

## SELECCIÓN DE VARIABLES EN MODELOS DEA USANDO SERIES DE TIEMPO

**Marcela C. González-Araya**

Departamento de Modelación y Gestión Industrial, Facultad de Ingeniería, Universidad de Talca.  
Merced N° 437, Curicó, Chile.  
mgonzalez@utalca.cl

**Gonzalo Campos Hernández**

Programa de Magíster en Gestión de Operaciones, Facultad de Ingeniería, Universidad de Talca.  
Merced N° 437, Curicó, Chile.  
gonzalo.campos.h@gmail.com

**Luis G. Acosta Espejo**

Departamento de Industrias, Universidad Técnica Federico Santa María  
Av. Santa María 6400, Vitacura, Santiago, Chile  
luis.acosta@usm.cl

### RESUMEN

En el Análisis y Encapsulamiento de Datos (DEA), una de las actividades iniciales es seleccionar variables de entrada y de salida a ser consideradas en el modelo. En general, en un modelo DEA, las variables utilizadas tienen un impacto directo en la puntuación de eficiencia. En esta investigación se propone un proceso de selección de variables, que busca entregar una buena discriminación entre las unidades eficientes y las ineficientes, y, además, una frontera que se ajuste de la mejor forma posible a los datos observados. Con este fin, se desarrollan tres métodos secuenciales de selección de variables, usando datos de series de tiempo. Los resultados obtenidos al aplicar los métodos a un caso de estudio, muestran la efectividad de éstos en la detección de variables discriminatorias con alto impacto en las puntuaciones de los modelos de DEA.

**PALABRAS CLAVES.** Análisis y Encapsulamiento de Datos, Análisis de Eficiencia, Selección de Variables.

### ABSTRACT

In DEA, one of the first activities is to select the inputs and outputs for the model. In general, in a DEA model, the used factors (inputs and outputs) have a direct impact in the efficiency score. In this research, a process that aims to establish a good discrimination among efficient and inefficient units, besides a best adjusted frontier according to the observed data, is proposed. For this purpose, three methods for variables selection using data series are developed. The obtained results applying the methods to a study case show the suitability of them to detect discriminatory variables that have high impact in the scores of DEA models.

**KEYWORDS.** Data Envelopment Analysis, Efficiency Analysis, Variables Selection.

## 1. Introdução

En el Análisis Envolvente de Datos, DEA, (propuesto por Charnes et al., 1978) se busca determinar la eficiencia de un grupo de unidades de toma de decisiones (DMU). Cada DMU tiene un conjunto de datos de entrada y de salida que se utilizan para obtener una única medida global del rendimiento (puntuación de eficiencia) para cada DMU. Esta medida de rendimiento lo realiza mediante la construcción de una base empírica de las mejores prácticas (o frontera eficiente) formado por el conjunto de DMU eficientes. Una DMU ineficiente no en la frontera eficiente.

Los resultados DEA dependen en gran medida del conjunto de variables de entrada y de salida que se utilizan en el análisis. Sin embargo, tal como tempranamente notaron Golany y Roll (1989), muy pocos estudios han sido desarrollados con el propósito de seleccionar de mejor forma estas variables. En este sentido, en este trabajo se propone un procedimiento para la selección de las variables de entrada y de salida. Muchos de los trabajos en esta línea tratan las variables de entrada y de salida como datos y cuando se realiza una selección de variables previa al modelo, éstas son de carácter de corte transversal, esto es no se aborda explícitamente la situación de cuando existen datos observados en un periodo de tiempo. En el método propuesto por Lins et al. (1999), se parte de un par input-output inicial que tenga alta correlación y el criterio de selección de las variables a ser incluidas depende de cual de las variables candidatas provoca mayor eficiencia media en el modelo DEA. El método se detiene al alcanzar un número de variables considerado ideal en función del mejor ajuste de la frontera eficiente. El método desarrollado por Soares de Mello et al. (2004) está basado en la selección de un número reducido de inputs y outputs, que representa adecuadamente la relación causal y que no atribuye eficiencia unitaria a un número muy grande de DMUs. Este método aborda el problema de cómo discriminar entre DMUs en el caso e que se utilice DEA como herramienta de Apoyo Multicriterio a la Decisión (MCDA). Los autores presentan un método basado en el uso de un número limitado de inputs y outputs, de modo a mejorar el poder de discriminación de DEA.

El método Stepwise de Wagner y Shimshak (2007) es un procedimiento que inicialmente estima el promedio de eficiencia de las DMUs evaluadas usando todas las variables disponibles y, a continuación, secuencialmente las va eliminando del conjunto de variables según el cambio promedio de las eficiencias de las DMUs. Este procedimiento se detiene cuando en el conjunto de variables sólo resta un *input* y un *output*. La desventaja de este método es que no toma en cuenta los criterios de “mejor ajuste a la frontera” y de “máxima discriminación” de las DMUs.

