

Nota sobre o papel da educação na diferenciação das rendas entre 1960 e 1970

RAMONAVAL AUGUSTO COSTA

1 — Introdução

Com o crescimento alcançado pela economia brasileira nos últimos anos, levantou-se uma discussão referente à situação da distribuição da renda pessoal no Brasil. A opinião dos autores que primeiramente fizeram uso dos dados de renda pessoal do Censo Demográfico de 1970 comparando-os com os do Censo de 1960, é de que entre 1960 e 1970 houve um aumento do grau de concentração da renda, conforme demonstrado por vários índices de concentração calculados para esses dois anos.

Entre os trabalhos que mais se destacaram na utilização dos dados dos Censos Demográficos de 1960 e 1970 encontramos primeiramente os trabalhos de Rodolfo Hoffmann,¹ Albert Fishlow² e João Carlos Duarte.³ Estes autores, principalmente Fishlow, são de opinião que grande parte deste aumento no grau de concentração da renda pode ser atribuído, parcialmente, às políticas econômicas adotadas pelo governo brasileiro, desde 1964. Em seguida temos a contribuição de Langoni que apareceu pela primeira vez em 1972, em forma de artigo e, em 1973, em forma de livro, numa versão mais completa.⁴ Tal trabalho foi

- 1 Rodolfo Hoffmann, "Contribuição à análise da distribuição da renda e da posse da terra no Brasil", Piracicaba — 1969, tese de livre docência para Eça-Luiz de Queiroz, não publicado.
- 2 Albert Fishlow, "Brazilian Size Distribution of Income", *American Economic Review*, maio 1972, página 391.
- 3 João Carlos Duarte, "Aspectos da distribuição da renda no Brasil em 1970", Piracicaba — 1972, tese de mestrado para Eça-Luiz de Queiroz, não publicado.
- 4 Carlos Geraldo Langoni, *Distribuição da renda e desenvolvimento econômico do Brasil*, editora Expressão e Cultura, 1973, tendo sido publicado com o mesmo título pela revista *Estudos Econômicos*, vol. 2, 1972 n.º 5, em forma de artigo.

amplamente divulgado através da revista *Veja* e imprensa em geral e, dado as circunstâncias em que foi realizado, sob o patrocínio do Ministério da Fazenda, representaria o ponto de vista oficial relativamente à situação da distribuição da renda no Brasil.

Considero o trabalho de Langoni muito interessante do ponto de vista metodológico.⁵ Isto, contudo, não o isenta de cometer erros de interpretação, descuidos e omissões no uso de um instrumental relativamente sofisticado para a média interessada no problema de distribuição de renda no Brasil.

Algumas críticas já apareceram dando ênfase, principalmente, às omissões de Langoni. Este tipo de crítica é muito importante, já que quaisquer trabalhos, por mais completos que sejam, sempre deixam de lado aspectos importantes a serem estudados. Porém, num País onde a pesquisa acadêmica é ainda incipiente, acredito ser mais eficiente do ponto de vista científico dar-se ênfase a críticas que dizem respeito ao que foi realizado em si, preocupando-se menos com o que se deixou de fazer. Tendo em vista estes elementos resolvi me ater a alguns problemas que penso serem relevantes para a validade das idéias defendidas pelo autor, tomando por base principalmente o que foi realizado.

Procurei concentrar minhas observações sobre o que está contido no capítulo 5, sob o título "As Causas das Mudanças na Distribuição da Renda entre 1960 e 1970". No meu ponto de vista este capítulo é o mais importante para a apresentação da educação como um dos principais fatores responsáveis pela concentração de renda no período 1960/1970.

Neste ensaio destacam-se três partes relevantes. Primeiro apresento, rapidamente, alguns problemas referentes à especificação do modelo usado por Langoni. Em seguida tento ressaltar algumas características do modelo que não foram devidamente explicadas. Por fim, procuro mostrar como as evidências sobre o aumento da contribuição marginal da educação na diferenciação das rendas não possuem bases tão sólidas como seria desejável.

2 — Alguns Problemas na Especificação do Modelo Usado

O capítulo 5 apresenta algumas imperfeições na descrição do sofisticado instrumental usado. Como o autor mesmo afirma na página 106, "optou-se por uma análise da variância com base numa regressão loglinear em que todas as variáveis são *dummies*". Neste método de análise de variância, a matriz *X* não deve ser chamada matriz de observações, pois neste modelo somente a matriz *Y* é constituída de observações, ao passo que a matriz *X* corresponde a uma matriz de variáveis classificadoras.

