Data Envelopment Analysis (DEA)

Data envelopment analysis (DEA) is a linear programming methodology to measure the efficiency of multiple decision-making units (DMUs) such as organizations or firms when the production process presents a structure of multiple inputs and outputs.

DEA has been used for both production and cost data. Utilizing the selected variables, such as unit cost and output, DEA software searches for the points with the lowest unit cost for any given output, connecting those points to form the efficiency frontier. Any company not on the frontier is considered inefficient. A numerical coefficient is given to each firm, defining its relative efficiency. Different variables that could be used to establish the efficiency frontier are: number of employees, service quality, environmental safety, and fuel consumption.

Although DEA has a strong link to production theory in economics, the tool is also used for benchmarking in operations management, where a set of measures is selected to benchmark the performance of manufacturing and service operations. The framework has been adapted from multi-input, multi-output production functions and applied in many industries. DEA develops a function whose form is determined by the most efficient producers. DEA identifies a "frontier" which is characterized as an extreme point method that assumes that if a firm can produce a certain level of output utilizing specific input levels, another firm of equal scale should be capable of doing the same. The most efficient producers can form a 'composite producer', allowing the computation of an efficient solution for every level of input or output. Where there is no actual corresponding firm, 'virtual producers' are identified to make comparisons".

Reference: Banker R.D. (1984). "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis". *Management Science*. **30**: 1078–1092. *doi*:*10.1287/mnsc.30.9.1078*.

PROBLEMS

Group A

1 Use the dual simplex method to solve the following LP:

$$\begin{aligned} \max z &= -2x_1 - x_3 \\ \text{s.t.} \quad x_1 + x_2 - x_3 &\geq 5 \\ x_1 - 2x_2 + 4x_3 &\geq 8 \\ x_1, x_2, x_3 &\geq 0 \end{aligned}$$

2 In solving the following LP, we obtain the optimal tableau shown in Table 45.

$$\begin{aligned} \max z &= 6x_1 + x_2 \\ \text{s.t.} \quad x_1 + x_2 &\leq 5 \\ 2x_1 + x_2 &\leq 6 \\ x_1, x_2 &\geq 0 \end{aligned}$$

a Find the optimal solution to this LP if we add the constraint $3x_1 + x_2 \leq 10$.

b Find the optimal solution if we add the constraint $x_1 - x_2 \geq 6$.

TABLE 45

	Basic Variable
$z + 2x_2 + 3s_2 = 18$	$z_1 = 18$
$0.5x_2 + s_1 - 0.5s_2 = 2$	$s_1 = 2$
$x_1 + 0.5x_2 + 0.5s_2 = 3$	$x_1 = 3$

c Find the optimal solution if we add the constraint $8x_1 + x_2 \leq 12$.

3 Find the new optimal solution to the Dakota problem if only 20 board ft of lumber are available.

4 Find the new optimal solution to the Dakota problem if 15 carpentry hours are available.

6.12 Data Envelopment Analysis†

Often we wonder if a university, hospital, restaurant, or other business is operating efficiently. The **Data Envelopment Analysis (DEA) method** can be used to answer this question. Our presentation is based on Callen (1991). To illustrate how DEA works, let's consider a group of three hospitals. To simplify matters, we assume that each hospital "converts" two inputs into three different outputs. The two inputs used by each hospital are

Input 1 = capital (measured by the number of hospital beds)

Input 2 = labor (measured in thousands of labor hours used during a month)

The outputs produced by each hospital are

Output 1 = hundreds of patient-days during month for patients under age 14

Output 2 = hundreds of patient-days during month for patients between 14 and 65

Output 3 = hundreds of patient-days during month for patients over 65

Suppose that the inputs and outputs for the three hospitals are as given in Table 46.

To determine whether a hospital is efficient, let's define t_r = price or value of one unit of output r and w_s = cost of one unit of input s. The *efficiency* of hospital i is defined to be

$$\frac{\text{value of hospital } i\text{'s outputs}}{}$$

For the data in Table 46, we find the efficiency of each hospital to be as follows:

$$\text{Hospital 1 efficiency} = \frac{9t_1 + 4t_2 + 16t_3}{5w_1 + 14w_2}$$

$$\text{Hospital 2 efficiency} = \frac{5t_1 + 7t_2 + 10t_3}{8w_1 + 15w_2}$$

$$\text{Hospital 3 efficiency} = \frac{4t_1 + 9t_2 + 13t_3}{7w_1 + 12w_2}$$

†This section may be omitted without loss of continuity.

