

Tipos de Especialização na Agricultura Brasileira*

Fernando Luís Garagorry**

Eliseu Alves***

Geraldo da Silva e Souza****

Sumário: 1. Introdução; 2. Procedimento geral; 3. Resultados; 4. Considerações finais.

Palavras-chave: especialização na agricultura; dominância esto-cástica.

Códigos JEL: Q12; C44.

Este artigo introduz uma metodologia para definir tipos de especialização de estabelecimentos agropecuários, a partir da estrutura das receitas de origem agrícola. Com base nos dados de um levantamento realizado em colaboração entre a Fundação Getulio Vargas e a Embrapa, que abrangeu nove regiões do Brasil, foi possível alocar cada estabelecimento em um único tipo de especialização e estimar um índice de especialização para cada região, com o qual foi obtido um ordenamento de especialização entre as regiões. A seguir, são feitas algumas considerações sobre o número de atividades envolvidas e certos problemas que podem surgir quando se trabalha com valores médios regionais. Dentro de uma situação em que predominam níveis altos de especialização, são identificadas importantes diferenças inter-regionais nessa matéria, o que deve contribuir para a formulação de políticas agrícolas e o direcionamento de ações de pesquisa. Em termos mais amplos, entende-se que uma metodologia similar pode ser aplicada para estudar a especialização em outras situações (por exemplo, no nível de municípios).

This paper introduces a methodology to define types of farm specialization, taking into account the structure of incomes from agricultural activities. Using the data from a survey jointly conducted by the Getulio Vargas Foundation and Embrapa, which covered nine regions in Brazil, it was possible to allocate each farm into a unique specialization type and estimate a specialization index for each region. With this index, a specialization ordering among the regions was found. This is followed by some considerations on the

* Artigo recebido em ago. 2001 e aprovado em jun. 2002.

**SEA/Embrapa.

***PR/Embrapa.

****SEA/Embrapa.

number of agricultural activities and problems which may appear when regional mean values are used. Within a situation where high specialization levels prevail, important interregional differences are identified, which may contribute to the formulation of agricultural policies and research activities. In broader terms, it is believed that a similar methodology may be applied to the study of specialization in other situations (e.g., at the level of municipalities).

1. Introdução

Como parte de projeto de pesquisa, envolvendo a Fundação Getúlio Vargas e a Embrapa, foi realizado um levantamento de dados de produtores agropecuários e de seus estabelecimentos, abrangendo as seguintes regiões: 1) agricultura comercial no Centro-Oeste; 2) agricultura de pequeno produtor, do tipo colonial, no Paraná, Santa Catarina e Rio Grande do Sul; 3) áreas de monocultura em São Paulo; 4) norte e oeste do Paraná; 5) agricultura com plantio direto no Paraná; 6) agricultura tradicional no Ceará; 7) agricultura tradicional em Pernambuco; 8) área de pobreza rural no Espírito Santo e Rio de Janeiro e 9) zona da mata de Minas Gerais. Para os efeitos deste artigo, foi necessário considerar estabelecimentos em que houve receita positiva, advinda da venda de seus produtos; assim sendo, foi possível contar com os dados de 1.791 estabelecimentos.

Identificaram-se receitas provenientes de 99 produtos diferentes, que foram agrupados nas seguintes atividades: 1) café; 2) frutas; 3) hortaliças; 4) grãos; 5) outros vegetais; 6) gado de corte; 7) gado de leite; 8) aves; 9) suínos e 10) outros animais. No restante do artigo será mantida a numeração aqui apresentada para as regiões e para as atividades.

Os principais objetivos do projeto foram a tipificação de produtores e de estabelecimentos agropecuários, dentro de cada região, e o estudo de diferenças inter-regionais. As nove regiões foram escolhidas com base no conhecimento de que elas apresentavam diferentes condições, o que deveria justificar a formulação de recomendações específicas em termos de política agrícola e prioridades de pesquisa. No mesmo sentido, o questionário incluiu numerosas perguntas que se relacionam com itens não cobertos pelos censos agropecuários. Como se sabe, estes se concentram nos dados dos estabelecimentos, o que não permite extrair conclusões detalhadas sobre os produtores.

Em termos gerais, existia a intenção de realizar, inicialmente, duas tipificações: uma de produtores e uma de estabelecimentos. Posteriormente, seria realizado um cruzamento de ambas as classificações, para se chegar a uma tipificação da agricultura. Este artigo se limita ao estudo de um aspecto dos estabelecimentos:

a especialização. De fato, nas primeiras etapas do estudo chegou-se à conclusão de que não seria adequado aplicar algumas técnicas quantitativas (por exemplo, funções de produção ou métodos de construção de conglomerados) de maneira uniforme a estabelecimentos que mostravam estruturas muito diferentes, no que diz respeito às receitas provenientes da agricultura. Em particular, alguns eram muito mais especializados do que outros. Uma vez estabelecido o tipo de especialização de um estabelecimento, seriam incorporados outros elementos (tais como área e uso da terra, estrutura de custos e uso da mão-de-obra), para se obter uma avaliação mais detalhada sobre sua tecnologia e, principalmente, sobre suas perspectivas para se manter no negócio agrícola.

O enfoque geral adotado considera duas etapas: na primeira são definidos, mediante regras simples, seis tipos de especialização, que têm um ordenamento natural, independentemente das atividades envolvidas, e cada estabelecimento é alocado em um deles; na segunda são identificadas as atividades que foram responsáveis pela alocação do estabelecimento no respectivo tipo de especialização. O artigo se concentra na primeira etapa, particularmente no que se refere à identificação de diferenças de especialização entre as regiões.

2. Procedimento Geral

2.1 Vetores decrescentes

Suponha-se que o número máximo de atividades geradoras de receita seja N (no caso, como foi indicado anteriormente, $N = 10$). Para cada estabelecimento existe um vetor $x = (x_1, x_2, \dots, x_N)$ de componentes não negativos, que somam um, onde $x_i, i = 1, \dots, N$, corresponde à proporção da receita gerada pela i -ésima atividade.

A cada vetor como x , corresponde um único vetor $y = (y_1, y_2, \dots, y_N)$, que resulta do ordenamento dos componentes de x em ordem decrescente. Ou seja, cada y_i é igual a algum x_j , onde $i, j = 1, 2, \dots, N$, e $y_1 \geq y_2 \geq \dots \geq y_N$. Quando for o caso, tal vetor y será designado como o vetor decrescente que corresponde a x .

Exemplo Suponha-se que, para um conjunto de três estabelecimentos, designados por A, B e C, em uma situação em que se consideram até seis atividades geradoras de receita, foram estimados os seguintes vetores de proporções de receita:

$$x^A = (0,36; 0,00; 0,60; 0,00; 0,04; 0,00)$$

$$x^B = (0, 00; 0, 10; 0, 00; 0, 70; 0, 10; 0, 10)$$

$$x^C = (0, 00; 0, 04; 0, 00; 0, 60; 0, 00; 0, 36)$$

Os respectivos vetores decrescentes são:

$$y^A = (0, 60; 0, 36; 0, 04; 0, 00; 0, 00; 0, 00)$$

$$y^B = (0, 70; 0, 10; 0, 10; 0, 10; 0, 00; 0, 00)$$

$$y^C = (0, 60; 0, 36; 0, 04; 0, 00; 0, 00; 0, 00)$$

Neste artigo, entender-se-á que os estabelecimentos A e C são igualmente especializados; para comparar a especialização de B com a dos outros dois poderá ser utilizado o indicador que será introduzido mais adiante, que usa uma forma particular de dominância estocástica. Evidentemente, no caso, as especializações de A e C envolvem atividades diferentes. Mas, em termos do conceito de especialização que está sendo formulado, ambos os estabelecimentos são equivalentes. O importante, até este ponto, é que o recurso ao vetor decrescente correspondente a cada estabelecimento permite liberar-se de qualquer referência às atividades envolvidas. Com isso, será proposta uma classificação de tipos de especialização. Numa segunda etapa, as diferentes atividades poderão ser introduzidas, para refinar essa classificação.

2.2 Tipos de especialização

Na construção de uma classificação para a especialização dos estabelecimentos havia dois objetivos: 1) a classificação deveria ser simples, tanto no sentido de que as classes fossem diretamente interpretáveis quanto no que se refere ao processamento de dados requerido; 2) o interesse não era propriamente o de definir um nível de especialização para cada estabelecimento, mas o de indicar um número de atividades que predominavam e definir tipos de especialização com base num pequeno número de atividades.

De acordo com o procedimento indicado, encontrou-se o vetor decrescente correspondente a cada estabelecimento, mediante a execução das permutações adequadas; no que segue, todas as considerações sobre especialização utilizam esses vetores. À continuação, foram aplicados, a esses vetores, diversos algoritmos

de construção de conglomerados, que não se mostraram de utilidade, tanto pela dificuldade para interpretar os diferentes grupos resultantes quanto pelas inconsistências observadas. Em resumo, acontecia uma instância do conflito *data fit vs. mental fit* (Eijkel, G. C. van den, 1999), em que os métodos de conglomeração tinham uma razoável justificativa matemática, mas os resultados obtidos não eram interpretáveis nem aceitáveis. Nessa situação, optou-se pela formulação de uma classificação a priori, com regras simples para alocar cada estabelecimento em uma classe.

