

PERFIL DA DESIGUALDADE ENTRE AS PESSOAS OCUPADAS NA AGRICULTURA BRASILEIRA: UMA ABORDAGEM MULTIVARIADA

PROFILE OF THE INEQUALITY AMONG WORKERS IN BRAZILIAN AGRICULTURE: A MULTIVARIATE APPROACH

Idemauro Antonio Rodrigues LARA ¹
Angela Maria Cassavia Jorge CORRÊA²
Carlos Tadeu Santos DIAS ³

RESUMO

Este estudo visa analisar a importância relativa de algumas variáveis (econômicas e sociodemográficas) para explicar o perfil da desigualdade entre as pessoas ocupadas no setor agrícola do Brasil no ano de 2002. O estudo é desenvolvido a partir das informações disponíveis nas Pesquisas Nacionais por Amostra de Domicílios (PNAD – 2002), com ênfase nos estados da federação. A discussão dos resultados se dá a partir da aplicação de métodos da análise multivariada, por meio da qual se efetua uma análise exploratória, visando identificar padrões regionais de desigualdades, com auxílio dos gráficos biplots, conforme metodologia descrita por Gabriel (1971). Visando à redução do número de variáveis explicativas e tendo em vista a presença de multicolinearidade, utiliza-se a técnica da Análise de Componentes Principais.

Palavras-chave: trabalhador agrícola, PNAD, análise multivariada, biplot.

ABSTRACT

This study aims to analyze the relative role played by some economic and socio-demographic explanatory variables to characterize inequality among workers in Brazilian agriculture, based on 2000's Household Sample National Survey (PNAD 2002), with emphasis on the States of Federation. Methods of multivariate analysis were applied in an explanatory analysis, to identify regional inequality patterns,

⁽¹⁾ FACEN - Grupo área Métodos Quantitativos - UNIMEP. E-mail: idemauro@esalq.usp.br

⁽²⁾ FACEN - Grupo área Métodos Quantitativos - UNIMEP. E-mail: ajcorrea@unimep.br

⁽³⁾ Ciagri - ESALQ/USP. E-mail: ctsdias@carpa.ciagri.usp.br

using biplots graphics methodology, as described by Gabriel (1971). To reduce the number of the explanatory variables, and taking on account the presence of multicollinearity, Principal Components technique of analysis was used.

Key words: agriculture worker, PNAD, multivariate analysis, biplot.

1. INTRODUÇÃO

Apesar das melhorias nos patamares de pobreza nos anos 90 e início de 2000, o Brasil apresenta elevados indicadores de desigualdade e pobreza, enquanto pelo valor de seu Produto Interno Bruto (PIB) não pode ser classificado no rol dos países pobres. Por sua renda *per capita*, apresenta-se entre os países de renda média alta, sendo o rendimento médio por pessoa fortemente afetado pela elevada desigualdade da distribuição da renda que persiste no país. Esse aspecto é ratificado pelos resultados de relevante pesquisa de Barros, Henriques e Mendonça (2000), que mostra por meio de estudo empírico e descritivo que o Brasil não é um país pobre, mas sim injusto para com um grande contingente de pessoas pobres, sendo a desigualdade o fator explicativo de maior relevância para a determinação dos elevados níveis de pobreza do país.

O setor agrícola apresenta importância singular na economia brasileira, sendo o principal responsável pelo crescimento do PIB nos últimos três trimestres (Revista Veja, 02/06/2004), dada a grande expansão do agronegócio, entre outros aspectos. No entanto, entre as pessoas ocupadas neste setor da economia, conforme vários estudos sobre desigualdade de rendimentos no setor agrícola brasileiro, em diferentes momentos econômicos, constata-se que os elevados patamares de concentração de renda, associados a diferentes níveis de rendimento médio nas várias regiões geográficas, traduzem-se em elevados índices de pobreza, aprofundando os históricos desequilíbrios regionais que caracterizam o país. [Ver, entre outros, Hoffmann (2000); Corrêa (1998) e Silva (1998)].

