

# UMA APLICAÇÃO DE MÉTODOS MULTIVARIADOS NA PREVISÃO DE RENDIMENTOS DO MILHO<sup>1</sup>

José Roberto Vicente<sup>2</sup>

## RESUMO

Variáveis climáticas, tecnológicas e análise de conglomerados, fatorial, discriminante e de regressão foram utilizadas para modelar o rendimento da cultura do milho no Estado de São Paulo, de 1958 a 1988. Previsões foram efetuadas para os rendimentos de 1986, 1987 e 1988, com erros de, em média, 6%. Todos os métodos testados foram considerados úteis para subsidiar previsões de safras.

**Palavras-chaves:** milho, previsão de safras, métodos multivariados.

## ONE APPLICATION OF MULTIVARIATE METHODS TO CORN YIELD FORECAST

### SUMMARY

Climate and agricultural technological data were combined using factor, cluster, discriminant and regression analyses to estimate corn yields for the State of São Paulo, Brazil, over the 1958-88 period. One-year forecasts were generated for 1986, 1987 and 1988, using each of the estimation techniques, with average errors of estimates of 6 per cent. All estimation methods were deemed useful for harvest forecast.

**Key-words:** corn yields, harvest forecast, multivariate methods.

## 1 - INTRODUÇÃO

Variações nas colheitas de produtos agrícolas podem exercer grande influência nas decisões do poder público e de vários setores da economia. A produção agrícola é fortemente condicionada pelos ciclos biológicos das culturas e depende, portanto, das condições do tempo atmosférico. O planejamento na agricultura, devido a essas peculiaridades, deve considerar essa fonte de incertezas, importante para todo o conjunto da economia cujo desempenho pode ser afetado por choques de oferta agrícola.

O abastecimento interno, as exportações e a estabilidade dos preços podem ser afetados por variações do tempo. Medidas destinadas a minorar esses efeitos dependem, em grau elevado, de boas estimativas sobre a magnitude das safras agrícolas. Por esse motivo, grandes esforços são efetuados,

principalmente pelo poder público, para elaborar e divulgar previsões e estimativas de safras; em São Paulo, o Instituto de Economia Agrícola (IEA) e a Coordenadoria de Assistência Técnica Integral (CATI) efetuam levantamentos cinco vezes por ano, com questionários em nível de município (levantamento subjetivo) e em nível de imóvel rural, com amostra probabilística atualmente constituída de 3.262 elementos (levantamento objetivo). Apesar dos custos elevados, os retornos provenientes de tais levantamentos aparentam ser compensadores, conforme estudo de NEGRI NETO et ali (1988).

Esses esforços são complementados com estudos de projeções de oferta, a maioria dos quais emprega modelos tipo Nerlove<sup>3</sup>. Nesse caso, a área cultivada é normalmente empregada como variável dependente, obtendo-se estimativas de produção a partir de rendimentos "médios", "otimistas" ou

"pessimistas". Para evitar essa limitação, alguns estudos vêm procurando considerar explicitamente variáveis representativas das condições do tempo, empregando as condições meteorológicas (basicamente temperatura, precipitação e, em alguns casos, deficiência hídrica, brilho solar, umidade do ar, geada, etc.) como explicativas do rendimento e/ou da produção. Comumente, análise de regressão vem sendo empregada em tais tentativas, o que limita o número de variáveis passíveis de serem incluídas nos modelos; o número de observações disponível nessas séries praticamente inviabiliza o uso de modelos ARIMA<sup>4</sup>.

Por essas características, e dada a importância das previsões de safras, justificam-se novos estudos, especialmente com metodologias capazes de permitir o uso de maior número de variáveis explicativas dos rendimentos.

O objetivo do presente estudo foi o de obter modelos abrangentes e operacionais capazes de subsidiar estimativas de rendimentos de culturas. Para tanto foram utilizados métodos multivariados, aplicados à série de rendimentos da cultura do milho no Estado de São Paulo. Os modelos foram estimados procurando-se, inicialmente, cobrir todo o ciclo da cultura, do plantio (que ocorre em São Paulo principalmente nos meses de outubro e novembro, FUNDAÇÃO GETÚLIO VARGAS, 1980) à colheita (concentrada nos meses de abril e maio); em seguida, esse período foi abreviado, visando estimativas antecipadas (previsões).

## 2 - METODOLOGIA

O rendimento pode ser considerado como uma função do tipo de solo, do clima, das técnicas ou do processo de produção utilizado e das condições meteorológicas. O solo e o clima determinam o rendimento potencial de cada técnica, sendo as condições do tempo responsáveis por flutuações anuais. As técnicas podem ser caracterizadas pela quantidade e pela qualidade de fatores de produção, estando entre os principais determinantes da produtividade da terra, sementes, fertilizantes e defensivos<sup>5</sup>. No presente estudo, em que se pretendeu obter modelos operacionais, apenas variáveis representativas das condições do tempo foram inicialmente considera-

das: temperaturas médias mensais, precipitações pluviométricas mensais e deficiências hídricas mensais do período outubro-abril<sup>6</sup>. As deficiências hídricas, definidas como as diferenças entre as evapotranspirações potenciais e as evapotranspirações reais, têm como determinantes as temperaturas, comprimentos dos dias, precipitações pluviométricas e quantidades de água armazenada nos solos. As deficiências foram obtidas pelo cálculo de balanços hídricos, através do método de Thorntwaite & Mather<sup>7</sup> e do roteiro proposto por ORTOLANI et alii (1970), assumindo-se uma capacidade de armazenamento de água no solo de 50 mm, conforme resultado anterior obtido por SILVA; VICENTE; CASER (1986). Os postos meteorológicos utilizados foram os localizados nos municípios de Araçatuba, Bauru, Campinas, Limeira, Mococa, Pindorama, Ribeirão Preto e Tietê, cujos dados tiveram como fonte o Instituto Agrônomo de Campinas, o Instituto Nacional de Meteorologia, o IEA e a CATI. De posse dos dados de cada posto, foram obtidos, através do cálculo de médias, os das Divisões Regionais Agrícolas (DIRAs) por eles representadas. A representação das condições médias a que esteve exposta a cultura no Estado, foi obtida por médias ponderadas, tendo como base de ponderação as áreas cultivadas com milho em cada uma das DIRAs. Os dados de áreas cultivadas necessários para essa ponderação, bem como os rendimentos da cultura do milho, tiveram como fonte o IEA. O período analisado entendeu-se de 1958 a 1988.

Os métodos multivariados utilizados foram análise fatorial, análise de conglomerados e análise discriminante.

A análise fatorial têm como propósito principal descrever as relações entre um número elevado de variáveis em termos de poucos fatores subjacentes e inobserváveis (JOHNSON & WICHERN, 1982). Procura agregar variáveis altamente correlacionadas e descrever cada variável como uma combinação de fatores latentes. Um modelo básico de análise fatorial pode ser expresso como (DILLON & GOLDSTEIN, 1984):

$$X = \Lambda f + e \quad (1)$$

onde:

X é o vetor (px1) das observações,

$\Lambda$  é a matriz (pxm) de cargas ou pesos fatoriais,

$f$  é o vetor (px1) de fatores comuns (variáveis inobserváveis), e  $e$  é o vetor de erros específicos.

Para a extração escolheram-se os métodos de componentes principais, máxima verossimilhança e alpha, com as variáveis sendo padronizadas. Os dois primeiros métodos são os mais comumente utilizados, enquanto que o último pode ser superior ao método de componentes principais em pequenas amostras (Acito & Anderson, 1980, citados por DILLON & GOLDSTEIN, 1984)<sup>8</sup>. Após a rotação (Varimax) dos fatores, foram estimados os escores fatoriais associados a cada observação, posteriormente empregados como variáveis explicativas do rendimento do milho em modelos de regressão múltipla<sup>9</sup>, obtendo-se previsões passo a passo dos rendimentos para o período 1986-88. A forma geral utilizada nesses modelos foi

$$R = \alpha + \beta_1 T + \beta_2 F_1 + \dots + \beta_n F_m + \varepsilon \quad (2)$$

onde  $R$  é o rendimento do milho,  $\alpha$  é a constante de regressão,  $\hat{\alpha}_1$  a  $\hat{\alpha}_n$  são os coeficientes angulares,  $T$  é a variável tendência,  $F_1$  a  $F_m$  são escores fatoriais e  $\varepsilon$  é o erro aleatório.

Na análise de conglomerados, o objetivo é a formação de grupos (ou conglomerados) tendo como características homogeneidade dentro desses grupos e heterogeneidade entre eles (ou, pequenas variações dentro do grupo em relação às variações entre os grupos, DILLON & GOLDSTEIN, 1984). A análise fundamenta-se na escolha de uma medida de distância (ou de similaridade) entre as observações e de um procedimento de formação de grupos baseado na medida de distância (ou similaridade) adotada. A medida de distância utilizada foi o Quadrado da Distância Euclidiana, definido como:

$$\text{Distância}(X, Y) = \sum (X_i - Y_i)^2 \quad (3)$$

Alternativamente, empregou-se uma medida de similaridade, o cosseno do vetor de variáveis, que é:

$$\text{Similaridade}(X, Y) = \frac{\sum (X_i Y_i)}{\sqrt{\sum (X_i^2) \sum (Y_i^2)}} \quad (4)$$

Como a distância Euclidiana não é invariante à escala, quando essa medida foi empregada, além das variáveis na forma original trabalhou-se, em

seguida, com as mesmas padronizadas.

Para formação de conglomerados foram utilizados, inicialmente, três procedimentos que podem ser classificados como métodos de enlace (NORUSIS, 1988). No método do enlace simples, os grupos são fundidos a partir de seus membros mais próximos:

$$d(A, B) = \min d(X, Y), X \in A e Y \in B \quad (5)$$

A menor distância (ou maior similaridade) entre os membros mais distantes é o critério de fusão no método do enlace completo:

$$d(A, B) = \max d(X, Y), X \in A e Y \in B \quad (6)$$

O método do enlace médio funde grupos a partir da menor distância (ou maior similaridade) considerada a média de seus membros:

$$d(A, B) = \frac{\sum d(X, Y)}{(N_A)(N_B)}, X \in A e Y \in B \quad (7)$$

onde  $N_A$  e  $N_B$  representam o número de membros nos conglomerados<sup>10</sup>.

