Data Envelopment Analysis with Maple in Operations Research and Modeling Courses

William C. Bauldry
Appalachian State University
Boone, NC 28608
BauldryWC@appstate.edu

Introduction

Data envelopment analysis (DEA), invented by Charnes, Cooper, and Rhodes in 1978, is a technique that assesses performance of disparate units, called *decision management units* (DMU's), in an organization relative to a set of input and output measures. Their initial nonlinear formulation in terms of "efficiency" was modeled on the definition from "combustion engineering, 'efficiency is the ratio of the actual amount of heat liberated . . . to the maximum amount which could be liberated.'" (Charnes, Cooper, and Rhodes, 1978). Subsequently, they developed linear models, optimizing via the simplex algorithm giving a connection to operations research or mathematical modeling courses.

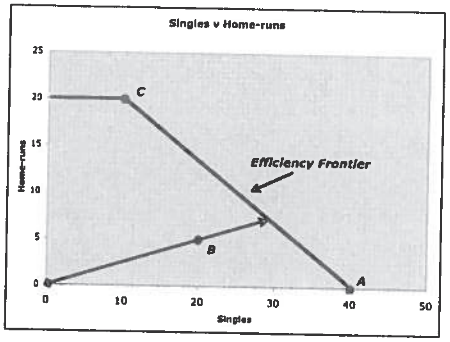
The strengths of DEA include: Multiple input and multiple output models; comparisons are against combinations of peers; both inputs and outputs can have very different units. The weaknesses of DEA include: Data noise can cause significant errors; estimates relative, not absolute efficiency; computationally intensive. Carefully applied, DEA is a very powerful tool for operations research.

A Standard First Example

Suppose we have three baseball players who are eligible to be traded, but can only choose one to release. Batting data appears in the table of Figure 1(a). Looking at the data we see that no affine combination of B and C can equal A. Also, no affine combination of A and B can equal C. However, $B = 44\%A + 25\%C$ for a 69% "DEA efficiency index." We'll trade B. This is the essence of DEA: compare the outputs of each unit relative to its peers. Graphically, DEA corresponds to computing an *efficiency frontier* (approximately a convex hull) and measuring a DMU's distance from the frontier. DEA has been called "frontier analysis." Figure 1(b) shows that B is 69% of the way to the frontier.

Player	Singles	Home Runs
A	40	0
B	20	5
C	10	20

(a) Baseball Data Table



(b) Graphical Analysis

Figure 1: Standard First Example

Formal Definition of Data Envelopment Analysis

We give the model definitions and lexicon for two forms of DEA. The first representation optimizes the ratio of output to input; the nonlinear form was the first developed by Charnes et al. The second is linear and maximizes closeness to the frontier relative to optimal use of inputs.

Nonlinear Representation (*'I-O Ratio'*)

Let:

n = number of *DMU*s
n_{in} = number of input measures
u_i = weight factor for input i
x_{ik} = input i for DMU_k
n_{out} = number of output measures
v_i = weight factor for output i
y_{ik} = output i for DMU_k
e_k = efficiency of DMU_k

For each DMU_j, $j = 1..n$, define the nonlinear program:

$$\textit{Choose } \vec{u}, \vec{v} \textit{ to maximize } e_j \textit{ subject to}$$

$$\begin{array}{rcl} e_k = \sum_{i=1}^{n_{\text{out}}} v_i y_{ik} \Big/ \sum_{i=1}^{n_{\text{in}}} u_i x_{ik} & & k = 1..n \\ 0 \le e_k \le 100\% & & k = 1..n \\ u_k \ge 0 & & k = 1..n_{\text{in}} \\ v_k \ge 0 & & k = 1..n_{\text{out}} \end{array}$$

Linear Representation (*'Input Oriented'*)

For each DMU, $i = 1..n$, define the linear program

$$\textit{Choose } \vec{\lambda} \geq \vec{0} \textit{ to minimize } \theta \textit{ subject to}$$

$$\vec{X}_{\text{Inputs}} \cdot \vec{\lambda} - \vec{X}_i \theta \le 0$$

$$\vec{M}_{\text{Outputs}_j} \cdot \vec{\lambda} - \vec{M}_{\text{Outputs}_{j,i}} \ge 0 \quad j = 1..n$$