El abordaje propuesto por González-Araya y Valdés (2009) propone una metodología secuencial de selección de variables que considera los casos cuando el número de DMUs es mucho mayor que el número de variables, ya que en estos casos, el poder discriminatorio del modelo DEA usado continúa siendo un criterio relevante para los tomadores de decisiones en el momento de validar los resultados. Para los casos donde el número de variables disponibles para la evaluación es mucho menor que el número de DMUs evaluadas, el método propone un procedimiento basado en el método propuesto por Soares de Mello et al. (2004), mediante el establecimiento de un nuevo criterio de selección de variables para lidiar mejor con esta situación.

En el presente trabajo se aborda la selección de variables para datos recolectados a través de una serie de tiempo. Se presentan tres métodos de selección de variables se muestran los resultados obtenidos un estudio de caso, que considera datos recolectados durante nueve años (desde 2004 hasta 2012).

Este artículo ha sido estructurado de la forma que se describe a continuación. En la Sección 2 se describe el método propuesto, en la Sección 3, se describe el caso de estudio. En la Sección 4, se discuten los resultados obtenidos al utilizar estos métodos con los datos del caso de estudio. Finalmente, en la Sección 5 se entregan las principales conclusiones obtenidas.

## 2. Metodología

En este trabajo se propone un método de selección de variables que considera la evolución en el tiempo de las DMUs. A continuación se detallan los métodos propuestos y se desarrolla un algoritmo para la selección de las variables.

### Método 1: Año a Año

El método desarrollado por Soares de Mello et al. (2004) expone una metodología de apoyo multicriterio a las decisiones para resolver una problemática de selección, según clasificación de las problemáticas básicas de apoyo a las decisiones de Roy y Bouyssou (1993). Dicho método considera como alternativas a las variables candidatas a ser incluidas en el modelo DEA y, como criterios, al mejor ajuste a la frontera y la máxima discriminación entre las DMUs. El ajuste a la frontera es proporcionado a través de la eficiencia media, y en base a su normalización, se obtiene la creación de la variable  $S_{EF}$ , la cual describe el impacto del criterio mejor ajuste a la frontera, asignándole un valor 10 en la eficiencia máxima y un valor 0 en la eficiencia mínima. Posteriormente, la máxima discriminación es calculada por el número de DMUs eficientes resultantes del modelo. Al igual que la variable anterior, la normalización de esta medida genera la variable  $S_{DIS}$ , la que describe el impacto del criterio máxima discriminación, asignándole un valor 10 para el menor número de DMUs en la frontera y 0 para el mayor. Finalmente, para seleccionar la variable que produce la mejor solución de compromiso, y que debiera entrar en el modelo DEA, se agregan los criterios mejor ajuste a la frontera y máxima discriminación en un criterio síntesis, a través de la suma ponderada de los respectivos descriptores de impacto,  $S_{EF}$  y  $S_{DIS}$ . El descriptor de impacto de este criterio síntesis es la variable  $S$ , tal que  $S = \omega \times S_{EF} + (1-\omega) \times S_{DIS}$ , con  $\omega$  es un escalar entre 0 y 1. Para que no exista predominancia de uno de los dos criterios, a menos que existan fuertes razones para estimar lo contrario, se debe hacer  $\omega = 0,5$ .

Este método es tomado como punto de partida para el Método 1 propuesto, ya que fue aplicado al caso real del análisis de la eficiencia de los Programas de Post Graduación en Ingeniería de la Universidad Federal de Río de Janeiro (Soares de Mello *et al.*, 2004) obteniendo satisfactorios resultados.

A continuación se detallan los pasos a seguir para lograr el desarrollo del método propuesto.

**Paso 1.** Aplicar a cada año del panel de datos, el método de selección de variables Multicriterio (Soares de Mello et al., 2004).

**Paso 2.** Con las variables ya seleccionadas para cada año del panel, calcular indicadores estadísticos como la Frecuencia Absoluta ( $n_i$ ), Frecuencia Absoluta Acumulada ( $N_i$ ), la Frecuencia Relativa ( $f_i$ ) y Frecuencia Relativa Acumulada ( $F_i$ ). Lo anterior se aplica a cada variable existente, lo que permite crear un ranking de acuerdo a la Frecuencia Relativa ( $f_i$ ) de cada variable en el horizonte de tiempo en estudio. Con la Frecuencia Relativa ( $f_i$ ) de cada variable, y con el fin de exista al menos una variable de salida y otra de entrada, se escogen las variables que posean mayor Frecuencia Relativa a través del tiempo.

**Paso 3.** Posteriormente se eligen gradualmente aquellas variables que posean Frecuencia Relativa inmediatamente menores al par seleccionado con anterioridad, sin importar si estas selecciones son en variables de entrada o variables de salida.