Especificamente, decidiu-se definir os tipos de especialização mediante regras de decisão numa estrutura de árvore (Eijkel, G. C. van den, 1999, Berry e Linoff, 2000, cap. 5). Tendo o respectivo vetor decrescente $y = (y_1, y_2, \dots, y_{10})$, cada estabelecimento foi colocado em um dos seguintes tipos de especialização:

- ESP-1, designado como “especializado em uma atividade”, se $y_1 > 0,8$;
- ESP-2, designado como “especializado em duas atividades”, se $0,8 \geq y_1 > 0,5$ e $y_1 + y_2 > 0,9$; ou seja, a segunda atividade contribui com mais de 10% da receita;
- ESP-3, designado como “especializado em três atividades”, se $0,8 \geq y_1 > 0,5$, $y_1 + y_2 \leq 0,9$ e $y_1 + y_2 + y_3 > 0,95$; ou seja, não é dos tipos anteriores, e a terceira atividade contribui com mais de 5% da receita;
- DOM, designado como “dominante simples”, se $0,8 \geq y_1 > 0,5$ e não é dos tipos anteriores;
- DIV-1, designado como “diversificado 1”, se $0,5 \geq y_1$ e $y_1 + y_2 + y_3 > 0,95$;
- DIV-2, designado como “diversificado 2”, se $0,5 \geq y_1$ e não é DIV-1.

Cabe assinalar os seguintes aspectos sobre a tipologia proposta:

- Considerou-se essencial registrar a presença ou não de dominância (isto é, se um estabelecimento tinha ou não $y_1 > 0,5$); evidentemente, os algoritmos usuais de conglomeração não se prestam para manipular esse tipo de restrição.
- Estabelecida a dominância, foram adicionadas umas poucas regras de refinamento, baseadas em um pequeno número de atividades; os tipos ESP-1, ESP-2 e ESP-3 representam casos particulares de dominância.

- A tipologia proposta para a especialização dos estabelecimentos aceita uma ordem natural, que coincide com a ordem em que os tipos foram apresentados. Essa ordem dos tipos de especialização será utilizada no que segue. Por exemplo, algumas estatísticas utilizarão o fato de que se tem uma escala ordinal, e não simplesmente nominal, sobre as classes que formam a tipologia.
- Foram encontrados 15 casos em que $0,5 \geq y_1$ e $y_1 + y_2 > 0,9$; como o interesse principal centrava-se no estudo dos casos com dominância (isto é, nos quatro primeiros tipos de especialização) e esses 15 casos de diversificação também satisfaziam a condição $y_1 + y_2 + y_3 > 0,95$, eles foram incluídos no tipo DIV-1.
- No conjunto dos vetores decrescentes, da dimensão escolhida, com componentes não negativos e que somam um, os tipos propostos definem subconjuntos disjuntos e exaustivos, e todos resultaram não vazios quando foram alocados os estabelecimentos da amostra; além disso, cada um desses subconjuntos é convexo, já que está definido por desigualdades lineares. Em particular, dado um conjunto finito de vetores pertencentes a um dos tipos de especialização, tanto a sua média aritmética quanto a de razões (considerando que cada componente de um desses vetores é o quociente entre a receita de uma atividade e a receita total do estabelecimento respectivo) resultam ser um vetor decrescente que pertence ao mesmo tipo.

2.3 Indicador de especialização

Suponha-se dada uma escala ordinal finita, correspondente a K classes ($K > 1$), C_1, C_2, \dots, C_K , onde aceita-se que, com respeito a um atributo em consideração, há um ordenamento tal que $C_1 > C_2 > \dots > C_K$ (o qual pode interpretar-se como “ C_1 tem mais que C_2 ”, etc). Por exemplo, no caso dos tipos de especialização, $K = 6$ e aceita-se que $ESP-1 > ESP-2 > \dots > DIV-2$. Suponha-se, ainda, que as observações realizadas determinaram as freqüências relativas f_1, f_2, \dots, f_K , nas respectivas classes. Denotando com f o vetor formado por essas freqüências, pode-se definir, então, a seguinte estatística:

$$D(f) = \sum_{k=1}^K (K - k) f_k = \sum_{k=1}^{K-1} (K - k) f_k$$

Nota-se, em particular, que nessa definição as freqüências relativas recebem ponderações decrescentes. Como a escala é ordinal, podem ser definidas as freqüências relativas acumuladas, mediante as fórmulas usuais: $F_k = f_1 + f_2 + \dots + f_k, k = 1, 2, \dots, K$. É fácil ver que $D(f) = \sum_{k=1}^{K-1} F_k$. Ou seja, D é o indicador de dominância estocástica de segundo grau (Anderson, J. R. and Dillon, J. L. e Hardaker, 1977), para uma distribuição sobre uma escala ordinal finita. Em termos intuitivos, se as classes fossem representadas pelos números $1, 2, \dots, K$, e a distribuição acumulada fosse interpretada como uma escada sobre esses números, então D corresponde à área que fica sob a escada.

Se bem que aqui o indicador D foi introduzido como uma estatística associada a uma distribuição de freqüências relativas (ou, analogamente, a uma distribuição de probabilidades), na apresentação usual ele aparece como uma estatística para comparar duas distribuições. Para evitar maiores formalismos, isso vai ser ilustrado com um exemplo.

Exemplo Dados os vetores decrescentes y^A e y^B , da seção 2.1, tem-se que:

$D(y^A) = 3,00 + 1,44 + 0,12 = 4,56$ e $D(y^B) = 3,50 + 0,40 + 0,30 + 0,20 = 4,40$
ou seja, $D(y^A) - D(y^B) = 0,16$.

De outra forma, considerando as distribuições acumuladas desses vetores, designadas por F^A e F^B , respectivamente, tem-se que:

$$D(y^A) - D(y^B) = \sum_{k=1}^5 (F_k^A - F_k^B) = (-0,10) + (0,16) + (0,10) = 0,16$$

ou seja, a massa da distribuição de y^A que “fica à esquerda” da de y^B (isto é, 0,26) é maior que no caso contrário (0,10). Nesse sentido, diz-se que y^A tem dominância estocástica de segundo grau sobre y^B ou, em termos mais intuitivos, que a primeira fica à esquerda da segunda. No contexto deste artigo, entende-se que o estabelecimento A tem mais especialização que B , já que o seu vetor decrescente favorece os tipos mais altos de especialização, em maior medida do que o vetor correspondente a B .

O indicador D é o que aparece com mais freqüência no restante do artigo. Quando a distribuição é crescente ($f_1 \leq f_2 \leq \dots \leq f_K$), há uma relação muito simples entre D e o índice de Gini; quando a distribuição é decrescente, o mesmo acontece com respeito ao índice de Rosenbluth (Hoffmann, 1991), utilizado, por exemplo, em estudos de concentração industrial. Como os valores de D variam entre zero, quando a distribuição está totalmente concentrada na classe mais baixa, e $K - 1$, quando ela está totalmente concentrada na primeira classe, pode ser utilizado o índice normalizado definido por $F = D/(K - 1)$, que varia entre zero e um. Quando a distribuição é uniforme, tem-se que $F = 1/2$; em forma mais geral, esse valor corresponde a qualquer distribuição simétrica, como será visto mais adiante. Quando o índice de Gini, G , é definido diretamente a partir dos desvios com respeito a uma distribuição uniforme discreta (Reis, 1991), e sendo crescente a distribuição observada, tem-se que $G = 1 - 2F$. No entanto, neste artigo prefere-se estudar a especialização como um conceito independente das formulações usuais sobre concentração; nesse sentido, não é difícil dar exemplos em que uma distribuição fica mais à esquerda que outra (isto é, a primeira tem um valor para D maior que o da segunda), ao mesmo tempo que a segunda é mais concentrada que a primeira, segundo algum dos indicadores usuais de concentração. De fato, D tem uma propriedade que o aproxima de um indicador de assimetria, como mostra a proposição seguinte.

Proposição 1 Se a distribuição f for simétrica, com respeito ao ordenamento adotado, então $D(f) = (K - 1)/2$ (e, portanto, $F = 1/2$).

Demonstração Seja $L = [K/2]$, onde $[x]$ denota a parte inteira de x . Como a distribuição é simétrica, tem-se que $f_j = f_{K-j+1}$, $j = 1, \dots, L$. No que segue, convém distinguir dois casos.

(a) K é par. Então, $L = K/2$. Portanto,

$$\begin{aligned} D(f) &= (K - 1)f_1 + (K - 2)f_2 + \dots + f_{K-1} + 0f_K \\ &= (K - 1)(f_1 + \dots + f_L) = (K - 1)/2 \end{aligned}$$

(b) K é ímpar. Então, $2L = K - 1$. Portanto,

$$\begin{aligned}
 D(f) &= (K - 1)(f_1 + \dots + f_L) + (K - L - 1)f_{L+1} = \\
 &\quad (K - 1)(1 - f_{L+1})/2 + (K - L - 1)f_{L+1} = \\
 &\quad (K - 1)/2 + \left(K - L - 1 - \frac{K - 1}{2} \right) f_{L+1} = (K - 1)/2
 \end{aligned}$$

porque o coeficiente do segundo termo é nulo.

Essa propriedade mostra que, se bem D pode assinalar algumas situações de concentração, particularmente quando assume valores próximos dos extremos, ele está numa classe separada dos indicadores habituais de concentração.