Salienta-se, entretanto, que no decorrer da década de 1990 ocorreu um aumento na remuneração dos trabalhadores agrícolas, acompanhado de declínio da pobreza rural, simultâneo a um declínio na ocupação agrícola [Conforme Paes de Barros, Carvalho e Franco (2004)]. Nesse contexto, este estudo visa analisar a importância relativa de algumas variáveis explicativas (econômico e socio-demográficas) para a desigualdade entre as pessoas ocupadas no setor agrícola do Brasil, no ano de 2002. O estudo é desenvolvido a partir das informações disponíveis nas Pesquisas Nacionais por Amostra de Domicílios (PNAD- 2002), com ênfase nos estados da federação.

2. BASE DE DADOS E PROCEDIMENTOS METODOLÓGICOS

Para efetuar uma análise das disparidades regionais bem como identificar grupos, subgrupos e variáveis associadas, considera-se uma amostra com as pessoas ocupadas na agricultura brasileira a partir das informações disponíveis nas PNAD-2002.

A PNAD é a maior fonte de informações para pesquisas e estudos até o momento desenvolvidos na área agrícola e urbana do país, de qualidade reconhecida e confiável. Tem sido utilizada tanto para subsidiar o planejamento e a elaboração de políticas públicas de diversos setores, como em estudos sobre a situação socioeconômica geral da população brasileira, por cientistas sociais de várias áreas, mas especialmente por economistas e sociólogos. Sendo assim, os dados são obtidos a partir da base da PNAD-2002.

As amostras constituídas consideram os indivíduos: a) pertencentes à população economicamente ativa (PEA), ocupados no setor agrícola; b) com renda mensal do trabalho declarada positiva; c) com 15 anos ou mais de idade e, d) trabalhando 15 horas ou mais por semana. Ao definir essa amostra, busca-se uma aproximação com o mercado de trabalho agrícola.

As variáveis consideradas neste estudo, que procuram aproximar características do trabalho na agricultura, são: posição na ocupação (com seis categorias ou grupos ocupacionais), local de residência (com duas categorias), educação (com quatro níveis educacionais) e rendimento (três faixas salariais). Assim, as pessoas ocupadas na agricultura do Brasil e unidades da federação são agregadas em seis grupos ocupacionais (PO): (1) trabalhador com carteira, (2) trabalhador sem carteira, (3) trabalhador doméstico, (4) trabalhador por conta própria, (5) empregador e (6) trabalhador do próprio consumo e construção. São consideradas também duas categorias quanto ao local de residência (LR): (1) urbano e (2) rural; quatro níveis educacionais (ED): (1) sem instrução ou menos de um ano de estudo, (2) de 2 a 4 anos de estudo, (3) de 5 a 8 anos de estudo, (4) de 9 ou mais anos de estudo; e três classes de renda (RD): (1) até 1 salário-mínimo, (2) mais de 1 a 5 salários-mínimos, (3) mais de 5 salários-mínimos.

Sob esse enfoque, a presente análise considera simultaneamente vários mecanismos distintos associados à produção da desigualdade no setor agrícola brasileiro, como: i) a contribuição de treinamento, representada por educação; ii) os efeitos da distribuição da riqueza, conforme representado por posição na ocupação. Para a seleção da amostra de acordo com os critérios especificados, bem como para a análise estatística desenvolvida, considerou-se a ponderação dos dados pelos fatores de expansão fornecidos pela PNAD, com auxílio do SAS.

A discussão dos resultados se dá a partir da aplicação de métodos da análise multivariada, por meio da qual se efetua uma análise exploratória com auxílio dos gráficos *biplots*, conforme metodologia descrita por Gabriel (1971). De acordo com o referido autor, qualquer matriz de posto r pode ser representada como um *biplot*, que é um recurso gráfico para apresentar vetores referentes a cada linha e a cada coluna dessa matriz. Isso é de grande interesse prático, do ponto de vista para análise de grandes matrizes de dados, pois além de resumir as informações, permite a inspeção da estrutura entre as unidades amostrais e as variáveis. Nesse contexto, a metodologia proposta alicerça-se no fato de que qualquer matriz Y ($n \times m$) de posto r pode ser fatorada como:

$$Y = GH'$$

em que G é uma matriz $n \times r$ e H uma matriz $m \times r$, ambas necessariamente de posto r . Essa fatoração não é única. Um modo de fatorar Y é escolher as r colunas de G como uma base ortonormal do espaço coluna de Y , e calcular H como $Y'G$. Na prática, busca-se uma aproximação de posto 2 para esta matriz, visando sobretudo uma melhor representação gráfica. Para aproximar qualquer matriz retangular ${}_n Y_m$ de posto r para uma matriz $n \times m$ de posto menor, pode-se usar a decomposição por valores singulares [Eckart & Young (1939)]. Isto é,

