

MÉTODO DE SELECCIÓN DE VARIABLES PARA MEJORAR LA DISCRIMINACIÓN EN EL ANÁLISIS DE EFICIENCIA APLICANDO MODELOS DEA*

VARIABLES SELECTION METHOD FOR IMPROVING DISCRIMINATION IN DEA EFFICIENCY ANALYSIS

Marcela C. González-Araya¹, Nelson G. Valdés Valenzuela²

¹Dpto. de Modelación y Gestión Industrial, Facultad de Ingeniería. Universidad de Talca. Merced N° 437. Curicó. Chile.

²Escuela de Ingeniería Civil Industrial, Facultad de Ingeniería. Universidad de Talca. Merced N° 437. Curicó. Chile. guillermo.valdes.valenzuela@gmail.com

RESUMEN

El Análisis y Encapsulamiento de Datos (DEA) es una técnica basada en programación lineal, que permite determinar la eficiencia relativa de un conjunto de unidades de toma de decisiones, las cuales consumen múltiples entradas para producir múltiples salidas. Por esta razón, una etapa importante cuando se realiza un análisis de eficiencia aplicando modelos de DEA, corresponde a la selección de las variables de entrada y de salida a utilizar en el análisis, pues éstas tienen un impacto directo en la puntuación de eficiencia. Al seleccionar las variables, se busca obtener una buena discriminación entre las unidades eficientes e ineficientes, y una frontera que se ajuste de la mejor forma posible a los datos observados. En este trabajo se propone un método secuencial de selección de variables de entrada y salida que considere estos criterios y que sea adecuado en los casos donde el número de observaciones, o unidades de toma de decisiones, es mucho mayor que el número de variables potenciales para realizar el análisis de eficiencia. Este método es aplicado a un caso de estudio donde se muestra su efectividad en la detección de variables con alto impacto en los resultados de los modelos de DEA.

Palabras clave: Análisis y encapsulamiento de datos, análisis de eficiencia, programación lineal.

ABSTRACT

Data Envelopment Analysis (DEA) is a linear programming technique that allows determining the relative efficiency of a set of decision making units (DMUs), which consume multiple inputs to produce multiple outputs. For this reason, the selection of inputs and outputs used in DEA models is an important stage in the efficiency analysis. The selection of these variables has a direct impact in the DMUs' efficiency scores. The main criteria considered for selecting inputs and outputs are to obtain a good discrimination among the efficient and inefficient DMUs and to obtain an efficient frontier adjusted the best way as possible to observed data. In this work, we propose a sequential method for selecting inputs and outputs variables that considers the mentioned criteria and it is suitable for cases when the number of observations is much larger than the number of potential inputs and outputs available for carrying out the efficiency analysis. This method is applied to a study case where its effectiveness for detecting variables with high impact in the efficiency scores is demonstrated.

Keywords: Data envelopment analysis, efficiency analysis, linear programming.

*Este trabajo fue presentado en el VIII Congreso Chileno de Investigación Operativa, realizado por la Universidad del Bio-Bio en la ciudad de Chillán en Octubre del año 2009

Autor para correspondencia: mgonzalez@utalca.cl

Recibido: 08.04.2009 Aceptado: 13.11.2009

INTRODUCCIÓN

Las organizaciones, en forma permanente, enfocan sus esfuerzos en mejorar los procesos y actividades que realizan habitualmente; es decir, intentan operar en forma eficiente, de modo que logren maximizar sus beneficios utilizando la menor cantidad de recursos. Por lo tanto, buscan mejorar sus procesos ineficientes para que operen permanentemente en forma eficiente. Esta tarea se complica en la medida en que es necesario utilizar múltiples recursos (*inputs*) para obtener múltiples productos (*outputs*). El Análisis y Encapsulamiento de Datos, conocido como DEA (abreviación del inglés *Data Envelopment Analysis*) fue propuesto para abordar este problema (Charnes *et al.*, 1978).