De um modo geral, o indicador D pode ser calculado para qualquer vetor com componentes não negativos e que somem um. Mas isso pode não ser diretamente interpretável, se não existir uma justificativa para dar pesos decrescentes aos sucessivos componentes. No entanto, com os vetores decrescentes correspondentes aos diferentes estabelecimentos isso pode justificar-se, o que permite alocar um valor de D a cada um deles (logicamente, na fórmula para o cálculo de D , o número K deverá ser substituído pelo número de atividades consideradas). Em geral, isso não será feito, porque não existe uma relação simples entre os valores de D e os tipos de especialização. No entanto, vale a pena mencionar algumas propriedades do indicador D , que decorrem do fato de ele ser definido por uma transformação linear. Suponha-se um caso como no presente artigo, em que há amostras em R regiões, onde os respectivos tamanhos são N^r , $r = 1, \dots, R$, sendo N o tamanho total da amostra. Sejam $f^r = (f_1^r, \dots, f_K^r)$ e $f = (f_1, \dots, f_K)$, respectivamente, os vetores com a distribuição das freqüências relativas dos estabelecimentos na região r e no total da amostra. Dado um conjunto de valores ou de vetores, a sua média aritmética é indicada com um traço superior; por exemplo, \bar{y}_k^r ; \bar{y}^r e \bar{y} denotam, respectivamente, as médias dos vetores decrescentes dos estabelecimentos no tipo k da região r , no conjunto da região r e no total da amostra. Finalmente, denotam-se com y_i^r e y_i os vetores decrescentes correspondentes a estabelecimentos genéricos da região r ou do conjunto da amostra, respectivamente.

Proposição 2 Com a notação introduzida, valem as seguintes propriedades:

- (a) $D(\bar{y}^r) = \sum_{k=1}^K f_k^r D(\bar{y}_k^r) = \bar{D}(y_i^r)$, onde, para simplificar a notação, o último termo indica a média aritmética dos valores de D para todos os estabelecimentos da região r ;

(b) $D(\bar{y}) = \sum_{r=1}^R (N^r/N)D(\bar{y}^r) = \bar{D}(y_i)$, onde o último termo denota a média aritmética dos valores de D para todos os estabelecimentos da amostra;

(c) $D(f) = \sum_{r=1}^R (N^r/N)D(f^r)$.

Demonstração (a) e (b) por definição, $\bar{y}^r = (1/N^r) \sum_{i \in I^r} y_i^r = \sum_{k=1}^K f_k^r \bar{y}_k^r$, onde I^r denota o conjunto de estabelecimentos da região r , e $\bar{y} = (1/N) \sum_{i=1}^N y_i = \sum_{r=1}^R (N^r/N) \bar{y}^r$. Os resultados seguem do fato de que D é uma transformação linear.

(c) Tem-se que $f_k = \sum_{r=1}^R (N^r/N) f_k^r$, para $k = 1, \dots, K$. Novamente, o resultado decorre da linearidade de D . Logicamente, por simplicidade, usa-se a mesma notação para dois operadores diferentes; nos casos (a) e (b), D opera sobre vetores com dimensão igual ao número de atividades, e no caso (c) sobre vetores de dimensão K .

Resultados análogos podem ser deduzidos para médias de razões onde, no caso do presente artigo, apareceriam termos com quocientes entre receitas. Neste artigo, o uso mais importante do indicador D estará centrado na comparação da especialização entre regiões, a partir das frequências (isto é, número de casos) obtidas mediante a alocação de cada estabelecimento em um tipo de especialização e, portanto, valem as propriedades enunciadas na proposição anterior. Apenas, vai ser apresentado um argumento relacionado com o uso de médias de razões quando forem comentados os resultados para valores médios regionais.

Em muitas aplicações (por exemplo, relacionadas com as distribuições aleatórias dos lucros correspondentes a diversas alternativas de investimentos), entende-se, naturalmente, que uma distribuição de frequências relativas domina outra se a sua distribuição acumulada fica “mais à direita” que a da segunda (por exemplo, a primeira levaria, em geral, a maiores lucros que a segunda), e uma forma de avaliar isso consiste em ver se o valor de D para a primeira é menor que o da segunda; obviamente, nesses casos, usa-se o ordenamento usual dos números reais, e preferem-se os valores mais altos. Isso não acontece no caso presente, em virtude do ordenamento adotado; aqui, uma distribuição domina outra, no sentido estocástico, se a primeira favorece mais que a segunda os tipos de especialização mais altos, os quais foram colocados nos primeiros lugares.

3. Resultados

3.1 Tipos de especialização

A tabela 1 mostra as freqüências obtidas para os tipos de especialização, por região e total da amostra. Vê-se, sem necessidade de cálculos adicionais, que, em todas as regiões há uma forte concentração nas duas classes de maior especialização. Cabe, ainda, estabelecer um ordenamento entre as regiões, no sentido de que uma possa ser considerada como mais especializada do que outra.

Tabela 1
Freqüência dos tipos de especialização, por região

Região	Tipo de especialização						Total
	ESP-1	ESP-2	ESP-3	DOM	DIV-1	DIV-2	
1. Comercial C-O	86	11	1	0	0	0	98
2. Colonial PR/SC/RS	155	130	42	10	33	22	392
3. Monocultura SP	98	33	6	0	2	1	140
4. Noroeste PR	147	50	13	1	11	1	223
5. Plantio direto PR	92	45	13	2	7	4	163
6. Tradicional CE	88	94	22	13	22	14	253
7. Tradicional PE	131	63	9	3	8	1	215
8. Agric. pobre ES/RJ	140	32	2	1	4	1	180
9. Zona da Mata MG	83	32	6	2	4	0	127
Total da amostra	1020	490	114	32	91	44	1791

Fonte: dados da amostra.

A tabela 2 apresenta alguns resultados que podem ser utilizados para definir um ordenamento entre as regiões. No caso das freqüências relativas acumuladas, não foi incluída a coluna correspondente ao tipo DIV-2, que é formada, exclusivamente, pelo valor um. Com base na ordem estabelecida entre os tipos de especialização, foi determinada a mediana de cada região. Vê-se que só aparecem os valores ESP-1 e ESP-2; embora isso confirme a concentração nos tipos superiores de especialização, não é satisfatório para determinar um ordenamento entre as regiões.

A seguir, foi calculado o indicador de dominância estocástica de segundo grau; no caso, bastou somar os valores das freqüências relativas acumuladas, o que aparece na coluna *D* da tabela 2.

Tabela 2
Tipos de especialização, por região: mediana, freqüências relativas acumuladas e índices de dominância estocástica, de Theil e de Simpson

Região	Mediana	Tipo de especialização					D	T	S
		ESP-1	ESP-2	ESP-3	DOM	DIV-1			
1. Comer. C-O	ESP-1	0,878	0,990	1,000	1,000	1,000	4,87	2,00	4,72
2. CO. PR/SC/RS	ESP-2	0,395	0,727	0,834	0,860	0,944	3,76	0,51	2,54
3. Monoc. SP	ESP-1	0,700	0,936	0,979	0,979	0,993	4,59	1,40	4,17
4. Noroeste PR	ESP-1	0,659	0,883	0,942	0,946	0,996	4,43	1,18	3,96
5. PL. DIR. PR	ESP-1	0,564	0,840	0,920	0,933	0,975	4,23	0,91	3,52
6. Tradic. CE	ESP-2	0,348	0,719	0,806	0,858	0,945	3,68	0,46	2,43
7. Tradic. PE	ESP-1	0,609	0,902	0,944	0,958	0,995	4,41	1,14	3,83
8. A. pob. ES/RJ	ESP-1	0,778	0,956	0,967	0,972	0,994	4,67	1,58	4,43
9. Z. Mata MG	ESP-1	0,654	0,906	0,953	0,969	1,000	4,48	1,22	3,98
Total da amostra	ESP-1	0,570	0,843	0,907	0,925	0,975	4,22	1,12	3,54

Fonte: dados da amostra.

Como já foi mencionado, maiores valores na coluna D indicam maior especialização. Eles variam entre o mínimo de zero, quando toda a distribuição está concentrada no tipo DIV-2 (não incluído na tabela 2), e o máximo de cinco, quando toda a distribuição está concentrada no tipo ESP-1. Cabe assinalar que os valores da coluna D servem tanto para a dominância estocástica de primeiro quanto de segundo grau. Por exemplo, os valores da distribuição acumulada (tabela 2), correspondentes à região 1 – Comercial C-O, são maiores que os respectivos valores para as outras regiões e, portanto, há dominância estocástica de primeiro grau dessa região sobre qualquer outra. Em outros casos, há “cruzamentos” entre os valores de duas regiões (alguns valores das freqüências acumuladas para uma região são maiores e outros são menores que os respectivos valores para outra região); nessa situação, é necessário considerar a dominância estocástica de segundo grau, o que leva a somar as freqüências acumuladas. No entanto, se uma distribuição domina outra no primeiro grau, forçosamente a soma das freqüências acumuladas da primeira será maior do que a da segunda e haverá, também, dominância estocástica de segundo grau.

