a x_i.

A esperança matemática de X, $EX = \mu$, seria então, neste caso, o número médio de empregados por empresa. Se existissem N empresas no setor industrial em estudo o produto de média μ por N daria o número total de empregados no setor. Como a percentagem de empresas que empregam x_i empregados é igual a p_i, segue-se, então que o número de empregados naquelas empresas cujo tamanho é x_i é igual a $p_i x_i N$. Portanto, a razão

$$\frac{p_i x_i N}{\sum\limits_{i=1}^{n} p_i x_i N} = \frac{p_i x_i}{\sum\limits_{i=1}^{n} p_i x_i}$$

mede a percentagem de empregados, em relação ao total de empregados no setor, nas empresas cujo tamanho é igual a x_i. Consequentemente, a função de distribuição do primeiro momento

$$F_1(t) = \frac{\sum\limits_{i=1}^{k} p_i x_i}{\sum\limits_{i=1}^{n} p_i x_i} \quad , \; x_k \leqslant t$$

mede a percentagem de empregados que trabalham nas empresas

cujos tamanhos são menores ou iguais a t. Observe que a função de distribuição de X, F(t), é igual a percentagem de empresas cujos tamanhos são menores ou iguais a t.

A razão de concentração C_r é definida como a proporção de emprego, em relação ao total, das r maiores empresas do setor industrial. Isto é:

$$C_r = 1 - F_1(x_{n-r}) = 1 - \frac{\sum_{i=1}^{n-r} p_i x_i}{\sum_{j=1}^{n} p_j x_j} \tag{1}$$

A escolha do número r de empresas é arbitrária. Em geral, em trabalhos de organização industrial, considera-se r igual a 3, 4 ou 5.

Outra medida de concentração é obtida quando se fixa a razão de concentração ao invés do número r de empresas. Assim η seria o número de empresas que absorvem $100\alpha\%$ do emprego na indústria:

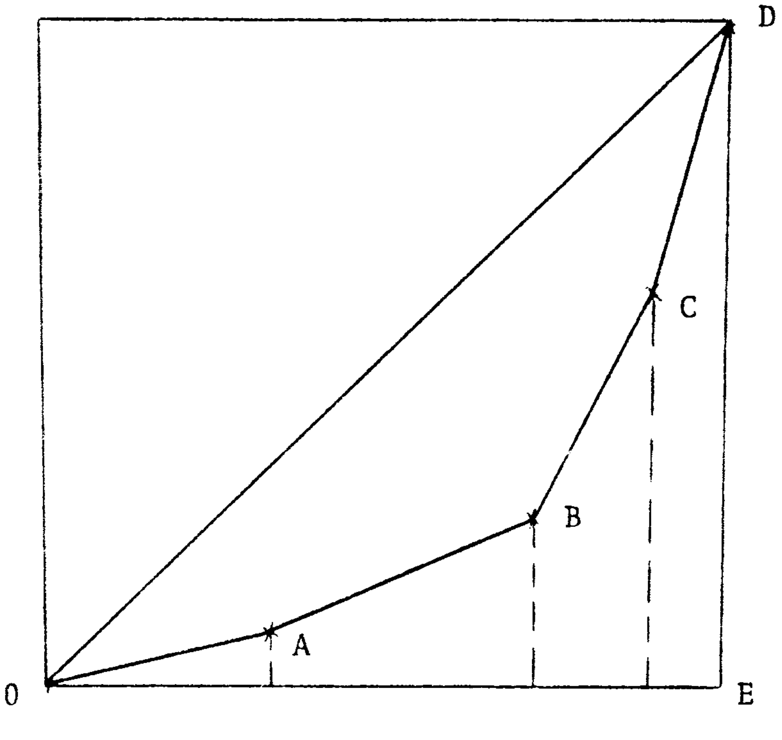
$$C_\eta = 1 - \frac{\sum_{i=1}^{\eta} p_i x_i}{\sum_{j=1}^{n} p_j x_j} = \alpha \tag{2}$$

Obviamente, neste caso, a escolha do parâmetro α é arbitrária. Assim, para um valor de α igual a 0,80 η representaria o número de empresas responsável por 80% do emprego na indústria.

IV - A CURVA DE LORENZ E O COEFICIENTE DE GINI

A curva de Lorenz é obtida quando se mede ao longo do eixo vertical os valores da função de distribuição do primeiro momento $F_1(t)$ e no eixo horizontal os valores da função de distribuição F(t). A Figura 1 mostra a curva de Lorenz, OABCD, para uma distribuição da Tabela 1 que contém apenas quatro valores (n = 4).