Outra deficiência na descrição deste modelo consiste numa admissão tácita, por parte do autor, na página 105, da não inclusão de certas variáveis, como *status* migratório, *back ground* familiar, em virtude de uma série de impossibilidades, ressaltando que o modelo apresenta problemas de especificação com a omissão de algumas variáveis. Como as variáveis omitidas no modelo estão certamente correlacionadas com pelo menos uma das variáveis presentes, o modelo adotado está sujeito a viés de especificação. Na página 107, entretanto, o autor tentou justificar o uso do método dos mínimos quadrados ao invés dos mínimos

⁵ Consideração feita por mim em "Reflexões a respeito do trabalho do Langoni sobre distribuição de renda no Brasil", não publicado, 1973.

quadrados generalizados, como seria mais apropriado devido à heterocedasticidade, alegando que o método dos mínimos quadrados não acarreta viés, mas somente perda de eficiência. Entretanto, nota-se que o viés já era de se esperar em razão da omissão de variáveis. Portanto, usar o método dos mínimos quadrados nestas condições não elimina o problema do viés.

Uma outra deficiência na apresentação do modelo refere-se às hipóteses a respeito do termo erro. O autor parece sugerir que a hipótese de normalidade não é necessária para se utilizar o método dos mínimos quadrados, o que é verdade. Mas para falar de significância dos parâmetros, requerem-se necessárias as hipóteses relativas à distribuição dos erros, principalmente em se tratando de análise de variância.⁶

O que se pode concluir desta primeira parte é que, em virtude da omissão de variáveis, os coeficientes obtidos estão viesados. Mais ainda, sem a hipótese de normalidade torna-se ilegítima a pretensão de usar os valores de "t" ou "F" para falar da significância dos parâmetros.⁷

3 — Peculiaridades do Modelo de Variáveis Dummies

Todo economista iniciado no estudo de econometria certamente já entrou em contacto com o exemplo clássico de uso de variáveis *dummies*, a estimação da função consumo dos EUA quando o conjunto de dados é dividido em dois subconjuntos, um referente aos dados de antes da II Guerra e o outro referente aos dados de depois dela. No entanto, a utilização de um modelo classificatório, onde todas as variáveis independentes são *dummies*, com o intuito de se fazer uma análise da variância, usando-se o método de regressão linear, é de uso pouco frequente entre os economistas.⁸

Langoni, ao usar este método, deveria esclarecer melhor suas propriedades, pois, apesar de adotar o método de regressão linear, as propriedades dos parâmetros diferem das de um modelo usual de regressão, onde as variáveis independentes constituem também um conjunto de observações. Como se salientou antes, a matriz das variáveis independentes não é constituída de observações. As únicas observações existentes são as da variável dependente. Em modelos desta natureza existe uma certa arbitrariedade nos valores dos coeficientes, fato que diminui um pouco a relevância da afirmação de Langoni de que "a magnitude

6 Para melhor esclarecimento sobre as hipóteses necessárias numa análise de variância pode-se consultar qualquer compêndio de Estatística ou Econometria, como por exemplo o livro *Design and Analysis of Experiments* de Oscar Kempthorne, publicado por John Wiley & Sons, Inc. página 72, capítulo 6, edição de 1952.

7 No entanto, Langoni afirma que "O que é efetivamente relevante é a significância estatística das variáveis consideradas no modelo e não necessariamente o seu poder de explicação". Dando tanta ênfase à significância estatística, é de se estranhar que o autor não tenha feito nenhuma menção sobre a forma da distribuição dos erros, vide página 168, *Pesquisa e Planejamento Econômico*, vol. 4, fevereiro 1974, n.º 9.

8 O exemplo mais antigo que tenho conhecimento é o artigo de T. P. Hill publicado na revista *Econométrica*, vol. 27, julho de 1959, "An Analysis of the Distribution of Wages and Salaries in Great Britain", onde o autor usa o método de variáveis *dummies*, já que o método comum de análise de regressão fica prejudicado, pois os fatores em termos dos quais a renda varia são em grande parte atributos não quantificáveis. Apesar de justificar adequadamente a conveniência do uso do método em discussão, chama atenção para o fato de que as classificações escolhidas são, na sua maioria, subjetivas, envolvendo um certo grau de arbitrariedade.

dos coeficientes já dá uma primeira idéia do impacto de cada variável".⁹

Para ilustrar melhor o que se passa com o modelo utilizado no capítulo 5, vamos admitir que apenas educação e setor de atividades sejam as variáveis relevantes para a especificação do modelo. Fazendo uso da notação encontrada neste capítulo tem-se:

$$Y_n = B_o E_o + \sum_{i=1}^2 B_{1i} E_{ih} + \sum_{j=1}^2 B_{2j} A_{jh} + E_h$$

onde $E_o = 1$, E_{ih} e A_{jh} representam, respectivamente, a variável educação e setor de atividade. Resumindo o modelo numa forma geral, tem-se $Y = XB + E$, onde Y é o vetor das observações e X é a matriz cujos elementos são constituídos de valores 0 (zero) e 1 (um) e E é o vetor das perturbações aleatórias.