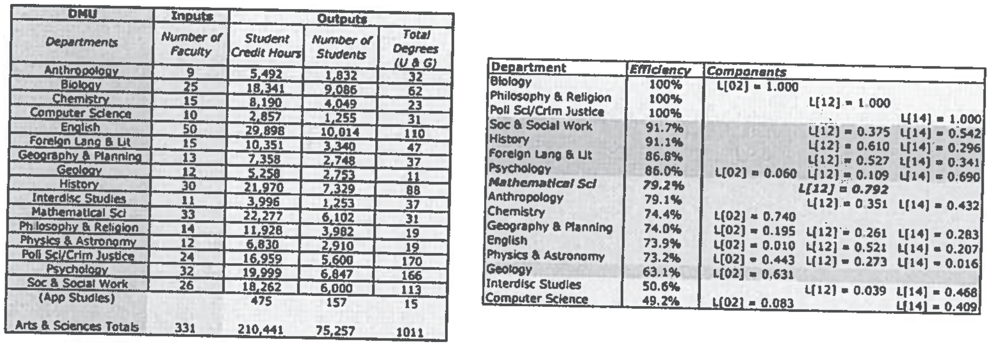
The value of θ gives the efficiency ranking of the DMU.

Naturally, the linear approach is both much easier to work with and to compute. An alternate linear approach emphasizes output measures so is called *output oriented.*

Class Project: DEA of ASU's College of Arts & Sciences using Maple

There have been a number of studies of academic units using DEA (see Tavares, 2002). Most papers present analyses of a specific discipline's departments, such as mathematics, across a selection of universities (see, e.g., Beasley, 1990). Some have attempted to analyze dissimilar departments within a division (see, e.g., Tayagi et al, 2009). Our class considered Appalachian's College of Arts & Sciences.

Task: Analyze the 16 disparate departments of Appalachian's College of Arts & Sciences.
We will analyze the college using a single input *number of faculty lines* and three outputs *student credit hours generated, number of majors*, and *number of degrees awarded.* The data used was obtained from Appalachian's Office of Institutional Research, Assessment & Planning's *ASU Fact Book* for Fall, 2006. The input and output data collected is shown in Figure 2(a).



DMU	Inputs	Outputs		
Departments	*Number of Faculty*	*Student Credit Hours*	*Number of Students*	*Total Degrees (U & G)*
Anthropology	9	5,492	1,832	32
Biology	25	18,341	9,086	62
Chemistry	15	8,190	4,049	23
Computer Science	10	2,857	1,255	31
English	50	29,898	10,014	110
Foreign Lang & Lit	15	10,351	3,340	47
Geography & Planning	13	7,358	2,748	37
Geology	12	5,258	2,753	11
History	30	21,970	7,329	88
Interdisc Studies	11	3,996	1,253	37
Mathematical Sci	33	22,277	6,102	31
Philosophy & Religion	14	11,928	3,982	19
Physics & Astronomy	12	6,830	2,910	19
Poli Sci/Crim Justice	24	16,959	5,600	170
Psychology	32	19,999	6,847	166
Soc & Social Work	26	18,262	6,000	113
(App Studies)		475	157	15
Arts & Sciences Totals	331	210,441	75,257	1011

(a) Initial Data

Department	*Efficiency*	*Components*		
Biology	100%	L[02] = 1.000		
Philosophy & Religion	100%		L[12] = 1.000	
Poli Sci/Crim Justice	100%			L[14] = 1.000
Soc & Social Work	91.7%		L[12] = 0.375	L[14] = 0.542
History	91.1%		L[12] = 0.610	L[14] = 0.296
Foreign Lang & Lit	86.8%		L[12] = 0.527	L[14] = 0.341
Psychology	86.0%	L[02] = 0.060	L[12] = 0.109	L[14] = 0.690
Mathematical Sci	***79.2%***		***L[12] = 0.792***	
Anthropology	79.1%		L[12] = 0.351	L[14] = 0.432
Chemistry	74.4%	L[02] = 0.740		
Geography & Planning	74.0%	L[02] = 0.195	L[12] = 0.261	L[14] = 0.283
English	73.9%	L[02] = 0.010	L[12] = 0.521	L[14] = 0.207
Physics & Astronomy	73.2%	L[02] = 0.443	L[12] = 0.273	L[14] = 0.016
Geology	63.1%	L[02] = 0.631		
Interdisc Studies	50.6%		L[12] = 0.039	L[14] = 0.468
Computer Science	49.2%	L[02] = 0.083		L[14] = 0.409