DEA es un método basado en programación lineal, que permite medir la eficiencia relativa de unidades organizacionales, denominadas “unidades de toma de decisión” (Decision Making Units - DMU), que utilizan múltiples recursos para producir múltiples productos. El objetivo fundamental de DEA es encontrar una frontera de eficiencia, formada por aquellas observaciones o DMUs que presentan las mejores prácticas, y, a partir de ella, evaluar la eficiencia relativa de las DMUs que no pertenezcan a la misma, intentando maximizar la eficiencia de éstas DMUs (Coelli *et al.*, 2005, González-Araya, 2006).

El primer desafío, al aplicar los modelos de DEA, consiste en seleccionar las variables (*inputs* y *outputs*) que describan con mayor precisión el proceso de transformación de las DMUs evaluadas y permita discriminar claramente las DMUs eficientes de las ineficientes. Sin embargo, tal como tempranamente notaron Golany & Roll (1989), muy pocos estudios han sido desarrollados con el propósito de seleccionar de mejor forma estas variables. Algunos autores que han propuesto métodos para seleccionar *inputs* y *outputs* han sido Lins & Moreira (1999), Soares de Mello *et al.* (2004), Senra *et al.* (2005), Wagner & Shimshak (2007). Todos ellos han enfocado sus esfuerzos en ordenar y seleccionar aquellas variables que permitan discriminar de la mejor forma posible el desempeño obtenido por cada DMU evaluada, en los casos cuando el número de *inputs* y *outputs* potenciales es mucho mayor que el número de observaciones (DMUs observadas). Estos esfuerzos se deben a que es reconocido en la literatura de DEA que mientras mayor es el número de *inputs* y *outputs* considerados en el análisis de eficiencia, menores serán las restricciones a los pesos o multiplicadores asignados por el modelo a estas variables y, por lo tanto, menor será el poder discriminatorio de la evaluación realizada a través de DEA (Jenkins & Anderson, 2003).

En este trabajo proponemos una metodología secuencial de selección de variables (*inputs* y *outputs*) que considere los casos cuando el número de DMUs sea 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.

Este artículo ha sido estructurado de la forma que se describe a continuación. En la sección 2 se presenta una revisión de la literatura sobre los métodos de selección de variables. En la sección 3 se describe el método propuesto en este trabajo y, en la sección 4, éste es aplicado a un caso de estudio. Finalmente, en la última sección se mencionan las principales conclusiones sobre la aplicación del método propuesto.

Métodos de selección de variables en DEA

Los métodos de selección de variables persiguen, primordialmente, ordenar un conjunto de variables (*inputs* y *outputs*) definidas para evaluar a las DMUs usando modelos DEA, con el propósito de seleccionar un número restringido de variables, de manera de mantener la

relación causal en los modelos y que éstos indiquen adecuadamente el desempeño de las DMUs evaluadas.

En el abordaje propuesto por Lins & Moreira (1999), se parte de un par *input-output* inicial que tenga alta correlación. El criterio de selección de nuevas variables a ser incluidas consiste en verificar cuál de las variables candidatas provoca mayor eficiencia media en el modelo DEA. El método se detiene al alcanzar un número de variables considerado ideal. El inconveniente de este procedimiento es que no toma en cuenta el número de DMUs eficientes obtenido en la evaluación de cada promedio, lo cual aísla el criterio de selección solamente al ámbito de mejor ajuste de la frontera eficiente.