Denotando com $[a] > [b]$ o fato de que a região a tem mais especialização que a região b , obtém-se o seguinte ordenamento, a partir dos valores na coluna D (tabela 2):

$$[1] \underset{E}{>} [8] \underset{E}{>} [3] \underset{E}{>} [9] \underset{E}{>} [4] \underset{E}{>} [7] \underset{E}{>} [5] \underset{E}{>} [2] \underset{E}{>} [6]$$

Todavia, com base nos comentários anteriores sobre a concentração das freqüências registradas, nos tipos superiores de especialização, considerou-se adequado estudar o comportamento de alguns indicadores de concentração. Nesse

sentido, foram considerados o índice de concentração de Theil, definido a partir do conceito de entropia de uma distribuição de frequências relativas (Theil, 1967, Andrade e Rathie, 1998), e uma forma modificada do índice de diversidade de Simpson (Hill, 1973). Especificamente, sendo f_1, f_2, \dots, f_6 as frequências relativas de uma região ou do total da amostra, foram utilizadas as seguintes fórmulas:

$$T = \log_2 6 - H$$

onde $H = - \sum_{i=1}^6 f_i \log_2 f_i$ é a entropia de Shannon; como sempre é feito na teoria matemática da informação (Rényi, 1962, Khinchin, 1957), estende-se a definição da função $x \log_2 x$, aceitando-se que ela assume o valor zero quando $x = 0$. O uso da base dois para os logaritmos é o mais comum na teoria da informação; no caso deste artigo, qualquer outra das bases usuais leva às mesmas conclusões, já que significa, apenas, multiplicar os valores encontrados por uma constante positiva, e o que interessa é o ordenamento desses valores.

$$S = 6 - 1 / \sum_{i=1}^6 (f_i)^2$$

Para ambos os índices, se a distribuição fosse uniforme (isto é, se houvesse o mínimo de concentração), o valor encontrado seria zero. Nos demais casos, valores mais altos indicam maior concentração. Se a distribuição estivesse concentrada em um único tipo de especialização (isto é, se houvesse o máximo de concentração), os valores desses índices seriam $T \cong 2,58$ e $S = 5$. Como se pode ver a partir dos valores nas colunas T e S da tabela 2, os dois índices levam ao mesmo ordenamento das regiões, que foi encontrado mediante a dominância estocástica. No entanto, se houvesse uma discrepância entre os ordenamentos encontrados para as regiões, o correto seria preferir o que foi obtido mediante a dominância estocástica.

Como ocorre, em geral, com os índices de concentração, os aqui apresentados requerem, apenas, a existência de uma escala nominal, e não utilizam o ordenamento que existe entre os tipos de especialização; eles podem indicar que uma distribuição é mais concentrada do que outra, mas não onde ocorre a concentração. Para poder interpretar o ordenamento obtido mediante um desses índices, no sentido de que uma região é mais especializada do que outra, em geral é necessário fazer alguma consideração adicional sobre os valores das frequências relativas; no caso, basta a constatação visual, indicada anteriormente, de que a maior concentração acontece nos tipos mais altos de especialização.

Todavia, cabe assinalar o seguinte:

- A aplicação do método da dominância estocástica permite que o ordenamento seja interpretado diretamente no sentido de especialização das regiões, sem necessidade de se recorrer ao uso combinado de medidas de concentração da especialização e do exame – mesmo que seja simplesmente visual – dos valores envolvidos. De fato, como já foi indicado, os conceitos de concentração e dominância estocástica são independentes. No caso do presente artigo, em que se quer estabelecer um ordenamento das regiões, com base nos tipos de especialização, que têm uma escala ordinal finita, os métodos de comparação de distribuições categóricas, com essa classe de escala, devem ser considerados como mais apropriados do que os que recorrem a outros tipos de estatísticas, tais como as de concentração.
- Os cálculos requeridos, nesse caso, para calcular o indicador D são extremamente simples, limitando-se à realização de somas sucessivas, a partir das frequências relativas. Há outros métodos de comparação de distribuições categóricas que, em geral, exigem cálculos mais elaborados.

Portanto, em ordem decrescente de especialização, tem-se o seguinte ordenamento das regiões estudadas, que será utilizado como base de comparação na discussão de outros resultados: 1^o) 1 – Comercial C-O; 2^o) 8 – Agric. pobre ES/RJ; 3^o) 3 – Monocultura SP; 4^o) 9 – Zona da Mata MG; 5^o) 4 – Noroeste PR; 6^o) 7 – Tradicional PE; 7^o) 5 – Plantio direto PR; 8^o) 2 – Colonial PR/SC/RS; 9^o) 6 – Tradicional CE.

Os indicadores construídos podem ser, também, aplicados ao total da amostra. É lógico que, nesse caso, espera-se um nível de especialização menor que o de algumas das nove regiões cobertas pelo estudo. Por um lado, vê-se, na tabela 2, que a mediana para o total da amostra corresponde ao tipo mais alto de especialização (ESP-1); mas, como foi mencionado anteriormente, esse indicador não se mostrou satisfatório para estabelecer um ordenamento da especialização por abrangência geográfica. Por outro lado, os índices de Theil e Simpson dão um nível de especialização entre as regiões 7 (Tradicional PE) e 5 (Plantio direto PR), enquanto o método da dominância estocástica dá uma posição entre as regiões 5 e 2 (Colonial PR/SC/RS). Ou seja, se bem que as frequências dos tipos de especialização, no total da amostra, se concentram nos tipos mais altos, o fato é que, em termos relativos, o seu nível de especialização encontra-se entre os mais baixos das regiões estudadas. Incidentalmente, o valor de D para o total da amostra pode ser conferido mediante a propriedade (c) da Proposição 2.

Os resultados anteriores foram obtidos, exclusivamente, mediante a utilização de estatísticas descritivas. No entanto, alguns valores dos diversos índices são muito próximos, e surge a necessidade de estudar a significância estatística do ordenamento resultante. Inicialmente, foi aplicado o teste de qui-quadrado (Siegel, 1975) às frequências da tabela 1. Para isso, seguindo as recomendações usuais para esse teste, foram somadas as frequências das colunas DIV-1 e DIV-2, de modo a se obter menos de 20% de celas com frequência esperada menor que cinco, e nenhuma cela em que essa frequência é menor que um. Resultou, assim, uma tabela com nove linhas e cinco colunas, o que deu um número de graus de liberdade igual a 32. O teste resultou ser significativo, com um valor da estatística calculada muito maior que o valor crítico correspondente ao nível de significância de 1%. Ou seja, o teste indica que, no seu conjunto, há diferenças entre as distribuições correspondentes às diferentes regiões. Entretanto, o teste de qui-quadrado serve para uma escala meramente nominal, e não explora o fato de que os tipos de especialização estão numa escala ordinal.

À continuação, foi utilizada a técnica de modelo linear generalizado, para dados com resposta multinomial e função de ligação logística (McCullagh e Nelder, 1989). Ela foi implementada com o sistema SAS, mediante o procedimento GENMOD, e o modelo foi aceito pelo teste de qui-quadrado. O teste global indica que há diferença significativa entre as distribuições regionais, com um nível de significância inferior a 0,0001. Outros resultados do mesmo procedimento indicam valores que permitem estabelecer um ordenamento entre as regiões, assim como as diferenças por pares de regiões, com uma estimativa do teste de qui-quadrado (com um grau de liberdade) e o seu nível de significância. O ordenamento resultante para as regiões foi o mesmo obtido antes, com base em estatísticas descritivas.

A aplicação do procedimento GENMOD, incluindo o total da amostra como se fosse mais uma região, mantém o ordenamento anterior entre as nove regiões do estudo e coloca o total da amostra entre as regiões 7 e 5. Ou seja, o resultado é consistente com o que foi mencionado anteriormente.

3.2 Agrupamento de regiões

O fato de que os valores de D (tabela 2) mostram uma diferença pequena entre algumas regiões sugeriu a realização de um exercício de agrupamento. Para isso, pode-se usar o valor absoluto da diferença nos respectivos valores de D como uma distância entre duas regiões. A natureza do problema indica que o mais adequado seja usar um tipo de agrupamento de ligação completa, utilizado com certa frequência na construção de conglomerados, onde os pontos de um conglomerado

devem ser mutuamente próximos. Tratando-se, apenas, de nove pontos (regiões) e com uma distância linear (aditiva) sobre a reta, é viável fazer os cálculos manualmente ou mediante uma planilha eletrônica, colocando os pontos na ordem que foi encontrada antes. Procedendo desse modo, foram encontradas as partições indicadas a seguir, onde os números contidos entre chaves identificam as regiões que formam cada grupo.

- Partição com quatro grupos: $G1 = \{1\}$; $G2 = \{3, 8\}$; $G3 = \{4, 5, 7, 9\}$; e $G4 = \{2, 6\}$.
- Partição com cinco grupos: $G1 = \{1\}$; $G2 = \{3, 8\}$; $G3 = \{4, 7, 9\}$; $G4 = \{5\}$; e $G5 = \{2, 6\}$.

O mesmo resultado é obtido trabalhando com os índices T ou S , em lugar de D . Novamente, o argumento anterior está baseado, exclusivamente, em medidas de distância fornecidas por estatísticas descritivas. No entanto, os resultados do método apresentado por McCullagh e Nelder permitem formular uma técnica diferente para a formação de grupos, com base no nível de significância. De fato, escolhendo um nível de significância, por exemplo, de 5%, o método indica que nem todas as diferenças, entre duas regiões, são significativas. Surge, desse modo, a possibilidade de definir alguns agrupamentos com base num nível de significância. Inicialmente, adotou-se, no que segue, o nível de significância de 5%. Essencialmente, o método utilizado consta dos seguintes passos: colocar as regiões na ordem encontrada anteriormente; e deixar no mesmo conglomerado as regiões que, consideradas em pares, não diferem significativamente, ao nível de 5%, de acordo com os valores de qui-quadrado fornecidos pelo procedimento GENMOD (ou seja, usa-se a distância de qui-quadrado, que não é aditiva, entre as regiões). A tabela 3 dá as distâncias utilizadas (os pontos suspensivos indicam que os valores seguintes, em cada linha, são crescentes, e não precisam ser considerados) e mostra quatro grupos resultantes. Vê-se que o maior diâmetro de um grupo é 3,82, correspondente a $G3$. O valor crítico, para o nível de significância de 5%, é 3,84. Adotando um nível de significância ligeiramente maior, por exemplo, de 6%, a região 5 fica num grupo separado e o diâmetro máximo da partição cai para 3,08. Isto é, seja no caso de quatro ou de cinco grupos, os resultados são os mesmos obtidos anteriormente com o uso de distâncias baseadas em estatísticas descritivas; no entanto, agora, há um critério de significância estatística para a formação dos grupos. Em resumo, pode considerar-se que as regiões que pertencem a cada grupo não diferem

significativamente, para o nível de significância escolhido, no que se refere à sua estrutura de especialização.

