

ANÁLISE FATORIAL E ANÁLISE DE ENVOLTÓRIA DE DADOS NA CONSTRUÇÃO DE INDICADORES E DETERMINAÇÃO DE PESOS EM MODELOS DE PRODUÇÃO COM RESPOSTAS MÚLTIPLAS

Geraldo da Silva e Souza

Eliane Gonçalves Gomes

Empresa Brasileira de Pesquisa Agropecuária – Embrapa
Parque Estação Bilógica, Av. W3 Norte final, 70770-901, Brasília, DF, Brasil
gerald.souza@embrapa.br
eliane.gomes@embrapa.br

RESUMO

Faz-se uso da análise fatorial e da análise de envoltória de dados (DEA) para definir uma medida de desempenho para aos centros de pesquisa da Empresa Brasileira de Pesquisa Agropecuária. Respostas múltiplas são agrupadas em três categorias de produção e suas ordenações são estudadas, via análise fatorial, dentro de cada categoria. Um sistema de pesos para cada categoria é definido de forma independente de rotações fatoriais, o que conduz a um conjunto de escores de produção de dimensão três. As medidas de desempenho têm por base um modelo de produção DEA-CCR com três produtos e três insumos. É também proposto um índice de alcance de metas gerenciais para as dimensões de produção.

Palavras-chave: Análise fatorial, Análise de envoltória de dados, Alcance de metas, Desempenho.

ABSTRACT

In this paper we used factor analysis and data envelopment analysis (DEA) to define a performance measure for the research centers of the Brazilian Agricultural Research Corporation. The multiple indicators were grouped into three categories of production and their rankings were studied via factor analysis within each category. A weighting system for each category was defined independently from factor rotations, which led to a set of three dimensional production scores. The performance measures were based on a DEA-CCR production model with three outputs and three inputs. We also proposed a goal achievement index for the production dimensions.

Key-words: Factor analysis, Data envelopment analysis, Goal achievement, Performance.

1. Introdução

Desde 1996, a Empresa Brasileira de Pesquisa Agropecuária (Embrapa) monitora a produção de seus centros de pesquisa com a utilização de um sistema de avaliação com base em um produto e três insumos. Os insumos do processo de produção são capital, custeio e mão de obra. O produto (*output*), unidimensional, é uma média ponderada de 28 atributos de produção, classificados em quatro categorias de produção: (a) produção técnico-científica; (b) produção de publicações técnicas; (c) transferência de tecnologias e promoção da imagem; (d) desenvolvimento de tecnologias, produtos e processos. Uma medida de eficiência é, então, calculada via modelos de análise de envoltória de dados (DEA). Veja Souza et al. (1999) para mais detalhes, notadamente no contexto DEA. O sistema de pesos utilizado para tornar os centros de pesquisa (DMUs) comparáveis é complexo. Toma por base a Lei dos Julgamentos Categóricos de Thurstone (Torgerson, 1958; Souza, 2002) e é dependente de um sistema de normalização que torna as grandezas de produção adimensionais. A subjetividade dos pesos utilizados e a redução dimensional drástica têm sido questionadas no âmbito da organização.

O problema de agregação observado na Embrapa é comum em qualquer instância onde se requer a redução substancial da dimensionalidade do espaço das variáveis de produção. A utilização de sistemas de pesos com base na análise de componentes principais e na análise fatorial visando à redução da dimensão do espaço-resposta tem bastante apelo, mas não raramente leva a situações difíceis de aceitar na prática como, por exemplo, o uso de rotações e de cargas fatoriais negativas. Nesse contexto, estendendo a abordagem de Lopes et al. (2008), sugere-se neste artigo o uso de um sistema de pesos com base em análise fatorial, formado por comunalidades relativas calculadas via métodos de máxima verossimilhança. Utilizam-se *ranks* como uma transformação que gera grandezas adimensionais e robustas nas componentes de produção (produtos e insumos). O sistema de pesos assim caracterizado não tem propriedades ótimas conhecidas. No entanto, é independente de rotações ortogonais e as variáveis de avaliação têm sinal correto. As comunalidades absolutas têm como *proxies* os coeficientes de correlação múltipla obtidos com regressões de cada atributo com todos os demais.

Nossa discussão procede como segue. Na Seção 2 apresenta-se o sistema de produção da Embrapa. Na Seção 3 apresentam-se os modelos multivariados sugeridos na redução da dimensão do espaço de produção. Na Seção 4 apresenta-se a medida de eficiência DEA utilizada na caracterização do desempenho. Na Seção 5 apresentam-se os resultados estatísticos obtidos e a medida de avaliação resultante. Na Seção 6 propõe-se um índice de alcance de metas institucionais. Finalmente, na Seção 7 apresentam-se um resumo e as conclusões do estudo, seguida das referências bibliográficas.

2. O Sistema de Produção da Embrapa

O sistema de produção da Embrapa é composto por 42 centros de pesquisa. Cinco desses centros foram recentemente criados. Têm-se dados históricos de produção para 37 unidades. Um conjunto de indicadores potenciais de produção está disponível na empresa desde 1991. Foram definidos por meio de planos anuais de trabalho e refletem as várias instâncias de produção dos centros de pesquisa. Estão detalhados no documento Embrapa (1996).

