

Composição probabilística de critérios na avaliação de cursos

*UFF
annibal@producao.uff.br

Annibal Parracho Sant'Anna*

Resumo

Este artigo aplica a composição probabilística de critérios na avaliação de cursos de pós-graduação em Engenharia de Produção. São confrontadas análises baseadas apenas em resultados com outras que contrabalançam medidas de resultados com medidas do volume de recurso docente empregado. São construídos, também, indicadores de evolução ao longo do tempo. As variáveis de resultado básicas medem a formação de mestres e a publicação de resultados de pesquisa. A adição de outras variáveis não modifica substancialmente as conclusões. Verifica-se que, no conjunto considerado, a abordagem probabilística oferece um instrumento objetivo para a avaliação global do desempenho dos cursos e, em especial, da evolução desse desempenho com o tempo.

Palavras-chave: apoio à decisão multicritério; avaliação; probabilidades; pós-graduação.

Abstract

In this article the probabilistic composition of criteria is applied to the evaluation of graduate courses in the area of Production Engineering (Industrial Engineering in some regions). The difference between analyses based only on outputs and those that counterbalance output measurements with measurements of the volume of faculty resources applied is investigated. Indicators of the evolution through time are also constructed. The basic output variables measure formation of M. Sc. and publication of research findings. Adding other variables does not substantially change the conclusions. It is demonstrated that, in the data set studied, the probabilistic approach offers an objective tool for the global evaluation of course performance and, especially, of the evolution of such performance through time.

Keywords: multicriteria decision aid – evaluation – probabilities – graduate courses.

1. Introdução

Este trabalho emprega a composição probabilística de critérios na avaliação da produção de cursos de pós-graduação. São calculadas medidas de eficiência relativa desses cursos e sua evolução ao longo do tempo. Os dados são extraídos dos relatórios de 1998 a 2003 dos cursos de doutorado e mestrado acadêmico em Engenharia de Produção das universidades brasileiras, divulgados pela Capes (2005).

A composição probabilística, conforme desenvolvida em Sant'Anna (2002), permite combinar objetivamente, de diversas maneiras, classificações segundo diferentes critérios. Caso se deseje levar em conta preferências entre os critérios, pode-se tratar a probabilidade de escolha segundo cada critério como probabilidade condicional. Preferências entre os critérios são, entretanto, difíceis de quantificar e a avaliação probabilística propicia, também, variadas formas de combinar os critérios sem atribuir antecipadamente prioridades entre eles. Por exemplo, permite adotar um ponto de vista otimista, segundo o qual uma opção é boa se é boa segundo pelo menos um dos critérios disponíveis, ou uma abordagem pessimista, que exige, para a opção ser considerada boa globalmente, que seja boa segundo todos os critérios.

Na presente análise da produção dos cursos de pós-graduação, dois atributos são considerados básicos, a quantidade de alunos formados e a qualidade da formação oferecida. Como não se deseja, por objetivos de gestão, perseguir o crescimento da quantidade à custa de relaxamento da qualidade ou vice-versa, a avaliação probabilística é orientada por uma ótica pessimista, de avaliar pela probabilidade de simultaneamente otimizar quanto aos dois atributos.

Quantidade e qualidade são inicialmente medidas por meio de duas variáveis objetivas de fluxo. A medida de quantidade é o número de mestres formados em um dado período e a de qualidade é o número de membros do corpo docente apresentando novos resultados de pesquisa nesse mesmo período. Em uma segunda análise, são considerados, também, o número de doutores formados e o número de artigos publicados em periódicos de circulação internacional qualificados pelos comitês de avaliação da Capes nos níveis Qualis A ou B.

Variações dessas medidas são consideradas extensivamente nos indicadores empregados na avaliação dos cursos de pós-graduação pela Capes, conforme indicado nos documentos de avaliação das diversas áreas. O mesmo ocorre em estudos anteriores específicos dos dados dos cursos de Engenharia de Produção, como o Estellita Lins et al. (2004) e Almeida et al. (2001). A opção por essas formulações está associada a aspectos como a sua universalidade e facilidade de interpretação. O que não significa que se possa desprezar a importância de variáveis diferentes que afetam a qualidade da formação, tais como a adequação das práticas empregadas no processo de ensino e aprendizagem, a variedade de interesses do corpo docente e a intensidade da sua participação nas atividades do curso ou os benefícios percebidos pelos egressos do curso e trazidos por esses ao contexto social em que venham a inserir-se, por

exemplo. É importante destacar que todas essas variáveis podem ser medidas e aproveitadas em instrumentos como o aqui desenvolvido. Ao contrário, se não forem adequadamente medidas e aproveitadas por instrumentos de avaliação objetiva, é difícil que a sua importância seja corretamente levada em conta.