FIGURA 1 - CURVA DE LORENZ



Quando uma das probabilidades for igual a um e as demais forem iguais a zero, $p_i = 1$ e $p_j = 0$ para $j \neq i$, a curva de Lorenz seria dada por OED. Por outro lado, quando todas as empresas tiverem o mesmo tamanho e todas as probabilidades forem iguais, a curva de Lorenz seria igual a diago-

nal OD. A razão entre as áreas OABCDO e OEDO mede, portanto, o grau de concentração da distribuição. Esta razão é o coeficiente de Gini:

$$G = \frac{L}{1/2} = 2L = 1 - 2A \qquad (3)$$

onde L representa a área OABCDO, a área OEDO é igual a 1/2, a área OABCDEO é igual a A e, obviamente, $L + A = 1/2$.

Quando G for igual a 1 a concentração será total, enquanto se G for igual a zero a igualdade será completa. Portanto, o coeficiente de Gini é um número compreendido entre zero e um:

$$0 \leq G \leq 1$$

O coeficiente de Gini, de um ponto de vista prático, pode ser medido através do seguinte procedimento. Denominemos o valor de $F_1(t)$ por Q_ℓ, ou seja:

$$F_1(t) = \frac{\sum_{i=1}^{\ell} p_i x_i}{\sum_{j=1}^{n} p_j x_j} = Q_\ell$$

A área A é igual a soma das áreas de n-1 trapézios e de um triângulo. Fazendo-se $Q_0 = 0$, a área A pode ser escrita como:

$$A = \sum_{i=1}^{n} \frac{(Q_i + Q_{i-1})}{2} p_i$$

Portanto, o coeficiente de Gini pode ser obtido através da seguinte fórmula:

$$G = 1 - \sum_{i=1}^{n} (Q_i + Q_{i-1}) p_i \tag{4}$$

O coeficiente de Gini é também definido através da seguinte expressão:

$$G = \frac{\sum_{i=1}^{n} \sum_{i=1}^{n} |x_i - x_j| p_i p_j}{2\mu} = \frac{\Delta}{2\mu} \tag{5}$$

onde Δ é a diferença média da distribuição. Esta definição não é muito útil de um ponto de vista prático para o cálculo de G mas ela torna mais fácil o estudo de decomposição do coeficiente de Gini que será visto logo adiante. Além disso,ela mostra claramente que o coeficiente de Gini é uma medida de dispersão relativa pois, segundo (5), G é igual a razão entre uma medida de dispersão, a metade da diferença média, e a média μ da distribuição.

Com a finalidade de mostrar que as fórmulas de G expressas por (4) e (5) são equivalentes, a diferença média Δ pode ser escrita do seguinte modo:

$$\Delta = \sum_{i} \sum_{j<i} p_i p_j (x_i - x_j) + \sum_{i} \sum_{j>i} p_i p_j (x_j - x_i)$$

$$= 2 \sum_{i} \sum_{j<i} p_i p_j (x_i - x_j)$$

Substituindo-se este valor de Δ em (5), resulta:

$$G = \frac{\sum_{i} \sum_{j<i} p_i p_j (x_i - x_j)}{\mu} \qquad (6)$$

Denominando-se por q_j a seguinte razão

$$q_j = \frac{p_j x_j}{\sum_{i=1}^{r} p_i x_i} = \frac{p_j x_j}{\mu}$$

o valor de Q_i será igual a:

$$Q_i = \sum_{j=1}^{i} q_j$$

Substituindo-se este valor em (4), o coeficiente de Gini passa, então, a ser dado por:

$$G = 1 - \sum_{i=1}^{n} p_i \left(\sum_{j=1}^{i} q_j + \sum_{j=1}^{i-1} q_j \right)$$

$$= 1 - \sum_{i=1}^{n} p_i \left(q_i + 2 \sum_{j=1}^{i-1} q_j \right)$$

ou:

$$G = 1 - \sum_{i=1}^{n} p_i q_i - 2 \sum_{i=1}^{n} p_i \sum_{j=1}^{i-1} q_j$$

Em virtude de

$$\sum_{i=1}^{n} p_i \sum_{j=1}^{i-1} q_j = \sum_{i} \sum_{j<i} p_i q_j$$

o coeficiente de Gini na expressão anterior pode ser escrito como:

$$G = 1 - \sum_{i} p_i q_i - 2 \sum_{i} \sum_{j<i} p_i q_j$$

Levando-se em conta a definição de q_j, a expressão acima para G transforma-se em:

$$G = 1 - \sum_{i} p_i \frac{p_i x_i}{\mu} - 2 \sum_{i} \sum_{j<i} p_i \frac{p_j x_j}{\mu}$$

ou, alternativamente:

$$G = \frac{\sum_{i} p_i x_i (1-p_i) - 2 \sum_{i} \sum_{j<i} p_i p_j x_j}{\mu}$$

Como

$$1 - p_i = \sum_{j \neq i} p_j = \sum_{j<i} p_j + \sum_{j>i} p_j \qquad ,$$

segue-se que:

$$\sum_i p_i x_i (1-p_i) = \sum_i p_i x_i \left(\sum_{j<i} p_j + \sum_{j>i} p_j \right)$$

$$= \sum_i \sum_{j>i} p_i p_j x_i + \sum_i \sum_{j<i} p_i p_j x_i$$

Substituindo-se o resultado acima na expressão (7), efetuando-se algumas simplificações, obtém-se:

$$G = \frac{\sum_i \sum_{j<i} p_i p_j (x_i - x_j)}{\mu}$$

Como esta fórmula é igual ao valor de G dado por (6), concluimos que as definições do coeficiente de Gini através das expressões (4) e (5) são equivalentes.

IV.1 - Decomposição do Coeficiente de Gini

Em algumas situações o analista está interessado em classificar, segundo algum critério, os valores que a variável aleatória pode tomar. Admita-se que os valores de x_i foram classificados em S grupos. Para facilitar a exposição introduziremos a seguinte notação:

- p_{si} = probabilidade da iésima observação do grupo s,
- x_{si} = valor da iésima observação do grupo s,
- p_{rj} = probabilidade da jésima observação do grupo r,
- x_{rj} = valor da jésima observação do grupo r.

A diferença média total, Δ, será, então, igual a:

$$\Delta = \sum_{s=1}^{S} \sum_{r=1}^{S} \sum_{i=1}^{n_s} \sum_{j=1}^{n_r} p_{si}\, p_{rj}\, |x_{si} - x_{rj}| \qquad (8)$$

onde n_s e n_r indicam o número de observações nos grupos $\underline{s}$ e $\underline{r}$, respectivamente.

A probabilidade de que uma observação pertença ao grupo s é igual a:

$$P_s = \sum_{i=1}^{n_s} p_{si}$$

e ao grupo r é dada por:

$$P_r = \sum_{j=1}^{n_r} p_{rj}$$

A diferença média dentro de cada grupo s, que representaremos por Δ_s, é igual a:

$$\Delta_s = \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \frac{p_{si}}{P_s} \frac{p_{sj}}{P_s} |x_{si} - x_{sj}|.$$

As médias dos grupos s e r são iguais a:

$$\mu_s = \sum_{i=1}^{n_s} \frac{p_{si}}{P_s} x_{si}$$

$$\mu_r = \sum_{i=1}^{n_r} \frac{p_{rj}}{P_r} x_{rj}$$

A diferença média entre os vários grupos é definida por:

$$\Delta_0 = \sum_{s=1}^{S} \sum_{r=1}^{S} P_s P_r \, |\mu_s - \mu_r|$$

O Coeficiente de Gini total, G, é igual a razão entre Δ e 2μ. Por sua vez, o coeficiente de Gini intra-grupo G_s, que mostra o grau de concentração dentro de cada grupo, e o coeficiente de Gini inter-grupo, G_0, que mostra o grau de concentração entre os grupos, são definidos através das seguintes expressões:

$$G_s = \frac{\Delta_s}{2\mu_s}$$

$$G_0 = \frac{\Delta_0}{2\ \mu}$$

Com a finalidade de obter a decomposição do coeficiente total de Gini nos coeficientes intra-grupos, G_s, $s = 1, \ldots, S$, e no coeficiente de Gini inter-grupo G_0, procederemos do seguinte modo. A diferença média total Δ (8) pode ser escrita como:

$$\Delta = \sum_{s} \sum_{r=s} \sum_{i} \sum_{j} p_{si} p_{rj} \, |x_{si} - x_{rj}| \, +$$

$$+ \sum_{s} \sum_{r \neq s} \sum_{i} \sum_{j} p_{si} p_{rj} \, |x_{si} - x_{rj}| \qquad (9)$$

Somando-se e subtratindo-se a esta expressão, a diferença média entre as médias dos grupos,

$$\sum_s \sum_r P_s P_r |\mu_s - \mu_r|$$

e levando-se em conta que:

$$\sum_s \sum_{r=s} \sum_i \sum_j p_{si} p_{rj} |x_{si} - x_{rj}| =$$

$$= \sum_s P_s^2 \sum_i \sum_j \frac{p_{si}}{P_s} \frac{p_{sj}}{P_s} |x_{si} - x_{sj}| = \sum_s P_s^2 \Delta_s$$

a expressão (9) transforma-se em:

$$\Delta = \Delta_0 + \sum_{s=1}^{S} P_s^2 \Delta_s + D \tag{10}$$

onde D é igual a:

$$D = \sum_s \sum_{r \neq s} \sum_i \sum_j p_{si} p_{rj} \{|x_{si} - x_{rj}| - |\mu_s - \mu_r|\} \tag{11}$$