Todos esses procedimentos são métodos hierárquicos aglomerativos (ou no sentido ascendente) onde, inicialmente, cada observação é um conglomerado, e o grupo, após formado, é indissolúvel. Para previsão dos rendimentos do milho, após a seleção das medidas e dos métodos considerados mais adequados, foi efetuada nova análise interrompendo-se a série em 1986, obtendo-se a classificação para esse ano; a seguir, a série estendeu-se até 1987 e 1988, para propiciar a classificação dos rendimentos desses anos.

Por último, empregou-se análise discriminante, uma técnica estatística para classificar indivíduos ou observações em grupos exaustivos e mutuamente exclusivos, com base num conjunto de variáveis independentes (DILLON & GOLDSTEIN, 1984). É formada uma combinação linear dessas variáveis independentes, que serve como base para alocar as observações nos grupos. Uma equação linear discriminante pode ser representada similarmente a um modelo de regressão linear múltipla<sup>11</sup> (NORUSIS, 1988):

$$D = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_p X_p \quad (8)$$

Supondo-se  $k$  grupos especificados *a priori*, têm-se  $k-1$  funções discriminantes. No caso presente,

foram especificados, alternativamente, três e cinco grupos<sup>12</sup>, efetuando-se análise classificatória para alocar da forma mais eficiente a observação seguinte (1986 ou 1987 ou 1988), a partir das variáveis independentes anteriormente descritas. Como critérios de seleção das variáveis utilizaram-se, além da entrada forçada (todas as variáveis), a minimização do lambda de Wilks<sup>13</sup> e a maximização da distância de Mahalanobis<sup>14</sup> entre os grupos (NORUSIS, 1988).

Todos os métodos citados são normalmente empregados em cortes seccionais; entretanto, apesar de terem sido utilizadas séries temporais das variáveis independentes, as mesmas podem ser consideradas variáveis aleatórias sem correlações temporais elevadas. Essa hipótese foi testada através da estatística Q de Box-Pierce (VANDAELE, 1983), em testes conjuntos para verificar a existência de autocorrelações estatisticamente significativas até os sétimos anos precedentes (cerca de 25% das séries): nenhum dos valores de  $\pm^2$  obtidos foi significativo a 5%.

### 3 - RESULTADOS E DISCUSSÃO

As temperaturas médias mensais dos meses de outubro a março, no período 1958-88, estiveram entre 20,41°C e 26,39°C; apenas o mês de abril apresentou maior variabilidade, com limites entre 16,76°C e 24,08°C (Tabela 1). As precipitações pluviométricas, como esperado apresentam maiores variações que as temperaturas; fevereiro, por exemplo, de 1958 a 1988 teve marcas entre 59 mm e 384,45 mm, enquanto que abril, o mês mais seco do período, mostrou índices entre 13,45 mm e 150,15 mm. Abril, seguido de perto por outubro, também foi o mês de, em média, diferenças entre as evapotranspirações potenciais e as evapotranspirações reais (deficiências hídricas) mais elevadas. As variâncias associadas a estas últimas variáveis são relativamente maiores do que as das anteriores.

#### 3.1 - Resultados da Análise Fatorial

Alguns testes foram efetuados para determinar se as pressuposições da Análise Fatorial (AF) eram atendidas pela amostra em questão<sup>15</sup>. Um deles foi o de Bartlett, usado para verificar se a matriz de correlações

é uma matriz identidade<sup>16</sup>; o teste requer que os dados sejam provenientes de uma população normal multivariada. Encontrando-se um valor alto, associado a um nível de significância baixo, pode-se inferir que é improvável que a matriz de correlações seja uma matriz identidade; caso essa hipótese não possa ser rejeitada, o uso de AF deve ser reconsiderado. O valor encontrado com as variáveis temperaturas mensais, precipitações pluviométricas mensais e deficiências hídricas mensais para o período outubro-abril foi de 447,86; para o mesmo período, excluídas as temperaturas, o valor foi 208,13. Considerando-se temperaturas, precipitações e deficiências para o período outubro-fevereiro, o valor foi 284,97; com essas mesmas variáveis para o período outubro-janeiro, encontrou-se 202,92 e, para o período outubro-dezembro, 136,92. Todos esses valores são significativos a 0,00000% de probabilidade.

Outra medida de adequação da amostra é a matriz de anti-imagem das correlações, que é construída pelos negativos dos coeficientes de correlações parciais (que, por sua vez, são estimados pelas correlações entre os fatores singulares), e devem ser próximos de zero, se as pressuposições da AF estiverem sendo respeitadas. Na matriz formada pelas temperaturas, precipitações e deficiências, período outubro-abril, apenas 1,4% dos elementos estava acima de 0,09; para o mesmo período, excluídas as temperaturas, 29,7%. Para o período outubro-fevereiro, 18,1%, para o período outubro-janeiro, 25,8% e para o período outubro-dezembro, 47,2%. Portanto, apenas a amostra para o período outubro-abril com todas as variáveis, seria adequada para o emprego de AF, segundo esse indicador.

O índice de Kaiser-Meyer-Olkin (KMO) é mais uma dessas medidas<sup>17</sup>: quando a soma dos quadrados dos coeficientes de correlação parcial entre todos os pares de variáveis é pequena quando comparada com a soma dos quadrados dos coeficientes de correlação, o índice se aproxima de um. Valores pequenos do índice (menores que 0,5) indicam que o uso de AF pode ser inadequado, uma vez que a correlação entre os pares de variáveis não pode ser explicada por outras variáveis. Os valores de KMO obtidos para as duas especificações do período outubro-abril foram inferiores a 0,5 (0,44 e 0,47, respectivamente); para o período outubro-fevereiro obteve-se 0,52, com o período outubro-janeiro 0,66

TABELA 1 - Estatísticas Básicas das Variáveis Utilizadas

Variável	Média	Desvio-padrão	Mínimo	Máximo
Temperatura média (°C)				
Outubro	22,58	1,09	20,44	24,88
Novembro	23,22	0,94	21,53	24,87
Dezembro	23,49	0,88	21,45	25,17
Janeiro	23,97	0,90	22,15	26,39
Fevereiro	24,07	1,02	22,39	26,06
Março	23,48	1,01	20,41	25,12
Abril	21,69	1,42	16,76	24,08
Precipitação pluviométrica (mm)				
Outubro	134,27	50,64	29,35	221,05
Novembro	145,69	45,68	82,90	294,15
Dezembro	225,97	75,74	76,85	386,40
Janeiro	228,10	72,31	105,00	377,25
Fevereiro	189,12	82,81	59,00	384,45
Março	142,16	59,15	45,10	300,60
Abril	67,79	36,22	13,45	150,15
Deficiência hídrica (mm)				
Outubro	12,90	20,47	0	92,75
Novembro	5,76	6,74	0	26,40
Dezembro	4,07	9,75	0	42,10
Janeiro	3,95	8,06	0	35,65
Fevereiro	5,58	9,43	0	32,75
Março	8,39	9,61	0	30,05
Abril	13,60	12,06	0	45,35

Fonte: Elaborada a partir de dados básicos do Instituto de Economia Agrícola, Instituto Nacional de Meteorologia, Instituto Agrônomo de Campinas e Coordenadoria de Assistência Técnica Integral.

e com outubro-dezembro, 0,59. Portanto, o período que melhor atendia as condições do indicador anterior, é o pior segundo este último teste. Como o objetivo do trabalho é o de obter as melhores previsões de rendimento da cultura, optou-se por aplicar o método, deixando a avaliação para os resultados fornecidos pelas diferentes especificações.

Para o período outubro-abril, com temperaturas, precipitações e deficiências, e através do método de extração  $\alpha$ <sup>18</sup>, obtiveram-se sete fatores com raízes características, inicialmente, maiores do que um; esse foi o critério de determinação do número de fatores, uma vez que aqueles com raízes características menores do que um não contribuem para a explicação da variância total mais do que uma variável isolada. Esses fatores explicaram 73,7% da variância total. As comunalidades<sup>19</sup> indicaram que, em relação às variáveis individuais, os percentuais de explicação da variância estiveram entre 0,67 e 0,95 (Tabela 2). A adequação dos resultados obtidos pode ser testada observando-se quão bem o modelo ajustado reproduz as correlações observadas, através da reprodução da matriz de correlações, construída tendo na diagonal principal as comunalidades, abaixo dela as correlações estimadas pelo modelo e, acima, os resíduos entre as correlações observadas na amostra e as reproduzidas. A magnitude desses resíduos é um indicador de ajuste: caso sejam muito grandes, o modelo não está bem ajustado. No presente caso, existiam 34 (apenas 16%) dos resíduos superiores a 0,05 em valor absoluto.

Pelos pesos dos fatores, após a rotação Varimax (JOHNSON & WICHERN, 1982), observou-se que o primeiro deles estava positivamente relacionado às temperaturas de outubro, dezembro e janeiro e às deficiências hídricas de dezembro e janeiro e, negativamente às precipitações pluviométricas desses mesmos meses. Pode, portanto, ser associado à deficiência de água desde o desenvolvimento vegetativo até - dependendo da época do plantio e do cultivar - à polinização, à frutificação ou ao início da maturação (FANCELLI & LIMA, 1982). O segundo fator apareceu positivamente associado às temperaturas de janeiro a abril (frutificação-maturação-início da colheita), o terceiro positivamente à temperatura e à deficiência e negativamente associado à precipitação do mês de fevereiro, representando, provavelmente, deficiência de

água na maturação. O quarto fator esteve positivamente associado à chuva e negativamente à deficiência hídrica do mês de abril (maturação-colheita); o quinto fator esteve positivamente relacionado à temperatura e à deficiência e negativamente à precipitação, todas do mês de outubro (pré-plantio e plantio). O sexto fator esteve relacionado, respectivamente positiva e negativamente, à precipitação pluviométrica e à deficiência hídrica em março (maturação-início da colheita). Por último, o sétimo fator exibiu correlações elevadas com a deficiência hídrica (positiva) e com chuvas (negativa) ocorridas em novembro (plantio-desenvolvimento inicial).