$$Y = \sum_{\alpha=1}^r \lambda_{\alpha} p_{\alpha} q'_{\alpha}$$

em que para cada $\alpha = 1, \dots, r$, λ_{α} é valor singular, p_{α} coluna singular e q'_{α} linha singular. A partir dessa decomposição, pode-se tomar os dois primeiros componentes para obter uma aproximação de posto 2. Registra-se, adicionalmente, que essa aproximação ou redução dos dados é especialmente útil na análise de componentes principais, na qual por meio dos gráficos *biplots*, pode-se mostrar as interdependências entre as unidades amostrais,

indicar a existência de agrupamentos de unidades e também mostrar as variâncias e correlações das variáveis.

3. RESULTADOS E DISCUSSÃO

Para identificar com clareza quais são as unidades da federação (UF) que têm melhores ou piores padrões de trabalho agrícola, bem como visualizar como estão agrupadas as unidades e quais são as variáveis que explicariam tal agrupamento, apresenta-se, na Figura 1, o *biplot* obtido a partir da decomposição por valores singulares da matriz ${}_{21}X_{14}$, na qual as linhas representam as 21 unidades da federação e as colunas desta matriz, as várias categorias das variáveis consideradas neste estudo, exceto da categoria PO3, trabalhador doméstico, o qual não houve declarante, de acordo com a PNAD (2002), segundo os critérios de filtro adotados nesta pesquisa.

Considerou-se uma aproximação para essa matriz,

$${}_{21}Y_{14} = \sum_{i=1}^2 \lambda_i p_i q_i'$$

a partir da decomposição por valores singulares da matriz X , de posto 2, tomando-se os dois maiores componentes, em que nenhuma ênfase foi dada ao efeito linha ou coluna, ou seja, a matriz ${}_{21}G_2$ (que carrega os efeitos das linhas) e ${}_{14}H_2$ (que carrega o efeito das colunas), foram igualmente ponderadas pelos valores singulares. O *biplot* aproximado de Y é definido por:

$$Y_2 = [p_1, p_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} q_1' \\ q_2' \end{bmatrix} = GH'$$

As condições em relação ao trabalho na agricultura no Brasil tendem a manter um modelo de diferencial entre as unidades da federação e regiões, na medida em que os estados do Nordeste do país aparecem num lócus oposto aos estados das regiões Sudeste, Sul e Centro-Oeste, aos quais estão

associadas variáveis que dão indicativos de melhores condições de trabalho (Figura 1). As variáveis que justificam esse distanciamento da região Nordeste do país são: educação 1, o que indica a prevalência de um baixo nível de instrução dos trabalhos dessa região; residência 1, o que caracteriza que predominantemente nessa região os trabalhadores moram na zona rural e posição na ocupação 4, indicando o grande contingente de pessoas ocupadas por conta própria, que têm no trabalho agrícola sua fonte própria e única de sustento. Observa-se, adicionalmente, que há uma possível existência de colinearidade entre algumas variáveis consideradas. Basta observar os ângulos pequenos entre algumas das variáveis.

Para destacar quão explicativo pode ser esse *biplot*, ajustado para resumir a matriz de dados ${}_{21}X_{14}$, considera-se o quadrado dos valores singulares associados a essa matriz, ou seja, seus autovalores não nulos (isto é, se s_i é um valor singular, $\lambda_i = s_i^2$ é autovalor). Assim, o percentual de explicação obtido por meio da redução dos dados, isto é, o quanto da informação da variabilidade dos dados é captado pela aproximação de posto 2 (representada no plano pelos vetores das unidades da federação e variáveis) é de 97,47%, indicando quão bom foi esse ajuste.