**Paso 4.** En el caso de exista empate en la frecuencia de las DMU's a seleccionar, se procederá a estimar el promedio a través del tiempo de la eficiencia técnica media de las respectivas DMU's, seleccionado aquella que posea un promedio mayor. El cálculo de la eficiencia técnica media de cada año, se obtiene al realizar el método de selección propuesto por Soares de Mello *et al.* (2004) para cada año.

**Paso 5.** El procedimiento se repite para las variables de entrada y salidas, incrementado gradualmente la selección de una variable a la vez, teniendo como criterio de parada la condición de que el número de DMU's exceda el triple de el número de variables (Soares de Mello et al., 2004).

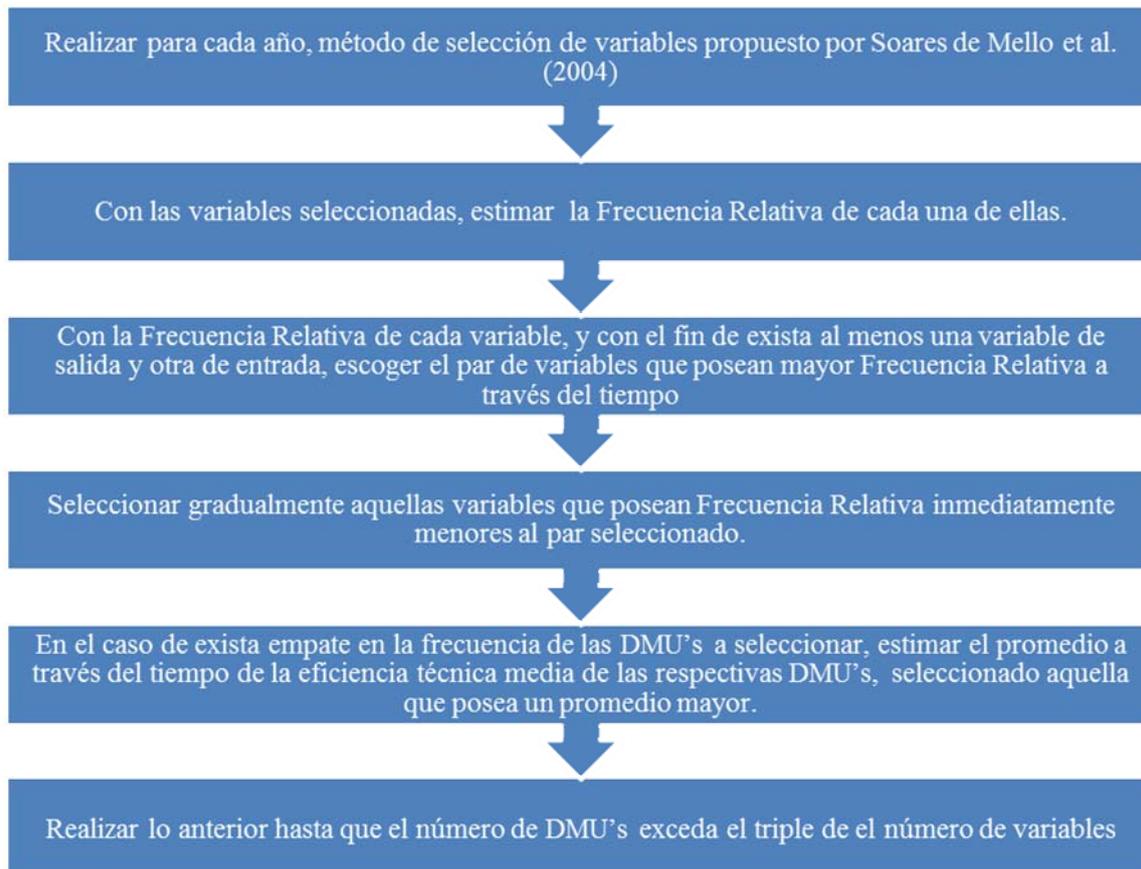


FIGURA N° 1: ALGORITMO PROPUESTO PARA MÉTODO DE SELECCIÓN DE VARIABLES PARA MODELOS DEA (MÉTODO 1)

### Método 2: Panel de Datos 1

Este método es similar al Método 1, con la salvedad de que el método propuesto por Soares de Mello et al. (2004), se aplica una sola vez a la totalidad de los datos del horizonte de tiempo. Los pasos a seguir son los siguientes:

**Paso 1.** Ordenar los datos de tal manera de disponer de un panel de datos ordenados por periodos de la forma  $T = 0, \dots, t$ , es decir:  $DMU_1 T_0, DMU_2 T_0, \dots, DMU_n T_0, DMU_1 T_1, DMU_2 T_1 \dots DMU_n T_t$

**Paso 2.** Aplicar método de selección de variables propuesto por Soares de Mello et al.(2004)

### Método 3: Panel de Datos 2

Este método de selección de variables, combina el método realizado por González-Araya y Valdés (2009), con el paso 1 del Método 2 propuesto. Los pasos a seguir son los siguientes:

**Paso 1.** Ordenar los datos de tal manera de disponer de un panel de datos ordenados por periodos de la forma  $T= 0, \dots, t$ , es decir:  $DMU_1 T_0, DMU_2 T_0, \dots, DMU_n T_0, DMU_1 T_1, DMU_2 T_1 \dots DMU_n T_t$

**Paso 2.** Aplicar método de selección de variables González-Araya y Valdés (2009)

A modo de resumen, se han expuesto tres métodos de selección de variables, todos evaluando las variables a través de un horizonte de tiempo.