A partir dos planos anuais de trabalho chegou-se a um conjunto de 28 variáveis de produto e três de insumo representativos do processo produtivo da Embrapa. Os 28 atributos de produto foram divididos em quatro categorias de produção: (a) produção técnico-científica; (b) produção de publicações técnicas; (c) transferência de tecnologia e promoção da imagem; (d) desenvolvimento de tecnologias, produtos e processos. Reconhecem-se com essas categorias as dimensões múltiplas do trabalho na Embrapa: a dimensão acadêmica, através da produção técnico-científica; a dimensão da extensão rural e do marketing, por meio das categorias de transferência de tecnologia e promoção da imagem e de produção de publicações técnicas; a dimensão de pesquisa e desenvolvimento (P&D), com a categoria de desenvolvimento de tecnologias, produtos e processos.

De forma sucinta, são descritos abaixo os indicadores de produção presentes no sistema de produção da Embrapa. Embora tenha havido alterações no sistema ao longo do tempo, a descrição seguinte é bem representativa das categorias de produção e do que representaram ao longo do período

1996-2009. A referência institucional básica é Embrapa (2009), na qual cada item de produção é descrito com detalhes.

- (a) Produção Técnico-Científica (PTC): artigo em periódico indexado; nota técnica; capítulo em livro técnico-científico; artigo em anais de congresso; resumo em anais de congresso; orientação de dissertação ou tese de pós-graduação.
- (b) Produção de Publicações Técnicas (PPT): sistema de produção; circular técnica; comunicado técnico/recomendações técnicas; boletim de pesquisa e desenvolvimento; documentos; organização/edição de livros; artigo de divulgação na mídia.
- (c) Desenvolvimento de Tecnologias, Produtos e Processos (DTPP): cultivar gerada/lançada e evento elite; cultivar testada/indicada; prática/processo agropecuário; raça/tipo; insumo agropecuário; processo agroindustrial; metodologia científica; máquina, equipamento e instalação; estirpes; monitoramento/zonamento; software.
- (d) Transferência de Tecnologia e Promoção da Imagem (TTPI): dia de campo; organização de eventos e participação da unidade em exposição ou feira; palestra; curso oferecido; folder/folheto/cartilha produzido; vídeo/dvd produzido; unidade demonstrativa e de observação; campanha interna corporativa; veículo interno de comunicação; veículo externo de comunicação; matéria jornalística.

Os insumos do processo de produção são definidos por despesas com trabalho, custeio e depreciação do capital.

Relativamente ao vetor de produção pretende-se utilizar uma medida de desempenho com base em componentes de produtos e de insumos. As componentes de insumos estão definidas e são três (custos com pessoal, custeio e capital). As componentes de produção são, em princípio, quatro (PTC, PPT, DTPP, TTPI), mas podem ser reduzidas pela análise multivariada. De fato, a análise multivariada que foi levada a efeito nos atributos de produção sugere uma redução no número de dimensões de produção e do número de variáveis, notadamente nas categorias PPT, TTPI e DTPP. Deste modo, considera-se para a produção (*outputs*) as categorias: PTC (artigo em periódico indexado, capítulo de livro, artigo e resumo em anais de congressos, orientação de teses); PPT, com as variáveis circular técnica, comunicado técnico/recomendações técnicas, boletim de pesquisa e desenvolvimento e documentos; outras atividades de produção – POUT, com as variáveis dias de campo, organização de eventos, unidade demonstrativa e de observação, estágios de pós-graduação, metodologia científica e monitoramento. A obtenção dessas três componentes de produção está descrita na Seção 5.

3. Modelos de Análise Fatorial

No estudo da consistência das variáveis em termos das dimensões envolvidas no sistema de produção da Embrapa fez-se uso da Análise Fatorial com opção pela técnica de estimação de máxima verossimilhança. Veja Mardia et al. (1979), Johnson e Wichern (2007) e Corrar et al. (2007) para maiores detalhes.

Seja x um vetor de dimensão p com média μ e matriz de covariâncias Ω . Diz-se que x satisfaz ao modelo com k fatores se puder ser escrito na forma: $x - \mu = \Lambda f + u$, sendo $\Lambda (p \times k)$ uma matriz de constantes. As componentes $f (k \times 1)$ e $u (p \times 1)$ são vetores aleatórios. As componentes de f são denominadas fatores comuns e os elementos de u fatores específicos. Adicionalmente, supõe-se que $E(f) = 0$, $Var(f) = I$, $E(u) = 0$, $Cov(u_i, u_j) = 0 \quad i \neq j$, $Cov(f, u) = 0$.

Denote a matriz de covariâncias do vetor de fatores específicos por $\Phi = diag(\varphi_{11}, \dots, \varphi_{pp})$.

Resulta que os fatores comuns são ortogonais com variância unitária. Impõe-se adicionalmente que os fatores comuns e específicos têm distribuição normal conjunta multivariada. Note-se que:

$$x_i - \mu_i = \sum_{j=1}^k \lambda_{ij} f_j + u_i, \text{ de modo que a variância } \sigma_i^2 \text{ da variável } x_i \text{ vem dada por:}$$

$$\sigma_i^2 = \sum_{j=1}^k \lambda_{ij}^2 + \varphi_{ii}.$$

A componente $h_i^2 = \sum_{j=1}^k \lambda_{ij}^2$ denomina-se comunalidade de x_i e representa a componente da variância de x_i que é compartilhada com as outras variáveis por meio dos fatores comuns. Em particular $\lambda_{ij} = Cov(x_i, f_j)$ representa a intensidade da associação entre x_i e o fator f_j .