Um outro *biplot* é apresentado na Figura 2. Neste caso, a atenção foi focada sobre os resíduos a partir do segundo e terceiro componentes principais, seguindo os procedimentos de fatoração com autoponderação. Assim, na sua construção retirou-se da matriz X os efeitos relativos ao maior componente principal, obtendo-se uma matriz residual, $X^{(1)}$, tal que $X^{(1)} = X - s_1 U^{(1)} V'^{(1)}$, em que $U^{(1)}$ é a primeira coluna da matriz U , $V'^{(1)}$ é a primeira linha de V , e X é a matriz original dos dados. Alternativamente, conforme descrito por Gabriel (1971), a matriz residual

$$X^{(1)} = \sum_{i=2}^{14} \lambda_i p_i q_i'$$

Figura 1. Biplot a partir das informações originais.

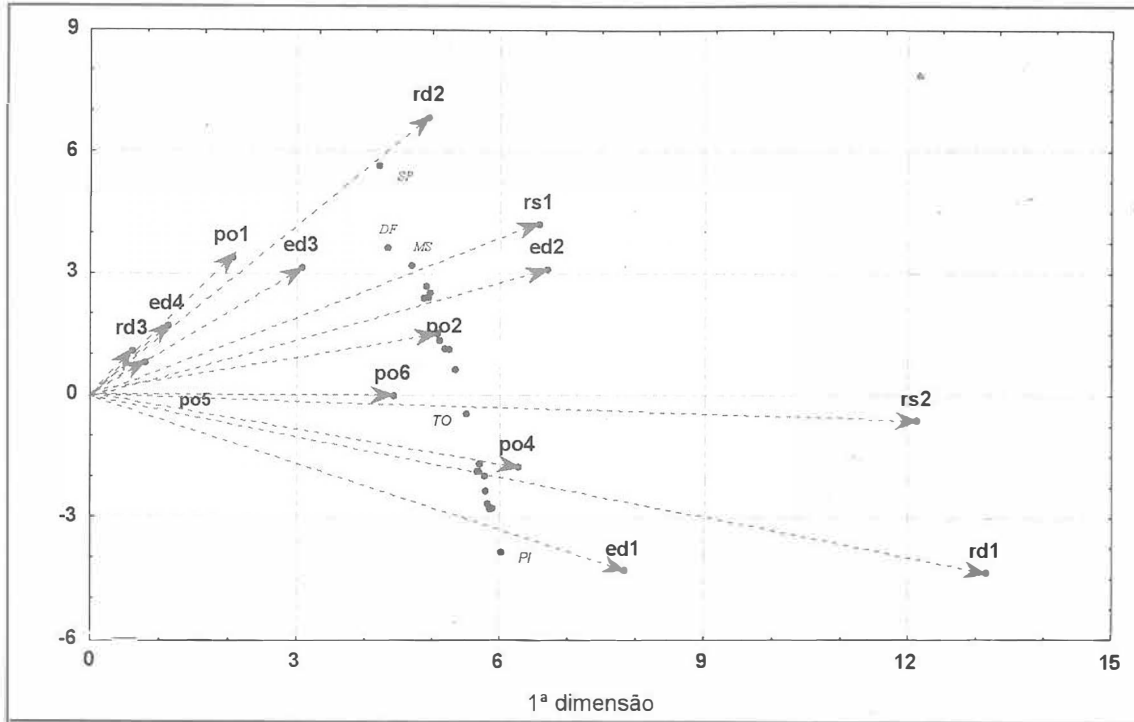


Figura 2. Biplot a partir da matriz residual $X^{(1)}$.