Foi utilizado o procedimento MDS (para construção de escalas multidimensionais), do SAS, o que permitiu visualizar, em duas dimensões, os pontos correspondentes às nove regiões (que estão em um espaço de seis dimensões).

Tabela 3
Distâncias qui-quadrado entre regiões ordenadas e quatro conglomerados resultantes

Região	Região								
	1	8	3	9	4	7	5	2	6
1. Comercial C-O	0,00	4,17	9,73					
8. Ag. pobre ES/RJ		0,00	3,08	5,82				
3. Monocult. SP			0,00	0,78	1,11	3,08	7,38	
9. Zona Mata MG				0,00	0,00	0,56	3,07	30,53
4. Noroeste PR					0,00	0,65	3,82	44,47
7. Tradicional PE						0,00	1,47	34,26
5. Pl. Direto PR							0,00	16,08
2. Colon. PR/SC/RS								0,00	0,92
6. Tradicional CE									0,00

Fonte: resultados do procedimento GENMOD.

Como resultado, tem-se que as partições encontradas, seja com quatro ou com cinco grupos, são aceitáveis.

3.3 Atividade principal

As mesmas permutações, que levaram, em cada estabelecimento, do vetor inicial de componentes da renda ao correspondente vetor decrescente, foram aplicadas às respectivas atividades. Designou-se como atividade principal a que ficou no primeiro lugar. Houve, apenas, dois casos de empates (ou seja, $y_1 = y_2$), ambos no tipo de especialização DIV-1, o que não tem importância para os resultados obtidos. A tabela 4 indica as freqüências das atividades principais. No total da amostra, os três primeiros lugares são ocupados por grãos, gado de leite e gado de corte; entre elas, essas atividades principais abrangem 74% dos estabelecimentos da amostra. Foram estudados outros cruzamentos, entre tipo de especialização, região e atividade principal, para se ter um conhecimento mais detalhado da orientação e da especialização da agricultura em cada região. Alguns desses resultados serão apresentados mais adiante.

Tabela 4
 Frequência dos tipos de especialização, por atividade principal

Atividade	Tipo de especialização						Total
	ESP-1	ESP-2	ESP-3	DOM	DIV-1	DIV-2	
Café	99	34	2	1	2	1	139
Frutas	21	12	2	0	1	1	37
Hortaliças	31	16	3	1	3	2	56
Grãos	379	190	43	9	39	9	669
Outros vegetais	30	27	9	0	3	4	73
Gado de corte	118	43	13	4	9	7	194
Gado de leite	289	116	22	12	20	4	463
Aves	7	9	3	1	5	10	35
Suínos	39	39	15	4	9	3	109
Outros animais	7	4	2	0	0	3	16
Total da amostra	1020	490	114	32	91	44	1791

Fonte: dados da amostra.

3.4 Número de atividades

Como já foi dito, as dez atividades consideradas neste artigo constituem agregações de operações elementares. Intuitivamente, pensa-se que seria um caso excepcional se um estabelecimento obtivesse renda das dez atividades. De fato, segundo os dados da amostra, o máximo número de atividades exercidas por um estabelecimento foi de seis, sendo que apenas sete, do total de 1.791 casos, estão nessa situação. As tabelas 5 e 6 apresentam a distribuição do número de atividades, por tipo de especialização e por região, respectivamente. Tem-se que, considerando até três atividades, há 1.576 estabelecimentos, o que representa 88% do total.

Tabela 5
 Frequência do número de atividades, por tipo de especialização

Tipo de especialização	Número de atividades						Mediana
	1	2	3	4	5	6	
ESP-1	543	339	108	22	7	1	1
ESP-2	0	301	146	37	5	1	2
ESP-3	0	0	70	36	7	1	3
DOM	0	0	0	24	6	2	4
DIV-1	0	1	68	20	1	1	3
DIV-2	0	0	0	28	15	1	4
Total da amostra	543	641	392	167	41	7	2

Fonte: dados da amostra.

Na tabela 5, a mediana reflete o ordenamento estabelecido entre os tipos de

especialização. No tipo DIV-1 há um caso com duas atividades; pela definição dos tipos de especialização, ele corresponde à única situação possível (isto é, $y_1 = y_2 = 0, 5$). Se bem que ele seja irrelevante do ponto de vista estatístico, serve para ilustrar a diferença entre a forma de construção da classificação adotada neste artigo e o que acontece com os métodos de conglomeração mais tradicionais, baseados em distâncias; de fato, qualquer pequena perturbação nos valores originais da renda das duas atividades envolvidas, que resultasse em $y_1 > y_2$, teria colocado o estabelecimento no tipo ESP-2; ou seja, as métricas usuais teriam indicado que esse estabelecimento está a uma distância zero do conjunto de alternativas possíveis que caracteriza o tipo ESP-2.

Pela definição do tipo de especialização ESP-1, um estabelecimento nessa categoria tem uma atividade que contribui com mais de 80% da renda, mas pode ter outras atividades com menor participação. Todavia, pode chamar a atenção que 543 dos 1.020 estabelecimentos classificados em ESP-1 (tabelas 1 e 5), ou seja, 53% do total nesse tipo de especialização, tenham só uma atividade. Nos outros tipos de especialização, a situação, quanto ao número de atividades, também está ligada à definição adotada em cada caso. Decorre das definições que os tipos ESP-2, ESP-3 e DOM exigem, pelo menos, duas, três e quatro atividades, respectivamente. Também o tipo DIV-2 exige, pelo menos, quatro atividades. No tipo DIV-1, desconsiderando o caso-limite, realmente excepcional, comentado anteriormente, o número mínimo de atividades tem que ser três. Esses números são, precisamente, os que aparecem na coluna da mediana (tabela 5). Ou seja, sem considerar aquela exceção, pode-se dizer que os estabelecimentos se dedicam, em sua maioria, ao menor número de atividades compatível com seu tipo de especialização.

Na tabela 6 aparecem as frequências dos números de atividades, por região. Já foi estabelecido um ordenamento das regiões, com respeito à sua especialização, a partir dos dados sobre os componentes da renda. Em outros levantamentos não se dispõe dessa informação, mas pode-se obter o número de atividades exercidas nos estabelecimentos. Cabe, então, discutir a possibilidade de que o número de atividades possa ser empregado como um “proxy” para comparar os níveis de especialização de diferentes regiões, na ausência de dados sobre os componentes da renda. Novamente, a mediana do número de atividades (não incluída na tabela 6) tampouco resultou adequada para separar as regiões, tal como aconteceu com os tipos de especialização. No caso, a média e o índice de dominância estocástica levaram ao mesmo ordenamento das regiões. Usando a notação utilizada anteriormente, mas agora no sentido de que uma maior especialização está relacionada com um menor número de atividades, o resultado obtido foi o seguinte:

$$[1] \underset{E}{>} [8] \underset{E}{>} [3] \underset{E}{>} [5] \underset{E}{>} [4] \underset{E}{>} [9] \underset{E}{>} [7] \underset{E}{>} [6] \underset{E}{>} [2]$$

Tabela 6
Frequência do número de atividades, por região

Região	Número de atividades						Média	D
	1	2	3	4	5	6		
1. Comercial C-O	63	26	6	3	0	0	1,48	4,52
2. Colonial PR/SC/RS	48	123	112	82	23	4	2,80	3,20
3. Monocultura SP	61	56	18	3	1	1	1,79	4,21
4. Noroeste PR	70	100	44	7	1	1	1,98	4,02
5. Plantio direto PR	59	66	28	6	4	0	1,96	4,04
6. Tradicional CE	37	84	80	40	11	1	2,63	3,37
7. Tradicional PE	55	94	52	14	0	0	2,12	3,88
8. Agric. Pobre ES/RJ	98	62	17	3	0	0	1,58	4,42
9. Zona da Mata MG	52	30	35	9	1	0	2,03	3,97
Total da amostra	543	641	392	167	41	7	2,19	3,81

Fonte: dados da amostra.

Esse ordenamento viu-se confirmado mediante a aplicação da mesma técnica de modelo linear generalizado que foi usada para ordenar as regiões, a partir das frequências dos tipos de especialização.

As três primeiras regiões estão nas mesmas posições que no primeiro ordenamento. Para as outras regiões aparecem diversas inversões de posição. As regiões 2 e 6 estão nas posições mais baixas, como no primeiro ordenamento, mas com inversão na ordem. Há alguns argumentos que permitiriam sustentar que os dois ordenamentos não são “muito diferentes”. Por exemplo, em alguns casos as diferenças entre os valores utilizados (média ou D) são pequenas, ou há diversas consistências parciais (por exemplo, em ambos casos, a região 4 domina as regiões 7, 2 e 6). Para avaliar essa situação, foi calculado o coeficiente de correlação por postos de Spearman (Siegel, 1975), que resultou no valor $r_S = 0,867$, o qual é significativo no nível de 1% e leva a rejeitar a hipótese de independência entre os dois ordenamentos.