TABLE 46
Inputs and Outputs for Hospitals

	Inputs		Outputs		
Hospital	1	2	1	2	3
1	5	14	9	4	16
2	8	15	5	7	10
3	7	12	4	9	13

The DEA approach uses the following four ideas to determine if a hospital is efficient.

1 No hospital can be more than 100% efficient. Thus, the efficiency of each hospital must be less than or equal to 1. For hospital 1, we find that $(9t_1 + 4t_2 + 16t_3)/(5w_1 + 14w_2) \leq 1$. Multiplying both sides of this inequality by $(5w_1 + 14w_2)$ (this is the trick we used to simplify blending constraints in Section 3.8!) yields the LP constraint $5w_1 + 14w_2 - 9t_1 - 4t_2 - 16t_3 \geq 0$.

2 Suppose we are interested in evaluating the efficiency of hospital i. We attempt to choose output prices (t_1, t_2, and t_3) and input costs (w_1 and w_2) that maximize efficiency. If the efficiency of hospital i equals 1, then it is efficient; if the efficiency is less than 1, then it is inefficient.

3 To simplify computations, we may scale the output prices so that the cost of hospital i's inputs equals 1. Thus, for hospital 2 we add the constraint $8w_1 + 15w_2 = 1$.

4 We must ensure that each input cost and output price is strictly positive. If, for example, $t_i = 0$, then DEA could not detect an inefficiency involving output i; if $w_j = 0$, then DEA could not detect an inefficiency involving input j.

Points (1)–(4) lead to the following LPs for testing the efficiency of each hospital.

Hospital 1 LP

$$
\begin{array}{llr}
\max z = 9t_1 + 4t_2 + 16t_3 & & (1)\\
\text{s.t.} & -9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0 & (2)\\
& -5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0 & (3)\\
& -4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0 & (4)\\
& 5w_1 + 14w_2 = 1 & (5)\\
& t_1 \geq .0001 & (6)\\
& t_2 \geq .0001 & (7)\\
& t_3 \geq .0001 & (8)\\
& w_1 \geq .0001 & (9)\\
& w_2 \geq .0001 & (10)
\end{array}
$$

Hospital 2 LP

$$
\begin{array}{llr}
\max z = 5t_1 + 7t_2 + 10t_3 & & (1)\\
\text{s.t.} & -9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0 & (2)\\
& -5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0 & (3)\\
& -4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0 & (4)\\
& 8w_1 + 15w_2 = 1 & (5)\\
& t_1 \geq .0001 & (6)
\end{array}
$$

$$
\begin{aligned}
t_2 \qquad\qquad\qquad &\geq .0001 \quad (7)\\
t_3 \qquad\qquad &\geq .0001 \quad (8)\\
w_1 \qquad &\geq .0001 \quad (9)\\
w_2 &\geq .0001 \quad (10)
\end{aligned}
$$

Hospital 3 LP

$$
\begin{aligned}
\max z &= 4t_1 + 9t_2 + 13t_3 && (1)\\
\text{s.t.} \quad & -9t_1 - 4t_2 - 16t_3 + 5w_1 + 14w_2 \geq 0 && (2)\\
& -5t_1 - 7t_2 - 10t_3 + 8w_1 + 15w_2 \geq 0 && (3)\\
& -4t_1 - 9t_2 - 13t_3 + 7w_1 + 12w_2 \geq 0 && (4)\\
& 7w_1 + 12w_2 = 1 && (5)\\
& t_1 \geq .0001 && (6)\\
& t_2 \geq .0001 && (7)\\
& t_3 \geq .0001 && (8)\\
& w_1 \geq .0001 && (9)\\
& w_2 \geq .0001 && (10)
\end{aligned}
$$

Let's see how the hospital 1 LP incorporates points (1)–(4). Point (1) maximizes the efficiency of hospital 1. This is because Constraint (5) implies that the total cost of hospital 1's inputs equal 1. Constraints (2)–(4) ensure that no hospital is more than 100% efficient. Constraints (6)–(10) ensure that each input cost and output price is strictly positive (the .0001 right-hand side is arbitrary; any small positive number may be used).