Após o cálculo dos escores fatoriais associados às observações, modelos de regressão foram ajustados utilizando-os como variáveis explicativas do rendimento do milho, juntamente com uma tendência, empregada para representar mudanças tecnológicas. As variáveis não significativas (a pelo menos 10%) foram excluídas do modelo, sendo mantidos, na especificação final, os escores referentes ao primeiro, quarto e sexto fatores (Tabela 3). Portanto, o rendimento seria afetado negativamente pela deficiência de água desde o desenvolvimento vegetativo até a polinização/frutificação (ou florescimento, dependendo do cultivar) e por chuvas na maturação (provavelmente de cultivares de ciclo médio ou tardios), e positivamente por deficiências hídricas no período de colheita (provavelmente de cultivares mais precoces). Naturalmente, como todas as variáveis são representadas nos fatores, os efeitos citados não são os únicos, embora dominantes<sup>20</sup>. Os ajustes obtidos (tanto com os escores desse método de extração, como com os obtidos por componentes principais<sup>21</sup>) podem ser considerados muito bons em termos de coeficientes de determinação (ao redor de 0,9).

Ainda para o período outubro-abril, excluídas as temperaturas, os seis fatores também explicaram 73,7% da variância total. As comunalidades indicaram que, em relação às variáveis individuais, os percentuais de explicação da variância estiveram entre 0,52 e 0,88 (Tabela 4). Na reprodução da matriz de correlações, houve apenas 8% dos resíduos superiores a 0,05 em valor absoluto.

O primeiro fator também apareceu relacionado às deficiências hídricas e à pluviosidade dos meses de dezembro e janeiro, porém com sinais contrários aos obtidos na especificação anterior.

TABELA 2 - Principais Resultados da Análise Fatorial Com as Variáveis Temperatura, Precipitação Pluviométrica e Deficiência Hídrica Mensal, do Período Outubro-Abril, Estado de São Paulo, 1958-88

Variáveis <sup>1</sup>	Pesos dos fatores <sup>2</sup>							Comunalidades
	F1	F2	F3	F4	F5	F6	F7	
TOUT.	0,63	0,04	0,03	0,06	0,56	-0,05	-0,10	0,90
TNOV.	0,44	0,52	-0,01	0,16	0,06	0,18	0,17	0,89
TDEZ.	0,67	0,25	0,07	0,05	0,08	0,04	0,11	0,67
TJAN.	0,50	0,61	0,31	0,04	0,15	-0,01	0,06	0,81
TFEV.	0,10	0,66	0,70	-0,13	0,02	0,09	0,20	0,95
TMAR.	-0,04	0,68	0,08	-0,04	0,27	-0,15	0,08	0,79
TABR.	0,01	0,83	-0,21	0,01	0,10	0,14	0,12	0,92
POUT.	-0,16	-0,22	0,01	-0,04	-0,82	0,00	-0,18	0,92
PNOV.	0,26	-0,10	-0,07	-0,04	-0,04	-0,01	-0,78	0,69
PDEZ.	-0,70	0,00	0,05	-0,02	0,05	0,03	0,11	0,76
PJAN.	-0,64	0,07	-0,20	0,07	-0,29	0,13	0,08	0,84
PFEV.	-0,17	0,01	-0,90	0,02	0,00	0,00	-0,12	0,88
PMAR.	-0,14	0,11	0,11	-0,02	0,02	0,93	0,00	0,76
PABR.	-0,18	0,40	0,07	0,67	-0,08	-0,13	-0,12	0,95
DOUT.	0,36	0,33	-0,06	0,38	0,73	0,05	0,00	0,95
DNOV.	0,36	0,23	-0,13	0,14	0,12	0,12	0,72	0,87
DDEZ.	0,67	0,20	-0,26	0,40	0,23	0,00	0,13	0,89
DJAN.	0,71	-0,10	0,20	0,29	0,22	0,00	0,04	0,88
DFEV.	-0,03	-0,05	0,74	0,13	-0,02	-0,19	-0,21	0,84
DMAR.	-0,07	0,07	0,34	0,01	0,04	-0,72	-0,10	0,77
DABR.	0,11	0,21	0,15	-1,00	0,12	-0,14	0,04	0,95
Raiz característica	5,63	2,55	2,47	1,53	1,43	0,97	0,91	
Variância explicada (%)	26,70	12,10	11,80	7,30	6,80	4,60	4,30	

<sup>1</sup>A primeira letra representa o tipo de variável (T = temperatura; P = prec. pluviométrica e D = def. hídrica); as três letras seguintes representam o mês (OUT. = outubro, etc.).

<sup>2</sup>Após rotação (Varimax); método de extração: alpha.

Fonte: Resultados da pesquisa.

TABELA 3 - Principais Resultados da Análise de Regressão Utilizando Escores Fatoriais Provenientes das Condições Climáticas<sup>1</sup> de Vários Períodos Como Variáveis Explicativas do Rendimento do Milho, Estado de São Paulo, 1958-88

Variável <sup>2</sup>	Período				
	Out.-abr. <sup>3</sup>	Out.-abr. <sup>4</sup>	Out.-fev. <sup>5</sup>	Out.-jan. <sup>6</sup>	Out.-dez. <sup>7</sup>
Intercepto	- 1.660,31 a	1.704,35 a	-1.562,28 a	-1.604,28 a	-1.541,56 a
Tendência	49,91 a	50,24 a	48,29 a	48,87 a	48,01 a
Fator 1	- 85,40 a	100,72 a	-93,42 a	- 88,95 a	- 57,32 d
Fator 3		- 55,14 b			
Fator 4	- 41,11 b				
Fator 5		- 53,99 c			
Fator 6	64,11 b				
R <sup>2</sup>	0,90	0,91	0,88	0,87	0,86
F <sup>8</sup>	57,28 a	69,08 a	100,00 a	96,22 a	82,87 a
Durbin-Watson	1,97	2,19	2,11	2,03	1,97

<sup>1</sup>Temperaturas, precipitações pluviométricas e deficiências hídricas mensais.

<sup>2</sup>Níveis de significância: a = 1%; b = 5%; c = 10%; d = 20%.

<sup>3</sup>Método de extração de fatores: alpha.

<sup>4</sup>Método de extração de fatores: alpha. Não foram incluídas temperaturas médias.

<sup>5</sup>Método de extração de fatores: máxima verossimilhança.

<sup>6</sup>Método de extração de fatores: máxima verossimilhança.

<sup>7</sup>Método de extração de fatores: alpha.

<sup>8</sup>Estatística F. Nível de significância: a = 1%.

Fonte: Resultados da pesquisa.

Pode, então, ser associado à disponibilidade de água do desenvolvimento vegetativo até a polinização, frutificação ou o início da maturação. Os demais fatores estão associados, respectivamente, às condições dos meses de outubro, abril, fevereiro, março e novembro.

No modelo de regressão correspondente, os escores referentes ao primeiro, terceiro e quinto fatores apareceram afetando significativamente o rendimento do milho (Tabela 3), confirmando os resultados do modelo anterior, já que as variáveis principais, em ambos os casos, são referentes aos mesmos meses.

Com as variáveis temperaturas, pluviosida-

des e deficiências hídricas mensais do período outubro-fevereiro, foram obtidos quatro fatores, que explicaram 63,2% da variância total; as comunalidades ficaram entre 0,57 e 0,92 (Tabela 5). Na reprodução da matriz de correlações, 38% dos resíduos foram superiores a 0,05 em valor absoluto. O primeiro fator - o único cujos escores foram significativos no modelo de regressão subsequente (Tabela 3) - apresentou correlações positivas com as temperaturas de outubro e dezembro e com as deficiências de dezembro e janeiro e, correlações negativas com as chuvas de dezembro e janeiro. Pode ser associado, no caso do milho, à deficiência de água no desenvolvimento vegetativo e florescimento. As principais

TABELA 4 - Principais Resultados da Análise Fatorial Com as Variáveis Precipitação Pluviométrica e Deficiência Hídrica Mensal, do Período Outubro-Abril, Estado de São Paulo, 1958-88

Variáveis <sup>1</sup>	Pesos dos fatores <sup>2</sup>						Comunalidades
	F1	F2	F3	F4	F5	F6	
POUT.	0,16	-0,65	-0,01	0,01	0,00	0,23	0,77
PNOV.	-0,27	-0,05	-0,03	0,05	0,02	0,72	0,52
PDEZ.	0,73	0,00	-0,05	-0,05	-0,02	-0,07	0,64
PJAN.	0,66	-0,21	0,02	0,18	-0,11	-0,05	0,63
PFEV.	0,17	0,05	0,02	1,01	0,03	0,17	0,73
PMAR.	0,16	0,05	-0,01	-0,07	-0,79	-0,01	0,59
PABR.	0,17	-0,04	-0,62	-0,05	0,12	0,06	0,61
DOUT.	-0,22	1,06	0,28	0,04	-0,03	0,00	0,88
DNOV.	-0,25	0,22	0,14	0,13	-0,09	-0,76	0,57
DDEZ.	-0,62	0,38	0,35	0,23	0,02	-0,19	0,86
DJAN.	-0,70	0,24	0,34	-0,20	-0,01	-0,04	0,75
DFEV.	0,02	0,01	0,09	-0,69	0,21	0,22	0,71
DMAR.	0,07	0,04	-0,03	-0,32	0,82	0,10	0,65
DABR.	-0,04	0,22	1,03	-0,16	0,18	-0,08	0,75
Raiz característica	3,65	2,05	1,43	1,08	1,18	0,93	
Varição explicada (%)	26,10	14,70	10,20	7,70	8,40	6,60	

<sup>1</sup>A primeira letra representa o tipo de variável (T = temperatura; P = prec. pluviométrica e D = def. hídrica); as três letras seguintes representam o mês (OUT. = outubro, etc.).