Em resumo, o uso do número de atividades, para estudar as diferenças de especialização entre regiões, pode não ser totalmente inadequado, como uma primeira aproximação, no caso da agricultura brasileira; no entanto, isso poderia ser aceitável só quando não se dispusesse dos componentes da renda dos estabelecimentos. Por exemplo, tomando como referência as frequências da tabela 5, a informação de que um estabelecimento trabalha com três atividades pode significar que ele pertence a qualquer um dos tipos de especialização designados com ESP-1, ESP-2, ESP-3 ou DIV-1; em particular, a principal atividade, com respeito à sua contri-

buição na renda, pode participar com qualquer proporção entre 1/3 e um, de modo que essa informação não é suficiente para identificar adequadamente o tipo de especialização. Simplesmente, quando o número de atividades presentes é maior que um, não se pode avaliar se uma é mais importante do que outra, em termos de sua contribuição na renda.

3.5 Frequência de atividades por região

Já foi comentada a distribuição da atividade principal entre os diferentes tipos de especialização (tabela 4). Cabem alguns comentários para os níveis superiores de especialização (ESP-1, ESP-2 e ESP-3), quando se considera a desagregação por região. A tabela 7 apresenta os resultados para o tipo ESP-1; a soma das frequências em cada linha dessa tabela dá os valores na coluna ESP-1 da tabela 4. Como foi assinalado anteriormente, as atividades grãos, gado de leite e gado de corte ocupam, nessa ordem, os primeiros lugares, no total da amostra; o mesmo acontece dentro do tipo ESP-1, onde essas atividades perfazem cerca de 77% dos 1.020 casos observados. No entanto, a tabela 7 mostra grandes diferenças regionais na presença das diversas atividades. Por exemplo, na região 8 – Agricultura pobre ES/RJ predomina Café, com 86 casos, e só há um caso com a atividade Grãos.

Tabela 7
Frequência, por região, dos casos especializados em uma atividade

Atividade	Região (*)									%
	1	2	3	4	5	6	7	8	9	
Café	0	0	8	0	0	0	0	86	5	9,71
Frutas	0	0	12	1	0	3	3	2	0	2,06
Hortaliças	0	0	16	1	2	0	7	2	3	3,04
Grãos	41	112	6	121	48	15	35	1	0	37,16
Outros vegetais	0	1	14	4	9	1	0	1	0	2,94
Gado de corte	26	3	16	10	18	15	9	12	9	11,57
Gado de leite	18	12	26	5	10	53	72	35	58	28,33
Aves	0	1	0	2	1	0	1	1	1	0,69
Suíños	0	25	0	3	2	0	2	0	7	3,82
Outros animais	1	1	0	0	2	1	2	0	0	0,69
Total da amostra	86	155	98	147	92	88	131	140	83	100,00

Fonte: dados da amostra. (*) Identificação das regiões: 1 = Comercial C-O; 2 = Colonial PR/SC/RS; 3 = Monocultura SP; 4 = Noroeste PR; 5 = Plantio direto PR; 6 = Tradicional CE; 7 = Tradicional PE; 8 = Agricultura pobre ES/RJ; 9 = Zona da Mata MG.

Para os casos especializados em duas atividades (ESP-2), em princípio, poderia haver até 90 arranjos de uma primeira e uma segunda atividades (ou seja, por exemplo, considerando que Grãos/Gado de leite, onde Grãos é a primeira e Gado de leite é a segunda atividade, é diferente de Gado de leite/Grãos). No entanto, entre os 490 casos da amostra que têm esse tipo de especialização, só foram encontrados 55 arranjos. A tabela 8 apresenta os 12 casos mais freqüentes, os quais reúnem 73,27% das observações que estão nessa categoria. Cerca de 49% dos casos observados estão entre os quatro arranjos mais freqüentes.

Tabela 8
 Freqüência, por região, de casos especializados em duas atividades
 12 casos mais freqüentes

1ª / 2ª atividade	Região (*)									%
	1	2	3	4	5	6	7	8	9	
Grãos / G. leite	3	50	1	18	5	26	5	0	1	22,24
G. leite / Grãos	4	17	1	7	2	33	18	0	1	16,94
Grãos / G. corte	0	7	0	4	6	6	2	0	1	5,30
G. corte / grãos	1	1	0	5	5	7	4	0	0	4,69
Suínos / grãos	0	15	0	6	1	0	0	0	0	4,49
Café / G. leite	0	0	2	0	0	0	0	10	9	4,29
Grãos / suínos	0	7	0	3	0	3	1	0	0	2,86
Suínos / G. leite	0	11	1	0	0	0	0	0	2	2,86
Grãos / hortaliças	0	0	0	0	1	4	7	0	0	2,45
Grãos / aves	0	6	0	1	3	2	0	0	0	2,45
Outros veg. / grãos	0	1	4	1	5	1	0	0	0	2,45
Grãos / outros veg.	0	2	4	2	3	0	0	0	0	2,24
Total dos 12 casos	8	117	13	47	31	82	37	10	14	73,27
Total da amostra	11	130	33	50	45	94	63	32	32	100,00

Fonte: dados da amostra. (*) Identificação das regiões: 1 = Comercial C-O;

2 = Colonial PR/SC/RS; 3 = Monocultura SP; 4 = Noroeste PR;

5 = Plantio direto PR; 6 = Tradicional CE; 7 = Tradicional PE;

8 = Agricultura pobre ES/RJ; 9 = Zona da Mata MG.

Para os casos especializados em três atividades (ESP-3), em princípio, poderia haver até 720 arranjos de uma primeira, uma segunda e uma terceira atividades, se o número de observações fosse, pelo menos, igual a esse número. No entanto, entre os 114 casos da amostra que têm esse tipo de especialização só foram encontrados 63 arranjos. A tabela 9 apresenta os nove casos mais freqüentes, os quais reúnem 38,6% das observações que estão nessa categoria. Cerca de 21% dos casos observados estão entre os quatro arranjos mais freqüentes.

Tabela 9
 Frequência, por região, de casos especializados em três atividades
 9 casos mais freqüentes

1ª / 2ª / 3ª atividade (**)	Região (*)									%
	1	2	3	4	5	6	7	8	9	
GR / GL / SU	0	7	0	1	0	0	0	0	0	7,02
GR / SU / GL	0	6	0	0	0	0	0	0	0	5,26
GR / GL / AV	0	1	0	2	0	2	0	0	0	4,39
GR / AV / GL	0	3	0	2	0	0	0	0	0	4,39
GR / AV / SU	0	0	0	0	0	4	0	0	0	3,51
OV / GL / GR	0	3	0	1	0	0	0	0	0	3,51
GL / GR / AV	0	2	0	0	0	1	0	0	1	3,51
SU / GR / GL	0	3	0	0	1	0	0	0	0	3,51
SU / GL / GR	0	3	0	1	0	0	0	0	0	3,51
Total dos nove casos	0	28	0	7	1	7	0	0	1	38,60
Total da amostra	1	42	6	13	13	22	9	2	6	100,00

Fonte: dados da amostra. (*) Identificação das regiões: 1 = Comercial C-O;

2 = Colonial PR/SC/RS; 3 = Monocultura SP; 4 = Noroeste PR;

5 = Plantio direto PR; 6 = Tradicional CE; 7 = Tradicional PE;

8 = Agricultura pobre ES/RJ; 9 = Zona da Mata MG.

(**) Abreviaturas das atividades: GR = grãos; OV = outros vegetais;

GL = gado de leite; AV = aves; SU = suínos.

3.6 Presença de atividades

Interessa, também, o conhecimento da frequência com que aparecem as diferentes atividades, qualquer que seja a sua importância, no total da amostra e dentro de cada tipo de especialização. Por exemplo, dentro do tipo ESP-1 um estabelecimento especializado em café (isto é, onde essa atividade contribui com mais de 80% da receita) pode trabalhar em outras atividades, que terão uma participação bem menor na receita. A tabela 10 resume a frequência com que as diferentes atividades aparecem nos tipos de especialização e no total da amostra. Por exemplo, na tabela 4, dentro de ESP-1, há 99 estabelecimentos especializados em Café; esse número foi desagregado por região na tabela 7. No entanto, a tabela 10 indica que, entre os estabelecimentos que estão em ESP-1, há 117 que exercem a atividade Café em alguma medida; ou seja, em 99 deles essa é a atividade principal, mas há outros 18 estabelecimentos em ESP-1 onde também foi registrada a presença dessa atividade. As percentagens que aparecem na última coluna da tabela 10 referem-se ao total da amostra; por exemplo, a atividade Grãos foi observada em cerca de 66% dos 1.791 estabelecimentos.

A tabela 11 mostra a frequência da presença das atividades nas diferentes regiões. Vê-se, por exemplo, a presença da atividade Café em cerca de 25% dos

127 estabelecimentos da região 9. Em alguma medida, isso se reflete nos dois tipos mais altos de especialização (tabelas 7 e 8). No entanto, na amostra predominam, nessa região, estabelecimentos pequenos. Possivelmente por essa razão, no tipo ESP-1 de especialização (tabela 7) há um número muito maior de casos com a atividade Gado de leite (58) do que com Café (5).