### 3. Caso de Estudio

Para la aplicación de los métodos se utilizaron datos de una entidad de educación superior, específicamente, una universidad privada nacional. Dicha universidad posee un total de 16.700 alumnos y actualmente se encuentra acreditada por la Comisión Nacional de Acreditación (CNA) de Chile por un periodo de 5 años. Las DMU's (carreras de pregrado) consideradas en el estudio son aquellas que poseen datos para la totalidad del periodo de tiempo analizado. El detalle de las 10 DMUs analizadas se muestra en la Tabla N° 1.

TABLA N° 1. DMUS ANALIZADAS SEGÚN FACULTAD

|      | Facultad Cs. Empresariales | Facultad Educación            | Facultad Cs. Jurídicas y Sociales |
|------|----------------------------|-------------------------------|-----------------------------------|
| DMUS | Contador Público Auditor   | Pedagogía en Artes Visuales   | Trabajo Social                    |
|      | Ingeniería Comercial       | Pedagogía Historia y Geog.    | Psicología                        |
|      | Ingeniería Informática     | Pedagogía General Básica      | Derecho                           |
|      |                            | Pedagogía en Matemáticas      |                                   |
|      |                            | Pedagogía en Inglés           |                                   |
|      |                            | Pedagogía en Educación Física |                                   |
|      |                            | Pedagogía en Castellano       |                                   |
|      | Pedagogía Ed. Parvularia   |                               |                                   |

El horizonte de tiempo considerado para el presente caso práctico, es del año 2004 al año 2012, es decir 9 años. Las variables de entrada (*inputs*) que son utilizadas en los procesos en cada año son:

- N° de horas de clases: variable discrecional; corresponde a la cantidad de horas a la semana que posee cada DMU en cada año.
- N° de Docentes: variable discrecional; corresponde a la cantidad de docentes que realizan clases en la determinada carrera, abarca la modalidad de contrato y part-time.
- Presupuesto: variable discrecional; corresponde a la cantidad de dinero anual presupuestada a cada carrera. Estos valores fueron ajustados al presente año (2012) de acuerdo al Índice de Precios del Consumidor (IPC) respectivo.
- Tasa de deserción: variable discrecional; corresponde al porcentaje de deserción de alumnos del primer año de cada carrera.

Las variables de salida (*outputs*) identificados en el estudio son:

- Puntaje Prueba de Selección Universitaria (PSU): variable no discrecional; corresponde al puntaje promedio de las pruebas de Matemáticas y Lenguaje rendidas para el ingreso de los alumnos a una determinada carrera.
- Matrículas Totales: variable discrecional; corresponde al número de alumnos matriculados anualmente en cada carrera.

- Notas de Enseñanza Media (NEM): variable no discrecional; corresponde al promedio de las calificaciones de enseñanza media (secundaria) del total de alumnos de una determinada carrera.
- Promedio de Notas de la Carrera: variable discrecional; corresponde al promedio general anual de notas de cada carrera. Esto involucra la totalidad de los ramos pertenecientes a dicha DMU.
- Tasa promedio de Aprobación de la Carrera: variable discrecional; corresponde a la tasa promedio de la totalidad de los ramos efectuados en una determinada carrera.

En relación a estas variables iniciales se trabaja para determinar cuáles son las que tienen mayor impacto en la evaluación de eficiencia.

A continuación se presenta un resumen de los principales resultados obtenidos luego de aplicar los métodos de definidos en la Sección 2.

#### 4. Principales Resultados

Los pasos de los métodos propuesto fueron realizados utilizando el modelo DEA CCR, debido a que los modelos CCR son más discriminatorios que los modelos BCC (Banker *et al.*, 1984). Además, estos modelos han sido orientados a las salidas, porque la institución evaluada desea mejorar la captación de alumnos y desempeño de las carreras, manteniendo constante los recursos utilizados. Las formulaciones de estos modelos DEA (CCR y BCC) puede ser encontrada en Coelli *et al.* (2005) y en Cooper *et al.* (2007).