En el método desarrollado por Soares de Mello *et al.* (2004) se propone una metodología de Apoyo Multicriterio a las Decisiones para resolver una problemática de selección $P\alpha$, según clasificación de las problemáticas básicas de apoyo a las decisiones de Roy & Bouyssou (1993). Este 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 medido a través de la eficiencia media, y su normalización crea la variable *SEF*, que describe el impacto del criterio “mejor ajuste a la frontera”, alcanzando el valor 10 en la eficiencia máxima y 0 en la eficiencia mínima. La máxima discriminación es medida por el número de DMUs eficientes que se desea minimizar. La normalización de esta medida genera la variable *SDIS*, la que describe el impacto del criterio “máxima discriminación”, que alcanza el valor 10 para el menor número de DMUs en la frontera y 0 para el mayor. 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, *SEF* y *SDIS*. El descriptor de impacto de este criterio síntesis es la variable *S*, tal que $S = \omega \times SEF + (1-\omega) \times SDIS$, con ω 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 $\omega = 0,5$.

El método propuesto Senra *et al.* (2005) selecciona variables (*inputs* y *outputs*) a partir de un procedimiento exhaustivo usando todas las combinaciones posibles de variables, a las que denomina escenarios. De esta manera, aquella variable que presenta el mayor valor de *S*, tal como se describió anteriormente en la propuesta de Soares de Mello *et al.* (2004), es seleccionada para entrar en el modelo. Sin embargo, y pese a que este método es exhaustivo, se puede criticar que al comparar modelos con distinto número de variables, el término *SEF* perjudica a aquellos modelos que poseen menor número de variables, ya que, según demostraciones empíricas, al aumentar el número de variables utilizadas en la evaluación de modelos DEA, aumenta en promedio la eficiencia relativa del mismo.

El método Stepwise de Wagner & 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.

En la literatura de DEA, ningún método de selección de variables considera explícitamente la situación cuando el número de DMUs es mucho mayor que el número de variables disponibles, involucrando diferentes criterios de parada de los métodos propuestos y un tratamiento diferente de las variables que debieran usarse en el análisis de eficiencia.

A continuación se presenta el método de selección de variables para el caso, cuando el

número de DMUs observadas es mucho mayor que el número de variables disponibles para la evaluación aplicando modelos DEA.

Método de selección de variables, cuando el número de DMUs es mucho mayor que el número de variables

Existen aplicaciones en DEA, como la presentada en este estudio, en que el número de variables disponibles para la evaluación es mucho menor que el número de DMUs evaluadas. Por tal motivo, se 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.

El aporte que entrega el método de Soares de Mello *et al.* (2004) a esta propuesta es un orden de selección de variables y la contribución a la eficiencia técnica media del modelo evaluado. El procedimiento de selección de variables propuesto en este trabajo, se inicia estableciendo la contribución técnica media de cada variable ingresada al modelo. Por lo tanto, este procedimiento comienza exactamente después del orden de selección y contribución de la eficiencia técnica media descrito por Soares de Mello *et al.* (2004).

Después de establecer la contribución técnica media de cada variable ingresada, se debe calcular, también para cada variable, la contribución porcentual de eficiencia técnica (C_p) y la contribución porcentual normalizada de eficiencia técnica (C_N). El valor de C_p es el incremento de la eficiencia técnica media producido por el ingreso de una variable, dividido por la suma de contribuciones de todas las variables ingresadas. El valor de C_N corresponde a la normalización de C_p , la cual se realiza multiplicando el término recién mencionado por un factor de normalización. Este factor corresponde a la cantidad de variables del modelo evaluado dividida por la cantidad máxima de variables ingresada para la evaluación de modelos DEA.

A continuación, se debe calcular el término que será utilizado como criterio para seleccionar las variables a ser usadas en los modelos DEA, el cual corresponde a un "límite de contribución porcentual normalizada aceptable" y se denotará por N^* . Este término corresponde al promedio de C_N , representado por \bar{x}_N , menos la mitad de la desviación de C_N , denotada por $\frac{\sigma_N}{2}$.

El siguiente paso del método consiste en comparar el valor C_N de cada variable con N^* , y según el criterio de selección establecido, seleccionar a las variables que cumplan con la condición: $C_N \geq N^*$. El método concluye cuando la contribución porcentual normalizada de la eficiencia técnica de cada variable, C_N , ha sido comparada con el "límite de contribución porcentual normalizada aceptable", N^* .