Usando-se uma notação que ressalta melhor a composição de cada valor da variável observada Y , apresenta-se o modelo na seguinte forma:

$$Y_{ij} = B_o E_o + B_{11} E_1 + B_{12} E_2 + B_{21} A_1 + B_{22} A_2 + E_{ij}$$

onde $E_o = 1$ para todo Y_{ij} ; $E_1 = 1$ para todo Y_{ij} onde $i = 1$ e $E_1 = 0$ para qualquer outro Y_{ij} ; $E_2 = 1$ para todo Y_{ij} onde $i = 2$ e $E_2 = 0$ para qualquer outro Y_{ij} ; $A_1 = 1$ para todo Y_{ij} onde $j = 1$ e $A_1 = 0$ para qualquer outro Y_{ij} ; $A_2 = 1$ para todo Y_{ij} onde $j = 2$ e

$A_2 = 0$ para qualquer outro Y_{ij} . Assim observa-se que as variações explicativas (E, A) assumem valores 0 e 1, funcionando como fatores de classificação da variável dependente efetivamente observada. Vejamos agora como se poderia interpretar este modelo num sistema de classificação de duas entradas, geralmente denominado tábua de contingência nos estudos de análise de variância:

		ATIVIDADE		
		A_j	1	2
EDUCAÇÃO	E_i			
	1	Y_{11}	Y_{12}	$Y_{1.}$
	2	Y_{21}	Y_{22}	$Y_{2.}$
	Σ_i	$Y_{.1}$	$Y_{.2}$	$Y_{..}$

Através deste quadro evidencia-se que as únicas observações existentes no modelo em discussão são devidas à variável Y .

9 Num modelo desta natureza, a obtenção de valores para os parâmetros envolve um certo grau de arbitrariedade que pode ser evidenciada pela seguinte afirmação de Oscar Kempthorne: "Unique solutions for the original parameters can only be obtained by imposing conditions on the parameters", vide página 69, capítulo 6, do seu livro *The Design and Analysis of Experiments*. O mesmo problema evidencia-se também quando H. P. Hill no seu artigo "An Analysis of the Distribution of Wages and Salaries in Great Britain" afirma: "A characteristic feature of a model of this type is that it is not fully determinate as it stands".

Sabe-se que o estimador de mínimos quadrados de B , num modelo linear do tipo $Y = XB + E$, pode ser obtido a partir do seguinte sistema $X'XB = X'Y$ de equações normais. Vejamos como nos modelos lineares com variáveis *dummies* a matriz $X'X$ é singular. Na ilustração de tal fato vamos utilizar, mais uma vez, o pequeno exemplo empregado até aqui. Examinemos a matriz X :

$$\begin{array}{rcccccc}
 & E_0 & E_1 & E_2 & A_1 & A_2 \\
 Y_{11} & 1 & 1 & 0 & 1 & 0 \\
 Y_{12} & 1 & 1 & 0 & 0 & 1 \\
 Y_{21} & 1 & 0 & 1 & 1 & 0 \\
 Y_{22} & 1 & 0 & 1 & 0 & 1
 \end{array}$$

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Conhecido X , observa-se que para o exemplo utilizado-se tem-se:

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix} =$$

$$X'X = \begin{bmatrix} 4 & 2 & 2 & 2 & 2 \\ 2 & 2 & 0 & 1 & 1 \\ 2 & 0 & 2 & 1 & 1 \\ 2 & 0 & 2 & 1 & 1 \\ 2 & 1 & 1 & 0 & 2 \end{bmatrix}$$

Neste exemplo verifica-se que $|X'X| = 0$, visto que a 4.^a e 5.^a linhas somadas igualam a primeira. Este exemplo justifica a diferença entre o modelo apresentado no capítulo 5 e os modelos usuais de regressão múltipla, onde $|X'X| \neq 0$, isto é, vale a condição de *full rank*. No caso extremo em que $|X'X| = 0$, tem-se o problema de perfeita multicoli-

nearidade.¹⁰ O importante é que se o modelo não for de *full rank* as características dos parâmetros serão afetadas. Mais precisamente, neste tipo de modelo o número de parâmetros independentes é menor que o número de parâmetros a serem obtidos.¹¹ Isto talvez explique porque Langoni, na página 107, diz: “A fim de evitar a singularidade da matriz $X'X$ de produtos cruzados em cada conjunto de *dummies*, um dos valores é deixado no termo constante”.

Aqui reside precisamente o elemento de arbitrariedade do modelo usado. O valor de B_0 pode ser escolhido arbitrariamente. No trabalho em questão, B_0 representa a média geométrica (em logs) da renda dos indivíduos analfabetos, no setor primário, com mais de 70 anos de idade, na Região VI (Norte e Centro-Oeste) e do sexo feminino.

