Lecture – 1

DATA ENVELOPMENT ANALYSIS - I
Learning objective

- To demonstrate the concept of data Envelopment Analysis to compare the performance of service units
11.1 Data Envelopment Analysis

Data envelopment analysis (DEA) is a technique used to compare the performances of several units. These units in the context of services can be various service organizations like banks, hospitals, schools etc. This technique is used in places where a relative performance of different units is to be compared and evaluated.

- DEA can be used to analyse the performance of several units to set a benchmark.
- The analysis can be used to discover the inefficient operations or units even for the most profitable organizations.
- DEA has an advantage over other analysis techniques as it can handle complex relation between multiple inputs and multiple outputs and the units are non-commeasurable.
- DEA techniques are based on linear algebra and are related to linear programming concepts. The technique is similar to mathematical duality relations in linear programming.

11.1.1 DEA process

DEA can be used to measure performance and evaluating the activities of organizations such as business firms, government agencies, hospitals, educational institutions. DEA measures the performance in the form of ratio of Output and Input to some processes like efficiency and productivity are measured.

Productivity can be measure by two methods.
- Partial productivity measures
- Total factor productivity measures

Partial productivity measures does not consider all output and input factors, whereas, total factor productivity measure can take into account of all the outputs and inputs. Therefore the mistake of imputing gains to one output that are attributable to another output in partial productivity measures can be avoided using total factor productivity measures.
To consider all inputs and outputs, a tool is required which can manage following challenges.

- How to get a single ratio considering multiple outputs and multiple inputs?
- How to give importance to one attribute over other: Weights!
- How to handle large number of variables and constraints?

To understand DEA methodology we will first consider single input and single output case.

### 11.2 DEA: Single input and single output

Let us start with an example with one input and one output.

Consider an example of a chain of coffee shops ‘Coffees and more’ located in eight locations from A to H as shown in Table 11.1. The owner of the coffee shop wants to evaluate the efficiency of the shops in one city. He considers the cups of coffee sold (per day) as output and the number of employees in the store as inputs. The owner wants to know which store is efficient and which store is inefficient. He is also interested to benchmark the best store so that he can suggest improvements for the inefficient stores by comparing inefficient stores with efficient store.

<table>
<thead>
<tr>
<th>Store location</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cups of coffee sold (thousands)</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>No. Of employees</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>16</td>
</tr>
</tbody>
</table>

The efficiency of each store is determined with the help of ratio of number of cups sold and Number of employees as shown in Table 11.2. We can see from the Table 11.2 that most efficient store is B with an efficiency of 1 and the least being F, with an efficiency of 0.4.
<table>
<thead>
<tr>
<th>Store location</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cups of coffee sold (thousands)</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>4</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>No. Of employees</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Cups sold/number of employees</td>
<td>0.5</td>
<td>1</td>
<td>0.667</td>
<td>0.75</td>
<td>0.8</td>
<td>0.4</td>
<td>0.5</td>
<td>0.625</td>
</tr>
</tbody>
</table>

If we plot input and output variables on a graph, eight stores can be located on a graph as shown in Figure 11.1. The line segment touching all the store points and origin represents the slopes all the store points have. It can be seen from the graph in Figure 11.1, that the store B has the maximum efficiency. If we extend the line segment from origin to store B, it will envelope all other points because it has maximum slope and efficiency. This line segment is called efficient frontier as shown in Figure 11.2.

![Comparisons of Branch Stores](image_url)

Figure 11.1: The input and output variables plotted on a graph
**Efficient Frontier**

The line connecting the point (0,0) and the most efficient point is the efficient frontier called ‘frontier line’. Efficient Frontier envelopes other data points, that’s why it is called Data Envelopment Analysis (DEA). Here, in the example point B is the efficient frontier.

- The frontier line displays the performance of the best store in the comparison.
- The efficiency of other stores can be measured by the deviation of the points from the frontier line. Efficiency of other stores is measured relative to the efficient frontier.
- Efficient frontier serves as Benchmark.

![Comparisons of Branch Stores](image)

**Figure 11.2: Efficient frontier for ‘coffees and more’**

**Relative efficiency of all stores with respect to most efficient store**

The relative efficiency of stores is measured by taking the ratio of efficiency of each store and the efficiency of most efficient store as shown below.

\[
0 \leq \frac{\text{Number of cups of coffee sold per number of employees for each store}}{\text{Number of cups of coffee sold per number of employees for most efficient store}} \leq 1
\]
The relative efficiency of each store lies between 0 and 1 as determined in Table 11.3 for the ‘Coffees and more’.