<sup>2</sup>Após rotação (Varimax); método de extração: alpha.

Fonte: Resultados da pesquisa.

variáveis relacionadas ao segundo fator foram a temperatura, a precipitação e a deficiência do mês de outubro e a deficiência hídrica em dezembro; o terceiro fator esteve associado às condições do tempo em fevereiro e o quarto às temperaturas dos meses de novembro, janeiro e fevereiro e à deficiência em novembro.

Para o período outubro-janeiro, os três fatores extraídos por máxima verossimilhança explicaram 60,3% da variância total, as comunalidades estiveram entre 0,46 e 0,90 (Tabela 6) e 33% dos

resíduos da matriz de correlações reproduzida ficaram acima de 0,05. O primeiro fator esteve associado às mesmas variáveis presentes no primeiro fator do período outubro-fevereiro, e também foi o único significativo no modelo de regressão ajustado com os escores fatoriais obtidos com o período outubro-janeiro (Tabela 3). O segundo fator também apresentou correlações com as variáveis presentes no correspondente da especificação anterior, enquanto que o terceiro fator teve como variáveis principais pluviosidade e deficiências hídricas em novembro.

TABELA 5 - Principais Resultados da Análise Fatorial Com as Variáveis Temperatura, Precipitação Pluviométrica e Deficiência Hídrica Mensal, do Período Outubro-Fevereiro, Estado de São Paulo, 1958-88

Variáveis <sup>1</sup>	Pesos dos fatores <sup>2</sup>				Comunalidades
	F1	F2	F3	F4	
TOUT.	0,52	0,59	0,07	0,05	0,81
TNOV.	0,24	0,30	0,04	0,68	0,85
TDEZ.	0,51	0,27	0,12	0,42	0,57
TJAN.	0,29	0,33	0,42	0,54	0,75
TFEV.	-0,07	0,09	0,75	0,54	0,86
POUT.	-0,05	-0,72	-0,04	-0,16	0,87
PNOV.	0,33	0,03	-0,05	-0,41	0,65
PDEZ.	-0,79	0,04	0,03	-0,14	0,71
PJAN.	-0,63	-0,19	-0,26	-0,02	0,68
PFEV.	-0,15	0,05	-0,85	-0,06	0,80
DOUT.	0,21	0,95	-0,02	0,22	0,92
DNOV.	0,17	0,13	-0,14	0,65	0,84
DDEZ.	0,57	0,51	-0,26	0,35	0,84
DJAN.	0,67	0,41	0,16	0,08	0,78
DFEV.	0,05	0,02	0,72	-0,22	0,67
Raiz característica	3,71	2,51	1,91	1,34	
Variância explicada (%)	24,70	16,80	12,80	9,00	

<sup>1</sup>A primeira letra representa o tipo de variável (T = temperatura; P = prec. pluviométrica e D = def. hídrica); as três letras seguintes representam o mês (OUT. = outubro, etc.).

<sup>2</sup>Após rotação (Varimax); método de extração: máxima verossimilhança.

Fonte: Resultados da pesquisa.

Por último, tomando-se as condições do tempo do período outubro-dezembro, também foram obtidos três fatores, que explicaram 64,2% da variância total (Tabela 7). As comunalidades, nessa especificação, ficaram entre 0,45 e 0,87, e na matriz de correlações reproduzida houve 30% de resíduos acima de 0,05 em valor absoluto. O primeiro fator teve como variáveis principais (e positivamente correlacionadas) as temperaturas de novembro e dezembro e a

deficiência hídrica de dezembro, além da pluviosidade em dezembro (negativamente correlacionada). Esse fator, representativo de deficiência de água no período de desenvolvimento vegetativo do milho, foi o único com escores afetando<sup>22</sup> (negativamente) o rendimento da cultura (Tabela 3). O segundo fator esteve relacionado às condições do tempo em outubro e o terceiro às de novembro.

Reajustando os modelos de regressão uti-

TABELA 6 - Principais Resultados da Análise Fatorial Com as Variáveis Temperatura, Precipitação Pluviométrica e Deficiência Hídrica Mensal, do Período Outubro-Janeiro, Estado de São Paulo, 1958-88

Variáveis <sup>1</sup>	Pesos dos fatores <sup>2</sup>			Comunalidades
	F1	F2	F3	
TOUT.	0,51	0,58	0,00	0,72
TNOV.	0,37	0,39	0,41	0,70
TDEZ.	0,59	0,32	0,16	0,56
TJAN.	0,42	0,39	0,21	0,61
POUT.	-0,06	-0,72	-0,22	0,80
PNOV.	0,29	-0,03	-0,57	0,46
PDEZ.	-0,80	0,01	0,06	0,65
PJAN.	-0,64	-0,19	0,07	0,61
DOUT.	0,22	0,97	0,05	0,90
DNOV.	0,25	0,21	0,94	0,69
DDEZ.	0,57	0,57	0,20	0,80
DJAN.	0,63	0,42	0,03	0,72
Raiz característica	4,02	1,22	2,00	
Varição explicada (%)	33,50	10,10	16,70	

<sup>1</sup>A primeira letra representa o tipo de variável (T = temperatura; P = prec. pluviométrica e D = def. hídrica); as três letras seguintes representam o mês (OUT. = outubro, etc.).

<sup>2</sup>Após rotação (Varimax); método de extração: máxima verossimilhança.

Fonte: Resultados da pesquisa.

lizando os escores fatoriais do período outubro-dezembro e interrompendo a série no ano anterior, pode-se "prever" os rendimentos para a cultura do milho nos anos de 1986, 1987 e 1988. Nos modelos com os escores fatoriais extraídos pelo método de componentes principais os desvios ficaram entre -5,8% e -9,4% e, com fatores extraídos pelo método alpha, entre -1,3% e -10,6%. O método alpha é superior ao primeiro, para o período 1986-88, tanto considerando-se a raiz quadrada do desvio quadrático médio (RQDQM), quanto o desvio percentual absoluto médio (DPM) e o coeficiente de desigualdade de THEIL (1966) (Tabela 8). Esse último indicador mostra que, para esse período, o modelo de regressão

empregando como variável explicativa escores fatoriais obtidos por componentes principais leva aos mesmos resultados de previsões ingênuas<sup>23</sup>.

Para o período outubro-janeiro, o mesmo procedimento resultou em desvios entre -6,1% e -22,5%, com escores fatoriais obtidos por componentes principais, entre -2,5% e -11,4% com o método alpha, e entre 0,5% e -10,8% com escores obtidos por máxima verossimilhança. O último método de extração de fatores propiciou os melhores resultados segundo os três indicadores de acurácia de previsão empregados. Os resultados desses modelos foram superiores, também considerando esses indicadores, aos de regressão múltipla utilizando deficiên-

TABELA 7 - Principais Resultados da Análise Fatorial Com as Variáveis Temperatura, Precipitação Pluviométrica e Deficiência Hídrica Mensal, do Período Outubro-Dezembro, Estado de São Paulo, 1958-88

Variáveis <sup>1</sup>	Pesos dos fatores <sup>2</sup>			Comunalidades
	F1	F2	F3	
TOUT.1	0,46	0,67	-0,17	0,61
TNOV.1	0,56	0,25	0,24	0,57
TDEZ.1	0,63	0,25	0,06	0,49
POUT.1	-0,09	-0,76	-0,18	0,73
PNOV.1	0,16	-0,01	-0,69	0,45
PDEZ.1	-0,71	-0,03	0,24	0,62
DOUT.1	0,32	0,91	0,08	0,87
DNOV.1	0,46	0,19	0,81	0,66
DDEZ.1	0,69	0,42	0,12	0,78
Raiz característica	3,66	1,24	0,88	
Varição explicada (%)	40,70	13,70	9,80	

<sup>1</sup>A primeira letra representa o tipo de variável (T = temperatura; P = prec. pluviométrica e D = def. hídrica); as três letras seguintes representam o mês (OUT. = outubro, etc.).

<sup>2</sup>Após rotação (Varimax); método de extração: alpha.

Fonte: Resultados da pesquisa.

cias hídricas do período outubro-janeiro, anteriormente ajustado por SILVA; VICENTE; CASER (1986). Na mesma ocasião em que seria possível obter essas estimativas, o IEA estaria divulgando a previsão de safras de novembro<sup>24</sup>. Para o período 1986-88 os desvios dessas previsões, em relação à estimativa final do IEA, ficaram entre -0,1% e -10,5%. Nesses anos, essas previsões foram melhores do que as dos modelos segundo o RQDQM, e piores que as do modelo com escores fatoriais extraídos por máxima verossimilhança, segundo o DPM e o U de Theil.

Com escores fatoriais do período outubro-fevereiro, extraídos por componentes principais, o modelo de regressão levou a previsões de rendimento com desvios de -3,4% a -16,1%; pelo método alpha, de -1,0% a -11,4%; e por máxima verossimilhança, de +1,2% a -9,6%. Novamente, os fatores extraídos por

máxima verossimilhança levaram, em média, aos melhores resultados, segundo os três indicadores, e o modelo de regressão empregando diretamente as deficiências hídricas desse período forneceu previsões piores do que as oriundas de análise fatorial.