A arbitrariedade de um modelo desta natureza é uma de suas características mais importantes. A infinidade de soluções (não estimativas) deve-se ao fato de $X'X$ ser singular e não possuir, portanto, nenhuma inversa. Sendo assim, o pesquisador, ao usar este modelo, deve deixar mais clara a técnica utilizada para levantar a singularidade de $x'x$.¹²

Outra relevante propriedade deste tipo de modelo é que apesar de ter infinitas soluções, o valor R^2 permanece constante, ou seja, existem infinitas combinações lineares que dão a estimativa da variável dependente (no caso em questão, renda).¹³ Finalmente, a mais relevante característica deste tipo de modelo, a despeito das infinitas soluções para os B , é a possibilidade de obtenção de estimativas através de estimadores de funções estimáveis dos parâmetros, estimativas essas que são invariantes às infinitas soluções das equações normais para o mesmo grupo de dados. Deve-se lembrar, no entanto, que mesmo essas funções estimáveis não são únicas.¹⁴

10 Para se entender melhor a diferença entre este modelo e os modelos comuns de regressão múltipla, nos quais se exige que $|X'X| \neq 0$ para a estimação dos parâmetros, veja-se página 123, capítulo 5 do livro *Econometric Methods*, 2.^a edição de autoria de J. Johnston, para uma completa análise dos modelos de *rank* completo (*models of full rank*). Para a análise dos modelos de *rank* incompleto (*models of not full rank*) vide os capítulos 5 e 6 do livro *Linear Models* de autoria de S. R. Searle. Este último livro discute todos os modelos lineares, apresentando os modelos de *rank* completo, onde $|X'X| \neq 0$ e os modelos de *rank* incompleto, onde $|X'X| = 0$, considerando-se casos especiais de um modelo linear generalizado. Porém, não se deve pensar que, pelo fato de se ter generalizado os modelos lineares, deixam de existir diferenças entre os modelos de regressão múltipla que o economista está acostumado a encontrar e o modelo que Langoni utiliza no capítulo 5. Para melhor ilustrar o porquê de nossa posição veja-se o que Searle diz no capítulo 4, página 143: “Now, however, X does not have full column rank — as seen in (5) the sum of its 3 columns equals its first. Thus is a model described as a “model not of full rank”. Its property is that X does not have full column rank, with the important consequence that $(X'X)^{-1}$ does not exist and so $X'X\delta = X'y$ can not be solved as $\delta = (X'X)^{-1}X'y$. However, by using generalized inverse of $X'X$ solutions can be found; but before discussing them, in chapter 5, we give another example and then describe other aspects of linear models.”

11 De fato, como se pode verificar através da exposição de Oscar Kempthorne, no caso do modelo em questão, tem-se $1 + i + j$ parâmetros para serem estimados e apenas $i + j - 1$ parâmetros independentes. A mesma observação é feita por H. T. Hill na página 137 do seu artigo já citado.

12 Para evitar discussões sobre este ponto, já que alguns economistas e mesmo estagiários têm relatado em entender o que pretendo chamar atenção, seria interessante repetir o que Searle diz no capítulo 5, página 168: “Whenever $X'X$ is not of full rank, as in (3), the normal equations $X'Xb = X'y$ cannot be solved with one solitary solution $b = (X'X)^{-1}X'y$ as in chapter 3. Many solutions are available”.

13 Para a constatação desta propriedade veja-se Searle, *op. cit.*, capítulo 5, página 172.

14 Vide Searle no capítulo 4, página 162 para esclarecimento: “Functions such as these are known as *estimators of estimable functions*. They all have the property to they are invariante to whatever solution is obtained to the normal equations. Because of this invariance property they are the only functions that can be of interest, so far as estimation of the parameters of a linear model is concerned”.

Nestas circunstâncias, a afirmativa de que B_0 foi mantido como o intercepto, no modelo, é uma informação insuficiente para caracterizar se os valores divulgados dos coeficientes na tabela 5.1 são estimativas obtidas através de estimadores de alguma função estimável dos parâmetros, ou somente uma das infinitas soluções que o modelo oferece.¹⁵ Pela afirmativa presente no livro de Langoni tem-se a impressão de que se trata de uma das várias maneiras de se eliminar a singularidade de $X'X$. Mesmo que a introdução de B_0 consiga, por si só, eliminar a singularidade de $X'X$, tornando possível a solução do modelo, ainda assim ela é insuficiente para garantir que a solução obtida seja um estimador de uma função estimável dos parâmetros do modelo. Se os valores dos coeficientes divulgados na tabela 5.1 não são resultados de estimadores de funções estimáveis dos parâmetros do modelo, então seriam apenas uma das infinitas soluções das equações normais $X'y = X'XB$.