En la Figura N° 1 se representa el algoritmo descrito para la selección de variables, cuando el número de DMUs es mucho mayor que número de variables.

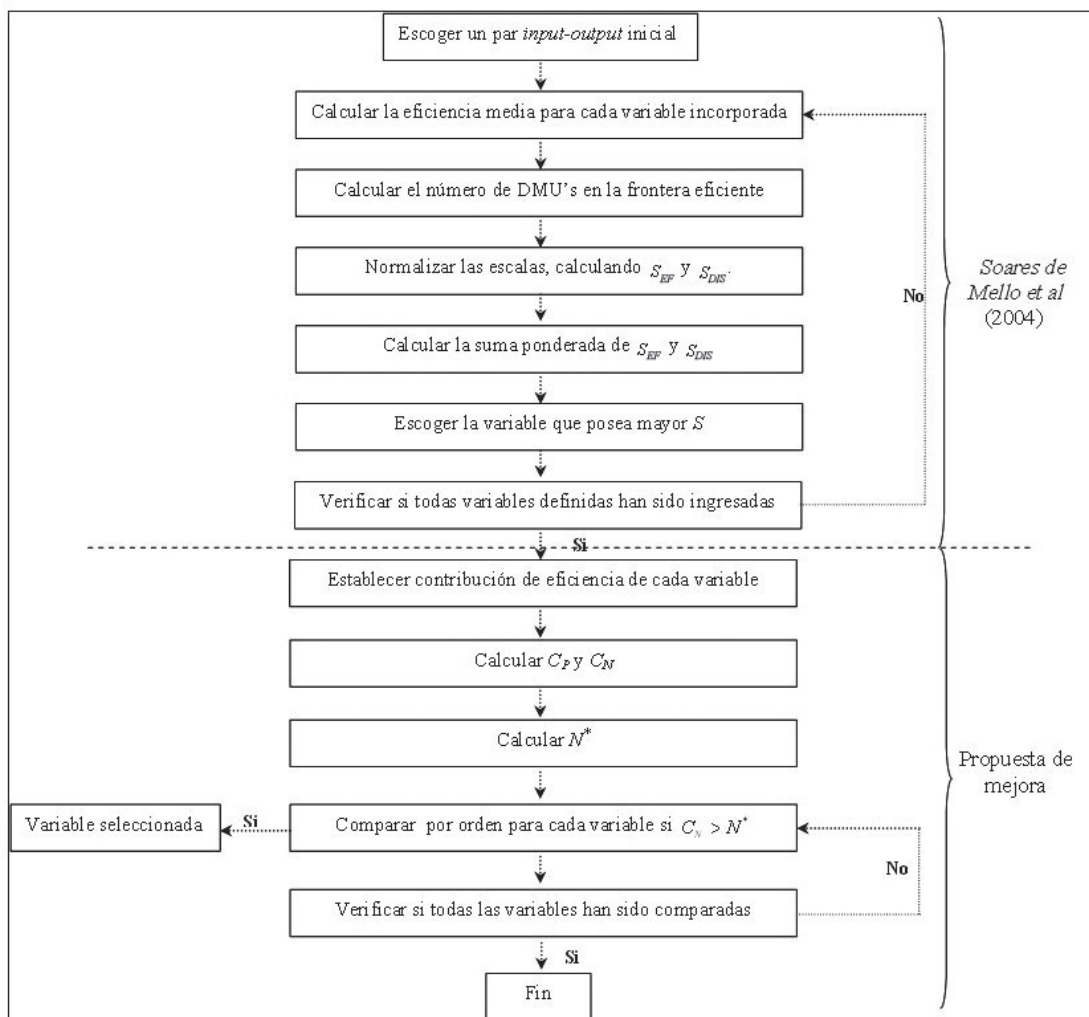


Figura N° 1: Algoritmo propuesto para seleccionar variables en modelos DEA cuando el número de DMUs mucho mayor que número de variables

METODOLOGÍA

Caso de estudio: aplicación para evaluar el proceso productivo de un packing de manzanas

El caso de estudio corresponde a la selección de las variables que permitirán evaluar la eficiencia técnica del proceso productivo de un packing de manzanas. En este estudio son consideradas como DMUs, los turnos de día y de noche en que son procesadas las manzanas de la variedad Gala, durante las temporadas 2004-2005 y 2005-2006.