Substituindo-se o valor da diferença média Δ obtido acima na definição (5) do coeficiente de Gini, resulta:

$$G = \frac{\Delta_0}{2\mu} + \sum_{s=1}^{S} \frac{P_s^2 \mu_s}{\mu} \frac{\Delta_s}{2\mu_s} + \frac{D}{2\mu}$$

ou, alternativamente:

$$G = G_0 + \sum_{s=1}^{S} w_s G_s + \frac{D}{2\mu} \qquad (12)$$

onde os pesos w_s são iguais a:

$$w_s = \frac{P_s^2 \mu_s}{\mu} \; , \quad \sum_{s=1}^{S} w_s < 1$$

A expressão (12) mostra que a decomposição do coeficiente de Gini contém três componentes: i) o coeficiente de Gini inter-grupo, ii) uma soma ponderada dos coeficientes de Gini intra-grupos, em que a soma dos pesos é inferior à unidade, e iii) um termo que reflete o grau pelo qual os vários grupos se superpõem.

Com o objetivo de entender o significado deste último termo iremos mostrar em que situação o valor de D é igual a zero. Comecemos por notar que a diferença entre as médias μ_s e μ_r pode ser escrita como:

$$\mu_s - \mu_r = \sum_i \frac{p_{si}}{P_s} x_{si} - \sum_j \frac{p_{rj}}{P_r} x_{rj}$$

$$\doteq \frac{P_r \sum_i p_{si} x_{si} - P_s \sum_j p_{rj} x_{rj}}{P_s \, P_r}$$

Levando-se em conta as definições de P_r e P_s a expressão anterior transforma-se em:

$$\mu_s - \mu_r = \frac{\sum_j p_{rj} \sum_i p_{si} x_{si} - \sum_s p_{si} \sum_j p_{rj} x_{rj}}{P_s \; P_r}$$

$$= \frac{\sum_i \sum_j p_{si} p_{rj} \; (x_{si} - x_{rj})}{P_s \; P_r}$$

O valor absoluto da diferença entre μ_s e μ_r é igual a:

$$|\mu_s - \mu_r| = \frac{|\sum_i \sum_j p_{si} p_{rj} (x_{si} - x_{rj})|}{P_s \; P_r}$$

Se $x_{si} > x_{rj}$ para todo i ε s e todo j ε r, ou se $x_{si} < x_{rj}$ nas mesmas condições, o que significa dizer que não existe superposição entre os grupos s e r, o valor absoluto da diferença entre μ_s e μ_r torna-se igual a:

$$|\mu_s - \mu_r| = \frac{\sum_i \sum_j p_{si} p_{rj} \; |x_{si} - x_{rj}|}{P_s \; P_r} \, .$$

É fácil verificar que neste caso o valor de D em (11) é igual a zero. Consequentemente, a decomposição do coeficiente de Gini passa a conter apenas dois termos de acordo com:

$$G = G_0 + \sum_{s=1}^{S} w_s G_s \qquad (13)$$

É interessante observar que na hipótese de não se dispor dos dados originais da distribuição mas sim de dados agrupados, classificados segundo os próprios valores de x_i, e se calcular o coeficiente de Gini baseado nos valores médios de cada classe, o coeficiente assim calculado, G_0, subestima o verdadeiro valor do coeficiente de Gini G, pois de (13) é fácil concluir que $G \geqslant G_0$.

V - O ÍNDICE DE HIRSCHMAN - HERFINDAHL

O índice de Hirschman-Herfindahl é definido por:

$$H = \sum_{i=1}^{n} p_i^2 \tag{14}$$

O valor máximo deste índice é igual a um quando $p_i = 1$ para algum i e p_j igual a zero para todo $j \neq i$.

O valor mínimo de H é igual a 1/n quando todos os valores de p_i forem iguais. Com efeito, a solução do problema de minimização condicionada de $H = \Sigma p_i^2$ com a condiçãode que $\Sigma p_i = 1$ é obtida com o auxílio da expressão de Lagrange:

$$\ell = \sum_i p_i^2 - \lambda(\sum_i p_i - 1)$$

Igualando-se a derivada parcial de ℓ com respeito a p_i a zero,

$$\partial\ell/\partial p_i = 2p_i - \lambda = 0$$

chega-se ao valor de p_i igual a $\lambda/2$, que é constante e independente do índice i. Logo, $p_i = 1/n$ e o valor mínimo de H é igual a:*

$$H_{min} = \sum_{i=1}^{n} (1/n)^2 = 1/n$$

O índice de Hirschman-Herfindahl está, então, compreendido entre um e 1/n, isto é:

$$1/n \leqslant H \leqslant 1$$

É fácil concluir que o índice H é um índice de concentração pois quanto maior o seu valor maior é a concentração da distribuição. Observe, também, que na hipótese do número de elementos n na distribuição crescer, $n \to \infty$, o valor mínimo do índice de Hirschman-Herfindahl tende para zero.