Se os valores dos coeficientes divulgados referem-se a uma das infinitas soluções do modelo, a interpretação de cada um não comporta nenhum sentido, a não ser que contribuam para estimar a variável dependente. Nestas circunstâncias a relevância do modelo se restringiria à obtenção de estimativas para a variável dependente, ficando a significância dos coeficientes obtidos destituída de base para qualquer afirmação a respeito dos valores teóricos dos parâmetros. Além disso, mesmo que os coeficientes divulgados sejam estimativas obtidas através de estimadores que são funções estimáveis dos parâmetros, o pesquisador para falar em significância deveria investigar qual a forma da distribuição da variável dependente (pelo menos para preencher os requisitos do instrumental, já que não temos poderes para fazer as variáveis se distribuírem como o instrumental exige) e conseqüentemente das perturbações aleatórias. E a fim de ter uma idéia sobre a possível forma da distribuição das estimativas. Só assim Langoni poderia se basear na significância das estimativas para defender as evidências obtidas.¹⁶

Em função da importância das conclusões do trabalho, acredito que seria imprescindível que se explicasse melhor se os valores divulgados para os coeficientes referem-se a uma das infinitas soluções ou estimativas a partir de estimadores de funções estimáveis dos parâmetros do modelo. Esta é uma grande deficiência na apresentação deste modelo, por parte do pesquisador, fazendo com que a maioria dos leitores fique pensando que se trata de um modelo de regressão linear múltipla de *rank* completo.

15 Dado o modelo linear $y = XB + E$, onde X é uma matriz de variáveis *dummies*, é de grande interesse que se registre as palavras de Searle sobre as infinitas soluções das equações normais: "In a general discussion of linear models that are not of full rank, it is essential to realize that what is obtained as a solution of the normal equations is just that a solution and *nothing more*. It is misleading and in most cases quite wrong for b^0 to be termed an estimator, particularly an estimator of B . It is true that b^0 is, as shall be shown, an estimator of something, but not of B , and indeed the expression it estimates depends entirely upon such generalized inverse of $X'X$ is used in obtaining b^0 ".

16 Mas Langoni parece não perceber isto quando fala em sua defesa: "O que é efetivamente relevante é a significância estatística das variáveis consideradas no modelo e não necessariamente o seu poder de explicação." Vide página 168, em Pesquisa e Planejamento — op. cit.

4 — A Contribuição Marginal da Educação, uma Evidência com Bases Muito Tênues

O propósito desta parte do trabalho é ressaltar que a evidência do aumento da contribuição marginal da educação entre 1960 e 1970 não foi apresentada em bases suficientemente sólidas para daí caracterizar a educação com o fator crucial na determinação do aumento de concentração entre 1960 e 1970. Outro aspecto importante que esta parte pretende ressaltar consiste na possibilidade de que a tênue evidência obtida se deva, em grande parte, à peculiar disponibilidade dos dados de 1960.

Acredito ser mais importante o papel da educação na diferenciação das rendas em 1960 do que em 1970. Isto porque em 1960 o Brasil saía de uma fase de grande desenvolvimento na qual se iniciava a indústria automobilística com toda a sua dinâmica e conseqüência; além do mais, o sistema educacional melhorou entre 1960 e 1970, podendo hoje oferecer oportunidades educacionais a uma maior parcela da população. Tendo em mente estas assertivas e deixando de lado as questões levantadas sobre a arbitrariedade dos valores dos coeficientes, não é absurdo se admitir que os coeficientes referentes aos vários níveis educacionais para o ano de 1960 pudessem se apresentar maiores do que os mesmos coeficientes para 1970, contrariando assim a evidência empírica encontrada por Langoni.

Antes de prosseguir seria interessante lembrar o modo pelo qual os dados de renda para 1960 diferem dos de 1970. Os dados de renda para 1960 foram obtidos por classes de renda. Para 1970, os dados disponíveis se constituem de informações de renda individual, não apresentando portanto nenhuma limitação para efeito de aplicação do modelo de análise da variância dos logaritmos das rendas. Em 1960 a pergunta foi feita ao indivíduo para que ele indicasse a que classe de renda pertencia. Logo, as únicas informações sobre a renda existente para 1960 são apenas os pontos médios das classes de renda e duas classes que são abertas e a última.¹⁷ É possível, no entanto, estimar-se a renda média para estas classes. Portanto, a informação de renda existente para 1960 se resume em seis pontos médios e duas estimativas das classes de renda. A fim de se obter informações de renda média para grupos de indivíduos de acordo com o nível educacional, idade, sexo, região e setor de atividade, utiliza-se estes 8 (oito) pontos, pois cada indivíduo terá sua renda vinculada a uma das oito classes de renda. Portanto, o resultado da reconstituição das informações para se aplicar o modelo de análise da variância dos logaritmos das rendas para o ano de 1960 devem ser médias dos pontos médios das classes de renda.¹⁸

Tendo em vista essas observações sobre os dados, pode-se concluir definitivamente que a variância dos dados de 1960 está subestimada, pelo simples fato de que a variância dos dados agrupados é sempre menor que a variância dos dados individuais. Portanto, as evidências de aumento de contribuição marginal da educação e do maior impacto

17 O correto seria dizer que a única informação sobre a renda são os limites dos intervalos de classe (com exceção da primeira e última classe), pois os pontos médios são utilizados pelo princípio de *insufficient reason*.