En el estudio, inicialmente se tenían disponibles ocho variables, siendo éstas cuatro *inputs* y las cuatro *outputs*. Las variables *inputs* corresponden a “número de personas”, “costo variable unitario”, “costo de la mano de obra” y “flujo de kilogramos ingresados a proceso”; mientras que las variables *outputs* corresponden a “porcentaje de tiempo efectivo utilizado”, “porcentaje de embalado de manzanas”, “porcentaje de embalado de manzanas en formato bolsa” y “flujo de cajas embaladas por persona”.

Las variables *inputs* involucran los recursos necesarios para realizar el proceso productivo y, en este caso, se refiere a los recursos utilizados para efectuar los procesos en cada turno. De esta manera, los *inputs* del proceso son:

- Número de personas: corresponde a la cantidad total de personas que participan directamente en la ejecución de las distintas etapas del proceso de la manzana en un turno, e incluye a operarios tales como operadores de grúas horquilla, operarios packing, seleccionadoras packing, entre otros.
- Costo variable unitario: corresponde a la valorización de procesar un kilogramo de manzana, es decir, es la razón entre el costo total del turno evaluado, dividido por la cantidad total de kilogramos procesados en el mismo.
- Costo de la mano de obra: corresponde a la valorización de una persona durante una hora de proceso; es decir, es la razón entre el costo total fraccionado por la cantidad total de personal (“números de personas”) en el total de horas que considera el turno evaluado.
- Flujo de kilogramos ingresados a proceso: corresponde al flujo de kilogramos de manzana que ingresan durante una hora al proceso productivo; es decir, es la razón entre la cantidad total de kilogramos procesados, dividida por el total de horas que considera el turno evaluado.

Las variables *outputs* se refieren a los resultados operacionales obtenidos en el proceso productivo y, en el caso de estudio, estas variables tienen el propósito de evaluar el desempeño de los turnos realizados en las temporadas 2004-2005 y 2005-2006. Por lo tanto, los *outputs* obtenidos en el proceso son:

- Porcentaje de tiempo efectivo utilizado: es la evaluación del porcentaje de tiempo efectivamente utilizado durante un proceso productivo; es decir, es la razón entre la diferencia del tiempo total que posee un turno menos el tiempo total de detenciones del mismo, dividido por el tiempo total del turno. Esta variable se expresa en unidad porcentual.
- Porcentaje de manzanas embaladas: corresponde a la evaluación del porcentaje de manzanas que efectivamente son embaladas; es decir, es la razón entre el total de kilogramos embalados en un turno, fraccionado por el total de kilogramos procesados en el mismo. Esta variable se expresa en unidad porcentual.
- Porcentaje de embalado de manzanas en formato bolsa: corresponde a la evaluación del porcentaje de embalaje que es realizado en bolsas; es decir, es la razón entre el total de kilogramos embalados en formato bolsa, fraccionado por el total de kilogramos embalados en un turno. Esta variable se expresa en unidad porcentual.
- Flujo de cajas embaladas por persona: corresponde a la cantidad de cajas embaladas por una persona en una hora; es decir, es la razón entre la cantidad total de cajas embaladas en un turno, dividido por la cantidad total de personal (“números de personas”) en el total de horas que considera el turno evaluado. Se debe mencionar que una caja se encuentra estandarizada en 16 kilogramos de producto.