Tomando-se os escores fatoriais obtidos com dados do período outubro-abril, os desvios dos modelos estimados a partir da extração por componentes principais estiveram entre -1,3% e -17,4% e, a partir do método alpha, entre 0,3% e -8,5%. O método alpha (excluídas as temperaturas médias mensais) forneceu, em média, os melhores resultados segundo os três indicadores anteriormente citados. Novamente, todos os métodos de extração de fatores levaram a estimativas melhores do que as obtidas a partir da inclusão direta de deficiências hídricas como variáveis explicativas, caso em que os desvios estive-

TABELA 8 - Previsões de Rendimentos de Milho Obtidas dos Modelos de Regressão Utilizando como Variáveis Explicativas Escores Fatoriais Associados a Variáveis Climáticas, Estado de São Paulo

Período/método	1986 <sup>1</sup>		1987 <sup>2</sup>		1988 <sup>3</sup>		RQDQM <sup>4</sup>	DPM <sup>5</sup>	U <sup>6</sup>
	Prev.	Des. (%)	Prev.	Des. (%)	Prev.	Des. (%)			
Out.-abr.									
Comp. prin. <sup>7</sup>	2.207	-8,7	2.583	-3,5	2.614	-8,8	197	7,0	0,98
Comp. prin. <sup>8</sup>	1.996	-17,4	2.641	-1,3	2.552	-11,0	304	9,9	1,63
Alpha <sup>9</sup>	2.319	-4,1	2.574	-3,8	2.623	-8,5	163	5,5	0,77
Alpha <sup>10</sup>	2.232	-7,7	2.684	0,3	2.703	-5,7	142	4,5	0,72
Regr. mul. <sup>11</sup>	1.536	-36,5	2.634	-1,6	2.730	-4,7	515	14,3	3,32
Out.-fev.									
Comp. prin. <sup>7</sup>	2.028	-16,1	2.585	-3,4	2.551	-11,0	294	10,2	1,55
Alpha <sup>9</sup>	2.351	-2,7	2.649	-1,0	2.538	-11,4	194	5,1	0,91
Máx. veros. <sup>12</sup>	2.374	-1,8	2.709	1,2	2.592	-9,6	161	4,2	0,75
Regr. mul. <sup>11</sup>	1.751	-27,6	2.663	-0,5	2.681	-6,5	399	11,5	2,40
Out.-jan.									
Comp. prin. <sup>7</sup>	1.873	-22,5	2.513	-6,1	2.513	-12,3	386	13,7	2,14
Alpha <sup>9</sup>	2.356	-2,5	2.588	-3,3	2.540	-11,4	198	5,7	0,93
Máx. veros. <sup>12</sup>	2.385	-1,3	2.691	0,5	2.556	-10,8	180	4,2	0,84
Regr. mul. <sup>11</sup>	1.712	-29,2	2.620	-2,1	2.630	-8,2	430	13,2	2,59
Out.-dez.									
Comp. prin. <sup>7</sup>	2.276	-5,8	2.484	-7,2	2.598	-9,4	207	7,5	1,00
Alpha <sup>9</sup>	2.386	-1,3	2.579	-3,7	2.562	-10,6	185	5,2	0,87

<sup>1</sup>Estimativa do rendimento obtido: 2.417 kg/ha.

<sup>2</sup>Estimativa do rendimento obtido: 2.677 kg/ha.

<sup>3</sup>Estimativa do rendimento obtido: 2.866 kg/ha.

<sup>4</sup>Raiz quadrada do desvio quadrático médio.

<sup>5</sup>Desvio percentual médio absoluto.

<sup>6</sup>Coefficiente de desigualdade de Theil.

<sup>7</sup>Fatores extraídos pelo método de componentes principais.

<sup>8</sup>Fatores extraídos pelo método de componentes principais, excluindo as temperaturas médias.

<sup>9</sup>Fatores extraídos pelo método alpha.

<sup>10</sup>Fatores extraídos pelo método alpha, excluindo as temperaturas médias.

<sup>11</sup>Modelo de regressão múltipla com deficiências hídricas mensais como variáveis explicativas.

<sup>12</sup>Fatores extraídos pelo método de máxima verossimilhança.

Fonte: Resultados da pesquisa.

ram entre -1,6% e -36,5%. A previsão de safras efetuada pelo IEA em fevereiro (que, como explicado anteriormente, estaria disponível no final de abril-início de maio), nesses mesmos anos, apresentou desvios entre 1,7% e -5,4%, melhores do que as dos modelos, de acordo com os três indicadores.

### 3.2 - Resultados da Análise de Conglomerados

Os resultados obtidos somente com as variáveis originalmente propostas não foram satisfatórios, com os grupos formados nem de longe sendo capazes de dividir as observações em consonância com alguma ordenação lógica dos rendimentos. Por esse motivo, foram incorporadas à análise mais duas variáveis. A primeira, representativa do estoque de progresso tecnológico, foi o número de artigos científicos publicados sobre tecnologia agrícola, referentes à cultura do milho e ao Estado de São Paulo. O período de influência desses estudos sobre o rendimento da cultura foi fixado, com base em trabalhos anteriores, entre o quinto e o vigésimo anos posteriores à publicação<sup>25</sup>. A segunda variável adicional foi o próprio rendimento estimado por tendência, que deve representar o fluxo de tecnologias e condições conjunturais a que a cultura esteve exposta no período em estudo.

Com essas duas variáveis, alguns resultados melhoraram, tanto com o método do enlace médio, quanto com o do enlace completo, ambos empregando todas as variáveis na forma original<sup>26</sup> (não padronizadas) (Tabela 9).

Para o período outubro-abril, os grupos obtidos foram idênticos com ou sem as temperaturas mensais. Adotando-se como critérios de seleção o menor número de grupos inteiramente contidos dentro de outros, e o menor intervalo de rendimentos, o método de formação de conglomerados mais adequado para os períodos outubro-abril, outubro-fevereiro e outubro-dezembro foi o do enlace completo. O método do enlace médio foi superior apenas para o período outubro-janeiro.

Foram efetuadas reclassificações com os dados cobrindo desde 1985 até os anos de 1986 e 1987, para obter estimativas sobre os rendimentos desses anos.

Dessa forma, para o ano agrícola 1985/86,

no começo de janeiro de 1986 teria sido possível obter uma previsão dos rendimentos desse ano, com os dados do período outubro-dezembro. Tanto pelo método do enlace médio como pelo do enlace completo, concluir-se-ia pela associação desse ano no grupo dos maiores rendimentos, variando de 2.378 a 2.709 kg/ha, o que efetivamente ocorreu, já que finda a colheita, os levantamentos de campo indicaram rendimento médio de 2.417 kg/ha (portanto, uma diferença de -5,6% em relação à média dos rendimentos dos elementos do grupo).

No começo de fevereiro, com os dados do período outubro-janeiro, esse tipo de análise levaria a concluir que os rendimentos seriam próximos a 2.570 kg/ha (diferença de 6,3% em relação à estimativa final da safra). Os dados do período outubro-fevereiro, com o método do enlace médio, levaram ao mesmo resultado do período outubro-dezembro. Já com o enlace completo, apontaram para o intervalo 2.378 a 2.570 kg/ha (com um erro de 3,4%, comparando-se a média do grupo com a estimativa final do rendimento do ano). Os do período outubro-abril, com ambos os métodos, repetiram esse último intervalo. Portanto, para o ano agrícola 1985/86 todas essas estimativas seriam adequadas para subsidiar o serviço de previsão de safras, com as obtidas com os dados dos períodos outubro-dezembro e outubro-abril sendo as mais próximas dos resultados estimados com o levantamento de campo efetuado em junho (estimativa final).

Para o ano agrícola subsequente (1986/87), pelos dois métodos e com dados do período outubro-dezembro, o rendimento também seria enquadrado no intervalo 2.570 a 2.709 kg/ha, de novo efetivamente contendo a estimativa final do rendimento médio efetuada em junho de 1987 (2.677 kg/ha, diferindo 5,7% da média do grupo). Com os dados do período outubro-janeiro, pelo enlace médio obteve-se o mesmo intervalo; já com o enlace completo, o grupo em que 1987 foi enquadrado, variou entre 2.417 e 2.709 kg/ha, com a diferença do ponto médio em relação à estimativa final da safra diminuindo para 4,3%.

Os dados do período outubro-fevereiro alteraram apenas os resultados obtidos do enlace completo, levando o intervalo para 2.549 a 2.709 kg/ha (com a diferença do ponto médio com a estimativa final caindo para 2,5%). No mês de maio,

TABELA 9 - Grupos de Rendimento do Milho Formados por Análise de Conglomerados Utilizando as Variáveis Temperaturas Médias, Precipitações Pluviométricas, Deficiências Hídricas, Artigos Científicos Publicados e Rendimentos do Milho Estimados por Tendência, Estado de São Paulo, 1958-88<sup>1</sup>

Outubro-abril		Outubro-fevereiro		Outubro-janeiro		Outubro-dezembro	
Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo	Mínimo	Máximo
Método do enlace médio							
1.121	1.663	1.121	1.121	1.121	1.663	1.121	1.709
1.396	2.000	1.200	1.663	1.396	2.037	1.396	2.000
1.709	1.821*	1.396	2.000	1.709	1.821*	1.750	2.340
1.750	2.709	1.621	1.821*	1.750	2.340	1.753	1.821*
1.899	2.222*	1.750	2.340	2.378	2.866	2.378	2.866
2.378	2.866	2.378	2.866				
Método do enlace completo							
1.121	1.753	1.121	1.753	1.121	1.663	1.121	1.663
1.200	1.663*	1.200	1.663*	1.396	1.910	1.396	2.000
1.396	2.000	1.396	2.000	1.709	1.821*	1.488	1.821*
1.750	2.340	1.750	2.340	1.750	2.340	1.750	2.340
2.378	2.866	2.378	2.866	1.998	2.222*	2.378	2.866
				2.378	2.866		

<sup>1</sup>Os grupos assinalados com asterisco estão inteiramente contidos no anterior.

Fonte: Resultados da pesquisa.

com os dados do período outubro-abril, o enlace completo manteria os resultados anteriores, enquanto o enlace médio levaria ao intervalo 2.378 a 2.570, pior do que os obtidos anteriormente, com diferença de 7,4% em relação à estimativa final.