18 A possibilidade de se obter as informações de renda média para grupos de indivíduos de acordo com o nível educacional, idade, sexo, região e setor de atividade, a partir das informações correspondentes para 1970, foi excluída porque julgamos tal alternativa inadequada nestas circunstâncias. Ou seja, quando se pretende comparar as contribuições marginais das variáveis nos dois anos (1960 e 1970) não se deve atribuir aos dados de 1960 nenhuma característica que seja dos dados de 1970, a fim de se evitar uma possível interferência pessoal nas evidências obtidas.

entre 1960 e 1970 devem ser encaradas com certa reserva, já que podem ter sido obtidas em função da existência desta fonte de subestimação da variância dos dados de renda em 1960.¹⁹

Outro fato vem de encontro à afirmação feita no início de que os coeficientes para 1960 poderiam ser maiores, caso os dados individuais se fizerem disponíveis. Tal fato se faz presente através das informações contidas na tabela 5.1, que reproduzimos parcialmente, a fim de melhor argumentação. Observa-se que os coeficientes dos dados individuais para 1970 são, em geral, superiores aos mesmos coeficientes com dados agrupados para 1970. Dos 20 coeficientes apresentados na tabela 5.1, referentes aos dados com informações individuais, 15 são superiores aos mesmos coeficientes correspondentes aos dados agrupados. Este fato ressalta que após o agrupamento dos dados de 1970 os valores dos coeficientes de regressão apresentaram reduzida magnitude. E no caso especial que ora discutimos, o da educação, todos os coeficientes obtidos com as informações individuais são maiores que os coeficientes obtidos com os dados agrupados (vide quadros 1 e 2).²⁰

Deve-se lembrar que os dados de renda para 1960 são dados agrupados e que, portanto, apresentam uma variância total subestimada e, além disso, os valores dos coeficientes de regressão para 1960, a exemplo do que se verifica para 1970, podem estar com suas magnitudes reduzidas. Estes fatos mostram que as informações de renda para 1960 não apresentam condições satisfatórias para a obtenção de informações quantitativas relevantes para uma comparação definitiva entre 1960 e 1970, no que se refere à situação da distribuição de renda do Brasil. Nestas circunstâncias é preciso cautela na aceitação das evidências sobre o aumento de contribuição marginal da variável educação, e de maior impacto de determinadas variáveis quando se compara os dois anos em questão.

Quando propomos a evidência do aumento da contribuição marginal da educação entre 1960 e 1970 não tendo sido apresentada em bases suficientemente sólidas, tomamos por base a argumentação de Langoni, que parece indicar a confirmação ou relevância deste aumento unicamente quando se usa a contribuição marginal com base nos dados agrupados para 1970. Julgamos que o maior \bar{R}^2 tenha sido a razão para se apresentar as evidências do aumento da contribuição marginal da variável educação só com dados agrupados, visto que os dados individuais de 1970 fornecem um $R^2 = 51,20\%$, enquanto que os dados agrupados permitem um $R^2 = 59,28\%$. A diferença entre os dois totaliza 8,08%, diferença esta que tão logo diminuída da contribuição marginal da educação na tabela 5.2 roubar-lhe-ia um importante trunfo da pesquisa, que consiste no aumento da contribuição marginal da educação entre 1960 e 1970 da ordem de 33,19%, conforme o quadro reproduzido parcialmente (vide quadro 3).

Outrossim, ainda que a diferença não seja totalmente diminuída da contribuição da variável educação, pode-se admitir que poderia diminuir de maneira tal que em lugar de se observar um aumento na

19 Não há dúvida sobre este fato, visto que é sabido que dada uma variável aleatória X , cuja distribuição apresenta uma variância σ^2 , a distribuição das médias das amostras retiradas de tal distribuição apresentará uma variância σ^2/N . Apesar deste fato por si só denunciar que a variância das rendas em 1960 está subestimada, a disponibilidade das informações em classes de renda também faz reduzir a variância. É fato notório de que a variância total é o resultado da soma da variância entre classes mais a variância dentro de cada classe. Sendo assim, em 1960 estaríamos considerando somente a variância entre as classes.

20 É preciso ter em mente que não estamos sugerindo que os valores dos coeficientes sejam viesados pelo simples fato de se agrupar os dados.

QUADRO 1

Coefficiente das Regressões Para 1970

N.º	INDIVIDUAIS	GRUPADOS	VARIAÇÃO
1	0,17	0,16	—
2	0,25	0,24	—
3	0,02	0,05	+
4	— 0,26	— 0,22	—
5	— 0,38	— 0,32	—
6	— 0,47	0,51	+
7	0,44	0,49	+
8	0,34	0,32	—
9	0,89	0,84	—
10	1,34	1,28	—
11	2,03	2,01	—
12	— 0,98	— 0,72	—
13	— 0,57	— 0,56	—
14	— 0,21	— 0,23	+
15	— 0,01	— 0,04	+
16	0,14	0,10	—
17	0,25	0,21	—
18	0,22	0,18	—
19	0,13	0,10	—
20	0,66	0,57	—

FONTE: Tabela 5.1 Langoni, página 109 op. cit.