Uma característica da composição probabilística que facilita a inclusão de maior número de variáveis é que ela é baseada no cálculo das probabilidades de cada opção ser a melhor, ou não ser a pior, segundo cada um dos critérios considerados. Essa referência aos desempenhos extremos resulta em que a incorporação de novas variáveis não traz distorções devidas à correlação, mesmo que os atributos envolvidos não sejam independentes. Caso se possa avaliar precisamente a correlação entre os atributos, essa pode ser levada em conta. Mas, em geral, podemos negligenciá-la porque a correlação se reduz fortemente quando se limita a análise aos indicadores de pertinência à fronteira (Sant'Anna, 2004).

É interessante observar que a composição probabilística adapta-se facilmente ao caso de se desejar estabelecer tetos ou patamares absolutos para as variáveis fora do campo de valores efetivamente observados. Isto foi empregado em Sant'Anna (2005). No caso oposto, de não se desejar atribuir importância aos acréscimos efetivamente observados que ultrapassem certos limites, pode-se recorrer ao truncamento de variáveis.

As duas últimas variáveis acima citadas, número de doutores formados e número de artigos publicados, entram em segundo lugar porque são menos adequadas à medição dos atributos que se deseja levar em conta. Pela sua complexidade e por estar mais sujeita à influência de características individuais dos alunos, a quantidade de doutores formados reflete, menos que a formação de mestres, a eficiência no gerenciamento dos recursos do curso. Por seu turno, a publicação em periódicos de circulação internacional e a avaliação dos mesmos como de alta qualidade é sujeita à influência da valoração das linhas de pesquisa que a torna menos confiável que a simples indicação de envolvimento em atividades de pesquisa.

A escolha das fronteiras de referência é um importante elemento da avaliação probabilística. A determinação de probabilidades de escolha tem o efeito de expandir as distâncias entre aquelas opções ou objetos de avaliação que estão no extremo a que se deseja atribuir maior importância. Um ponto de vista progressista dará maior importância às distâncias aos melhores desempenhos; já um ponto de vista conservador dará maior valor às distâncias aos piores. No caso presente, são usados como referência os valores mais baixos das variáveis de resultado. Isso se deve ao fato de que, para valores já elevados, o crescimento dessas variáveis perde, gradualmente, a importância. Passa a poder indicar a massificação da atribuição do título de pós-graduação ou da publicação de artigos. Assim, o esforço para elevar os valores dessas variáveis deve receber maior importância enquanto se trata de se afastar da fronteira inferior e ir perdendo importância à medida que crescem os valores observados.

É realizada, também, uma análise em termos de produtividade, comparando a produção com o volume de recurso docente empregado, medido pelo número de docentes no núcleo de referência docente (NRD6)

adotado pela área de Engenharia no período considerado. Também quanto a essa variável, adota-se uma ótica conservadora, que consiste, no caso de recursos, em considerar as probabilidades de atingir a fronteira de maiores valores e não a de menores valores. Isso se justifica porque, abaixo de certas dimensões, é maior a importância relativa da participação dos docentes em atividades que não são diretamente ligadas aos resultados.

Na verdade, a análise comparativa da produção relativamente à dimensão do corpo docente tem, como contrapartida do atrativo de transferir a avaliação do curso para a avaliação individual dos seus docentes, a desvantagem de opor-se ao natural estímulo à expansão das instituições bem-sucedidas. Para levar em conta o tamanho sem inibir o emprego de recursos, a avaliação probabilística pode ser dirigida ao acompanhamento da evolução dos resultados ao longo do tempo. Tal acompanhamento é aqui realizado adotando o ponto de vista implícito nos índices de Malmquist (1953), de quantificar a evolução de cada curso levando em conta o contexto, isto é, medindo a evolução da sua posição relativa aos outros cursos.

Em análises anteriores como as de Almeida (2001), Sant'Anna (2003) e Estellita Lins et al. (2004), demonstrou-se a grande homogeneidade dos cursos dessa área, sobretudo quanto à produtividade. Dessa homogeneidade resulta que a classificação dos cursos muda consideravelmente se muda o ponto de vista da avaliação. Verifica-se, também, nos dados ora disponíveis, que a evolução, seja das medidas de produtividade seja das medidas absolutas de produção, oscila muito se confrontamos dados de um ano com os do ano seguinte. Quando se consideram dados agregados de períodos mais longos, variações acidentais são absorvidas e esta flutuação se reduz. Por esse motivo, este artigo considera períodos trienais na análise da evolução ao longo do tempo.

Na seção seguinte, é apresentada a estrutura básica da composição probabilística. Na Seção 3, é apresentado o índice de Malmquist probabilístico. Na seção 4, desenvolve-se a análise dos cursos e a Seção 5 apresenta comentários finais.