##### Desarrollo Método 1: Año a año

En el presente estudio, existen en cada año del panel de datos 9 variables, las cuales no se comportan de la misma manera durante todos los años del estudio. Debido a esto, el nuevo método explicado a continuación, propone un procedimiento basado en el método formulado por Soares de Mello *et al.* (2004), mediante el establecimiento de un nuevo criterio de selección de variables a través del tiempo de tal manera de realizar una correcta y representativo selección de variables del horizonte de tiempo analizado. Se debe recordar que el método de Soares de Mello *et al.* (2004), se basa en la selección de un número reducido de inputs y outputs, que representen adecuadamente la relación causal y que no atribuye eficiencia unitaria a un número muy grande de DMU's. Además, considera que al reducir el número de variables con alta correlación, se aumenta la discriminación en los modelos DEA, con la consecuente disminución del ajuste a la frontera eficiente. De acuerdo con el algoritmo propuesto para selección de variables en un horizonte de tiempo, como primera etapa, se procede a realizar para cada año el método de selección de variables de Soares de Mello *et al.* (2004), obteniendo los pasos y resultados descritos a continuación:

- Para el primer año del panel, año 2004, y como primer paso, se establece la matriz de correlación de las variables en estudio, obteniendo como variables más correlacionadas “Matrículas Totales” y “Nº de horas de clases”, con un coeficiente de correlación de 0.904.

El resto de las correlaciones entre las variables de entrada y salida se detalla en la tabla siguiente:

TABLA 2. MATRIZ DE CORRELACIÓN DE VARIABLES AÑO 2004

|                       | NEM    | PSU    | Matrículas Totales | Promedio Notas | Tasa de Aprobación |
|-----------------------|--------|--------|--------------------|----------------|--------------------|
| Presupuesto           | -0,335 | 0,030  | 0,051              | 0,429          | 0,325              |
| Nº de Docentes        | 0,095  | -0,461 | 0,889              | -0,243         | -0,130             |
| Nº de horas de clases | 0,153  | -0,373 | <b>0,904</b>       | -0,162         | -0,050             |
| Tasa de Deserción     | -0,241 | 0,315  | -0,079             | -0,142         | -0,199             |

- Luego, para el análisis DEA, se utiliza el Software académico Efficiency Measurement System (EMS, Versión 1.3.0), con rendimiento a escala constante (CCR), debido a que posee una mayor discriminación, y orientación hacia las salidas. Al par inicial Input–Output

escogido en el paso anterior, se le va agregando en cada iteración una variable, almacenando su puntaje de eficiencia.

- Una vez, que se obtuvo las eficiencias respectivas de cada variable, se procede a calcular las variables Sef y Sdis, para con ambas, obtener la variable S. (ver Tabla 2)
- Ingresa al modelo en forma definitiva la variable con mayor valor de S. Para este caso, ingresa la variable Notas de Enseñanza Media (NEM).
- Ya se encuentran como variables base del modelo, tres variables, “Matrículas Totales”, “Nº de horas de clases” y “NEM”.
- Como se ha dicho en oportunidades anteriores, el criterio de parada del método en cuestión, es que el número de DMU's no debe exceder al triple del número de variables. En base a lo anterior y teniendo en cuenta que el número de DMU's es 14 y el número de variables es 9, se debe tener como máximo 4 variables finales. Por lo tanto se debe agregar una nueva variable al modelo.
- Para agregar una nueva variable al modelo, primeramente se calcula las eficiencias de las DMU's para el trío de variables ya seleccionada, almacenándose dichos valores.

Posteriormente, en una segunda iteración, se adiciona al modelo, una a una las restantes variables, almacenando su puntaje de eficiencia.

TABLA 3. RESULTADO PRIMERA ITERACIÓN MÉTODO SELECCIÓN VARIABLES SOARES DE MELLO ET AL. (2004)