Table 11.3: The relative efficiency of each store with respect to most efficient store

<table>
<thead>
<tr>
<th>Store</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Efficiency</td>
<td>0.5</td>
<td>1</td>
<td>0.667</td>
<td>0.75</td>
<td>0.8</td>
<td>0.4</td>
<td>0.5</td>
<td>0.625</td>
</tr>
</tbody>
</table>

The relative efficiency score are free from units of measure (units invariance) that’s why it can be used to compare different stores regardless of units. The comparison moves from Ratios to the ratio of ratios. Let us consider the example of improving the efficiency of store A. This can be done by increasing the number of coffees sold or by reducing the number of employees employed in the store A.

11.3 DEA: Multiple inputs and one output

Now, let us consider the case where the owner of ‘coffees and more’ likes to see the impact of one more input that is floor space along with number of employees on the performance of each store. Using the concept of constant return to scale, convert the values of input to the amount of input required to obtain a unit output as shown in Table 11.4. Here, in this case DEA will be used to evaluate stores which use less input resource to get one unit of output.

Table 11.4: Two inputs and one output of ‘Coffees and more’

<table>
<thead>
<tr>
<th>Store location</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cups of coffee sold (thousands)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>No. of employees</td>
<td>8</td>
<td>14</td>
<td>16</td>
<td>8</td>
<td>4</td>
<td>10</td>
<td>12</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Floor space (1000m²)</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>4</td>
<td>8</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
Since, the output is scaled to unit output, plot the graph by taking two input variables as x-y coordinates and locate all the stores on a graph as shown in Figure 11.3.

**Figure 11.3: Input resources utilized by all stores to produce one unit of output**

For multiple inputs case we will determine the efficient frontier by selecting the points or stores, for which one input cannot be improved without worsening the other input. We can see that for ‘Coffees and more’, the efficient frontier is the line connecting EDC in Figure 11.4. The points E, D and C cannot improve one of its input values without worsening the other input. This line envelops all the other units. This set of points is called as the **Production Possibility Set**.

The observed points are assumed to provide (empirical) evidence that production is possible at the rates specified by the coordinates of any point in this region.
Let us consider a unit and see how we can measure the amount of improvement needed for inefficient store or a unit to make to reach the efficient frontier. Consider the point A. Connect the point A to the origin. It would cross the efficient frontier at a point and let us call it P as shown in Figure 11.5. The efficiency of A can be computed by the lengths OP and OA. Efficiency of A = 0.8751. This value can be improved by dropping a line along the X-axis and the Y-axis from the point A to the efficient frontier. Thus the unit can strive to achieve efficiency equal to the ones on the efficient frontier.
11.4 DEA: Single input, multiple outputs

Let us consider the case when there is one input and two outputs. A retail giant has 7 stores in a metro city. The owner of retail giant wants to compare the performance of the retail stores. The owner would like to consider input resources to be the number of salesman (x) and two outputs; number of footfalls or customers visiting those stores (y₁) and revenue earned (y₂) by 7 stores. The values observed are presented in Table 11.5

Table 11.5: The input and output variables observed by a retail giant for 7 stores

<table>
<thead>
<tr>
<th>Store</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Salesmen</td>
<td>x</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Customers (100s)</td>
<td>y₁</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Revenue (100,000s)</td>
<td>y₂</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>
In this case, the data is plotted with Revenue/Number of Employees against Customers/Number of Employees. The efficient frontier obtained is presented in Figure 11.6.

Similar to the previous case, the efficiency of a particular store can be obtained by connecting a line from the origin to the point of the store, crossing the efficiency frontier. Let’s calculate the efficiency of D as presented in Figure 11.7. Efficiency of D = distance \((O,D)\) / distance \((O,P)\) = 0.75.
Figure 11.7: Reference set for inefficient units for single input and multiple output case

- This ratio 0.75 depicts the proportion of the output that ‘P’ shows was possible of achievement and the proportion of inefficiency present in both outputs by D.
- The store D can be made efficient (as efficient as the stores in the efficient frontier) by maintaining the same ratio of inputs to output by moving along the lone (O, P).
- The inefficiency which can be eliminated without changing the proportions is referred to as “Technical Inefficiency”.
- If we have to alter the proportion of outputs, it is referred as “Mix Inefficiency” similar to the case of unit A in the example.