2. Pontos de vista probabilísticos

Para comparar unidades produtivas de acordo com a produtividade, a abordagem mais empregada atualmente é a da Análise Envoltória de Dados (DEA), desenvolvida inicialmente por Charnes et al. (1978) com base no conceito de eficiência de Farrell (1957). A DEA compara as unidades pela eficiência em extrair o maior agregado de produtos possível do menor agregado de recursos utilizados. Algumas características tornam a DEA um instrumento atraente: a eficiência é medida realisticamente, em termos de distância aos melhores desempenhos efetivamente observados; para levar em conta que cada unidade pode ter o seu próprio nicho de atuação, a agregação dos recursos e a agregação dos produtos são efetuadas utilizando, para cada unidade avaliada, os pesos que lhe sejam mais favoráveis; algoritmos distintos podem ser empregados para analisar a situação em que as unidades de operação atendem a encomendas cujo volume é determinado fora do seu campo de decisão, de modo que seus esforços

para elevar a produtividade são orientados para a minimização do volume de recursos utilizados, e para analisar a situação oposta, em que os recursos disponíveis estão fora do campo de decisão da unidade produtiva e a produtividade é dada pelo volume de produção extraída.

A composição probabilística apresenta essas mesmas características, substituindo-se distâncias absolutas por probabilidades de afastamento da fronteira e usando o cálculo de probabilidades conjuntas em vez do "minimax". Mais importante nessa abordagem que a precisa atribuição de probabilidades é a forma como essas probabilidades são combinadas para gerar as medidas de eficiência. Nesse sentido, ainda que a indisponibilidade de dados nos impeça de modelar estatisticamente as distribuições, podemos chegar a medidas de eficiência mais confiáveis que as extraídas dos dados tratados como determinísticos.

A abordagem probabilística, se, por um lado, eleva fortemente a medida de eficiência de qualquer unidade de produção sempre que apresente desempenho extremo seja na minimização do volume de algum recurso seja na maximização do volume de algum produto, por outro lado, ameniza a influência desses pontos extremos, levando em conta o desempenho de mais variáveis e de mais unidades de observação. Enquanto a fronteira de excelência tende a ser formada por desempenhos raros, a comparação em variáveis em que a unidade não apresente desempenho extremo e a comparação com um conjunto de observações com valores mais frequentes torna o procedimento de avaliação resistente a erros aleatórios.

Com o reconhecimento da presença de erros de medida aleatórios, os volumes de entradas e saídas observados passam a ser tratados como estimativas de parâmetros de posição de distribuições de probabilidades. Pode-se derivar, também, do conjunto de valores observados, estimativas para outros parâmetros dessas distribuições.

É difícil dispor de informação *a priori* sobre a forma da distribuição das perturbações aleatórias e, nas primeiras aplicações, não é comum dispor de um número de observações, em cada unidade, suficiente para, mesmo assumindo as habituais hipóteses de normalidade e independência, estimar, com precisão satisfatória, seus parâmetros de dispersão. Mais difícil, ainda, seria testar essas hipóteses. Neste trabalho se assume a hipótese clássica de perturbações independentes e de valores esperados nulos, com distribuição normal.

É interessante notar que, embora as variáveis assumam valores inteiros e positivos, adota-se aqui uma aproximação contínua para suavizar as avaliações das probabilidades de maximizar. A influência dos empates no caso de se adotar uma distribuição discreta é exemplificada em Sant'Anna (2005). É interessante notar, também, que as variáveis empregadas nessa aplicação são variáveis de contagem, que, freqüentemente, têm uma distribuição de Poisson ou binomial negativa. Na situação aqui considerada, entretanto, os números contados são de partes de uma população mais ampla de alunos, professores, artigos, etc., o que torna as distribuições muito mais simétricas do que ocorre no caso de contagem em um Processo de Poisson.

Em algumas áreas de aplicação da Estatística, em que se modelam relações entre variáveis, houve nas últimas décadas um importante crescimento de técnicas baseadas na generalização dos modelos baseados na distribuição normal. Testes de normalidade (Jarque e Bera, 1980), para séries univariadas longas, e ajustamento de modelos lineares generalizados (McCullagh e Nelder, 1989), quando se podem supor relações entre parâmetros das distribuições das diferentes variáveis envolvidas, tornaram-se rotineiros nessas áreas. Já no campo de aplicação que aqui se explora, de comparação entre valores esperados de distribuições diferentes, a escolha da distribuição é determinada por conjecturas sobre a assimetria e o peso das caudas das distribuições que não podem ser testadas.

Para modelar a dispersão, toma-se por base a amplitude observada. Estima-se o desvio-padrão da perturbação que afeta cada variável dividindo a amplitude do conjunto das medidas obtidas para essa variável pela razão, conhecida, entre o valor esperado da amplitude de uma amostra normal do tamanho observado e o desvio-padrão da população com distribuição normal de que essa amostra tenha sido retirada. Vantagens desse estimador sobre o desvio-padrão amostral, como ausência de tendência e menor erro esperado, são discutidas em Duncan (1955) e Woodall e Montgomery (1999), por exemplo.