| DMU                        | Eficiencia del Par más Correlacionado | Presupuesto | Nº de Docentes | Tasa de Deserción | NEM            | PSU     | Promedio Notas | Tasa de Aprobación |
|----------------------------|---------------------------------------|-------------|----------------|-------------------|----------------|---------|----------------|--------------------|
| CONT. PUBLICO AUDITOR      | 292,98%                               | 195,41%     | 229,49%        | 233,15%           | <b>292,98%</b> | 292,98% | 255,59%        | 262,03%            |
| DERECHO                    | 100,00%                               | 100,00%     | 100,00%        | 100,00%           | <b>100,00%</b> | 100,00% | 100,00%        | 100,00%            |
| EDUCACION PARVULARIA       | 134,44%                               | 100,00%     | 110,54%        | 100,00%           | <b>134,44%</b> | 134,44% | 134,44%        | 134,44%            |
| INGENIERIA COMERCIAL       | 425,54%                               | 237,48%     | 407,29%        | 425,54%           | <b>425,54%</b> | 425,54% | 265,00%        | 270,12%            |
| INGENIERIA INFORMATICA     | 309,40%                               | 100,00%     | 193,16%        | 236,78%           | <b>309,40%</b> | 309,40% | 309,40%        | 309,40%            |
| PEDAGOGIA EN CASTELLANO    | 256,13%                               | 217,37%     | 256,13%        | 244,74%           | <b>100,00%</b> | 100,00% | 100,00%        | 100,00%            |
| PED. EN EDUCACIÓN FISICA   | 200,91%                               | 200,91%     | 184,71%        | 177,09%           | <b>200,91%</b> | 200,91% | 159,27%        | 166,28%            |
| PEDAGOGIA EN INGLES        | 177,50%                               | 145,55%     | 108,81%        | 134,59%           | <b>177,50%</b> | 177,50% | 177,50%        | 177,50%            |
| PED. EN MATEMATICAS        | 326,09%                               | 270,86%     | 326,09%        | 326,09%           | <b>240,77%</b> | 190,86% | 137,47%        | 139,05%            |
| PED. GENERAL BASICA        | 167,38%                               | 119,94%     | 100,00%        | 126,27%           | <b>167,38%</b> | 167,38% | 167,38%        | 167,38%            |
| PED. HISTORIA Y GEOGRAFIA  | 170,61%                               | 130,26%     | 102,17%        | 132,91%           | <b>170,61%</b> | 170,61% | 170,61%        | 170,61%            |
| PSICOLOGIA                 | 145,50%                               | 129,19%     | 106,32%        | 116,77%           | <b>145,50%</b> | 145,50% | 145,50%        | 145,50%            |
| TRABAJO SOCIAL             | 204,76%                               | 100,00%     | 142,64%        | 100,00%           | <b>204,76%</b> | 204,76% | 204,76%        | 204,76%            |
| PED. EN ARTES VISUALES     | 584,35%                               | 584,35%     | 491,61%        | 584,35%           | <b>514,82%</b> | 466,42% | 146,01%        | 149,31%            |
| <b>Eficiencia Promedio</b> | 249,69%                               | 187,95%     | 204,21%        | 217,02%           | <b>227,47%</b> | 220,45% | 176,64%        | 178,31%            |
| S ef                       | -                                     | 2,2256      | 5,4242         | 7,9439            | <b>10,0000</b> | 8,6186  | 0,0000         | 0,3295             |
| Nº de Unidades Eficientes  | 1                                     | 4           | 2              | 3                 | <b>2</b>       | 2       | 2              | 2                  |
| S dis                      | -                                     | 0           | 10             | 5                 | <b>10</b>      | 10      | 10             | 10                 |
| S                          |                                       | 1,1128      | 7,7121         | 6,4719            | <b>10,0000</b> | 9,3093  | 5,0000         | 5,1648             |

- Una vez, que se obtuvo las eficiencias respectivas de cada variable, se procede a calcular las variables Sef y Sdis, para con ambas, obtener la variable S. (ver Tabla 3)
- Ingresa al modelo en forma definitiva la variable con mayor valor de S. Para este caso, ingresa la variable PSU, completando la selección de variables para el periodo 2004 del panel de datos.
- Resumiendo, las variables seleccionadas para el año 2004 son: “Matrículas Totales”, “Nº de horas de clases”, “NEM” y “PSU”

El procedimiento de selección de variables descrito anteriormente, se realiza para la totalidad de años del panel de datos, es decir, desde el año 2004 a 2012, obteniendo los siguientes resultados expuestos en Tabla 4.

TABLA 4. SELECCIÓN DE VARIABLES PARA CADA AÑO DEL PANEL

| Año  | VARIABLES SELECCIONADAS MÉTODO SOARES DE MELLO |                    |                    |                    |
|------|--|--------------------|--------------------|--------------------|
| 2012 | Presupuesto                                    | Tasa de Deserción  | Número de docentes | Matrículas Totales |
| 2011 | Nº de horas de clases                          | PSU                | NEM                | Matrículas Totales |
| 2010 | Presupuesto                                    | Tasa de Deserción  | Nº de Docentes     | Matrículas Totales |
| 2009 | Nº de horas de clases                          | PSU                | Nº de Docentes     | Matrículas Totales |
| 2008 | Tasa de Deserción                              | Tasa de Aprobación | NEM                | PSU                |
| 2007 | Nº de horas de clases                          | Tasa de Deserción  | Nº de Docentes     | Matrículas Totales |
| 2006 | Nº de horas de clases                          | PSU                | NEM                | Matrículas Totales |
| 2005 | Nº de horas de clases                          | PSU                | NEM                | Matrículas Totales |
| 2004 | Nº de horas de clases                          | PSU                | NEM                | Matrículas Totales |

Ahora bien, como se aprecia en la Tabla 4, el grupo de variables seleccionadas no es el mismo durante el transcurso de los años, por lo tanto no es posible seleccionar de inmediato las cuatro variables más representativas al horizonte de estudio seleccionado.

Confeccionando una nueva tabla, a partir de la anterior, en relación a la frecuencia de selección de las variables, es posible obtener lo expuesto en la Tabla 5.