O modelo fatorial pode ser equivalentemente determinado pela condição $\Omega = \Lambda\Lambda' + \Phi$ e é invariante por transformações de escala no vetor x .

Se o modelo com k fatores se verifica para x , segue que para qualquer matriz ortogonal $G(k \times k)$ tem-se $x - \mu = (\Lambda G)(G' f) + u$ e $\Omega = (\Lambda G)(G' \Lambda') + \Phi$. Portanto, rotações não conduzem a formulações distintas da estrutura fatorial. As comunalidades permanecem inalteradas, mas as cargas fatoriais podem diferir. É claro que por sua invariância a transformações de escala o modelo fatorial pode ser equivalentemente formulado em termos da matriz de correlação. Por conveniência, manter-se-á a discussão com base em Ω .

Seja S a matriz de covariância amostral de x com base em n observações. Estimativas de máxima verossimilhança de Λ e Φ são obtidas maximizando a função log-verossimilhança $-\frac{1}{2} n \log |2\pi\Omega| - \frac{1}{2} \text{tr} \Omega^{-1} S$ em relação a Λ e Φ .

Uma das vantagens principais do uso do método de máxima verossimilhança em Análise Fatorial é que, quando se ajusta, fornece um teste da hipótese H_k de que a utilização de k fatores é suficiente para descrever a matriz de covariâncias contra alternativa de que Ω seja não estruturada (sem restrições). A estatística da razão de verossimilhança para este teste vem dada pela expressão $np(\hat{a} - \log \hat{g} - 1)$, onde \hat{a} e \hat{g} representam as médias aritmética e geométrica dos autovalores de $\hat{\Omega}^{-1} S$. Sob a hipótese H_k , a estatística distribui-se como χ^2 com $s = \frac{1}{2}(p-k)^2 - \frac{1}{2}(p+k)$ graus de liberdade. No caso trivial em que $k=0$ (independência das variáveis), a estatística pode ser calculada utilizando a fórmula $-n \log |R|$, sendo R a matriz de correlação. Bartlett (1954) mostra que a aproximação qui-quadrado melhora se n é substituído por $n' = n - 1 - \frac{1}{6}(2p+5) - \frac{2}{3}k$.

Os testes de independência e de números de fatores estão disponíveis, por exemplo, em Stata (2011) e SAS (2012) nos procedimentos de Análise Fatorial. Faz-se uso dessas versões neste artigo.

Um índice de adequabilidade muito utilizado como informação adicional sobre a adequabilidade da Análise Fatorial a um conjunto de variáveis é a medida de adequabilidade amostral de Kayser-Meyer-Olkin (KMO). É um indicador que compara o tamanho dos coeficientes de correlação relativamente aos coeficientes de correlação parcial. Para a variável i define-se

$$KMO_i = \frac{\sum_{j \neq i} r_{ij}^2}{\sum_{j \neq i} r_{ij}^2 + \sum_{k \neq i} b_{ij}^2}, \text{ onde } R = (r_{ij}) \text{ é a matriz de correlação entre os atributos e } B = (b_{ij}) \text{ é}$$

$$\text{a matriz de correlações parciais. Para o modelo define-se } KMO = \frac{\sum_i \sum_{j \neq i} r_{ij}^2}{\sum_i \sum_{j \neq i} r_{ij}^2 + \sum_i \sum_{k \neq i} b_{ij}^2}.$$

Sugere-se que valores de $KMO < 0,50$ são indicativos de que a matriz de correlação não se presta para a Análise Fatorial. Neste contexto, as medidas parciais são utilizadas, tipicamente, como critério para escolha de variáveis. Veja Corrar et al. (2007) para mais detalhes sobre este procedimento.

O uso de procedimentos que tomam por base a Análise Multivariada sob a hipótese de normalidade com *ranks* emprestam aos métodos propriedades robustas e não paramétricas (Conover, 1999). Na aplicação aqui em discussão, a utilização de *ranks* como medidas de resposta torna-se ainda mais conveniente, uma vez que definem grandezas adimensionais que permitem o cálculo de medidas

de agregação por meio de um sistema de pesos convenientemente escolhido. Previamente ao cálculo de *ranks*, todos os indicadores de produção são normalizados pelo quantitativo de pessoal da unidade em estudo. Este procedimento torna as unidades mais comparáveis e reduz variabilidade.

Os escores finais definidos com base na Análise Fatorial divergem da representação induzida pelas cargas fatoriais. Considere uma das três dimensões de produção contendo v variáveis com observações de *ranks* (c_1^i, \dots, c_v^i) . O escore da unidade i na dimensão j é definido por

$$y_i = \sum_{j=1}^v \theta_j c_j^i \quad \theta_j = \frac{h_j^2}{\sum_{\tau=1}^v h_\tau^2} \cdot$$

O sistema de pesos é, portanto, invariante por transformações ortogonais do modelo fatorial. A ideia da definição de um sistema de pesos com base em comunalidades relativas é uma extensão de Souza et al. (2004), que se utiliza dos coeficientes de correlação múltipla para a definição de um escore de crise no contexto de avaliações de comércio internacional.