Diferentes medidas podem ser construídas a partir da composição das probabilidades de atingir as fronteiras em cada variável, conforme utilizemos os conectivos “e” ou “ou” e conforme tomemos a fronteira inferior ou a fronteira superior como referência. O uso de formas diferentes de exigir que se atinja simultaneamente a maximização de saídas e a minimização de entradas permite avaliar os efeitos de pequenas diferenças. Uma classificação de pontos de vista em dois eixos subjetivos facilita a identificação dessas diferenças. Esses eixos são um eixo pessimista-otimista e um eixo conservador-progressista. Uma composição é otimista quando considera a probabilidade de atingir a fronteira de excelência (ou afastar-se da fronteira de pior desempenho) em pelo menos uma variável. E pessimista quando se baseia na exigência de aproximação da fronteira conjuntamente, em todas as variáveis. Por outro lado, uma composição é progressista quando toma como referência a fronteira de excelência e conservadora quando toma como referência a fronteira de pior desempenho.

Algebricamente, a abordagem acima desenvolvida pode ser formalizada considerando m unidades de produção U_1, \dots, U_m , nas quais medimos, com incerteza, volumes empregados de n entradas e volumes produzidos de p saídas. Representemos por (i_{j1}, \dots, i_{jn}) e (o_{j1}, \dots, o_{jp}) os vetores de valores observados para as n entradas e p saídas da unidade de produção U_j . Tratamos esses vetores como observações de vetores aleatórios (I_{j1}, \dots, I_{jn}) e (O_{j1}, \dots, O_{jp}) . Atribuímos às coordenadas desses vetores distribuições independentes. Ao volume empregado I_{j0i} da i -ésima entrada na unidade de produção U_{j0} atribuímos uma distribuição normal de valor esperado i_{j0i} e desvio-padrão $Rl_i/d_2(m)$ onde $Rl_i = \max_{j_{ji}} - \min_{j_{ji}}(1)$ é a amplitude observada para a i -ésima variável de entrada e $d_2(m)$ é a razão tabelada entre o valor esperado da amplitude da amostra normal de tamanho m e o desvio-padrão da distribuição amostrada. Analogamente, ao volume

produzido O_{js} da s -ésima saída em U_{jo} atribuímos uma distribuição normal de valor esperado o_{jos} e desvio-padrão $RO_s/d_2(m)$ onde $RO_s = \max_{j, js} o_{jos} - \min_{j, js} o_{jos}$ (2) é a amplitude observada em tal variável.

Para cada unidade de produção U_{jo} e cada variável de entrada i , sejam $M_{joi} = P[l_{joi} \leq l_{ji}'$, para todo j de 1 a m] (3) a probabilidade de a referida unidade de produção maximizar a utilização da referida variável e $m_{joi} = P[l_{joi} \geq l_{ji}'$, para todo j de 1 a m] (4) a probabilidade de tal unidade de produção minimizar a utilização da mesma variável. Analogamente, para as saídas.

Com essa notação, a avaliação probabilística conservadora e pessimista da unidade U_j será proporcional ao produto $\prod_s (1 - m_{js}) \prod_i (1 - M_{ji})$, para s denotando produtório com s variando ao longo de todas as saídas e i produtório com i variando ao longo de todas as entradas. Para facilitar a comparação, o resultado dessa operação é padronizado, dividindo por uma constante de modo que a avaliação da unidade de avaliação mais alta seja sempre igual a 1.

Essa abordagem probabilística constitui-se em um instrumento de composição de múltiplos critérios que se aplica tanto ao caso de dois grupos de critérios, dados, no presente contexto, respectivamente, pelos grupos das variáveis de resultados produzidos e das variáveis de recursos utilizados, quanto no caso de mais de dois grupos ou no caso de todas as variáveis em um mesmo grupo. Esse último caso será explorado neste trabalho, quando se concentre a avaliação nos resultados produzidos deixando de lado a relação com os volumes de recursos empregados.

3. Índice de Malmquist probabilístico

O índice de produtividade de Malmquist foi criado por Caves et al. (1982), inspirado na proposta de Malmquist (1953) de avaliar a evolução da produtividade de cada unidade de produção relativamente à evolução do conjunto de unidades em que se insere. O índice de Malmquist, além de controlar o efeito da evolução tecnológica, não leva em conta o valor atribuído a cada produto ou recurso. Färe et al. (1995) propuseram um índice com essas mesmas características, baseado na composição de escores de eficiência gerados pela Análise Envoltória de Dados, chamado índice de Malmquist-DEA.