(b) DEA Results

Figure 2: Arts & Sciences DEA. *Data Source: Inst. Research, Assess., & Planning, ASU*

The class used the Maple code shown in Table 1 below for the Arts & Sciences DEA; it is simple, short, and straightforward. We simply loop through the DMU's, solving the associated linear program. The 'minimize' function from Maple's 'simplex' package was chosen; we could have used the 'LPSolve' function from the 'Optimization' package. The results, sorted by efficiency ranking, are displayed in Figure 2(b). This code scales easily to larger problems.

```
DMU := ["Anthropology", "Biology", ..., "Soc & Social Work"]:
N := nops(DMU):
MO := Matrix([[610, 204, 3.56], ..., [702, 231, 4.35]]):
eq1 := sum(λ[i],i=1..N) - θ ≤ 0:
OV := Vector[row](N,symbol=λ).MO:
Results := NULL:
for n from 1 to N do
    eq2 := OV[1] - M[n,1] ≥ 0:   # credit hours
    eq3 := OV[2] - M[n,2] ≥ 0:   # number of students
    eq4 := OV[3] - M[n,3] ≥ 0:   # degrees awarded
    s := simplex[minimize](θ, {eq||(1..4)}, NONNEGATIVE);
    Results := Results, [DMU[n],s))];
  end do:
Results;
```

Table 1: Maple code for the Arts & Sciences DEA

Even this simple DEA model provides more depth than a university's typical comparison of *number of faculty* to *total student credit hours generated.*

A Fun Student Project

Have a group of students choose a college, division, or school. Then:

Identify DMUs Departments, academic areas, or groups

Define inputs Numbers of: faculty by rank; non-tenure-track faculty by rank; graduate GTAs, RAs, fellows; staff; operating budget; classrooms; laboratories; offices; total assignable square feet; etc. (14 *inputs*)

Define outputs Number of: majors; degrees awarded; sections offered; student credit hours produced; publications; number of conference presentations; grant proposals submitted, awarded; external committee service; professional organization offices held; etc. (10 *outputs*)

Perform a data envelopment analysis

Give a copy to the Dean. Run. Hide.

Conclusion

Data envelopment analysis gives a very powerful tool to decision makers in an organization. DEA is then a natural choice for an operations research or mathematical modeling course. Simple projects can give students practice and insight. A larger project, for instance, analyzing a college, makes for significant group work that can be useful beyond the classroom. A full report would require sensitivity analysis and interpretation of the results along with a discussion of the limitations imposed by the input and output measures chosen. A semester-long DEA project can ties the various parts of an operations research together very nicely. For a DEA project to be generally accepted by a college as a valid decision-informing tool, each stakeholder, that is, all the DMU's, must be involved in the choices of input and output measures and the weighting for each.

Bibliography

- Beasley (1990), "Comparing University Departments," *Omega* **18**, no. 2, 171–183.
- Charnes, Cooper, and Rhodes (1978), "Measuring the efficiency of decision making units," *Eur. J. Opl. Res.* **2**, 429–444.
- Charnes, Cooper, Lewin, and Seiford (Eds.) (1995), *Data Envelopment Analysis: Theory, Methodology and Applications*, Springer.
- Tavares (2002), *A Bibliography Of Data Envelopment Analysis (1978–2001),* RUTCOR Research Report, RRR 01-02. Rutgers University. (3203 entries)
  Available at http://rutcor.rutgers.edu/pub/rrr/reports2002/1_2002.pdf
- Tyagi, Yadav, and Singh (2009), "Relative performance of academic departments using DEA with sensitivity analysis," *Evaluation and Program Planning* **32**,168–177.