4. Medidas de Desempenho – Análise de Envoltória de Dados

Como medida de desempenho (performance) para o sistema de produção de pesquisa da Embrapa utiliza-se um modelo de envoltória de dados DEA-CCR (Cooper et al., 2011), com orientação para insumos. A orientação do modelo, embora irrelevante sob a hipótese de retornos constantes de escala, mantém consistência com modelos anteriores.

A normalização de insumos e produtos pelo quantitativo de pessoal e posterior transformação em *ranks* inibe o efeito escala e motiva a utilização de um conceito mais rigoroso de performance por parte dos gestores do processo. Deste modo, se $X (3 \times 37)$ representa a matriz de *ranks* de utilização de insumos normalizados e $Y (3 \times 37)$ a matriz de *ranks* médios ponderados calculados com a utilização da Análise Fatorial, a medida de performance para uma dada unidade o com vetor de produção (x_o', y_o') de dimensão seis vem dada pela solução γ_o do modelo de programação linear $\text{Max } \gamma$, sujeito a $X\lambda \leq \gamma_o, Y\lambda \geq y_o, \lambda \geq 0$. Portanto, um modelo tipo DEA na formulação do envelope.

5. Resultados Estatísticos e Medidas de Produção e Desempenho

Inicia-se a discussão com a Análise Fatorial e as medidas de produção (Seções 5.1, 5.2 e 5.3). Os resultados sobre as medidas de desempenho são apresentadas no item 5.4.

5.1. Produção Técnico-Científica (PTC)

As medidas iniciais de adequabilidade (KMO) da Análise Fatorial para *ranks* das variáveis artigo em periódico indexado (V101), capítulo em livro técnico-científico (V103), artigo em anais de congresso (V104), resumo em anais de congresso (V105) e orientação de dissertação ou tese de pós-graduação (V106) constam da Tabela 1. A componente nota técnica (V102) não consta da análise, por não ser considerada no sistema de avaliação desde o ano de 2000. As estatísticas sugerem boa representação para o modelo fatorial. A Tabela 2 apresenta os resultados de estimação considerando um fator do modelo. Os testes estatísticos evidenciam representação adequada do modelo. As variáveis dominantes na dimensão estimada são V101, V105 e V106, com comunalidades relativas de 38,5%, 31,4% e 22,4%, respectivamente.

Tabela 1: Medidas de Kaiser-Meyer-Olkin (KMO) para a dimensão PTC.

Variável	KMO
V101	0,6638
V103	0,6840
V104	0,5831
V105	0,6743
V106	0,8322
Total	0,6926

Tabela 2: Análise Fatorial pelo Método de Máxima Verossimilhança para a dimensão PTC. Listagem SAS para as variáveis V101, V103-V106.

Estimativas iniciais de comunalidades: Coeficiente de Correlação Múltipla				
V101	V103	V104	V105	V106
0.65427730	0.11505519	0.23688700	0.61021914	0.43643606

Autovalores preliminares: Total = 4,67289318; Média = 0,93457864				
	Autovalor	Diferença	Proporção	Cumulativo
1	4,91179695	4,39959767	1,0511	1,0511
2	0,51219928	0,61427042	0,1096	1,1607
3	-0,10207114	0,10331786	-0,0218	1,1389
4	-0,20538899	0,23825393	-0,0440	1,0949
5	0,44364292		-0,0949	1,0000

1 fator será retido pelo critério “número de autovalores > 1”.

Testes de significância baseados em 37 observações

Teste	DF	Chi-Square	Pr>ChiSq
H0: Não há fatores comuns	10	57,0753	<0,0001
HA: Há pelo menos um fator comum			
H0: 1 fator é suficiente	5	5,6348	0,3434
HA: São necessários mais fatores			

Estimativa de comunalidades finais e pesos para as variáveis
Comunalidade total: 2,158261

Variável	Comunalidade	Peso (%)
V101	0,82991608	38,4530
V103	0,04223360	1,9568
V104	0,12458285	5,7724
V105	0,67726077	31,3799
V106	0,48426744	22,4379

5.2. Produção de Publicações Técnicas (PPT)

As medidas iniciais de adequabilidade (KMO) da Análise Fatorial para os *ranks* das variáveis circular técnica (V201), comunicado técnico/recomendações técnicas (V202), boletim de pesquisa e desenvolvimento (V203) e documentos V(204) constam da Tabela 3. As estatísticas sugerem boa representação para o modelo fatorial. A Tabela 4 apresenta os resultados de estimação considerando um fator do modelo. Os testes estatísticos evidenciam representação adequada do modelo. As variáveis dominantes na dimensão estimada são V202, V203 e V204, com comunalidades relativas de 16,3%, 19,6% e 21,9%, respectivamente.

Tabela 3: Medidas de Kaiser- Meyer-Olkin (KMO) para a dimensão PPT.

Variável	KMO
V201	0,5850
V202	0,6332
V203	0,6110
V204	0,5773
Total	0,6014

Tabela 4: Análise Fatorial pelo Método de Máxima Verossimilhança para a dimensão PPT. Listagem SAS para as variáveis V201-V204.

Estimativas iniciais de comunalidades: Coeficiente de Correlação Múltipla			
V201	V202	V203	V204
0,15106205	0,19987571	0,23806904	0,25347411

Autovalores preliminares: Total = 1,07974136; Média = 0,26993534				
	Autovalor	Diferença	Proporção	Cumulativo
1	1,40902510	1,28899769	1,3050	1,3050
2	0,12002741	0,18614078	0,1112	1,4161
3	-0,06611337	0,31708441	-0,0612	1,3549
4	-0,38319778		0,3549	-1,0000

1 fator será retido pelo critério “número de autovalores > 1”.