V.1 - Decomposição do Índice H

Com a finalidade de tornar o estudo da decomposição do índice de Hirschman-Herfindahl menos abstrata suponhamos que estamos estudando a concentração em uma determinada indústria e que dividimos, segundo algum critério, a indústria em S grupos. Designando por p_{si} a proporção de em-

(*) É fácil verificar que para $p_i=1/n$ o valor de H é mínimo examinando-se as derivadas segundas da função ℓ.

presas de tamanho i do grupo s, no total de indústria, o índice H passa a ser escrito como:

$$H = \sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si}^2 \tag{15}$$

A percentagem de cada grupo no total é igual a

$$P_s = \sum_{i=1}^{n_s} p_{si}$$

Portanto, o índice de Hirschman-Herfindahl inter-grupo é dado por:

$$H_0 = \sum_{s=1}^{S} P_s^2 \tag{16}$$

O índice de Hirschman-Herfindahl dentro de cada grupo é definido por:

$$H_s = \sum_{i=1}^{n_s} \left(\frac{p_{si}}{P_s}\right)^2, \quad s=1, \ldots, S \tag{17}$$

A decomposição do índice total nos índices inter e intra-grupos é facilmente obtida a partir de (15). Com efeito, multiplicando-se e dividindo-se os termos p_{si}^2 daquela expressão por P_s^2, resulta:

$$H = \sum_{s=1}^{S} P_s^2 \sum_{i=1}^{n_s} \left(\frac{p_{si}}{P_s}\right)^2$$

ou, alternativamente:

$$H = \sum_{s=1}^{S} P_s^2 H_s$$

Multiplicando-se e dividindo-se esta expressão por H_0 chega-se finalmente a fórmula de decomposição do índice H:

$$H = H_0 \sum_{s=1}^{S} \frac{P_s^2}{H_0} H_s = H_0 \sum_{s=1}^{S} w_s H_s \qquad (18)$$

onde:

$$w_s = \frac{P_s^2}{H_0} \qquad e \qquad \sum_{s=1}^{S} w_s = 1$$

A decomposição do índice de Hirschman-Herfindahl, segundo (18), é do tipo multiplicativo. O índice total, H, é igual ao produto do índice inter-grupo, H_0, por uma média ponderada dos índices intra-grupos, $\Sigma w_s H_s$, com a soma dos pesos, w_s, igual a unidade.

VI - A ENTROPIA E A REDUNDÂNCIA

A entropia é definida por

$$I = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} = - \sum_{i=1}^{n} p_i \log p_i \qquad (19)$$

e seu valor mínimo é igual a zero enquanto seu valor máximo é igual a log n.

O mínimo da entropia ocorre quando $p_i = 1$ para algum i e $p_j = 0$, para todo $j \neq i$, lembrando-se que o limite de $x \log x$ é igual a zero quando x se aproxima de zero.

O valor máximo da entropia é obtido através da solução do seguinte problema de máximo condicionado:

Maximizar $- \Sigma p_i \log p_i$ com a condição de que $\Sigma p_i = 1$

A expressão de Lagrange para este problema é:

$$\ell = - \sum_{i=1}^{n} p_i \log p_i - \lambda\left(\sum_{i=1}^{n} p_i - 1\right)$$

onde λ é o multiplicador de Lagrange. Igualando-se a zero a derivada parcial de ℓ com respeito a p_i,

$$\frac{\partial \ell}{\partial p_i} = -1 - \log p_i - 1 = 0, \quad i=1, \ldots, n$$

obtém-se para a probabilidade p_i o seguinte valor:

$$p_i = e^{-\lambda-1}, \quad i=1, \ldots, n$$

que independe do índice i. Logo, todas as probabilidades p_i são iguais a $1/n$ e a entropia máxima é igual a:*

$$I_{max} = \sum_{i=1}^{n} \frac{1}{n} \log n = \log n$$

(*) A condição de segunda ordem confirma que para $p_i = 1/n$ o valor de I é máximo. Para isto, basta que se examine os sinais das derivadas segundas de ℓ com respeito a p_i.

A entropia I é, portanto, uma medida inversa de concentração, ou seja I é uma medida de dispersão, pois quando todas as probabilidades forem iguais a incerteza é máxima o mesmo ocorrendo com o valor de entropia.