No ano agrícola 1987/88, o rendimento do milho foi superior aos obtidos até então (2.866 kg/ha); portanto, impossibilitando a obtenção de grupos cujos intervalos de variação contivessem a marca efetivamente alcançada. O resultado proveniente dos dois métodos, nos quatro períodos, foi o mesmo: 2.378 a 2.709 kg/ha, com a média do grupo sendo 11% inferior à estimativa final. Seria possível com essa metodologia, ao menos, enquadrar o rendimento do ano de 1988 entre os maiores da série. Os indicadores de

acurácia das previsões utilizados para comparar os resultados da análise fatorial conjugados à análise de regressão, não são diretamente aplicáveis no caso da análise de conglomerados, onde não foi obtido um único ponto, mas um intervalo. Com essa ressalva, tomando-se os pontos médios dos grupos em que os rendimentos do período foram enquadrados, chega-se a 221 e 192 como valores das raízes quadradas dos desvios quadráticos médios (RQDQM), respectivamente para os métodos do enlace médio e enlace completo, com os dados do período outubro-abril. Nesse mesmo período, os desvios percentuais absolutos médios (DPM) foram 7,3% e 5,6% e os valores do coeficiente de desigualdade de Theil (U) foram 1,05 e 0,90, os primeiros associados ao enlace médio e os

segundos ao enlace completo. Com os dados do período outubro-fevereiro, para os dois métodos, os indicadores tiveram os valores de: 217 e 192 (RQDQM), 7,4% e 5,6% (DPM) e 1,03 e 0,90 (U). Os valores para outubro-janeiro foram: 221 e 213 (RQDQM), 7,7% e 7,2% (DPM) e, 1,05 e 1,01 (U). Por último, para outubro-dezembro, obtiveram-se os mesmos valores para os dois métodos: RQDQM de 217, DPM de 7,4% e U igual a 1,03. Por esses indicadores, é possível concluir que, em relação aos rendimentos médios dos grupos, os melhores resultados para o período 1986-88 foram os do enlace completo, com dados de outubro-fevereiro e outubro-abril.

### 3.3 - Resultados da Análise Discriminante

A análise discriminante é classificada como um método de dependência (DILLON & GOLDSTEIN, 1984). Por isso, as variáveis artigos científicos publicados e rendimentos estimados por tendência foram mantidas entre as independentes.

Inicialmente, com a série completa (1958-88), procurou-se verificar se os três ou cinco grupos formados *a priori* estariam estatisticamente separados. Com as temperaturas médias, precipitações pluviométricas e deficiências hídricas do período outubro-abril, os três métodos de seleção de variáveis - entrada forçada, lambda de Wilks e distância de Mahalanobis - indicaram 100% de casos corretamente classificados, tanto com três como com cinco grupos. Eliminando-se as temperaturas médias, apenas o método da entrada forçada repetiu esse resultado, considerando-se três grupos; os demais métodos, também com três grupos, e todos eles, no caso de cinco grupos, apontaram 93,66% de casos corretamente classificados.

Para o período outubro-fevereiro, com três grupos, todos os métodos resultaram em 100% de classificação correta. Com cinco grupos, e com a entrada forçada o percentual foi 96,77% e, de 93,44% com o lambda de Wilks e a distância de Mahalanobis.

Para três grupos, os três métodos apontaram para 100% de classificações corretas com dados do período outubro-janeiro. O método da entrada direta indicou 96,77% de casos corretamente classificados, com cinco grupos; os outros dois métodos resultaram em

90,32%.

Os dados do período outubro-dezembro possibilitaram 96,77% de classificação correta em três grupos, segundo os três métodos. Já com cinco grupos, a entrada forçada indicou 87,10% e os outros dois, 83,87%.

Portanto, apesar dos percentuais de classificações corretas serem sempre elevados, a análise discriminante aparentemente funcionou melhor trabalhando com três grupos do que com cinco.

Isso ficou confirmado na análise classificatória efetuada em seguida. Com a série 1958-85 distribuída em três grupos, o ano de 1986 foi alocado corretamente pelos três métodos no grupo de rendimentos altos, nas análises com dados dos períodos outubro-dezembro e outubro-janeiro. A distância de Mahalanobis também classificou esse ano corretamente com dados do período outubro-fevereiro. Com o lambda de Wilks, 1986 foi agrupado junto aos de baixos rendimentos e, pelo método da entrada forçada, entre os de rendimentos médios. Com os dados do período outubro-abril (tanto com as temperaturas médias quanto sem elas) a distância de Mahalanobis e o lambda de Wilks levaram a enquadrar 1986 no grupo de rendimentos médios, enquanto que a entrada forçada enquadrou-o no de baixos rendimentos.

O ano de 1987 só não foi classificado no grupo de altos rendimentos, pelos três métodos, com os dados do período outubro-abril excluídas as temperaturas médias. Já 1988 foi corretamente classificado no grupo de rendimentos altos em todos os períodos por todos os métodos.

Por esses resultados pode-se considerar ligeira superioridade da distância de Mahalanobis sobre os outros métodos empregados.

Se a análise classificatória com três grupos levou a resultados corretos em quase todos os casos, trabalhando-se com cinco grupos isso não ocorreu. O ano de 1986 somente foi classificado em seu grupo correto (o de número 4, rendimentos altos) com os dados do período outubro-dezembro e empregando o lambda de Wilks e a distância de Mahalanobis. Nas demais análises, esse ano foi alocado duas vezes no grupo 5 (rendimentos muito altos), duas vezes no grupo 3 (rendimentos médios), seis vezes no grupo 2 (rendimentos baixos) e três vezes no grupo 1 (rendimentos muito baixos). No ano seguinte os resultados

também foram ruins: nenhuma vez 1987 foi classificado em seu grupo correto (grupo 5). Foi enquadrado nove vezes no grupo 4 e seis vezes no grupo 3. Em compensação, 1988 foi classificado onze vezes no grupo 5, e quatro vezes, incorretamente, no grupo 4<sup>27</sup>.

Devido a esses maus resultados decidiu-se empregar os escores das funções discriminantes (oriundos da pré-classificação em cinco grupos) como variáveis explicativas dos rendimentos em modelos de regressão linear múltipla, a exemplo do efetuado anteriormente com os escores fatoriais.

Com os dados do período outubro-abril, inclusive temperaturas médias, os escores das três primeiras funções discriminantes, obtidas pelo método da entrada forçada<sup>28</sup>, apareceram no modelo de regressão afetando positivamente os rendimentos do milho (Tabela 10). A associação dessas funções às variáveis originais não foi tão clara como no caso dos escores fatoriais; entretanto, após a rotação, a primeira função discriminante apresentou correlações positivas mais elevadas com as variáveis artigos publicados e rendimentos estimados por tendência, podendo então, ser associada ao estoque e fluxo de conhecimentos técnicos. A segunda função esteve negativamente associada às temperaturas de outubro, novembro e janeiro, e às deficiências hídricas de outubro, janeiro, fevereiro e abril, e positivamente à pluviosidade de novembro e fevereiro; portanto, associada à disponibilidade de água em várias fases do ciclo da cultura. Por último, a terceira função discriminante apareceu positivamente relacionada aos artigos publicados, à estimativa do rendimento por tendência e às precipitações pluviométricas de novembro, fevereiro e março; as correlações negativas mais elevadas foram com as temperaturas e as deficiências hídricas de dezembro, associando essa função a um *mix* das duas anteriores.

Excluindo-se as temperaturas médias, com dados do mesmo período outubro-abril e o método da minimização do lambda de Wilks, novamente os escores das três primeiras funções discriminantes resultaram estatisticamente significativos (Tabela 10). A primeira função teve correlações positivas com artigos publicados, rendimento estimado por tendência e chuvas em dezembro. A segunda, correlações negativas com chuvas em dezembro e positivas com deficiências em novembro, dezembro, janeiro e abril. A terceira função exibiu correlações negativas, prin-

cipalmente, com as deficiências em novembro e janeiro, e positivas com a deficiência hídrica em março.

A primeira função discriminante construída através da maximização da distância de Mahalanobis e os dados do período outubro-fevereiro, foi a única significativa no modelo de regressão correspondente (Tabela 10), e as variáveis a ela positivamente associadas foram os artigos publicados, rendimento estimado por tendência, precipitação pluviométrica em dezembro e deficiência em janeiro, enquanto que a temperatura média em novembro e as deficiências de novembro e dezembro formam as variáveis com correlações negativas mais elevadas.

Com os dados do período outubro-janeiro e o método do lambda de Wilks, somente os escores da primeira função discriminante resultaram em coeficiente significativo (Tabela 10). As variáveis positivamente associadas a essa função foram artigos científicos publicados, rendimento estimado por tendência, precipitação pluviométrica em dezembro e temperatura média de janeiro; temperaturas e deficiências de novembro e deficiências de dezembro exibiram as maiores correlações negativas.

O modelo com dados do período outubro-dezembro, oriundos de escores de funções discriminantes obtidas pelo método da distância de Mahalanobis, apresentou coeficiente significativo para a primeira função (Tabela 10). As variáveis positivamente correlacionadas à mesma foram os artigos publicados, o rendimento estimado por tendência e as chuvas em novembro; as correlações negativas mais pronunciadas foram as associadas às deficiências em novembro e dezembro.

Similarmente ao efetuado com os modelos empregando escores fatoriais como variáveis explicativas, os modelos de regressão foram reestimados interrompendo a série no ano anterior àquele em que se desejava "prever" o rendimento. As previsões passo a passo assim obtidas, para o período 1986-88, a partir de funções discriminantes do período outubro-dezembro, tiveram erros entre -4,4% e -10,0% (Tabela 11). O lambda de Wilks e a distância de Mahalanobis tiveram desempenhos iguais segundo os três indicadores utilizados. Também com escores de funções com dados de outubro-janeiro, cujos desvios estiveram entre -0,7% e -7,6%, e do período outo-

TABELA 10 - Principais Resultados da Análise de Regressão Utilizando Escores de Funções Discriminantes Provenientes das Condições Climáticas de Vários Períodos<sup>1</sup>, Artigos Científicos Publicados e Rendimentos Estimados por Tendência como Variáveis Explicativas do Rendimento do Milho, Estado de São Paulo, 1958-88

Variável <sup>2</sup>	Período				
	Out.-abr. <sup>3</sup>	Out.-abr. <sup>4</sup>	Out.-fev. <sup>5</sup>	Out.-jan. <sup>6</sup>	Out.-dez. <sup>7</sup>
Intercepto	930,53 b	1.245,65 a	939,62 b	1.325,54 b	742,70 d
Tendência	14,15 a	9,83 c	14,02 b	8,73 d	16,72 b
Função 1	26,81 a	84,04 a	106,62 a	126,76 a	109,78 a
Função 2	22,01 a	-43,56 a			
Função 3	49,71 a	-31,39 b			
R <sup>2</sup>	0,95	0,96	0,93	0,93	0,90
F <sup>2</sup>	126,73 a	149,41 a	185,15 a	191,07 a	132,85 a
Durbin-Watson	2,08	1,62	2,05	2,05	2,13

<sup>1</sup>Temperaturas médias, precipitações pluviométricas e deficiências hídricas mensais.