Testes de significância baseados em 37 observações

Teste	DF	Chi-Square	Pr>ChiSq
H0: Não há fatores comuns	6	17,2555	0,0084
HA: Há pelo menos um fator comum			
H0: 1 fator é suficiente	2	2,7532	0,2524
HA: São necessários mais fatores			

Estimativa de comunalidades finais e pesos para as variáveis
Comunalidade total: 1,803447

Variável	Comunalidade	Peso (%)
V201	0,15518942	8,6052
V202	0,29466609	16,3390
V203	0,35355862	19,6046
V204	0,39580069	21,9469

5.3 Outras Atividades de Produção

As medidas iniciais de adequabilidade (KMO) da Análise Fatorial para os *ranks* das variáveis: dia de campo (V301), organização de eventos (V302), palestra (V303), estágios de pós-graduação (V308), unidades demonstrativas e de observação (V311), metodologia científica (V408), monitoramento (V411) constam da Tabela 5. As estatísticas sugerem boa representação para o modelo fatorial. O KMO de V411 é marginal, mas decidiu-se manter a variável dada sua importância prática para a análise. A Tabela 6 apresenta os resultados de estimação considerando dois fatores no modelo para representar esta dimensão. Os testes estatísticos evidenciam representação adequada do modelo com dois fatores. Nesta instância, foram adicionadas as cargas fatoriais que evidenciam correlações negativas com as dimensões fatoriais. As variáveis dominantes na dimensão estimada são V301, V302, V303 e V411, com comunalidades relativas de 10,0%, 8,0%, 7,7% e 7,1%, respectivamente.

Tabela 5: Medidas de Kaiser-Meyer-Olkin (KMO) para a dimensão POUT.

Variável	KMO
V301	0,6057
V302	0,6327
V303	0,6763
V308	0,6657
V311	0,5574
V408	0,6058
V411	0,4957
Total	0,6064

Tabela 6: Análise Fatorial pelo Método de Máxima Verossimilhança para a dimensão POUT. Listagem SAS para as variáveis V301, V302, V303, V308, V311, V408 e V411.

Estimativas iniciais de comunalidades: Coeficiente de Correlação Múltipla						
V301	V302	V303	V308	V311	V408	V411
0,49337297	0,46032254	0,46823826	0,16003060	0,32167975	0,25479316	0,36127693

Autovalores preliminares: Total = 4,27962124; Média = 0,61137446

	Autovalor	Diferença	Proporção	Cumulativo
1	3,21265362	1,64930085	0,7507	0,7507
2	1,56335277	1,00145537	0,3653	1,1160
3	0,56189740	0,60526695	0,1313	1,2473
4	-0,04336955	0,13056158	-0,0101	1,2372
5	-0,17393113	0,19724615	-0,0406	1,1965
6	-0,37117728	0,09862731	-0,0867	1,1098
7	-0,46980459		-0,1098	1,0000

2 fatores serão retidos pelo critério “número de autovalores > 1”.

Significance Tests Based on 37 Observations

Teste	DF	Chi-Square	Pr>ChiSq
H0: Não há fatores comuns	21	57,1874	<0,0001
HA: Há pelo menos um fator comum			
H0: 2 fatores são suficientes	8	8,6794	0,3701
HA: São necessários mais fatores			

Padrão dos fatores

	Fator1	Fator2
V301	0,79061	-0,31775
V302	0,65187	0,39513
V303	0,72486	0,19109
V308	-0,26371	0,18373
V311	0,44778	0,03070
V408	-0,13771	0,55789
V411	0,09280	0,71429

Estimativa de comunalidades finais e pesos para as variáveis

Comunalidade total: 7,258261.

Variável	Comunalidade	Peso (%)	Variável	Comunalidade	Peso (%)
V301	0,7260228	10,0027	V311	0,2014535	2,7755
V302	0,5810623	8,0055	V408	0,3302008	4,5493
V303	0,5619325	7,742	V411	0,5188184	7,148
V308	0,1032986	1,4232			

5.4. Medidas de Desempenho – DEA

A Tabela 7 apresenta a matriz dos dados de produção obtidos com o cálculo de *ranks* ponderados em cada dimensão de produção. São essas as informações que foram utilizadas no cálculo das medidas de desempenho DEA-CCR (TE), conforme a Seção 4, que também constam da Tabela 7.

Na Figura 1 apresentam-se o histograma e o gráfico de caixa (desenho esquemático) correspondente à distribuição amostral da eficiência técnica TE (medida de desempenho ou performance). O primeiro quartil da distribuição vale 0,393, a mediana vale 0,466 e o terceiro quartil 0,802. A distribuição amostral não apresenta observações atípicas e é não simétrica. O histograma evidencia superposição de populações. Das unidades do primeiro quartil, somente duas não são

eficientes (DMU14 e DMU19). Apenas uma solução de performance é Pareto ótima (DMU24). As demais apresentam multiplicadores nulos para alguma componente de insumo e/ou produto.

Tabela 7: *Ranks* das variáveis de insumo (Pessoal, Capital e Custeio), das dimensões de produção (PTC, PPT e POUT) e eficiência técnica TE (DEA-CCR).