Uma medida de concentração, denominada de redundância, é facilmente deduzida a partir da entropia. Com efeito, subtraindo-se do logaritmo (neperiano) de n o valor da entropia I obtém-se a redundância R:

$$R = \log n - \sum_{i=1}^{n} p_i \log \frac{1}{p_i} = \sum_{i=1}^{n} p_i \log np_i \qquad (20)$$

A redundância está compreendida entre zero e o logaritmo de n:

$$0 \leqslant R \leqslant \log n$$

pois quando I for igual a zero R é igual ao logarítmo de n, e quando I = log n, R é igual a zero.

VI.1 - Decomposição da Entropia e da Redundância

Usando a notação introduzida quando tratamos de decomposição do coeficiente de Gini e do índice de Hirschman-Herfindahl, a entopia inter-grupo é igual a:

$$I_0 = \sum_{s=1}^{S} P_s \log \frac{1}{P_s} \qquad (21)$$

e as entropias intra-grupos são definidas através de:

$$I_s = \sum_{i=1}^{n_s} \frac{p_{si}}{P_s} \log \frac{P_s}{p_{si}} , \quad s=1, \ldots, S \tag{22}$$

onde

$$P_s = \sum_{i=1}^{n_s} p_{si}$$

Como a entropia total é igual a:

$$I = \sum_{i=1}^{S} \sum_{i=1}^{n_s} p_{si} \log \frac{1}{p_{si}} \tag{23}$$

esta expressão pode também ser escrita do seguinte modo:

$$I = \sum_{s=1}^{S} P_s \sum_{i=1}^{n_s} \frac{p_{si}}{P_s} \left[\log \frac{P_s}{p_{si}} - \log P_s \right]$$

que depois de algumas manipulações algébricas transforma-se em:

$$I = \sum_{s=1}^{S} P_s \sum_{i=1}^{n_s} \frac{p_{si}}{P_s} \log \frac{P_s}{p_{si}} + \sum_{s=1}^{S} P_s \log \frac{1}{P_s}$$

Levando-se em conta as definições das entropias inter e intra-grupo a expressão acima resulta em:

$$I = I_0 + \sum_{s=1}^{S} w_s I_s \tag{24}$$

onde

$$w_s = P_s \quad e \quad \Sigma w_s = 1$$

A entropia total I é, então, igual a entropia inter-grupo I_0, adicionada a uma média ponderada das entropias intra-grupos, em que a ponderação de cada grupo é igual a sua participação no total. Este tipo de decomposição da entropia constitui-se em um dos principais atrativos na utilização desta medida (inversa) de concentração.

A decomposição da redundância é análoga à de entropia. Isto é:

$$R = R_0 + \sum_{s=1}^{S} P_s R_s \tag{25}$$

onde R_0 é a redundância inter-grupo e R_s é a redundância intra-grupo, que são definidas, respectivamente, por:

$$R_0 = \sum_{s=1}^{S} P_s \log \frac{P_s}{n_s/n} \tag{26}$$

e:

$$R_s = \sum_{s=1}^{S} \frac{p_{si}}{P_s} \log \frac{p_{si}/P_s}{1\ /n_s} \tag{27}$$

A dedução da fórmula da decomposição de redundância pode ser feita facilmente a partir da seguinte expressão de redundância total:

$$R = \sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si} \log n\, p_{si}$$

com o mesmo tipo de procedimento que seguimos no caso da entropia.

VI - A VARIÂNCIA LOGARÍTMICA

A variância logarítmica é uma medida bastante popular de dispersão relativa e sua definição é dada por:

$$V = \sum_{i=1}^{n} p_i \; (\log x_i - \log x_g)^2 = \sum_{i=1}^{n} p_i \; (\log \frac{x_i}{x_g})^2 \qquad (28)$$

onde x_g é a média geométrica da distribuição:

$$\log x_g = \sum_{i=1}^{n} p_i \; \log x_i \qquad (29)$$

ou, alternativamente:

$$x_g = \prod_{i=1}^{n} x_i^{p_i}$$

A variância logarítmica é sempre um número não negativo, igual a zero quando todos os valores de x_i forem iguais entre si, e é uma medida relativa de dispersão pois, de acordo com a sua definição, a variância logaritmica é igual ao valor esperado do quadrado do logaritmo da variável aleatória normalizada pela média geométrica.

VI.1 - Decomposição da Variância Logarítmica

Quando se classifica as observações em S grupos a variância total passa a ser expressa por:

$$V = \sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si} \; (\log x_{si} - \log x_g)^2 \qquad (30)$$

onde a notação é idêntica a usada anteriormente.