Para levar em conta a evolução da fronteira de melhores desempenhos ao longo do tempo, o índice de Malmquist-DEA é obtido pela média geométrica de dois escores de evolução da eficiência, o primeiro calculado relativamente à fronteira do instante inicial e o segundo relativo à fronteira final. O primeiro é obtido dividindo pelo escore de eficiência do instante inicial da unidade avaliada a eficiência de uma unidade de produção hipotética adicionada ao conjunto do instante inicial em substituição à unidade avaliada e com os valores de recursos e produtos dessa unidade no instante seguinte. O segundo é, analogamente, a razão entre o escore de eficiência do instante final da unidade avaliada e o escore de eficiência de uma unidade incluída no conjunto de dados desse ano em substituição à mesma e com os valores de seus produtos e recursos no instante inicial.

Aqui, a mesma idéia é aplicada substituindo-se os escores da DEA pelos da avaliação probabilística. Assim, denotando-se por $P(j,t,u)$ a avaliação probabilística da unidade j quando os valores observados nessa unidade no instante t são introduzidos na análise em conjunto com os observados nas demais unidades no instante u , o índice de evolução de Malmquist probabilístico é dado pela raiz quadrada do produto das razões $P(j,t+1,t)/P(j,t,t)$ e $P(j,t+1,t+1)/P(j,t,t+1)$.

4. Análises comparativas

A avaliação probabilística exige, essencialmente, da etapa de modelagem, apenas a escolha das variáveis consideradas relevantes. Não há necessidade de fixar padrões de desempenho em cada critério e pesos para os critérios podem ser usados, caso sejam disponíveis, mas não são necessários. A avaliação pode ser feita em termos relativos às fronteiras do conjunto de dados observados e a combinação em termos de probabilidade de atingir a fronteira permite tratar todos os critérios, em princípio, em pé de igualdade. No caso, adotamos um modelo simples com apenas duas medidas de resultado, inicialmente. Mais adiante são adicionadas mais duas medidas de resultado e uma contrapartida, para a avaliação em termos de produção relativa ao volume de recurso empregado.

As variáveis consideradas são:

- primeira variável de quantidade de produção (*mestres*): número total de mestres formados pelo curso no triênio;
- primeira variável de qualidade da produção (*autores*): número máximo de docentes do NRD6 do curso apresentando resultados de pesquisa publicados na íntegra em anais de congressos em um mesmo ano;
- segunda variável de quantidade de produção (*doutores*): número total de doutores formados pelo curso no triênio;
- segunda variável de qualidade de produção (*artigos*): número total de artigos publicados pelos docentes do NRD6 do curso no triênio em periódicos de circulação internacional classificados nos níveis A ou B pelos critérios dos avaliadores da Capes;
- variável de recurso (NRD6): número máximo de docentes no NRD6 do curso no triênio; mais precisamente, o número aqui considerado é o máximo entre os três números informados nos relatórios anuais do triênio de avaliação.

Depois de o sistema de avaliação da Capes vir funcionando há bastante tempo, com pouca variação nos indicadores e com muitos indicadores relativos, verifica-se, dentro de cada área, considerável homogeneidade. Isso conduziu, em estudos anteriores, à constatação de grande variabilidade nas posições dos cursos em anos consecutivos resultante de flutuações acidentais, sobretudo quando se utilizam modelos com, não apenas produtos, mas, também, contrapartidas de recursos empregados. Para evitar

oscilações devidas a variações logo corrigidas, são empregadas neste estudo variáveis trienais. Os triênios de referência são os de 1998 a 2000 e de 2001 a 2003.

A Tabela 1 apresenta os valores registrados para essas variáveis nos triênios de avaliação de 1998 a 2000 e de 2001 a 2003 pelos 15 cursos de pós-graduação em Engenharia de Produção avaliados durante esses anos. Na Tabela 2, são apresentadas, na forma de porcentagem, as probabilidades de cada curso minimizar cada variável de resultado e maximizar a variável de recurso, supondo que os valores observados registrados na Tabela 1 representam valores esperados de distribuições normais com desvio-padrão estimado da maneira descrita na Seção 2.

Tabela 1 – Valores observados

Curso	NRD6 98-00	NRD6 01-03	Mestres 98-00	Mestres 01-03	Doutores 98-00	Doutores 01-03	Autores 98-00	Autores 01-03	Artigos 98-00	Artigos 01-03
UFMS	32	27	70	161	0	0	10	17	0	3
UFRJ	28	23	203	197	113	174	19	21	1	37
USP	25	25	59	58	38	48	19	21	1	10
UFSCar	21	27	62	82	0	6	19	25	1	11
UFF	16	20	57	92	0	0	11	14	0	11
Unimep	11	13	33	53	0	1	9	12	0	11
UFPE	11	12	27	57	0	1	9	11	1	4
UFRN	11	11	0	39	0	0	7	11	0	0
PUC – Rio	11	10	74	60	10	9	10	9	1	7
USPSC	10	15	52	83	0	0	8	12	1	4
UFPB	10	8	35	63	0	0	8	7	0	2
UFMG	9	7	39	52	0	0	9	6	0	1
Unip	8	11	33	67	0	0	8	9	0	9
UFRGS	8	9	42	46	0	7	8	8	2	10
Unifei	8	9	24	48	0	0	7	9	0	3