Unidade	Pessoal	Capital	Custeio	PTC	PPT	POUT	TE
DMU1	37	35	36	31,5296	22,9612	13,6864	0,393
DMU2	22	6	7	22,2157	16,9093	8,4661	1,000
DMU3	16	29	25	34,2652	30,8803	21,5683	0,600
DMU4	4	9	3	11,8827	17,9911	12,6231	1,000
DMU5	32	30	33	33,8777	18,2345	13,1978	0,487
DMU6	11	15	30	10,8931	18,9579	18,1033	0,338
DMU7	34	37	35	36,6673	32,1385	23,4504	0,452
DMU8	30	11	16	27,2637	21,5523	27,6870	0,762
DMU9	9	34	19	23,7957	15,3251	11,8751	0,423
DMU10	14	23	1	20,1368	8,4595	26,8624	1,000
DMU11	5	16	8	14,8892	17,9517	11,8054	0,655
DMU12	33	36	34	10,7541	14,7759	14,8934	0,141
DMU13	29	28	14	23,0864	13,1275	27,6819	0,466
DMU14	23	10	24	30,5856	23,1518	25,9106	0,912
DMU15	13	19	23	13,4750	21,1951	14,8782	0,332
DMU16	35	24	15	18,8796	23,7228	27,3079	0,397
DMU17	12	22	13	11,1301	7,7609	22,9159	0,348
DMU18	15	25	32	27,2374	8,2204	25,2537	0,528
DMU19	10	12	9	18,6969	26,4724	23,6897	0,802
DMU20	36	32	27	24,3951	22,3511	21,0056	0,338
DMU21	7	31	26	20,9088	10,8519	19,8410	0,449
DMU22	17	3	10	3,3967	13,3891	11,1582	0,305
DMU23	21	13	20	23,2122	23,2009	14,4377	0,649
DMU24	1	17	6	14,4433	27,1055	25,9457	1,000
DMU25	28	20	37	19,1080	5,7224	11,3360	0,349
DMU26	31	8	28	7,5525	7,5643	9,4634	0,219
DMU27	8	1	2	3,6417	25,0265	17,2382	1,000
DMU28	26	27	22	21,9523	33,7149	21,7792	0,404
DMU29	3	2	5	6,0129	15,6037	13,6183	1,000
DMU30	19	14	29	9,0103	18,5145	16,7077	0,231
DMU31	27	18	11	16,6057	24,8239	18,6856	0,462
DMU32	24	5	21	24,6492	28,6282	24,9340	1,000
DMU33	2	7	4	12,1463	7,8870	18,2237	1,000
DMU34	20	4	12	5,7347	6,9819	19,8783	0,415
DMU35	18	26	18	27,2785	25,7300	23,7292	0,540
DMU36	6	21	17	11,9965	25,0867	33,3894	0,642
DMU37	25	33	31	29,6934	21,0293	9,7720	0,428

A Embrapa classifica seus centros de pesquisa em temáticos, de produtos e ecorregionais. Embora a dependência entre as medidas não viabilizem testes estatísticos formais entre os diferentes tipos, nota-se que os centros de produto têm melhor desempenho, seguido dos ecorregionais e temáticos. As medianas dos três tipos em ordem decrescente são 0,642 (produto), 0,462 (ecorregionais) e 0,428 (temáticos). A diferença entre as medianas dos centros ecorregionais e temáticos não parece significativa, mas as distribuições são distintas. A Figura 2 apresenta os gráficos de caixa por tipo de centro de pesquisa.

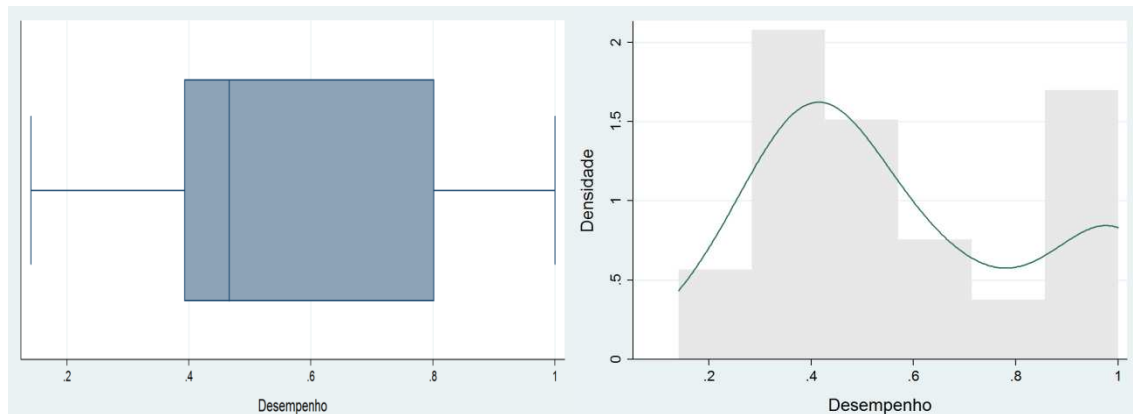


Figura 1: Histograma e desenho esquemático da distribuição do desempenho (TE).