A fin de abreviar el nombre de cada variable en la indicación de los resultados más adelante, se usará la siguiente notación:

Tabla N° 1. Abreviatura de la denominación de las variables definidas

Inputs (I)		Outputs (O)	
Nombre	Abreviatura	Nombre	Abreviatura
Número de personas	Personas (I.1)	Porcentaje de tiempo efectivo utilizado	% Tiempo (O.1)
Costo variable unitario	CVU (I.2)	Porcentaje de manzanas embaladas	% Emb (O.2)
Costo mano de obra	C.M.O (I.3)	Porcentaje de manzanas embaladas en formato bolsa	% Bolsas (O.3)
Flujo de kilogramos ingresados a proceso	Kg Proc./h (I.4)	Flujo de cajas embaladas por persona	Cajas/(pers*h) (O.4)

En este caso de estudio, además de las 8 variables mencionadas anteriormente (considerando los *inputs* y *outputs*), han sido observadas 64 DMUs (turnos) en cada temporada evaluada; por lo tanto, el método de selección de variables se aplicará a un conjunto de 128 DMUs. Claramente, en este caso, el número de DMUs (128) es mucho mayor que el número de variables (8). De esta forma, si se aplicara el método de Soares de Mello *et al.*, (2004), se deberían seleccionar todas las variables descritas anteriormente, pues este método se detiene cuando la cantidad de variables seleccionadas es, a lo más, la tercera parte del total de DMUs evaluadas.

Primer paso del método de Soares de Mello *et al.* (2004)

El primer paso del método consiste en escoger un par inicial *input-output*, considerando aquel par que posea el mayor valor de correlación absoluta. Este cálculo se realizó utilizando el software estadístico SPSS versión 10.0. En la Tabla N° 2 se muestran las correlaciones obtenidas:

Tabla N° 2. Resultados de correlación entre combinaciones input-output

	<i>Personas</i>	<i>CVU</i>	<i>C.M.O</i>	<i>Kg Proc./h</i>
<i>% Tiempo</i>	-0,002	-0,628	-0,084	0,510
<i>% Emb</i>	0,429	0,151	-0,316	-0,117
<i>% Bolsas</i>	0,189	0,228	-0,041	-0,261
<i>Cajas/(pers*h)</i>	-0,875	-0,382	0,742	0,245

En la Tabla N° 2 se observa que el par *input-output* con mayor correlación absoluta es la combinación de la variable *input* “número de personas” con la variable *output* “flujo de cajas embaladas por persona”, con una correlación de -0.875 .

Segundo paso del método de Soares de Mello *et al.*, (2004)

El siguiente paso consiste en escoger la siguiente variable que debe ser considerada en el análisis de eficiencia. Según el procedimiento de Soares de Mello *et al.* (2004), para cada variable, se debe calcular un valor normalizado que represente la eficiencia media, denotado por S_{EF} , y un valor normalizado que represente al número de DMUs eficientes, denotado por S_{DIS} . La variable escogida será aquella que obtenga el mayor valor de la suma ponderada de los términos antes mencionados, denotado por S .

Utilizando el software académico *Efficiency Measurement System – EMS* (Scheel, 2000), se evaluó las 128 DMUs. En particular, para este procedimiento se utilizó el modelo BCC (Banker *et al.*, 1984) con orientación *input*, es decir, el modelo DEA con rendimientos de escala variables. La justificación de la elección del modelo BCC orientación *input* se debe a que el proceso productivo presenta diferencias de escala significativas en las DMUs evaluadas. Usando el par inicial *input-output* anteriormente seleccionado, se agrega en cada evaluación una variable *input* u *output*, según corresponda. Los resultados de la primera iteración se muestran en la Tabla N° 3:

Tabla N° 3. Resultado parcial del método de Soares de Mello *et al.* (2004)