TABLA 5. SELECCIÓN DE VARIABLES PARA CADA AÑO DEL PANEL Y CÁLCULO DE FRECUENCIA

|          |                       | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | Frecuencia Absoluta (ni) | Frecuencia Acumulada (Ni) | Frecuencia Relativa (fi) | Frecuencia Relativa Acumulada (Fi) |
|----------|-----------------------|------|------|------|------|------|------|------|------|------|--------------------------|---------------------------|--------------------------|------------------------------------|
| Entradas | Presupuesto           |      |      |      |      |      |      | 1    |      | 1    | 2                        | 2                         | 6%                       | 6%                                 |
|          | Nº de Docentes        |      |      |      | 1    |      | 1    | 1    |      | 1    | 4                        | 6                         | 11%                      | 17%                                |
|          | Nº de horas de clases | 1    | 1    | 1    | 1    |      | 1    |      | 1    |      | 6                        | 12                        | 17%                      | 33%                                |
|          | Tasa de Deserción     |      |      |      | 1    | 1    |      | 1    |      | 1    | 4                        | 16                        | 11%                      | 44%                                |
| Salidas  | NEM                   | 1    | 1    | 1    |      | 1    |      |      | 1    |      | 5                        | 21                        | 14%                      | 58%                                |
|          | PSU                   | 1    | 1    | 1    |      | 1    | 1    |      | 1    |      | 6                        | 27                        | 17%                      | 75%                                |
|          | Matrículas Totales    | 1    | 1    | 1    | 1    |      | 1    | 1    | 1    | 1    | 8                        | 35                        | 22%                      | 97%                                |
|          | Promedio Notas        |      |      |      |      |      |      |      |      |      | 0                        | 35                        | 0%                       | 97%                                |
|          | Tasa de Aprobación    |      |      |      |      | 1    |      |      |      |      | 1                        | 36                        | 3%                       | 100%                               |

De acuerdo al procedimiento descrito en la metodología, se elige el par input-output con mayor frecuencia relativa, es decir, “Matrículas Totales” y “Nº de horas de clases”. Posteriormente, de acuerdo al Paso 3 de la metodología propuesta, se eligen gradualmente aquellas variables que posean *Frecuencia Relativa* inmediatamente menores al par seleccionado con anterioridad, sin importar si estas elecciones son en variables de entrada o variables de salida. De esta manera se selecciona las variables “PSU” y “NEM”, dando por terminado el método de selección. Por lo tanto las variables seleccionadas para el panel de datos analizados son: “Matrículas Totales”, “Nº de horas de clases”, “PSU” y “NEM”.

En el caso de exista empate en la frecuencia de las DMU’s a seleccionar, Paso 4, se procederá a estimar el promedio a través del tiempo de la eficiencia técnica media de las respectivas DMU’s, seleccionado aquella que posea un promedio mayor. El cálculo de la eficiencia técnica media de cada año, se obtiene al realizar el método de selección propuesto por Soares de Mello et al. (2004) para cada año. De manera hipotética, si aún faltasen variables a incluir, se debe optar entre las variables “Nº de Docentes” y “Tasa de Deserción”, ambas variables de entrada y con un frecuencia relativa de 11%. Por lo tanto, una manera o forma de seleccionar a una de ellas es mediante la eficiencia técnica media, seleccionando aquella que posea la mayor, debido a que posee un mayor grado de discriminación. Para este caso hipotético sería “Tasa de Deserción”. La Tabla 6 cita dichos valores para cada año.

TABLA 6. EFICIENCIA TÉCNICA MEDIA DE CADA VARIABLE

|          | 2004                  | 2005    | 2006    | 2007    | 2008    | 2009    | 2010    | 2011    | 2012    | Promedio |         |
|----------|-----------------------|---------|---------|---------|---------|---------|---------|---------|---------|----------|---------|
| Entradas | Presupuesto           |         |         |         |         |         | 143,07% |         | 154,31% | 148,69%  |         |
|          | Nº de Docentes        |         |         |         | 168,69% |         | 195,35% | 142,42% | 134,02% | 160,12%  |         |
|          | Nº de horas de clases | 249,69% | 163,34% | 177,98% | 163,82% |         | 255,27% |         | 158,55% | 194,77%  |         |
|          | Tasa de Deserción     |         |         |         | 159,14% | 209,28% |         | 143,07% |         | 130,56%  | 160,51% |
| Salidas  | NEM                   | 227,47% | 153,59% | 176,00% |         | 209,28% |         |         | 141,50% | 181,57%  |         |
|          | PSU                   | 220,45% | 153,91% | 175,72% |         | 209,28% | 181,95% |         | 144,13% | 180,91%  |         |
|          | Matrículas Totales    | 249,69% | 163,34% | 177,98% | 168,69% |         | 255,27% | 143,07% | 158,55% | 154,31%  | 183,86% |
|          | Promedio Notas        |         |         |         |         |         |         |         |         | -        |         |
|          | Tasa de Aprobación    |         |         |         |         | 209,28% |         |         |         | 209,28%  |         |

Es necesario recordar que el presente análisis seleccionada aquellas variables que hacen el modelo más discriminatorio a partir de los datos disponibles. Debido a lo anterior y a modo resumen, la Tabla 7 detalla las variables seleccionadas por cada uno de los tres métodos de selección expuestos en la metodología.