<sup>2</sup>Níveis de significância: a = 1%; b = 5%; c = 10%; d = 20%.

<sup>3</sup>Método de seleção de variáveis (para inclusão na função discriminante): entrada forçada.

<sup>4</sup>Método de seleção de variáveis: lambda de Wilks, sem incluir temperaturas médias.

<sup>5</sup>Método de seleção de variáveis: distância de Mahalanobis.

<sup>6</sup>Método de seleção de variáveis: lambda de Wilks.

<sup>7</sup>Método de seleção de variáveis: distância de Mahalanobis.

Fonte: Resultados da pesquisa.

bro-fevereiro (com desvios de -0,2% a -8,6%) esses dois métodos levaram aos mesmos resultados.

Com variáveis do período outubro-abril, cujas previsões tiveram desvios de -0,5% a -11,4%, o coeficiente de desigualdade de Theil e a raiz quadrada do desvio quadrático médio indicaram o método do lambda de Wilks, incluindo temperaturas, como o mais adequado<sup>29</sup>, enquanto que o menor desvio percentual absoluto médio foi fornecido tanto pelo lambda de Wilks, quanto pela distância de Mahalanobis, ambos sem empregar as variáveis temperaturas médias mensais.

#### 4 - CONCLUSÕES E CONSIDERAÇÕES FINAIS

Os modelos multivariados utilizados revelaram-se capazes de prever adequadamente os rendimentos da cultura do milho no Estado de São Paulo.

Por análise discriminante, geralmente foi possível classificar antecipada e corretamente os rendimentos do período 1986-88, utilizando três grupos definidos *a priori* (rendimentos baixos, médios e altos). Esses mesmos anos normalmente foram agrupados, por análise de conglomerados, junto a anos de níveis similares.

Previsões pontuais passo a passo empregando escores fatoriais e escores de funções discriminantes foram, no mais das vezes, bastante próximas das estimativas finais efetuadas pelo IEA.

As melhores previsões com dados climáticos do período outubro-dezembro, segundo todos os indicadores utilizados, foram as efetuadas com escores fatoriais extraídos pelo método alpha. Com dados do período outubro-janeiro, os escores das funções discriminantes ajustados pelos métodos da minimização do lambda de Wilks e pela maximização da distância de Mahalanobis levaram às melhores

TABELA 11 - Previsões de Rendimentos de Milho Obtidas dos Modelos de Regressão Utilizando como Variáveis Explicativas Escores de Funções Discriminantes Associadas a Variáveis Climáticas<sup>1</sup>, Estado de São Paulo

Período/método <sup>2</sup>	1986 <sup>3</sup>		1987 <sup>4</sup>		1988 <sup>5</sup>		RQDQM <sup>6</sup>	DPM <sup>7</sup>	U <sup>8</sup>
	Prev.	Des. (%)	Prev.	Des. (%)	Prev.	Des. (%)			
<b>Outubro-abril</b>									
Entrada forçada	2.331	-3,6	2.652	-0,9	2.540	-11,4	195	5,3	0,92
Entrada forçada <sup>9</sup>	2.354	-2,6	2.513	-6,1	2.691	-6,1	143	4,9	0,68
Lambda de Wilks	2.320	-4,0	2.598	-3,0	2.653	-7,4	143	4,8	0,67
Lambda de Wilks <sup>9</sup>	2.405	-0,5	2.499	-6,6	2.690	-6,1	145	4,4	0,68
Dist. Mahalanobis	2.396	-0,9	2.459	-8,1	2.618	-8,7	191	5,9	0,91
Dist. Mahalanobis <sup>9</sup>	2.405	-0,5	2.499	-6,6	2.690	-6,1	145	4,4	0,68
<b>Outubro-fevereiro</b>									
Entrada forçada	2.225	-7,9	2.619	-2,2	2.619	-8,6	184	6,2	0,91
Lambda de Wilks	2.291	-5,2	2.672	-0,2	2.665	-7,0	137	4,1	0,66
Dist. Mahalanobis	2.291	-5,2	2.672	-0,2	2.665	-7,0	137	4,1	0,66
<b>Outubro-janeiro</b>									
Entrada forçada	2.253	-6,8	2.657	-0,7	2.649	-7,6	157	5,0	0,77
Lambda de Wilks	2.304	-4,7	2.657	-0,7	2.669	-6,9	132	4,1	0,63
Dist. Mahalanobis	2.304	-4,7	2.657	-0,7	2.669	-6,9	132	4,1	0,63
<b>Outubro-dezembro</b>									
Entrada forçada	2.274	-5,9	2.548	-4,8	2.578	-10,0	200	6,9	0,96
Lambda de Wilks	2.287	-5,4	2.558	-4,4	2.578	-10,0	195	6,6	0,93
Dist. Mahalanobis	2.287	-5,4	2.558	-4,4	2.578	-10,0	195	6,6	0,93

<sup>1</sup>Temperaturas médias, precipitações pluviométricas e deficiências hídricas mensais.

<sup>2</sup>Método de seleção de variáveis para inclusão na função discriminante.

<sup>3</sup>Estimativa do rendimento obtido: 2.417 kg/ha.

<sup>4</sup>Estimativa do rendimento obtido: 2.677 kg/ha.

<sup>5</sup>Estimativa do rendimento obtido: 2.866 kg/ha.

<sup>6</sup>Raiz quadrada do desvio quadrático médio.

<sup>7</sup>Desvio percentual médio absoluto.

<sup>8</sup>Coefficiente de desigualdade de Theil.

<sup>9</sup>Não inclui temperaturas médias mensais.

Fonte: Resultados da pesquisa.

previsões, o que se repetiu com dados do período outubro-fevereiro. Segundo o coeficiente de desigualdade de Theil, as melhores previsões com dados do período outubro-abril foram as obtidas utilizando escores discriminantes obtidos pela minimização do lambda de Wilks; já segundo a raiz quadrada do desvio quadrático médio e o desvio percentual absoluto médio, as previsões provenientes de escores fatoriais extraídos pelo método alpha foram superiores.

O serviço de previsão de safras do Estado de São Paulo, dispondo de levantamentos periódicos baseados em amostras probabilísticas, pode ser considerado o melhor do Brasil; infelizmente tem sofrido os efeitos das dificuldades que vêm se abatendo sobre o País, em particular sobre o setor público. Amostras sorteadas com base em cadastros desatualizados e/ou levantados inadequadamente, atividades como levantamentos de campo e análises de dados comprometidas por pessoal insuficiente e desmotivado devido a questões salariais, etc.

Portanto, estudos com modelagens de área e rendimento, além da importância intrínseca como

subsídios aos levantamentos de campo, tornam-se essenciais para complementar essas atividades. No resto do País, onde geralmente esse serviço é bem mais precário, modelos adequados poderiam trazer maior credibilidade às estimativas de safras, diminuindo o grau de incerteza dos agentes econômicos ligados à agricultura e aprimorando os dados disponíveis para a elaboração de políticas dirigidas para o meio rural. Essas colocações valem mesmo para os estados em que o IBGE vem efetuando levantamentos com amostras baseadas em sistemas referenciais provenientes de imagens de satélite. Tais levantamentos são efetuados apenas uma vez por ano, e os rendimentos não são objetivamente medidos, mas informados pelos produtores. Obviamente, têm que ser complementados por várias outras estimativas durante o ano, que continuam sendo feitas de forma subjetiva e precária.

Os resultados obtidos neste estudo parecem ser bons o suficiente para incentivarem a inclusão dos métodos multivariados utilizados em esforços futuros destinados à previsão de rendimentos.

## NOTAS

<sup>1</sup>Trabalho referente ao projeto SPTC 16-027/89. Estudo iniciado durante estágio no International Food Policy Research Institute (IFPRI), de junho a julho de 1989, sob os auspícios do EMBRAPA/IFPRI Collaborative Research Contract. O autor agradece a colaboração de Stephen A. Vosti, do IFPRI, Luiz J. M. Irias, da EMBRAPA, e Adolpho W. P. Canton, da FEA/USP. Recebido em 05/01/94. Liberado para publicação em 16/03/94.

<sup>2</sup>Engenheiro Agrônomo, MS, Pesquisador Científico do Instituto de Economia Agrícola (IEA).

<sup>3</sup>No Estado de São Paulo são pioneiros os estudos de BRANDT (1966) e de TOYAMA & PESCARIN (1970).

<sup>4</sup>Como exemplos de trabalhos nessa área, com esse tipo de metodologia, podem ser citados: TOSELLO & ARRUDA (1962), INSTITUTO DE PLANEJAMENTO ECONÔMICO E SOCIAL (1972), JUNQUEIRA (1973), OMETTO (1974), MOTA & SILVA (1980), SANTOS (1983), WADSTED (1983), CASER & VICENTE (1984), CAMARGO et alii (1984) e SILVA; VICENTE; CASER (1986).

<sup>5</sup>Discussão detalhada sobre determinantes da produtividade da terra encontra-se em SILVA; VICENTE; CASER (1986).

<sup>6</sup>Variáveis referentes ao mês de maio foram eliminadas em análises preliminares.

<sup>7</sup>O balanço hídrico consiste em contabilizar a água no solo, considerando-se a chuva como o abastecimento e a evapotranspiração como o consumo, assumindo-se uma capacidade de armazenamento de água no solo capaz de ser utilizada pelas plantas.