Tabela 10
Presença das diferentes atividades, por tipo de especialização

Atividade	Tipo de especialização						Total	%
	ESP-1	ESP-2	ESP-3	DOM	DIV-1	DIV-2		
Café	117	49	4	3	8	2	183	10,22
Frutas	49	33	8	6	8	7	111	6,20
Hortaliças	55	46	18	11	17	10	157	8,77
Grãos	516	400	104	32	87	44	1183	66,05
Outros vegetais	51	47	23	1	10	14	146	8,15
Gado de corte	204	104	30	6	23	12	379	21,16
Gado de leite	440	304	74	26	60	29	933	52,09
Aves	78	101	55	18	34	30	316	17,64
Suínos	96	85	52	18	33	27	311	17,36
Outros animais	68	50	27	17	17	18	197	11,00
Total da amostra	1674	1219	395	138	297	193	3916	—

Fonte: dados da amostra.

Tabela 11
Presença das diferentes atividades, por região

Atividade	Região (*)									Total
	1	2	3	4	5	6	7	8	9	
Café	0	0	20	0	0	0	1	130	32	183
Frutas	0	20	26	2	2	30	9	19	3	111
Hortaliças	1	13	25	14	4	22	61	8	9	157
Grãos	49	365	44	200	112	208	146	18	41	1183
Outros vegetais	0	38	29	11	50	4	0	2	12	146
Gado de corte	34	34	38	65	74	52	32	28	22	379
Gado de leite	36	286	51	86	35	159	114	70	96	933
Aves	9	146	6	31	11	81	6	6	20	316
Suínos	10	156	8	28	14	37	36	3	19	311
Outros animais	6	39	3	4	17	73	50	1	4	197
Total da amostra	145	1097	250	441	319	666	455	285	258	3916

Fonte: dados da amostra. (*) Identificação das regiões: 1 = Comercial C-O;

2 = Colonial PR/SC/RS; 3 = Monocultura SP; 4 = Noroeste PR;

5 = Plantio direto PR; 6 = Tradicional CE; 7 = Tradicional PE;

8 = Agricultura pobre ES/RJ; 9 = Zona da Mata MG.

Há 3.916 casos registrados nas tabelas 10 e 11; dividindo esse número por 1.791 obtêm-se 2,19 atividades por estabelecimento, que é, precisamente, o valor dado na tabela 6 para o número médio de atividades no total da amostra.

A abertura da tabela 10, numa terceira dimensão, por região, permite obter uma visão regional detalhada por tipo de especialização e por atividade, que pode ser de utilidade para orientar diversos tipos de ações. A título de exemplo, são incluídas as tabelas 12 e 13, onde se mostra o cruzamento da presença de atividades com o tipo de especialização, nas regiões que ocuparam o primeiro e o último lugares, respectivamente, no ordenamento por especialização. Limitando-se a uns poucos comentários, essencialmente qualitativos, a comparação dessas duas tabelas permite ver que: a) trata-se de situações regionais muito diferentes, que devem receber tratamentos específicos; b) a região 6 é bem mais diversificada que a 1, não só em termos dos tipos de especialização e do número de atividades – o qual já se sabia – mas também em termos da presença de atividades (por exemplo, a atividade Outros animais, que inclui eqüinos, asininos, caprinos, etc, está presente em cerca de 6% dos estabelecimentos da região 1 e de 29% na região 6; ela aparece, apenas, em dois tipos de especialização na região 1, e nos seis tipos na região 6; c) algumas atividades, tais como Grãos ou Gado de leite, sobressaem em ambas as regiões, mas, certamente, como confirmam os dados da amostra, trata-se de sistemas de produção muito diferentes. Dentro do trabalho que está sendo desenvolvido, esse tipo de considerações é combinado com a utilização de outras variáveis (área, uso de insumos, grau de instrução do produtor, etc.), para se ter um conhecimento mais aprofundado sobre os diferentes tipos de agricultura.

Tabela 12
Presença das diferentes atividades, por tipo de especialização
Região 1 – Comercial Centro-Oeste

Atividade	Tipo de especialização			Total	%
	ESP-1	ESP-2	ESP-3		
Hortaliças	0	1	0	1	1,02
Grãos	41	8	0	49	50,00
Gado de corte	32	2	0	34	34,69
Gado de leite	26	9	1	36	36,73
Aves	6	2	1	9	9,18
Suínos	7	2	1	10	10,20
Outros animais	4	2	0	6	6,12

Fonte: dados da amostra.

Tabela 13
 Presença das diferentes atividades, por tipo de especialização
 Região 6 – Tradicional Ceará

Atividade	Tipo de especialização						Total	%
	ESP-1	ESP-2	ESP-3	DOM	DIV-1	DIV-2		
Frutas	7	8	3	4	3	5	30	11,86
Hortaliças	2	6	4	4	2	4	22	8,70
Grãos	46	92	21	13	22	14	208	82,21
Outros vegetais	1	2	0	0	0	1	4	1,58
Gado de corte	15	15	7	4	6	5	52	20,55
Gado de leite	56	63	10	9	15	6	159	62,85
Aves	15	28	12	7	10	9	81	32,02
Suínos	3	8	8	6	5	7	37	14,62
Outros animais	13	19	7	12	12	10	73	28,85

Fonte: dados da amostra.

O número total de atividades presentes em cada região e no conjunto da amostra, por tipo de especialização, aparece na tabela 14. Os quocientes dos valores na coluna “Total” pelos respectivos números totais de estabelecimentos (tabela 1) dão as médias do número de atividades que aparecem na tabela 6. Um cálculo análogo dá o número médio de atividades, por tipo de especialização, entre os estabelecimentos de cada região ou do total da amostra.

Tabela 14
 Presença total de atividades, por tipo de especialização e por região

Região	Tipo de especialização						Total	D
	ESP-1	ESP-2	ESP-3	DOM	DIV-1	DIV-2		
1 - Comerc. C-O	116	26	3	0	0	0	145	4,78
2 - Col. PR/SC/RS	342	351	153	43	114	94	1097	3,44
3 - Monocul. SP	142	75	21	0	6	6	250	4,32
4 - Noroeste PR	236	120	44	4	32	5	441	4,15
5 - Plant. dir. PR	129	100	42	8	21	19	319	3,79
6 - Tradic. CE	158	241	72	59	75	61	666	3,25
7 - Tradic. PE	232	151	32	12	24	4	455	4,19
8 - Ag. pob. ES/RJ	189	69	6	4	13	4	285	4,42
9 - Z. Mata MG	130	86	22	8	12	0	258	4,22
Total da amostra	1674	1219	395	138	297	193	3916	3,83

Fonte: dados da amostra.

Mediante a utilização dos valores na coluna D da tabela 14, obtém-se o seguinte ordenamento entre as regiões:

$$[1] >_E [8] >_E [3] >_E [9] >_E [7] >_E [4] >_E [5] >_E [2] >_E [6]$$

Ele mostra só uma inversão, entre as regiões 4 e 7, quando comparado com o ordenamento inicial, sendo muito mais próximo deste que o que foi obtido pela simples contagem do número de atividades; em particular, o coeficiente de correlação por postos de Spearman tem o valor $r_S = 0,983$, que leva a rejeitar a hipótese de independência entre os dois ordenamentos, no nível de 1%. Obviamente, no caso da tabela 14 cada estabelecimento foi colocado, inicialmente, no seu tipo de especialização, mediante a consideração de suas receitas. Em termos práticos, se num levantamento não foram registrados os valores das receitas das diferentes atividades, mas se fosse possível estimar os tipos de especialização indiretamente (por exemplo, utilizando preços de mercado e os volumes das vendas), esse resultado sugere que a contagem das atividades presentes permitiria estabelecer um ordenamento de especialização entre as regiões, que deveria estar próximo do obtido com as frequências dos tipos de especialização estimados.

3.7 Valores médios

O foco deste artigo, do ponto de vista metodológico, está no tratamento individual dos estabelecimentos. Cada um deles foi colocado em um tipo de especialização, de acordo com o procedimento apresentado, com base no uso do correspondente vetor decrescente. Esta seção foi incluída para ilustrar o tipo de distorções que podem acontecer quando se trabalha com valores médios.

Especificamente, para cada região, foram somadas as receitas advindas de cada atividade, considerando todos os estabelecimentos. Cada um desses totais por atividade foi dividido pelo total das receitas na região, obtendo-se o vetor médio das contribuições de cada atividade. Ou seja, não se trata de um vetor de médias aritméticas das contribuições, em cada atividade, dos diferentes estabelecimentos da região; pelo contrário, cada região foi considerada como um único estabelecimento, no qual foram totalizadas as receitas de cada atividade (isto é, usa-se a média de razões). Até esse ponto, o procedimento coincide com o que tem sido feito por outros autores (por exemplo, Souza, s/d). À continuação, foi encontrado o correspondente vetor decrescente para cada região. Com isso, foi determinado o tipo de especialização da região, segundo o procedimento já apresentado.

A tabela 15 mostra os principais resultados obtidos com base nos vetores decrescentes regionais.

Tabela 15
Principais resultados para a média dos estabelecimentos, por região,
com base nas receitas das 10 atividades

Região	Tipo de especialização	Principal atividade	Nº de atividades	D
1 - Comercial C-O	ESP-1	Grãos	7	8,82
2 - Colonial PR/SC/RS	DOM	Grãos	9	8,34
3 - Monocultura SP	DIV-2	Outros veg.	10	6,76
4 - Noroeste PR	DOM	Grãos	9	8,33
5 - Plantio direto PR	DOM	Grãos	9	8,11
6 - Tradicional CE	DIV-2	Gado/leite	9	8,01
7 - Tradicional PE	DOM	Gado/leite	9	8,41
8 - Agric. pobre ES/RJ	DOM	Café	10	8,07
9 - Zona da Mata MG	DOM	Suínos	10	8,10
Total da amostra	DOM	Grãos	10	7,27

Fonte: dados da amostra.