The LINDO output for these LPs is given in Figures 10(a)–(c). From the optimal ob-

```
MAX 9 T1 + 4 T2 + 16 T3
SUBJECT TO
       2)  - 9 T1 - 4 T2 - 16 T3 + 5 W1 + 14 W2 >=   0
       3)  - 5 T1 - 7 T2 - 10 T3 + 8 W1 + 15 W2 >=   0
       4)  - 4 T1 - 9 T2 - 13 T3 + 7 W1 + 12 W2 >=   0
       5)    W1 >=    0.0001
       6)    W2 >=    0.0001
       7)    T1 >=    0.0001
       8)    T2 >=    0.0001
       9)    T3 >=    0.0001
      10)    5 W1 + 14 W2 =    1
END

LP OPTIMUM FOUND AT STEP       6

        OBJECTIVE FUNCTION VALUE

        1)    1.00000000

VARIABLE          VALUE          REDUCED COST
      T1           .110889           .000000
      T2           .000100           .000000
      T3           .000100           .000000
      W1           .000100           .000000
      W2           .071393           .000000


     ROW     SLACK OR SURPLUS      DUAL PRICES
      2)           .000000         -1.000000
      3)           .515548           .000000
      4)           .411659           .000000
      5)           .000000           .000000
      6)           .071293           .000000
      7)           .110789           .000000
      8)           .000000           .000000
      9)           .000000           .000000
     10)           .000000          1.000000

NO. ITERATIONS=        6
```

FIGURE 10(a)
Hospital 1 LP

```
MAX   5 T1 + 7 T2 + 10 T3
SUBJECT TO
       2)  - 9 T1 - 4 T2 - 16 T3 + 5 W1 + 14 W2 >=    0
       3)  - 5 T1 - 7 T2 - 10 T3 + 8 W1 + 15 W2 >=    0
       4)  - 4 T1 - 9 T2 - 13 T3 + 7 W1 + 12 W2 >=    0
       5)    8 W1 + 15 W2 =    1
       6)    W1   >=  0.0001
       7)    W2   >=  0.0001
       8)    T1   >=  0.0001
       9)    T2   >=  0.0001
      10)    T3   >=  0.0001
END

LP OPTIMUM FOUND AT STEP        0

        OBJECTIVE FUNCTION VALUE

        1)    .773030000

VARIABLE          VALUE          REDUCED COST
      T1            .079821            .000000
      T2            .053275            .000000
      T3            .000100            .000000
      W1            .000100            .000000
      W2            .066613            .000000


     ROW     SLACK OR SURPLUS      DUAL PRICES
      2)            .000000          -.261538
      3)            .226970           .000000
      4)            .000000          -.661538
      5)            .000000           .773333
      6)            .000000          -.248206
      7)            .066513           .000000
      8)            .079721           .000000
      9)            .053175           .000000
     10)            .000000         -2.784615

NO. ITERATIONS=        0
```

FIGURE 10(b)
Hospital 2 LP

jective function value to each LP we find that

$$\text{Hospital 1 efficiency} = 1$$
$$\text{Hospital 2 efficiency} = .773$$
$$\text{Hospital 3 efficiency} = 1$$

Thus we find that hospital 2 is inefficient and hospitals 1 and 3 are efficient.

REMARK **1** An easy way to create the hospital 2 LP is to use LINDO to modify the objective function of the hospital 1 LP and the constraint $5w_1 + 14w_2 = 1$. Then it is easy to modify the hospital 2 LP to create the hospital 3 LP.

Using LINGO to Run a DEA

DEA.lng

The following LINGO program (see file DEA.lng) will solve our hospital DEA problem. When faced with another DEA problem, we begin by changing the numbers of inputs, outputs, and units. Next we change the resource usage and outputs for each unit. Finally, by changing number to (say) 1, we can evaluate the efficiency of unit 1. If the optimal objective function value for unit 1 is less than 1, then unit 1 is inefficient. Otherwise, unit 1 is efficient.

```
SETS:
INPUTS/1..2/:COSTS;
OUTPUTS/1..3/:PRICES;
```

```
MAX 4 TI + 9 + T2 + 13 T3
SUBJECT TO
       2) - 9 T1 - 4 T2 - 16 T3 + 5 W1 + 14 W2 >= 0
       3) - 5 T1 - 7 T2 - 10 T3 + 8 W1 + 15 W2 >= 0
       4) - 4 T1 - 9 T2 - 13 T3 + 7 W1 + 12 W2 >= 0
       5)   W1 >=   0.0001
       6)   W2 >=   0.0001
       7)   T1 >=   0.0001
       8)   T2 >=   0.0001
       9)   T3 >=   0.0001
      10)   7 W1 + 12 W2 = 1
END

LP OPTIMUM FOUND AT STEP 7

       OBJECTIVE FUNCTION VALUE

       1)     1.00000000

 VARIABLE        VALUE           REDUCED COST
       T1          .099815           .000000
       T2          .066605           .000000
       T3          .000100           .000000
       W1          .000100           .000000
       W2          .083275           .000000


      ROW   SLACK OR SURPLUS      DUAL PRICES
       2)          .000000           .000000
       3)          .283620           .000000
       4)          .000000         -1.000000
       5)          .000000           .000000
       6)          .083175           .000000
       7)          .099715           .000000
       8)          .066505           .000000
       9)          .000000           .000000
      10)          .000000          1.000000

NO. ITERATIONS=       7
```