TABLA 7. RESUMEN MÉTODOS DE SELECCIÓN DE VARIABLES

| Métodos de Selección de Variables a Través del Tiempo |                       |                    |
|---|-----------------------|--------------------|
| Método 1  | Método 2              | Método 3           |
| Nº de horas de clases                                 | Matrículas Totales    | Promedio de Notas  |
| NEM   | Nº de horas de clases | Tasa de Aprobación |
| PSU   | Nº de Docentes        | Presupuesto        |
| Matrículas Totales                                    | NEM                   | Tasa de Deserción  |

## 5. Conclusiones

Al observar las variables seleccionadas por cada método propuesto, es posible percatarse que son tres enfoques o miradas distintas. El Método 1, posee solamente una variable de entrada “Nº de horas de clases” y dos de sus tres salidas son No Discrecionales, “NEM” y “PSU”. Se atrae más matrículas y mejores promedios. Por otro lado, el Método 2, posee dos variables de entrada y dos de salida, siendo solamente una de estas últimas No Discrecional, “NEM”. Finalmente, el Método 3, al igual que el Método 2, posee dos variables de entradas y dos variables de salida, con la salvedad, que ninguna de las variables seleccionadas son No Discrecionales. Al observar las variables seleccionadas por los dos primeros métodos de selección, es posible apreciar que éstos poseen una mirada de atraktividad de las DMU’s (carreras), por otra parte el método último, se encuentra enfocado a la permanencia en el tiempo de los alumnos en las carreras, la cual depende directamente de las gestiones internas de cada carrera.

La mayor parte de las aplicaciones de los modelos DEA encontrada en la literatura no coloca atención en la selección de las variables mediante un periodo de tiempo. Por otro lado, cuando se coloca atención en este tema, la motivación principal es la comparación entre un periodo y otro, no considerando el panel de datos en su totalidad. Sin embargo, en este trabajo se propone realizar la selección de variables para la totalidad del tiempo estudiado, de una sola vez, analizando todas las variables involucradas en dicho periodo de tiempo. De esta manera, los métodos para la selección de variables a través del tiempo propuestos, abordan un vacío en la literatura DEA, permitiendo establecer las variables necesarias para alcanzar un “mejor ajuste a la frontera” y la “máxima discriminación” de las DMUs evaluadas, en los casos donde las variables van más allá de un simple corte transversal en los datos.

Finalmente, cabe destacar que al aplicar los métodos propuestos antes de realizar los análisis de eficiencia usando DEA, los administradores y los tomadores de decisiones obtendrían

información de mejor calidad, permitiéndoles obtener estimaciones más fidedignas de las puntuaciones de eficiencia de las DMUs y establecer metas para cada DMU ineficiente de las variables críticas seleccionadas por el método.

### Referencias

- Banker, R.D., Charnes, A. and Cooper, W.W.** (1984), *Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis*, Management Science, 30 (9), 1078-1092.
- Charnes, A., Cooper, W.W. and Rhodes, E.** (1978), *Measuring the Efficiency of Decision Making Units*, European Journal of Operational Research, 2, 429-444.
- Coelli, T.J., Prasada Rao, D.S., O'Donnell, Battese, G.** *An Introduction to Efficiency and Productivity Analysis*, 2nd Edition, Springer, USA, 2005.
- Cooper, W., Seiford, L.M., Tone, K.** *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2nd Edition, Springer, USA, 2007.
- Golany, B., y Roll, Y.** (1989), *An application procedure for DEA*, OMEGA, 17(3), 237-250.
- González-Araya, M., Valdés, G.** (2009). Método de selección de Variables para mejorar la discriminación en el análisis de eficiencia aplicando modelos DEA. Revista Ingeniería Industrial - Año 8, N° 2: 45 – 56.
- Lins, M.P.E., Moreira, M.C.B.** (1999). *Método I-O Stepwise para Seleção de Variáveis em Modelos de Análise Envoltória de Dados*, Pesquisa Operacional, Vol. 19, 39-50.
- Roy, B. y Bouyssou, D.** *Aide Multicritère à la Décision: Méthodes et Cas*, Economica, Paris, 1993.
- Soares de Mello, J.C.C.B., Gomes, E., Meza, L.A., & Lins, M.P.E.** (2004). Selección de variables para el incremento del poder de discriminación de los modelos DEA. Revista Escuela de Perfeccionamiento en Investigación Operativa, 24, 40-52.
- Wagner, J.M., & Shimshak, D.G.** (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. European Journal of Operational Research, 180, 57-67.