Como era de se esperar, em geral, há uma diminuição nos tipos de especialização. Por exemplo, comparando com as medianas apresentadas na tabela 2, a região 3 – Monocultura SP, que tinha a mediana ESP-1, quando se trabalhou com os seus estabelecimentos, em forma individual, caiu para o tipo DIV-2, quando foi considerado o vetor médio das receitas das 10 atividades; ou seja, caiu do tipo mais alto de especialização para o mais baixo. Também, como era esperado, há um aumento no número de atividades presentes, que se pode ver ao comparar com os valores das médias, na tabela 6.

No entanto, dentro dos objetivos do presente artigo, a diferença mais importante surge quando se considera o ordenamento das regiões, por especialização. Foi calculada a estatística D para cada um dos vetores decrescentes regionais. Com a mesma notação já utilizada, e com base nesse indicador, resulta o seguinte ordenamento, que é muito diferente do que foi obtido a partir do tipo de especialização de cada estabelecimento, considerado individualmente:

$$[1] \underset{E}{>} [7] \underset{E}{>} [2] \underset{E}{>} [4] \underset{E}{>} [5] \underset{E}{>} [9] \underset{E}{>} [8] \underset{E}{>} [6] \underset{E}{>} [3]$$

O coeficiente de correlação por postos, entre esse ordenamento e o inicial, tem o valor $r_S = 0,067$, que é muito próximo de zero e não permite rejeitar a hipótese de independência com níveis críticos razoáveis. Em termos intuitivos, isso indica que não há relação alguma entre os dois ordenamentos. Ou seja, quando se trabalha com agregados regionais, no estudo das variações de especialização entre regiões, podem ser obtidas algumas conclusões que se afastam totalmente do que acontece na realidade, dependendo da técnica de cálculo utilizada. Para ilustrar

esse ponto, foi introduzida uma “pequena” modificação no esquema de cálculo. Em cada estabelecimento, as receitas das diferentes atividades foram ordenadas em forma decrescente, obtendo-se um vetor de seis componentes (tendo em vista que o número máximo de atividades encontradas no nível de estabelecimento foi de seis), onde alguns dos últimos podem ser zero. Esses vetores foram somados em cada região e cada componente foi dividido pelo total da receita na região, o mesmo sendo feito para o total da amostra. Assim, foram obtidos os vetores médios (isto é, trata-se de médias de razões, e não de médias aritméticas) dos vetores decrescentes dos estabelecimentos, utilizados no início, independentemente das atividades que ocupassem cada posição. Ou seja, como se trata de estudar a especialização, com independência das atividades envolvidas, o vetor médio deve servir a esse objetivo, dentro de cada região ou no total da amostra. Sendo médias de vetores decrescentes, esses vetores médios também são decrescentes. Finalmente, cada um deles foi alocado no tipo de especialização correspondente. A tabela 16 resume os principais resultados, com base nessa forma de cálculo.

Tabela 16
Principais resultados para a média dos estabelecimentos, por região,
com base nos vetores decrescentes individuais

Região	Tipo de Especialização	Nº de atividades	D
1. Comercial C-O	ESP-1	4	4,963
2. Colonial PR/SC/RS	ESP-1	6	4,755
3. Monocultura SP	ESP-1	6	4,837
4. Noroeste PR	ESP-1	6	4,900
5. Plantio direto PR	ESP-1	5	4,862
6. Tradicional CE	ESP-2	6	4,700
7. Tradicional PE	ESP-1	4	4,882
8. Agric. pobre ES/RJ	ESP-1	4	4,888
9. Zona da Mata MG	ESP-1	5	4,864
Total da amostra	ESP-1	6	4,873

Fonte: dados da amostra.

O tipo de especialização de cada um dos vetores médios coincide com a mediana que aparece na tabela 2, salvo na região 2 – Colonial PR/SC/RS, onde ela está no tipo ESP-2. Por sua vez, o número de atividades reflete, exatamente, o que aparece na tabela 6, e não poderia ser diferente, em vista do algoritmo utilizado. Quanto ao ordenamento entre regiões fornecido pelo indicador D , obtém-se o seguinte resultado:

$$[1] \underset{E}{>} [4] \underset{E}{>} [8] \underset{E}{>} [7] \underset{E}{>} [9] \underset{E}{>} [5] \underset{E}{>} [3] \underset{E}{>} [2] \underset{E}{>} [6]$$

Como se pode ver, ele está muito mais próximo do ordenamento inicial do que o obtido anteriormente, com base nas médias dos componentes de receitas por cada uma das 10 atividades. Em particular, o valor do coeficiente de correlação por postos, entre esse ordenamento e o inicial, é $r_S = 0,733$, que resulta ser levemente inferior ao valor crítico no nível de 1% e bem maior que o valor crítico no nível de 5%, o que leva a rejeitar a hipótese de independência. Portanto, o ordenamento obtido para as médias regionais, calculadas mediante o método exposto, dá um resultado que é consistente com o que foi obtido inicialmente.

4. Considerações Finais

O enfoque seguido neste artigo, para o estudo da especialização na produção agropecuária, pode ser resumido nos seguintes pontos:

- obtenção de dados, no nível de estabelecimento, que permitam estimar as receitas provenientes da agricultura;
- definição, a priori, de tipos de especialização, mediante um procedimento para alocar cada estabelecimento em um único tipo, independentemente das atividades envolvidas;
- utilização de uma forma específica da dominância estocástica para comparar a especialização entre regiões.

Parece claro que um enfoque análogo pode ser utilizado em outros estudos; por exemplo, para avaliar a especialização de municípios ou microrregiões; em lugar das receitas, pode-se pensar na utilização de outras variáveis, como valor da produção, uso da terra, área colhida, etc., dependendo das atividades consideradas. Em regiões pequenas, pode-se trabalhar com microdados de censos agropecuários.

Inicialmente, foram consideradas 12 atividades, em lugar das 10 utilizadas neste artigo; por exemplo, em lugar de Frutas, considerou-se Citrus e Outras frutas. No entanto, isso não mostrou nenhum efeito significativo sobre as distribuições dos tipos de especialização, dentro de cada região. Também foram realizados diversos exercícios, nos quais se modificavam alguns dos limites que definem os tipos de especialização. Se bem que isso não deve ser considerado como um estudo exaustivo de sensibilidade, o fato é que, nesses exercícios, não foram observadas mudanças que levassem a alterar o ordenamento de especialização entre as regiões.

Também foi feito um exercício em que os estabelecimentos do tipo DIV-1 foram deixados livres, para serem realocados nos outros tipos mediante um procedimento

de discriminação. Nesse sentido, foram utilizados tanto algoritmos tradicionais da análise discriminante quanto outros mais recentes, com redes neurais artificiais. Obtiveram-se resultados similares com esses dois métodos, sendo que a maioria dos estabelecimentos foi realocada em algum dos tipos dominantes (particularmente em ESP-2 e ESP-3), e apenas uns poucos foram para DIV-2. Isso reforça a opinião de que o tipo DIV-1 corresponde a uma situação de transição e instável; na maioria dos casos, com pequenas modificações os estabelecimentos desse grupo poderiam passar para um tipo de especialização dominante.

Dentro de um quadro geral em que predomina uma orientação para os tipos mais altos de especialização, a metodologia utilizada permite detectar diferenças entre as regiões. Quando isso é complementado com a identificação das atividades mais freqüentes, dentro de cada região, tem-se uma descrição muito mais precisa para definir ações de pesquisa e, possivelmente, formular algumas medidas de política agrícola, particularmente no que se refere ao estabelecimento de uma agricultura sustentável.

Referências

- Anderson, J. R. and Dillon, J. L. & Hardaker, J. B. (1977). *Agricultural Decision Analysis*. The Iowa State Univ. Press, Ames, Iowa.
- Andrade, B. B. & Rathie, P. N. (1998). Medidas de informação e aplicações. Brasília: UnB, Depto. de Estatística. Trabalho de iniciação científica, mimeog, 87 p.
- Berry, M. J. A. & Linoff, G. S. (2000). *Mastering Data Mining. The Art and Science of Customer Relationship Management*. John Wiley & Sons, New York.
- Eijkel, G. C. van den (1999). Rule induction. In Berthold, M. & Hand, D. J., editors, *Intelligent Data Analysis. An Introduction*. Springer, New York.
- Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2):427–432. Durham (USA).
- Hoffmann, R. (1991). *Estatística Para Economistas*. Livraria Pioneira, São Paulo, 2 edition.
- Khinchin, A. I. (1957). *Mathematical Foundations of Information Theory*. Dover Publications, New York.

- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London, 2 edition.
- Reis, E. (1991). *Estatística Descritiva*. Edições Sílabo, Lisboa.
- Rényi, A. (1962). *Wahrscheinlichkeitsrechnung. Mit Einem Anhang über Informationstheorie*. VEB Deutscher Verlag der Wissenschaften, Berlin.
- Siegel, S. (1975). *Estatística Não-Paramétrica Para as Ciências Do Comportamento*. McGraw-Hill do Brasil, São Paulo.
- Souza, J. (s/d). *Métodos Estatísticos Nas Ciências Psicossociais*, volume I. The-saurus, Brasília.
- Theil, H. (1967). *Economics and Information Theory*. North-Holland Pub. Co., Amsterdam.