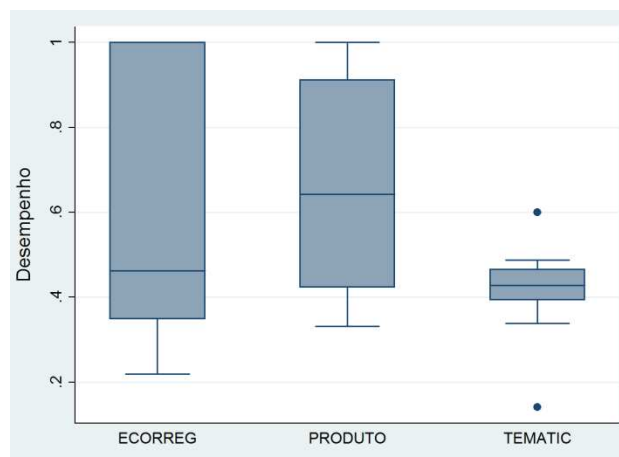


Figura 2: Desenho esquemático do desempenho (TE), por tipo de unidade de pesquisa.

6. Alcance de Metas Institucionais

O sistema de pesos derivado da Análise Fatorial com o uso das comunalidades relativas pode ser estendido para o acompanhamento da realização de metas de produção, negociadas entre os gestores da organização e os gestores locais da produção. Isto pode ser conseguido por meio de índices médios de alcance de objetivos. Para a dimensão ν , com variáveis x_i , $i = 1, \dots, m_\nu$ e pesos

w_i , $i = 1, \dots, m_\nu$, o índice de alcance de metas (IAM_ν) é dado por $IAM_\nu = \sum_{i=1}^{m_\nu} w_i \frac{x_i^o}{x_i^m}$, onde x_i^m

representa a meta negociada para o atributo i e x_i^o o valor observado. Casos óbvios, como metas nulas, são tratados por meio de eliminação do atributo e reavaliação de pesos. A Tabela 8 apresenta o resultado de uma simulação tomando como base as observações de 2009 e metas fictícias. Observações atípicas na tabela representam metas mal negociadas. Valores ausentes representam observações não realizadas. A categoria PTC não apresenta observações atípicas. Na categoria PPT são atípicas as unidades 9 e 29. A categoria POUT tem como atípicos 1, 4 e 24. A Figura 3 é representativa desses comentários.

Tabela 8: Resultado de uma simulação para o cálculo do IAM para cada dimensão de produção.

Unidade	IAM			Unidade	IAM		
	PTC	PPT	POUT		PTC	PPT	POUT
DMU1	1,0635	2,0378	8,1101	DMU20	0,9352	0,6600	0,9673
DMU2	1,8495	3,2729	0,7698	DMU21	0,8945	1,3440	1,4346
DMU3	1,1305	1,5337	1,0578	DMU22	1,0746	1,5930	1,4049
DMU4	1,1154	1,6386	3,8145	DMU23	1,7854	1,1226	1,6964
DMU5	1,1661	0,8605	1,3507	DMU24	0,8397	1,4489	3,0803
DMU6	0,6161	1,6508	1,0122	DMU25	1,7541	1,3107	0,9554
DMU7	1,8087	1,8244	1,9007	DMU26	0,7571	0,0000	0,6067
DMU8	0,8621	1,2428	0,9074	DMU27	1,0181	1,6583	2,0204
DMU9	0,8667	4,8104	0,8711	DMU28	1,1946	2,0141	0,6787
DMU10	0,8417	1,4531	1,6092	DMU29	0,9000	10,3056	0,8981
DMU11	1,5082	2,1109	0,8732	DMU30	1,1392	1,2914	1,3487
DMU12	0,6175	0,3013	0,7706	DMU31	0,9334	1,3848	1,3642
DMU13	1,5736	2,0259	1,4300	DMU32	1,5842	2,6945	1,9432
DMU14	0,8968	0,9493	0,9563	DMU33	0,9819	0,7543	1,9277
DMU15	1,1623	1,0350	0,8270	DMU34	0,9287	2,9613	1,0429
DMU16	1,3627	0,7669	1,1671	DMU35	1,0424	1,4825	1,1856
DMU17	1,0806	2,3167	1,3200	DMU36	1,5382	0,9096	1,1773
DMU18	1,3637	0,7613	1,8464	DMU37	0,6957	1,4609	0,7451
DMU19	1,6340	1,4060	0,9135				

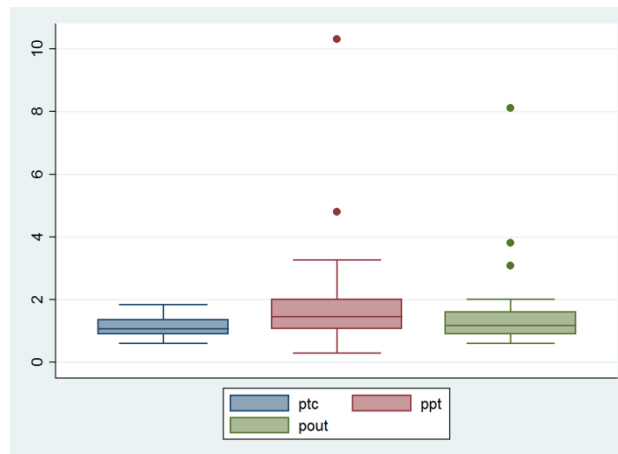


Figura 3: Desenho esquemático dos índices de alcance de metas para as três categorias de produção.

7. Considerações Finais

Neste artigo, com base no modelo de produção dos centros de pesquisa da Embrapa, estuda-se a redução da dimensão do espaço das variáveis de produção, a performance dos centros de pesquisa e o nível da consecução de metas de cada centro relativamente a objetivos negociados.