Tabela 2 – Probabilidades de apresentar o pior desempenho

Curso	NRD6 98-00	NRD6 01-03	Mestres 98-00	Mestres 01-03	Doutores 98-00	Doutores 01-03	Autores 98-00	Autores 01-03	Artigos 98-00	Artigos 01-03
UFSM	0%	0%	4%	0%	8%	8%	5%	1%	12%	9%
UFRJ	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%
USP	0%	0%	5%	8%	1%	2%	0%	0%	1%	4%
UFSCar	1%	0%	4%	4%	8%	7%	0%	0%	1%	3%
UFF	3%	1%	5%	3%	8%	8%	3%	3%	12%	3%
Unimep	7%	5%	8%	9%	8%	8%	7%	5%	12%	3%
UFPE	7%	6%	9%	8%	8%	8%	7%	6%	1%	8%
UFRN	7%	8%	16%	13%	8%	8%	14%	6%	12%	13%
PUC - Rio	7%	10%	3%	8%	5%	6%	5%	10%	1%	6%
USPSC	9%	3%	5%	4%	8%	8%	10%	5%	1%	8%
UFPB	9%	15%	8%	7%	8%	8%	10%	15%	12%	11%
UFMG	11%	18%	7%	9%	8%	8%	7%	18%	12%	12%
Unip	13%	8%	8%	6%	8%	8%	10%	10%	12%	5%
UFRGS	13%	12%	7%	11%	8%	7%	10%	12%	0%	4%
Unifei	13%	12%	10%	10%	8%	8%	14%	10%	12%	9%

Mais útil que a comparação em termos de produtividade pela inclusão do volume de recurso empregado é a comparação dinâmica em que cada curso tem a sua evolução ao longo do tempo avaliada, comparativamente à evolução do conjunto. Isto é feito, nesta análise, por meio do índice de Malmquist probabilístico definido na Seção 3. A estabilidade do sistema de avaliação facilita a realização desse tipo de comparação.

A Tabela 3 apresenta os valores das medidas de eficiência relativa calculadas em termos de otimização simultânea, considerando-se apenas dois produtos, mestres formados e docentes autores. Nas duas primeiras colunas estão os valores para os dois triênios das medidas derivadas da probabilidade de não minimizar nenhuma dessas duas variáveis. Nas duas colunas seguintes, as medidas extraídas das probabilidades de, além de não minimizar essas duas variáveis, simultaneamente não maximizar o volume de recurso medido pelo NRD6. E as duas últimas colunas apresentam os índices de Malmquist, primeiro para o caso de se considerarem apenas as duas variáveis de resultado e depois para o caso de se considerar, também, o volume de recurso docente. Finalmente, na Tabela 4, as mesmas medidas são apresentadas, acrescentando-se ao modelo as duas últimas variáveis de resultado, doutores formados e artigos publicados.

Tabela 3 – Avaliações com duas variáveis de resultado

Curso	98-00 só resultados	01-03 só resultados	98-00 com recurso	01-03 com recurso	Malmquist só resultados	Malmquist com recurso
UFSM	92%	99%	59%	80%	110%	133%
UFRJ	100%	100%	83%	97%	100%	117%
USP	95%	92%	88%	83%	100%	100%
UFSCar	96%	96%	96%	78%	103%	87%
UFF	92%	95%	98%	98%	107%	103%
Unimep	85%	88%	93%	96%	110%	110%
UFPE	84%	87%	91%	95%	111%	110%
UFRN	72%	82%	78%	90%	123%	123%
PUC - Rio	92%	84%	100%	93%	96%	96%
USPSC	85%	92%	93%	100%	113%	111%
UFPB	83%	80%	90%	88%	102%	102%
UFMG	86%	75%	94%	83%	93%	93%
Unip	83%	85%	90%	94%	110%	109%
UFRGS	84%	79%	91%	87%	101%	101%
Unifei	86%	75%	94%	83%	113%	102%

Tabela 4 – Avaliações com quatro variáveis de resultado

Curso	98-00 só resultados	01-03 só resultados	98-00 com recurso	01-03 com recurso	Malmquist só resultados	Malmquist com recurso
UFSM	76%	83%	51%	69%	120%	144%
UFRJ	100%	100%	87%	100%	111%	129%
USP	94%	87%	92%	81%	105%	105%
UFSCar	88%	87%	94%	73%	109%	93%
UFF	76%	84%	85%	90%	119%	115%
Unimep	70%	78%	80%	89%	123%	123%
UFPE	78%	73%	89%	83%	114%	113%
UFRN	59%	65%	68%	75%	123%	123%
PUC - Rio	88%	74%	100%	85%	99%	99%
USPSC	79%	78%	90%	87%	115%	113%
UFPB	68%	66%	78%	75%	110%	110%
UFMG	71%	60%	81%	69%	99%	100%
Unip	68%	75%	78%	85%	122%	121%
UFRGS	78%	71%	90%	81%	106%	106%
Unifei	71%	60%	81%	69%	109%	109%