<sup>8</sup>O método alpha, menos utilizado, considera as variáveis de uma particular análise como uma amostra de um universo de variáveis em potencial. Difere dos outros métodos citados, que consideram as observações como uma amostra de certa população e as variáveis como sendo fixas (NORUSIS, 1988).

<sup>9</sup>Análise fatorial e análise de componentes principais são freqüentemente empregadas associadas a modelos de regressão, em especial quando existe

multicolinearidade elevada (MADDALA, 1977 e JUDGE et alii, 1988).

<sup>10</sup>Também foram experimentados, na fase inicial, o método de Ward (que é um método de variância, buscando a cada passo o menor total para a soma de quadrados das distâncias nos conglomerados) e o método do centróide, em que a distância entre dois conglomerados é a distância entre as médias para todas as suas variáveis. Como os resultados obtidos por esses dois métodos não foram melhores que os dos anteriores, eles foram abandonados na segunda etapa da análise, em que novas variáveis foram necessárias, conforme discutido no próximo item.

<sup>11</sup>Entretanto, tanto as hipóteses como os objetivos dos dois tipos de análise são diferentes; ver, por exemplo, DILLON & GOLDSTEIN (1984).

<sup>12</sup>Os limites dos grupos foram definidos dividindo-se a diferença entre o maior e o menor rendimentos da série por três e por cinco, respectivamente.

<sup>13</sup>O lambda de Wilks representa a proporção da variância total dos escores discriminantes não explicada pelas diferenças entre os grupos.

<sup>14</sup>Formalmente,

$$D_{ij}^2 = (X_i - X_j)' S^{-1} (X_i - X_j)$$

onde  $X_i$  e  $X_j$  são vetores constituídos pelos centróides dos grupos  $i$  e  $j$ , e  $S$  é a matriz de covariâncias (ou de correlações, caso as variáveis estejam padronizadas).

<sup>15</sup>Os testes empregados foram os disponíveis na versão 3.0 do SPSS (ver NORUSIS, 1988).

<sup>16</sup>Formalmente,

$$\chi^2 = -[(n-1) - \frac{1}{6}(2p+1 + \frac{2}{p})][\ln |S| + p \ln(\frac{1}{p} \sum_{j=1}^p l_{(j)})]$$

onde  $n$  é o número de observações,  $p$  o número de variáveis,  $S$  é a matriz de covariâncias e  $l_{(j)}$  é a  $j$ -ésima raiz característica de  $S$  (DILLON & GOLDSTEIN, 1984).

<sup>17</sup>Formalmente,

$$KMO = \frac{\sum \sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum \sum_{i \neq j} a_{ij}^2}$$

onde  $r_{ij}$  é o coeficiente de correlação simples e  $a_{ij}$  é o coeficiente de correlação parcial entre as variáveis  $i$  e  $j$  (NORUSIS, 1988).

<sup>18</sup>Para cada período será apresentado o método de extração que forneceu, como será visto mais à frente, as melhores previsões de rendimentos para os anos de 1986, 1987 e 1988 (em média).

<sup>19</sup>Proporção da variância total da variável descrita pelos fatores comuns (DILLON & GOLDSTEIN, 1984).

<sup>20</sup>É possível mensurar os efeitos das variáveis individuais multiplicando-se os coeficientes angulares pelos pesos fatoriais associados a cada variável.

<sup>21</sup>Pelo método de máxima verossimilhança, a inexistência de mínimos locais durante as iterações para extração de fatores impediu a obtenção de valores, tanto para esse período como para o período outubro-dezembro.

<sup>22</sup>Observe-se, todavia, que o coeficiente estimado não é significativo a 10%.

<sup>23</sup>Considerar que o rendimento em determinado ano seria igual ao obtido no ano anterior; nesse caso,  $U$  seria igual a 1 (THEIL, 1966).

<sup>24</sup>Geralmente os dados são levantados, no campo, até o final de novembro. É necessário um período de mais ou menos dois meses para envio, recepção e seleção preliminar de questionários, digitação de dados, depurações, processamento e análise final.

<sup>25</sup>Em SILVA; VICENTE; CASER (1993) são apresentados detalhes sobre a construção dessa variável e comentados vários outros estudos com variáveis representativas do progresso técnico.

<sup>26</sup>Os modelos com dados padronizados e os provenientes do método de formação de conglomerados, denominado enlace simples, não formaram grupos consistentes: vários estavam contidos dentro de outros e muitos deles eram constituídos por apenas um elemento. Por esse motivo, foram abandonados.

<sup>27</sup>Com o lambda de Wilks e dados do período outubro-abril sem temperaturas e do período outubro-janeiro; pela distância de Mahalanobis com dados de outubro-abril sem temperaturas; e pela entrada forçada com o período outubro-abril (incluindo temperaturas médias).

<sup>28</sup>Para não alongar a discussão serão apresentados somente os modelos que propiciaram as melhores previsões, como será visto em seguida.

<sup>29</sup>Pela raiz quadrada do desvio quadrático médio, os resultados da entrada forçada sem as temperaturas médias, foram tão bons quanto os do método citado.

### LITERATURA CITADA

- BRANDT, Sérgio A. Estimativas de ofertas de produtos agrícolas no Estado de São Paulo. In: CONGRESSO BRASILEIRO DE ECONOMIA E SOCIOLOGIA RURAL, 4, 1966. **Anais...** São Paulo, SOBER, 1966. p.323-348.
- CAMARGO, Marcelo B. P. et alii. Relações entre a precipitação pluviométrica e a produtividade do cafeeiro. **Ecosistemas**, SP, 9:165-171, 1984.
- CASER, Denise V. & VICENTE, José R. **Projeções de áreas e previsões de rendimentos de feijão e de arroz nas DIRAs de Sorocaba e São José do Rio Preto**. São Paulo, IEA, 1984. 28 p. (Relatório de Pesquisa 15/84).
- DILLON, W. R. & GOLDSTEIN, Morris. **Multivariate analysis: methods and applications**. New York, John Wiley & Sons, 1984. 587 p.
- FANCELLI, Antonio L. & LIMA, Urgel A. **Mi-lho: produção, pré-processamento e transformação agroindustrial**. São Paulo, Secretaria da Indústria, Comércio, Ciência e Tecnologia, 1982. 112 p.
- FUNDAÇÃO GETÚLIO VARGAS. **Distribuição e sazonalidade da produção agrícola**. Rio de Janeiro, FGV, 1980. 112 p.
- INSTITUTO DE PLANEJAMENTO ECONÔMICO E SOCIAL. **Variações climáticas e flutuações de oferta agrícola no Centro-Sul do Brasil: relatório de pesquisa**. Brasília, IPEA/IPLAN, 1972. 419 p.
- JOHNSON, Richard A. & WICHERN, Dean W. **Applied multivariate statistical analysis**. Englewood Cliffs, Prentice-Hall, 1982. 594 p.
- JUDGE, George G. et alii. **Introduction to the theory and practice of econometrics**. New York, John Wiley & Sons, 1988.
- JUNQUEIRA, Antonio A. B. **Efeitos da precipitação pluviométrica e da modificação tecnológica sobre a produtividade física de quatro culturas na Região de Ribeirão Preto, 1947 a 1967**. Jaboticabal, Faculdade de Medicina Veterinária e Agronomia, 1973. 47 p. mimeo.
- MADDALA, G. S. **Econometrics**. New York, McGraw-Hill, 1977.
- MOTA, F. S. & SILVA, J. B. A weather-technology model for rice in southern Brazil. In: SYMPOSIUM ON THE AGROMETEOROLOGY OF THE RICE CROP, Manila. **Proceedings...** Manila, 1980. 254 p.
- NEGRI NETO, Afonso et alii. Custo e benefício

- social de previsões e estimativas de produção agrícola: o valor da informação. **Agricultura em São Paulo**, SP, **35**(T. único):37-50, 1988.
- NORUSIS, M. J. **SPSS/PC+ Advanced statistics v.3.0**. Chicago, SPSS, 1988.
- OMETTO, José C. **Uma equação para a estimativa de evapotranspiração potencial e suas aplicações no cálculo das necessidades hídricas e do rendimento agroindustrial da cana-de-açúcar na Região de Piracicaba, SP**. Piracicaba, ESALQ/USP, 1974. (Tese de Livre Docência).
- ORTOLANI, Altino A. et alii. **Parâmetros climáticos e a cafeicultura**. Rio de Janeiro, IBC, 1970. 27p.
- SANTOS, Maria J. Z. Análise econométrica da produção canavieira paulista. **Revista de Geografia**, **8**(15/16):75-104, 1983.
- SILVA, Gabriel L. S. P.; VICENTE, José R.; CASER, Denise V. Mudança tecnológica e rendimento do milho e da soja na região Centro-Sul. **Revista Brasileira de Economia**, **47**(2): 281-303, abr./jun. 1993.
- SILVA, Gabriel L. S. P.; VICENTE, José R.; CASER, Denise V. **Variações do tempo e produtividade agrícola: um subsídio à previsão de safras no Estado de São Paulo**. Campinas, Fundação Cargill, 1986. 148p.
- THEIL, Henri. **Applied economic forecast**. Amsterdam, North-Holland, 1966. 474 p.
- TOSELLO, Rino N. & ARRUDA, Hermano V. Correlação entre estimativas oficiais de produção de café e precipitação pluviométrica no Estado de São Paulo. **Bragantia**, Campinas, **21**(1):449-465, 1962.
- TOYAMA, Nelson K. & PESCARIN, Rosa M. C. Projeções da oferta agrícola do Estado de São Paulo. **Agricultura em São Paulo**, SP, **17**(9/10):1-97, set./out. 1970.
- VANDAELE, Walter. **Applied time series and Box-Jenkins models**. Orlando, Academic Press, 1983. 417p.
- WADSTED, Otto G. O clima e a economia: análise de algumas culturas no Estado de São Paulo. **Revista Brasileira de Economia**, RJ, **37**(2):225-244, abr./jun. 1983.