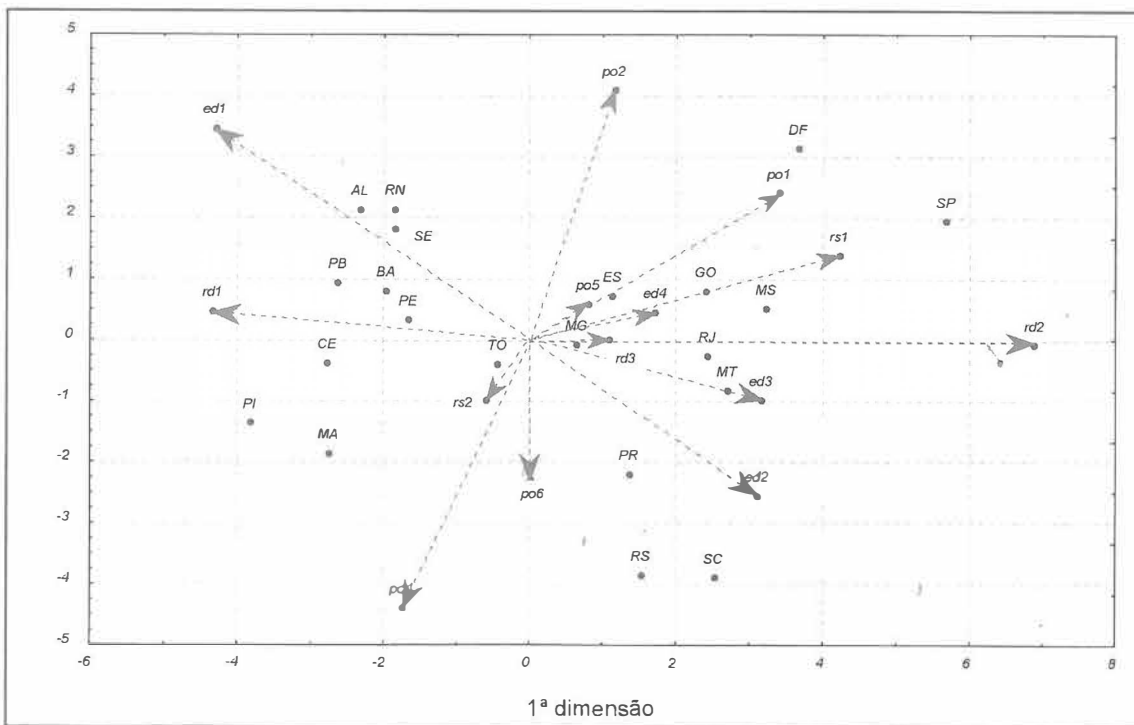


Tabela 1. Categorias de variáveis e componentes principais.

Variáveis	1ª Componente	2ª Componente	3ª Componente
PO 1	0.235642	-.275943	-.308771
PO 2	0.013040	-.471599	0.485470
PO 4	-.232207	0.531835	-.422258
PO 5	0.052400	-.068513	0.059459
PO 6	-.068174	0.282677	0.190599
LR 1	0.229810	-.140415	-.375850
LR 2	-.230247	0.141566	0.365689
ED 1	-.449338	-.405289	-.327175
ED 2	0.136120	0.328695	0.207481
ED 3	0.199080	0.130275	0.130485
ED 4	0.118056	-.049414	-.003443
RD 1	-.535471	-.035915	0.037386
RD 2	0.458866	0.035217	-.076148
RD 3	0.076631	0.000630	0.038684

Nota: Exceto PO3, no qual não houve declarantes.

corresponde à interação residual após ajustamento em modelos aditivos na análise de variância de classificação dupla.

A inspeção da Figura 2, embora ratifique as tendências mostradas pela Figura 1, é mais explicativa e revela também aspectos de correlação entre as variáveis. Percebe-se que os estados do Nordeste formam um *cluster*, em oposição aos demais estados da federação. As variáveis que explicam esse afastamento são: educação 1 (menos de um ano de estudo), renda 1 (renda de até um salário-mínimo, base de 2002), posição na ocupação 4 (trabalhador por conta própria) e residência 2 (zona rural), indicando a prevalência de padrão de condições de trabalho inferiores ao das demais áreas do país, caracterizado especialmente pela baixa renda.

Nota-se claramente que as variáveis que indicam melhores padrões de trabalho e rendimento estão em sentido oposto, indicando as melhores condições de trabalho agrícola nas unidades da federação das regiões Sul, Sudeste e Centro-Oeste. Pode-se notar ainda que os estados de São Paulo e o Distrito Federal se afastam dos demais. Em São Paulo, a presença de residência urbana é predominante, o que fica claramente refletido

na sua posição em relação à variável residência 1 (urbana), enquanto a posição do Distrito Federal é explicada, em parte, pela posição na ocupação 1 (trabalhador com carteira).

Esses indicadores ratificam resultados históricos, obtidos em estudos anteriores, sobre as condições de desigualdade e pobreza na agricultura brasileira, porém, por meio de outras metodologias quantitativas, como a utilização de modelos de regressão com o uso de variáveis *dummy* [ver, entre outros, Corrêa et al. (2003); Corrêa et al. (2004)]. A avaliação do quanto da informação é explicada nesse segundo *biplot*, agora considerando a matriz $X^{(1)}$, foi de 88,21%, indicando, assim, o quanto as duas primeiras componentes explicam a variabilidade dos dados da informação residual.