A variância inter-grupo é igual a

$$V_s = \sum_{s=1}^{S} P_s \,(\log x_{gs} - \log x_g)^2 \qquad (31)$$

onde x_{gs} é a média geométrica do grupo s:

$$\log x_{gs} = \sum_{i=1}^{n_s} \frac{p_{si}}{P_s} \log x_{si} \qquad (32)$$

A variância intra-grupo é definida através da seguinte fórmula:

$$V_s = \sum_{i=1}^{n_s} \frac{p_{si}}{P_s} (\log x_{si} - \log x_{gs})^2, \; s=1, \ldots, S \qquad (33)$$

A decomposição inter-grupo e intra-grupo é obtida com o seguinte procedimento. A variância total pode ser escrita como:

$$V = \sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si} \left[(\log x_{si} - \log x_{gs}) + (\log x_{gs} - \log x_g)\right]^2$$

que é igual a:

$$V = \sum_{s=1}^{S} P_s \sum_{i=1}^{n_s} \frac{p_{si}}{P_s} (\log x_{si} - \log x_{gs})^2 +$$

$$+ \sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si} (\log x_{gs} - \log x_g)^2$$

$$+ 2 \sum_{s=1}^{S} \sum_{i=1}^{n_s} p_{si} (\log x_{si} - \log x_{gs})(\log x_{gs} - \log x_g)$$

O último termo desta expressão é igual a zero em virtude da definição da média geométrica x_{gs}. Logo, temos que:

$$V = \sum_{s=1}^{S} P_s \sum_{i=1}^{n_s} \frac{p_{si}}{P_s} (\log x_{si} - \log x_{gs})^2 +$$

$$+ \sum_{s=1}^{S} P_s (\log x_{gs} - \log_g)^2$$

Alternativamente:

$$V = V_0 + \sum_{s=1}^{S} P_s V_s \tag{34}$$

A variância total é, portanto, igual a soma da variância inter-grupo com uma média ponderada das variâncias intra-grupos, cujos pesos são iguais a participação de cada grupo no total. A decomposição da variância é do mesmo tipo da decomposição da entropia e da redundância.

BIBLIOGRAFIA

- BHATTACHARYA, N. e B. MAHALANOBIS (1967). "Regional Disparities in Household Consumption in India", Journal of the American Statistical Association, Vol. 62, p. 143-161.
- GASTWIRTH, J. L. (1972). "The Estimation of the Lorenz Curre and Gini Index", the Review of Economics and Statistics , Vol. 54, p. 306-316.
- HART, P.E. (1971). "Entropy and Other Measures of Concentration", Journal of the Royal Statistical Society,Series A, vol. 134, p. 73-85.
- HART, P.E. (1975). "Moment Distribution in Economics: An Exposition", Journal of the Royal Statistical Society, Series A, vol. 138, p. 423 - 434.
- RAO, V. M. (1969). "Two Decompositions of Concentration Ratio", Journal of the Royal Statistical Society, Series A, Vol. 132, p. 418-425.
- PYATT, G. (1976). "On the Interpretation and the Disaggregation of Gini Coefficients", Economic Journal, 86, p. 243-255.
- STIGLER, G. J. (1968). "The Measurement of Concentration", in: The Organization of Industry, Capítulo 4. Homewood, Illinois: Richard D. Irwin, Inc.
- THEIL, H. (1967). Economics and Information Theory. Amsterdam: North-Holland.

ENSAIOS ECONÔMICOS DA EPGE

1. ANÁLISE COMPARATIVA DAS ALTERNATIVAS DE POLÍTICA COMERCIAL DE UM PAÍS EM PROCESSO DE INDUSTRIALIZAÇÃO - Edmar Bacha - 1970 (ESGOTADO)

2. ANÁLISE ECONOMÉTRICA DO MERCADO INTERNACIONAL DO CAFÉ E DA POLÍTICA BRASILEIRA DE PREÇOS - Edmar Lisboa Bacha - 1970 (ESGOTADO)

3. A ESTRUTURA ECONÔMICA BRASILEIRA - Mario Henrique Simonsen - 1971 (ESGOTADO)

4. O PAPEL DO INVESTIMENTO EM EDUCAÇÃO E TECNOLOGIA NO PROCESSO DE DESENVOLVIMENTO ECONÔMICO - Carlos Geraldo Langoni - 1972 (ESGOTADO)

5. A EVOLUÇÃO DO ENSINO DE ECONOMIA NO BRASIL - Luiz de Freitas Bueno - 1972

6. POLÍTICA ANTI-INFLACIONÁRIA - A CONTRIBUIÇÃO BRASILEIRA - Mario Henrique Simonsen - 1973 (ESGOTADO)

7. ANÁLISE DE SÉRIES DE TEMPO E MODELO DE FORMAÇÃO DE EXPECTATIVAS - José Luiz Carvalho - 1973 (ESGOTADO)

8. DISTRIBUIÇÃO DA RENDA E DESENVOLVIMENTO ECONÔMICO DO BRASIL: UMA REAFIRMAÇÃO - Carlos Geraldo Langoni - 1973 (ESGOTADO)