FIGURE 10(c)
Hospital 3 LP

```
UNITS/1..3/;
UNIN(UNITS,INPUTS):USED;
UNOUT(UNITS,OUTPUTS):PRODUCED;
ENDSETS
NUMBER=2;
@FOR(UNITS(J)|j#EQ#NUMBER:MAX=@SUM(OUTPUTS(I):PRICES(I)*PRODUCED(J,I)));
@FOR(UNITS(J)|J#EQ#NUMBER:@SUM(INPUTS(I):COSTS(I)*USED(J,I))=1);
@FOR(INPUTS(I):COSTS(I)>=.0001);
@FOR(OUTPUTS(I):PRICES(I)>=.0001);
@FOR(UNITS(I):@SUM(INPUTS(J):COSTS(J)*USED(I,J))>=@SUM(OUTPUTS(J):PRICES(J)*PRODUCED(I,J)
)
);
DATA:
USED=5,14,
     8,15,
     7,12;
PRODUCED=9,4,16,
         5,7,10,
         4,9,13;
ENDDATA
END
```

Dual Prices and DEA

The DUAL PRICES section of the LINDO output gives us great insight into Hospital 2's (or any organization's found inefficient by DEA) inefficiency. Consider all hospitals whose efficiency constraints have nonzero dual prices in the hospital 2 LP (Figure 10b). (In our example, hospitals 1 and 3 have nonzero dual prices.) If we average the output vectors and input vectors for these hospitals (using the absolute value of the dual price for each hospital as the weight) we obtain the following:

Averaged Output Vector

$$.261538\begin{bmatrix}9\\4\\16\end{bmatrix} + .661538\begin{bmatrix}4\\9\\13\end{bmatrix} = \begin{bmatrix}5\\7\\12.785\end{bmatrix}$$

Averaged Input Vector

$$.261538\begin{bmatrix}5\\14\end{bmatrix} + .661538\begin{bmatrix}7\\12\end{bmatrix} = \begin{bmatrix}5.938\\11.6\end{bmatrix}$$

Suppose we create a composite hospital by combining .261538 of hospital 1 with .661538 of hospital 3. The averaged output vector tells us that the composite hospital produces the same amount of outputs 1 and 2 as hospital 2, but the composite hospital produces $12.785 - 10 = 2.785$ more of output 3 (patient days for more than 65 patients). From the averaged input vector for the composite hospital, we find that the composite hospital uses less of each input than does hospital 2. We now see exactly where hospital 2 is inefficient!

By the way, the objective function value of .7730 for the hospital 2 LP implies that the more efficient composite hospital produces its superior outputs by using at most 77.30% as much of each input. Note that

$$\text{Input 1 used by composite hospital} < .7730 * (\text{Input 1 used by hospital 2}) = 6.2186$$

and

$$\text{Input 2 used by composite hospital} = .7730 * (\text{Input 2 used by hospital 2}) = 11.6$$

An explanation of why the dual prices are needed to find a composite hospital that is superior to an inefficient hospital is given in Problems 5–7.

PROBLEMS

Group A

1 The Salem Board of Education wants to evaluate the efficiency of the town's four elementary schools. The three outputs of the schools are defined to be

Output 1 = average reading score
Output 2 = average mathematics score
Output 3 = average self-esteem score

The three inputs to the schools are defined to be

Input 1 = average educational level of mothers (defined by highest grade completed—12 = high school graduate; 16 = college graduate, and so on).
Input 2 = number of parent visits to school (per child)
Input 3 = teacher to student ratio

The relevant information for the four schools is given in Table 47. Determine which (if any) schools are inefficient. For any inefficient school, determine the nature of the inefficiency.

TABLE 47

	Inputs			Outputs		
School	1	2	3	1	2	3
1	13	4	.05	9	7	6
2	14	5	.05	10	8	7
3	11	6	.06	11	7	8
4	15	8	.08	9	9	9

2 Pine Valley Bank has three branches. You have been assigned to evaluate the efficiency of each. The following inputs and outputs are to be used for the study.

Input 1 = labor hours used (hundreds per month)
Input 2 = space used (in hundreds of square feet)
Input 3 = supplies used per month (in dollars)