Combinación de DMU's	I.1-I.2-O.4	I.1-I.3-O.4	I.1-I.4-O.4	I.1-O.1-O.4	I.1-O.2-O.4	I.1-O.3-O.4
Promedio	81,26%	93,62%	81,28%	78,03%	79,55%	79,57%
SEF	2,30	10,00	2,32	0,00	1,08	1,10
N° de DMU's Efic.	4	6	8	6	8	5
SDIS	10,00	5,00	0,00	5,00	0,00	7,50
S	6,15	7,50	1,16	2,50	0,54	4,30

De la Tabla N° 3 se puede observar que la siguiente variable seleccionada debiera ser el *input* “costo mano de obra” (I.3), pues posee el mayor valor de S en relación a las demás combinaciones de variables, siendo este valor igual a 7,5. Este procedimiento se debe realizar para todas las variables, con el fin de obtener el orden de selección de éstas y conocer, además, la contribución a la eficiencia técnica media de cada variable.

Propuesta de mejora al método de Soares de Mello *et al.* (2004)

Debido a que la aplicación del método de selección de variables propuesto por Soares de Mello *et al.* (2004) impone para este estudio seleccionar todas las variables definidas, se propone una mejora a este método que permita una selección adecuada de las variables para la evaluación en modelos DEA.

En la Tabla N° 4 se presenta el orden de ingreso y la contribución a la eficiencia técnica media que produjo cada variable al ser ingresada al modelo DEA. Además, se indica la contribución porcentual (C_p) y contribución porcentual normalizada (C_N) de cada variable.

La contribución porcentual (C_p) indica la medida en que una variable seleccionada aporta a la contribución total de la eficiencia técnica media; es decir, es la razón entre el aporte individual de la variable seleccionada con la suma de las contribuciones de las variables seleccionadas. En la Tabla N° 4 se aprecia que la contribución de las variables seleccionadas fue cambiando a medida que al modelo DEA se iba integrando una mayor cantidad de variables. Se puede observar, también, que la variable con mayor contribución fue el *input* “costo mano de obra”, el cual aportó el 16,78% de la eficiencia técnica. Además, se observa que la variable *output* “porcentaje de tiempo efectivo utilizado” aportó el menor porcentaje de contribución a la eficiencia técnica media. La contribución a la eficiencia técnica media desde la incorporación de la variable “costo mano de obra” hasta la última variable incorporada, que corresponde al *input* “costo variable unitario”, es de 20,36%.

La contribución porcentual normalizada (C_N) es una medida estandarizada y comparable del nivel de contribución de cada variable, considerando el orden de ingreso que tuvo. Este valor servirá como medida de comparación para la selección de variables según se propone a continuación.

Para obtener el valor de la contribución porcentual normalizada, se debe multiplicar el valor de la contribución porcentual por un factor de normalización. Este factor de normalización corresponde a la cantidad de variables utilizadas en el modelo evaluado, dividido por la máxima cantidad de variables usadas para evaluar un modelo, que en este caso es ocho. Por ejemplo, cuando se produjo la selección de la variable “costo mano de obra”, el modelo estaba compuesto por tres variables, que corresponden al par inicial *input-output*, más la variable ingresada para realizar la evaluación correspondiente (tal como lo indica la Tabla N° 4). En este caso, el factor de normalización es 3/8, el cual debe multiplicarse por la contribución porcentual, que corresponde a 82,4%, para obtener la contribución porcentual normalizada, que en este caso es 30,90%.

Tabla N° 4. Resumen de cálculos para la aplicación del método propuesto

Variable	Nombre	Eficiencia técnica media	Contribución	Contribución porcentual (C _p)	Contribución Porcentual Normalizada (C _N)
Output	<i>Cajas/(Pers*h)</i>	76,844%	-	-	-
Input	<i>Personas</i>				
Input	C.M.O.	93,622%	16,778%	82,40%	30,90%
Output	% Bolsas	94,427%	0,805%	3,95%	1,98%
Output	% Tiempo	94,785%	0,358%	1,76%	1,10%
Input	Kg Proc /h	95,880%	1,095%	5,38%	4,03%
Output	% Emb	96,448%	0,568%	2,79%	2,44%
Input	C.V.U.	97,205%	0,757%	3,72%	3,72%