9. UMA NOTA SOBRE A POPULAÇÃO ÓTIMA DO BRASIL - Edy Luiz Kogut - 1973

10. ASPECTOS DO PROBLEMA DA ABSORÇÃO DE MÃO-DE-OBRA: SUGESTÕES PARA PESQUISAS - José Luiz Carvalho - 1974 (ESGOTADO)

11. A FORÇA DO TRABALHO NO BRASIL - Mario Henrique Simonsen - 1974 (ESGOTADO)

12. O SISTEMA BRASILEIRO DE INCENTIVOS FISCAIS - Mario Henrique Simonsen - 1974 (ESGOTADO)

13. MOEDA - Antonio Maria da Silveira - 1974 (ESGOTADO)

14. CRESCIMENTO DO PRODUTO REAL BRASILEIRO - 1900/1947 - Claudio Luiz Haddad 1974 (ESGOTADO)

15. UMA NOTA SOBRE NÚMEROS ÍNDICES - José Luiz Carvalho - 1974 (ESGOTADO)

16. ANÁLISE DE CUSTOS E BENEFÍCIOS SOCIAIS I - Edy Luiz Kogut - 1974 (ESGOTADO)

17. DISTRIBUIÇÃO DE RENDA: RESUMO DA EVIDÊNCIA - Carlos Geraldo Langoni - 1974 (ESGOTADO)

18. O MODELO ECONOMÉTRICO DE ST. LOUIS APLICADO NO BRASIL: RESULTADOS PRELIMINARES - Antonio Carlos Lemgruber - 1975

19. OS MODELOS CLÁSSICOS E NEOCLÁSSICOS DE DALE W. JORGENSON - Eliseu R. de Andrade Alves - 1975

20. DIVID: UM PROGRAMA FLEXÍVEL PARA CONSTRUÇÃO DO QUADRO DE EVOLUÇÃO DO ESTUDO DE UMA DÍVIDA - Clóvis de Faro - 1974

21. ESCOLHA ENTRE OS REGIMES DA TABELA PRICE E DO SISTEMA DE AMORTIZAÇÕES CONS~~TANTES: PONTO-DE-VISTA~~ DO MUTUÁRIO - Clóvis de Faro - 1975

22. ESCOLARIDADE, EXPERIÊNCIA NO TRABALHO E SALÁRIOS NO BRASIL - José Julio Senna - 1975

23. PESQUISA QUANTITATIVA NA ECONOMIA - Luiz de Freitas Bueno - 1978

24. UMA ANÁLISE EM CROSS-SECTION DOS GASTOS FAMILIARES EM CONEXÃO COM NUTRIÇÃO, SAÚDE, FECUNDIDADE E CAPACIDADE DE GERAR RENDA - José Luiz Carvalho 1978

25. DETERMINAÇÃO DA TAXA DE JUROS IMPLÍCITA EM ESQUEMAS GENÉRICOS DE FINANCIAMENTO: COMPARAÇÃO ENTRE OS ALGORÍTMOS DE WILD E DE NEWTON-RAPHSON - Clovis de Faro - 1978

26. A URBANIZAÇÃO E O CÍRCULO VICIOSO DA POBREZA: O CASO DA CRIANÇA URBANA NO BRASIL - José Luiz Carvalho e Uriel de Magalhães - 1979

27. MICROECONOMIA - Parte 1 - FUNDAMENTOS DA TEORIA DOS PREÇOS - Mario Henrique Simonsen - 1979

28. ANÁLISE DE CUSTOS E BENEFÍCIOS SOCIAIS II - Edy Luiz Kogut - 1979

29. CONTRADIÇÃO APARENTE - Octávio Gouvêa de Bulhões - 1979

30. MICROECONOMIA - Parte 2 - FUNDAMENTOS DA TEORIA DOS PREÇOS - Mario Henrique Simonsen - 1980 (Em fase de impressão)

31. A CORREÇÃO MONETÁRIA NA JURISPRUDÊNCIA BRASILEIRA - Arnold Wald - 1980

32. MACROECONOMIA - Parte A - TEORIA DA DETERMINAÇÃO DA RENDA E DO NÍVEL DE PREÇOS - José Julio Senna - 2 Volumes - 1980

33. ANÁLISE DE CUSTOS E BENEFÍCIOS SOCIAIS III - Edy Luiz Kogut - 1980

34. MEDIDAS DE CONCENTRAÇÃO - Fernando de Holanda Barbosa - 1981