QUADRO 2

Coefficientes da Variável Educação — 1970

NÍVEIS DE EDUCAÇÃO	INDIVIDUAIS	GRUPADOS
E ₂	0,34	0,32
E ₃	0,89	0,84
E ₄	1,34	1,28
E ₅	2,03	2,01

FONTE: Tabela 5.1 — Langoni, página 109 op. cit.

contribuição marginal da educação poder-se-ia constatar uma diminuição. Para tanto não seria necessário que a contribuição marginal da educação com os dados individuais (não agrupados) para 1970 diminuísse do total 8,08%, bastaria que diminuísse somente um pouco mais de 5,45%, que constitui a diferença entre a contribuição marginal referente aos dados agrupados de 1970 e a contribuição marginal referente aos dados de 1960 agrupados.

Com base nestes fatos, a evidência de se ter realmente verificado um aumento na contribuição marginal da educação entre 1960 e 1970 deveria ser ratificada pelas informações sobre a contribuição marginal da educação, quando se leva em consideração os dados individuais de renda (não agrupados) de 1970. Como se trata de uma evidência crucial, seria cientificamente louvável que se desse a público tal informação. Além do mais, esta divulgação provavelmente mostraria que, apesar de um menor R² obtido com os dados individuais (sem agrupar), a

QUADRO 3

Contribuição Marginal de Educação

VARIÁVEL	1960 CONTRI- BUIÇÃO MARGINAL (Agrupado)	1970 CONTRI- BUIÇÃO MARGINAL (Agrupado)	1970 CONTRI- BUIÇÃO MARGINAL (Individual)	DIFE- RENÇAS	AUMENTO DA CONTRI- BUIÇÃO MARGINAL PARA
Educação	9,98	15,43	?	5,45	33,19
Total	32,41	37,62	?	—	—
R ²	50,74	59,28	51,20	8,08	—
Multicolinearidade	18,33	21,66	?	—	—
Variação por Explicar	67,59	62,37	?	—	—

FONTE: Tabela 5.2 e 5.1 Langoni
op. cit.

educação continuaria apresentando uma taxa de crescimento relativamente grande. Caso contrário, confirmar-se-ia a não relevância da evidência divulgada, a qual estaria condicionada ao agrupamento dos dados de 1970, situação em que realmente se consolidaria a falta de base sólida para a apresentação da educação com a vedete das variáveis utilizadas no modelo, em virtude de seu tão desconcertante aumento de contribuição marginal entre 1960 e 1970.

Ainda que a divulgação do valor da contribuição marginal da variável educação para 1970 (com dados individuais) venha sanar as dúvidas levantadas a respeito do aumento da contribuição marginal da educação, ou se esquecermos por um momento estes problemas mesmo assim acreditamos que não se deveria impressionar com o fato de que a variável educação tenha apresentado uma maior contribuição marginal. Num país onde a possibilidade de diferenciar a renda dos indivíduos através do grau educacional é muito ampla, certamente não significaria muito o fato de que dos 100% de variação dos logaritmos da renda, em 1970 e 1960, somente 15,43% e 10%, respectivamente, sejam devidos à variável educação. Além disso, a variação total explicada, descontada a multicolinearidade, é de 32,41% para 1960, 37,62% para 1970 com os dados agrupados, conforme o quadro 3, onde também não se constata esta informação para 1970 com dados individuais de renda.

É necessário também não esquecer que 67,59% para 1960 e 62,38% para 1970, das variações dos logaritmos das rendas continuam sem uma explicação plausível dentro do modelo utilizado. Julgamos que o autor não pensa que toda esta parcela da variação seja devida a perturbações aleatórias. Apesar de um R² pequeno não ser razão suficiente para abalar as evidências referentes ao papel da educação na diferenciação das rendas individuais, tal fato, no entanto, não deve servir de motivo para não se dar a devida importância aos 67,59% e 62,38% das variações que ficaram sem explicações e apresentar a educação como a grande vedete para a solução dos problemas de desigualdade de renda no Brasil.²¹

21 Tal qualificação a respeito do R² só pode ser feita quando temos todas hipóteses do modelo utilizado especificados e os dados adequados. Pois para se aceitar um R² relativamente baixo é necessário testá-lo se não difere significativamente de 0 e isto só é possível quando se admite a hipótese de normalidade do termo E_n, omissão já apontada no início do trabalho. Pode-se investigar com mais detalhe a distribuição de E_n, com o objetivo de se fazer uso das propriedades de robustez dos testes t e F, mas as informações contidas no capítulo 5 não são suficientes para demonstrar tal preocupação por parte do autor.