O sistema de pesos usado na agregação de variáveis de produção é objetivo e toma como referência as comunalidades relativas calculadas via o método de máxima verossimilhança em análise fatorial. As quatro dimensões originais de produção (PTC, PPT, DTPP, TTPI) são reduzidas a três (PTC, PPT, POUT), com recomposição dos indicadores em cada dimensão. Essas três dimensões consideram *ranks* médios ponderados como medidas de produção, obtidos com os pesos derivados da análise fatorial. A vantagem do uso da análise fatorial na determinação de pesos em cada uma das dimensões é a não necessidade de emissão de julgamentos de valor por parte dos decisores em relação à importância relativa de cada indicador. A experiência na empresa mostra que os gestores são bastante relutantes em emitir juízos de valor sobre importâncias ou impor pesos. O uso de *ranks*

empresta propriedades não paramétricas à análise fatorial, permite a agregação e é robusto à presença de observações atípicas.

Para calcular a medida de desempenho dos centros de pesquisa da Embrapa é usado um modelo DEA-CCR, no qual os *inputs* são os *ranks* de três categorias de custo (pessoal, custeio, capital) e os *outputs* são os *ranks* médios ponderados nas três dimensões resultantes do estudo via análise fatorial. Entende-se que o modelo DEA aqui usado gera uma medida de desempenho, ao invés de uma medida de eficiência propriamente dita, calculada a partir de uma função de produção. O estudo da distribuição amostral dos escores de desempenho DEA mostra que não há observações atípicas (*outliers*), que a distribuição é não simétrica e que há evidências de superposição de populações. Das dez unidades pertencentes ao primeiro quartil de desempenho, apenas duas são não eficientes. Dentre as unidades eficientes, apenas uma é Pareto ótima. Ao considerar-se a categorização de centros de pesquisa por tipo de missão que executam, verifica-se que os centros de produto têm desempenho mediano superior aos ecorregionais, e estes aos centros temáticos. As distribuições por tipo, contudo, diferem.

A proposta de uma medida que reflita o alcance de metas institucionais vai ao encontro de objetivos gerenciais de acompanhamento da realização das metas de produção negociadas. A negociação de metas ocorre entre os gestores locais (chefes de centros de pesquisa) e a administração superior da empresa (presidente e diretoria executiva). O exercício de simulação do índice aqui proposto (IAM), calculado por dimensão de produção, mostra, além do alcance das metas, negociações mal conduzidas. IAM extremamente elevados (como os valores atípicos discutidos no item 6) representam valores negociados extremamente baixos, ou seja, possibilidade de falha no processo de negociação de metas entre os gestores centrais e os locais.

Cabe salientar que as abordagens aqui propostas, seja do uso de análise fatorial, de DEA ou o IAM, podem ser replicadas para outros conjuntos de indicadores e dimensões de produção que venham a ser propostos ou institucionalizados.

Referências

- Bartlett, M.S. A note on multiplying factors for various chi-squared approximations. *Journal of the Royal Statistical Society*, series B, v. 16, p. 296-298, 1954.
- Conover, M.J. *Practical nonparametric statistics*. 3rd ed. New York: Wiley, 1999.
- Cooper, W.W.; Seiford, L.M.; Zhu, J. *Handbook on data envelopment analysis*. 2nd ed. New York: Springer, 2011.
- Corrar, L.J.; Paulo, E.; Dias Filho, J.M. *Análise multivariada: para os cursos de administração, ciências contábeis economia*. São Paulo: Atlas, 2007.
- Embrapa. *Manual do Sistema de Avaliação e Premiação por Resultados*. Brasília: Embrapa - Gabinete do Presidente, 1996.
- Embrapa. *Manual dos Indicadores de Avaliação de Desempenho dos Centros de Pesquisa da Embrapa: 2008 a 2011*. Brasília: Secretaria de Gestão e Estratégia, 2009.
- Johnston, J.; Dinardo, J. *Econometric Methods*. 4th ed. Boston: McGraw-Hill, 1997.
- Lopes, M.R.; Souza, G.S.; Lopes, I.V.; Oliveira, M.S.; Bogado, P.R. Estradas rurais ou urbano-industriais: processo de escolha em regime de competição por fundos públicos. *Revista de Política Agrícola*, v. 4, p. 47-64, 2008.
- Mardia, L.V.; Keni, J.T.; Bibby, J.M. *Multivariate analysis*. London: Academic, 1979.
- SAS. *SAS/STAT® 12.1 User's Guide*. North Carolina: SAS Institute Inc. 2012.
- Souza, G.S. The law of categorical judgement revisited. *Brazilian Journal of Probability and Statistics*, v. 16, p. 123-140, 2002.
- Souza, G.S.; Avila, A.F.D.; Alves, E.R. Technical efficiency of production in agricultural research. *Scientometrics*, v. 46, n. 1, p. 141-160, 1999.
- Souza, G.S.; Moreira, T.B.S.; Pinto, M.B.P. Uma metodologia alternativa para mensuração de pressão sobre o mercado de câmbio. *Estudos Econômicos*. v. 34, n.1, p. 73-99, 2004.
- Stata. *Multivariate Statistics Reference Manual*. Texas: Stata Press. 2011.
- Torgerson, W.S. *Theory and Methods of Scaling*. New York: Wiley. 1958.