Uma conclusão que se pode extrair do exame das Tabelas 3 e 4 é quanto à estabilidade do índice de Malmquist probabilístico. A inclusão de mais produtos e a inclusão de recurso alteram pouco a classificação dos cursos quanto à sua evolução de um triênio para o outro segundo esse índice. Os coeficientes de correlação de Spearman entre os vetores de classificação derivados dos dados de duas e de quatro variáveis de resultado, para os casos com e sem recurso, são muito elevados, de 94% e 92%, respectivamente. Quando se fixa o número de variáveis de resultado e se calcula a correlação entre as classificações obtidas incluindo o volume de recurso docente e sem a inclusão do mesmo, os coeficientes de correlação são ainda elevados, de 73% para o caso de duas variáveis e de 81% para o caso de quatro.

As avaliações estáticas mudam consideravelmente quando se inclui o recurso docente como contrapartida dos resultados gerados, de tal sorte que podemos associar à inclusão do recurso como variável explicativa dos resultados um efeito de instabilidade dos índices. Por exemplo, o coeficiente de correlação entre os vetores de classificação com dois produtos com e sem recurso é de, apenas, 32% no primeiro triênio e 20% no segundo.

Do mesmo modo, a inclusão do recurso resulta em redução do coeficiente de correlação de Spearman entre as classificações dos dois triênios, seja no caso da análise com duas variáveis de resultado seja no caso da análise com quatro. No primeiro caso, a correlação entre os vetores dos dois triênios é de 68% sem recurso e de apenas 16% com recurso. No caso de quatro variáveis de resultado, a queda é de 70% para 29%.

Examinando os desempenhos curso a curso, constata-se que o maior progresso segundo o índice de Malmquist, é apresentado pela UFSM. O índice é mais alto quando se inclui a variável de recurso, devido à redução de 32 para 27 docentes, enquanto na área, como um todo, o total de docentes manteve-se estável (com um crescimento inferior a 4%), mas é alto também sem a inclusão dessa variável. Isso é facilmente explicável pela elevação do número de mestres formados por essa instituição, que mais que duplicou de um triênio para o outro, enquanto o crescimento da área como um todo foi de 43%. Outros valores altos do índice podem ser facilmente explicados pela redução do NRD6, no caso da UFRJ, ou, no caso da UFRN, pelo aparecimento, no segundo triênio, da primeira turma de mestres titulados.

5. Conclusão

A aplicação aqui realizada demonstra como a composição probabilística pode ser empregada na avaliação de cursos. A composição probabilística permite, uma vez definidas as variáveis de interesse e determinado um ponto de vista em termos de fronteira desejada e forma de composição, classificar objetivamente os cursos e avaliar a sua evolução ao longo do tempo. O emprego de um instrumento de composição dessa natureza não reduz a importância da correta escolha dos critérios de avaliação e da análise dos componentes não-quantificáveis da qualidade, mas, libera os

especialistas para a crítica dos dados, ao mesmo tempo em que facilita o emprego das informações geradas na tomada de decisões, seja de quem demanda um curso, seja dos gestores do sistema.

Foi adotada neste estudo uma ótica pessimista, no sentido de avaliar em termos de satisfação simultânea de todos os critérios, e conservadora, no sentido de atribuir maior importância às variações mais próximas à fronteira formada pelos valores menos satisfatórios. Outros pontos de vista conduziriam a diferentes classificações, sobretudo quando se cotejem resultados e recursos empregados. Esse fato, de as medidas envolvendo o volume de recurso docente empregado apresentarem mais inversões de posições que as baseadas apenas nos resultados finais, reflete uma maior proximidade entre os cursos entre si quanto à produtividade dos docentes. Em cada instituição, uns docentes mais produtivos compensam outros menos produtivos de tal modo que os cursos satisfazem de forma semelhante os padrões *per capita* da área.

A variação das conclusões da análise com o ponto de vista adotado reflete a homogeneidade do conjunto de cursos analisados, que pode ser considerada um efeito da estabilidade do sistema de avaliação da Capes, a que os cursos vêm sendo submetidos há um tempo considerável. A divulgação dos documentos de área, especialmente, contribuiu para a homogeneização dos padrões. O confronto dos resultados da aplicação de diferentes pontos de vista pode, nessas circunstâncias, ser usado para detectar diferenças entre estratégias gerenciais imperceptíveis de outra forma.

Com a análise limitada a uma área do conhecimento, verificou-se que as conclusões das comparações iniciais, considerando apenas duas variáveis de desempenho, uma medindo quantidade de formados e outra medindo qualidade da formação, não foram grandemente alteradas quando se incluíram outras variáveis no modelo. O instrumental aqui apresentado aplica-se, entretanto, igualmente, se um número maior de variáveis for necessário. Da mesma forma, pode-se aumentar o número de objetos de comparação, sem dificuldade.