La mejora al método de Soares de Mello *et al.* (2004) propone un nuevo criterio para la selección de variables mediante una medida denominada N^* , que corresponde a la diferencia entre el promedio de la contribución porcentual normalizada del conjunto de variables, denominada \bar{x}_N , y la mitad de la desviación estándar de la contribución porcentual normalizada del conjunto de variables, denominada $\frac{\sigma_N}{2}$. Una variable será seleccionada si su contribución porcentual normalizada es mayor que esta medida.

Para el caso analizado, en promedio se obtuvo una contribución porcentual normalizada de 7,36% y una desviación estándar de 11,58%, por lo tanto, la medida de selección es:

$$N^* = 7,36\% - \frac{11,58\%}{2} = 1,57\%$$

Las variables fueron seleccionadas aplicando este criterio, aportando en mayor medida al poder de decisión de posteriores análisis de eficiencia.

En este caso de estudio, las variables seleccionadas para evaluar las 128 DMUs (turnos) son los *inputs* “número de personas”, “costo variable unitario”, “costo de la mano de obra” y “flujo de kilogramos ingresados a proceso”; además de los *outputs* “porcentaje de manzanas embaladas”, “porcentaje de embalado en formato bolsas” y “flujo de cajas embaladas por persona”. Por lo tanto, sólo siete de las ocho variables disponibles han sido seleccionadas para participar en la evaluación de la eficiencia de las DMUs.

CONCLUSIONES

La mayor parte de las aplicaciones de los modelos DEA encontrada en la literatura no coloca atención en la selección de las variables usadas para el análisis de eficiencia. Por otro lado, cuando se coloca atención a este tema, la motivación principal es que el caso estudiado presenta un número de DMUs pequeño en relación al número de variables disponibles, ya que es sabido que en estos casos el poder discriminatorio de los modelos DEA es bajo (Jenkins & Anderson, 2003). Sin embargo, en este trabajo se propone realizar la selección de variables independientemente del número de DMUs observadas, pues no todas las variables aportan de igual manera al análisis de eficiencia. De esta manera, el método para la selección de variables propuesto aborda 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 el número de DMUs es mucho mayor que el número de variables.

Finalmente, cabe destacar que, al aplicar el método propuesto 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., & 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., & Rhodes E. (1978).** Measurement the efficiency of decision making units. *European Journal of Operational Research*, 2, 429-444.
- Coelli T., Prasada R., & Battese G.E. (2005).** An Introduction to Efficiency and Productivity Analysis, 2nd Edition. Springer, USA.
- Golany, B., & Roll, Y. (1989).** An application procedure for DEA. *OMEGA*, 17(3), 237-250.
- González-Araya, M.C. (2006).** Introducción al análisis y encapsulamiento de datos. Apuntes del Diplomado en Ingeniería Industrial, Facultad de Ingeniería, Universidad de Talca.
- Jenkins, L., & Anderson, M. (2003).** A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research*, 147, 51-61.
- Lins, M.P.E., & Moreira, M.C. (1999).** Método I-O stepwise para seleção de variáveis em modelos de análise envoltória de dados. *Pesquisa Operacional*, 19, 39-50.
- Roy, B., & Bouyssou, D. (1993).** Aide Multicritère à la Décision: Méthodes et Cas. Economica, Paris.
- Scheel, H. (2000).** EMS: Efficiency Measurement System version 1.3, Users Manual. University of Dortmund.
- Senra, L.F., Nanci, L.C., & Meza, L. (2005).** Revisão dos métodos totais de seleção de variáveis em DEA. In: Anais do XXXVII SBPO - Simpósio Brasileiro de Pesquisa Operacional (pp. 398-405). Gramado, Brasil.

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.