A identificação de fontes de subestimação da variância dos dados de renda para 1960, o problema da evidência de acréscimo da contribuição marginal da educação entre 1960 e 1970 ter sido divulgada só com os dados agrupados (para o caso do Brasil), a existência de uma grande parcela da variação não explicada pelo modelo, são fatos suficientes para justificar que as evidências sobre o verdadeiro impacto das variáveis e sobre o acréscimo da contribuição marginal da educação entre 1960 e 1970 não estão solidamente assentadas.

5 — Conclusões

Certamente existem outros problemas no trabalho de Langoni além daqueles aqui discutidos, visto que nos concentramos apenas no capítulo 5. Eles exigem novos ensaios e seguramente serão mostrados em outros trabalhos. Entre os problemas mais importantes destaque: o fato de admitir a variável educação como uma variável independente; a pouca importância dada ao papel da política salarial no processo de concentração que se verificou entre 1960 e 1970; o uso da teoria do capital humano para o tipo de renda individual divulgada pelo Censo; omissão sobre o papel da política de crescimento econômico voltada para o exterior no aumento de concentração verificado; o excessivo uso de raciocínios que implicam em hipóteses de concorrência perfeita. Muitos desses problemas constituem omissões e poderão futuramente ser estudados.

Na primeira parte concluiu-se que, teoricamente, os coeficientes obtidos devem apresentar viés, em virtude de omissões de variáveis. Em seguida tenta-se mostrar certas peculiaridades do tipo de modelo usado, no qual a matriz $X'X$ é singular; com a eliminação da singularidade desta matriz, através da escolha de B_0 introduz-se um elemento de arbitrariedade, ou seja o valor dos coeficientes dependeriam da escolha de B_0 . Finalmente, mesmo omitindo-se os problemas apontados nas duas primeiras partes, procura-se mostrar quão tênues são as evidências empíricas do maior impacto e aumento da contribuição marginal da educação entre 1960 e 1970, quando se examina com mais detalhe os dados utilizados e a maneira pela qual as evidências foram apresentadas.

Gostaria de lembrar ao Langoni, aos que leram seu livro e aos que não leram e no entanto tomam como verdadeiras suas conclusões sobre o papel da educação na diferenciação das rendas, como são tênues as evidências, como é relativa a importância da educação, como é preciso procurar entender melhor o processo de diferenciação das rendas no Brasil, antes de apresentar a educação como a principal variável para explicar o aumento de concentração entre 1960 e 1970, antes de se condicionar as soluções dos problemas de desigualdade às soluções dos problemas educacionais.

Julgamos ainda que uma ênfase muito grande ao papel da educação, apesar das evidências não serem convincentes, como se tentou mostrar neste ensaio, pode ser altamente nociva para a discussão séria e sem ideologia dos problemas de distribuição de renda. Não há dúvida de que, com a aceitação da educação como o fator mais relevante na diferenciação das rendas individuais, esvazia-se parcialmente a discussão a respeito da distribuição de renda, já que os problemas educacionais, na sua maioria, só podem ser solucionados a longo prazo, adiando, assim, discussões de soluções mais realísticas e humanas para os problemas de desigualdades de renda.

BIBLIOGRAFIA

1. LANGONI, Carlos G. — *Distribuição da Renda e Desenvolvimento Econômico do Brasil*, editora Expressão e Cultura, julho de 1973.
2. ———— “Distribuição da Renda e Desenvolvimento Econômico do Brasil”, revista *Estudos Econômicos*, n.º 5, 1972, IPE-USP.
3. HILL, H. P. — “An analysis of the distribution of wages and salaries in Great Britain” *Econometrica*, vol. 27, julho 1959.
4. KEMPTHORNE, Oscar — *The Design and Analysis of Experiments*, New York, John Wiley & Sons. Inc. 1952.
5. SPRENT, Peter — *Models in Regression and Related Topics*, Methuen & Co. Ltd., 1969.
6. JOHNSTON, J. — *Econometric Methods*, 2.^a edição, MacGraw-Hill Book Company, New York, 1972.
7. HOFFMAN, R — “Contribuição à análise da Distribuição da renda e da posse da terra no Brasil”, 1971, tese de livre docência, Eça-Luiz de Queiroz, Piracicaba, não publicado.
8. DUARTE, João C. — “Aspectos da Distribuição de Renda no Brasil em 1970”, 1972, tese de mestrado, Eça-Luiz de Queiroz, Piracicaba, não publicado.
9. FISHLOW, Albert — “Brazilian size distribution of income”, *American Economic Review*, maio de 1972.
10. COSTA, Ramonaval A — “Reflexões a respeito do trabalho de Langoni sobre distribuição de renda no Brasil”, maio de 1973, não publicado.
11. SEARLE, S. R. — *Linear Models*, John Wiley & Sons Inc. New York London Sidney — Toronto. Chapter 9. page 135.