Para a mais clara interpretação das conclusões é recomendável que o número de possíveis valores de cada critério seja mantido pequeno. Isso pode ser feito, facilmente, discretizando as avaliações ou considerando idênticos valores próximos na reta numérica. Se o número de objetos de avaliação é elevado, a transformação das avaliações iniciais em probabilidades de atingir a fronteira pode ser baseada no cálculo das probabilidades em uma pequena amostra representativa do universo de valores observados.

Um último esclarecimento pode ser útil. O emprego de modelo de poucas variáveis e a busca de resultados quantitativos não significa que se pretenda desconhecer a complexidade do problema da avaliação dos cursos. Ao contrário, é o reconhecimento de que a complexidade do problema exige instrumentos de análise solidamente constituídos que suscitem o desenvolvimento dessa proposta de aplicação e mesmo de uma nova forma de composição de critérios capaz de extrair, de dados incompletos, significados objetivos.

6. Agradecimentos

1. Sou grato aos árbitros da RBPG, que apontaram importantes oportunidades de aperfeiçoamento deste trabalho, aproveitadas pelo autor nos limites da sua competência.

2. Este trabalho é resultado de projeto de pesquisa do autor, apoiado por bolsa de produtividade do CNPq e pela Capes por meio do Proap.

7. Referências

ALMEIDA, A. T.; MIRANDA, C. M. G. e FERRER, F. C. Avaliação de Programas de Pós-graduação com Método Multicritério. *Anais do XXXIII SBPO*. Rio de Janeiro: Sobrapo. 2001

CAPES. Dados sobre programas de pós-graduação. 2005. Disponível em <http://www1.capes.gov.br/Scripts/Servicos/Indicadores/Dados/FiltraArquivos.idc>, acesso em 4 fev. 2005.

CAVES D. W.; CHRISTENSEN L. R.; DIEWERT W. E. The Economic Theory of Index Numbers and the Measurement of Inputs, Outputs and Productivity. *Econometrica*, n. 50, p. 1.393-1.414, 1982.

CHARNES, A.; COOPER, W. W. e RHODES. E. Measuring the Efficiency of Decision Making Units. *European Journal of Operations Research*, n. 2, p. 429-444, 1978.

DUNCAN, A. J. The Use of the Range in Computing Variabilities. *Industrial Quality Control*, n. 9, p. 18-22, 1955.

ESTELLITA LINS, M. P.; ALMEIDA, B. F. e BARTHOLO JR., R. Avaliação de desempenho na pós-graduação utilizando a Análise Envoltória de Dados: o Caso da Engenharia de Produção. *Revista Brasileira de Pós-Graduação*, n. 1, p. 41-56, 2004.

FÄRE, R.; GROSSKOPF, S.; LINDGREN, B. e ROOS, P. Productivity Developments. In: SWEDISH HOSPITALS: a Malmquist Output Index Approach. In: CHARNES, A.; COOPER, W. W.; LEWIN, A. Y. e SEIFORD, L. M. (eds.). *Data Envelopment Analysis: Theory, Methodology and Applications*. Boston: Kluwer, 1995.

FARRELL, M. J. The Measurement of Productive Efficiency. _____: *Journal of the Royal Statistical Society*, n. 120, p. 449-460, 1957.

JARQUE, C. M. e BERA, A. K. Efficient Tests for Normality, Heteroskedasticity and Serial Independence of Regression Residuals, *Economic Letters*, n. 6, p. 255-259, 1980.

MCCULLOUGH e NELDER. *Generalized Linear Models*. London: Chapman and Hall, 1989.

MALMQUIST S. Index Numbers and Indifference Surfaces. *Trabajos de Estadística*, n. 4, p. 209-242, 1953.

SANT'ANNA, A. P. Cálculo Probabilístico de Produtividades Globais. *Anais do XXXIV SBPO*. Rio de Janeiro: Sobrapo, 2002.

SANT'ANNA, A.P. Cálculo Probabilístico de Produtividades Globais no Ensino de Pós-Graduação em Engenharia de Produção. *Anais do IX Encontro de Educação em Engenharia*. Petrópolis, 2003.

_____. Rough Sets in the Probabilistic Composition of Preference Evaluations. In: DE BAETS, B.; DE CALUWE, R., DE TRÉ, G., FODOR, J., KACPRZYK, J. e ZADROZNY, S. (eds.) *Current Issues in Data and Knowledge Engineering*, Varsóvia: *Exit*, p. 407-414, 2004.

_____: Weighting Points of View in the Composition of Correlated Criteria of Variable Importance. *Anales de la XVI EPIO*. Cordoba, 2005.

WOODALL, W. H. e MONTGOMERY, D. C. Using Ranges to estimate Variability. *Quality Engineering*, n. 13, p. 211-217, 1999.