Também, visando à redução do número de variáveis explicativas e tendo em vista a presença de multicolinearidade, conforme indicativo das Figuras 1 e 2, utiliza-se a técnica da Análise de Componentes Principais, para obter variáveis ortogonais.

A seguir, apresentam-se as três primeiras componentes principais associadas às variáveis deste estudo, as quais explicam 90,80% da variabilidade dos dados.

4. CONSIDERAÇÕES FINAIS

Apesar de os resultados sinalizarem uma tendência histórica no tocante à distribuição de renda e pobreza no Brasil, verificou-se a eficiência da metodologia exploratória do gráfico *biplot* para explicar o perfil da desigualdade entre as pessoas ocupadas na agricultura brasileira, em particular, na associação das variáveis explicativas às unidades da federação. Ao trabalhar com uma grande massa de dados, como é a proveniente da PNAD, faz-se necessário encontrar técnicas que possam não somente resumir a matriz dos dados, mas explicá-la de forma consistente.

Dessa forma, identificou-se, claramente, a presença de dois grupos ou blocos com características adversas, de um lado os estados do Sul, Sudeste e Centro-Oeste, registrando os melhores indicadores, contrastando com os estados do Nordeste que, ironicamente, detêm mais de 50% da mão-de-obra nesse setor, mas registra os piores indicadores socioeconômicos. A técnica usada também permite diagnosticar a presença de multicolinearidade entre as variáveis explicativas. Essa é, sem dúvida, um dos grandes problemas no ajuste de modelos a dados socioeconômicos, pois, como se sabe, na presença da multicolinearidade, as inferências baseadas num modelo de previsão, por exemplo, podem ser enganosas.

REFERÊNCIAS BIBLIOGRÁFICAS

BARROS, R. P.; HENRIQUES, R.; MENDONÇA, R. A estabilidade inaceitável: desigualdade e pobreza no Brasil. In: HENRIQUES, R. *Desigualdade e pobreza no Brasil*. RJ, IPEA, p.21-48, 2000.

BARROS, R. P.; CARVALHO, M.; FRANCO, S. . Pobreza rural e trabalho agrícola no Brasil ao longo da década de noventa (versão não revista). IPEA - Seminários DIMAC 150. Rio de Janeiro, 2004.

CORRÊA, A. M. C. J. *Distribuição de renda e pobreza na agricultura brasileira*. Piracicaba, São Paulo: Ed. UNIMEP, 1998, 260 p.

CORRÊA, A.M.C.J.; MONTEBELLO, M.I.L.; LARA, I. A. R. Condicionantes da desigualdade de rendimentos entre as pessoas ocupadas na agricultura brasileira: uma análise *cross-section* através de modelos lineares gerais. *Pesquisa e Debate*, v.14 p. 111-134, 2003.

CORRÊA, A.M.C.J.; MONTEBELLO, M.I.L.; LARA, I. A. R.; FIGUEIREDO, N.M.S. Uma aplicação de modelos lineares gerais para o estudo da desigualdade e pobreza na agricultura brasileira em 2002. In: *Anais do 16º SINAPE – Simpósio Nacional de Probabilidade e Estatística*, v.1, 2004.

ESKART, C., YOUNG, G.A. A principal axis transformation for non-Hermitian matrices. *Am. Math. Soc. Bull*, 45, p.118-139, 1939.

GABRIEL, K. R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 58, s, p.453-467, 1971.

HOFFMANN, R. *Desigualdade e pobreza no Brasil no período 1979-1998*. ESALQ/USP e IE/UNICAMP (digitado). Piracicaba, São Paulo, 2000, 24 p.

PNAD, 2002. *Pesquisa Nacional por Amostra de Domicílios*. Rio de Janeiro, IBGE (Microdados).

SAS Institute Inc., SAS / STAT – The GLM Procedure, cap. 9. 2001.

SILVA, J. G. *A nova dinâmica da agricultura brasileira*. 2a ed. Campinas, São Paulo, IE/UNICAMP, 1998, 211 p.