



FACULTAD  
DE CIENCIAS  
ECONÓMICAS



Universidad  
Nacional  
de Córdoba

# REPOSITORIO DIGITAL UNIVERSITARIO (RDU-UNC)

## Análisis comparativo de tres métodos de selección de variables en DEA

Josefina Racagni, Hernán Pablo Guevel, Sofía Cortaberra

Ponencia presentada en XXIX Encuentro Nacional de Docentes en Investigación Operativa y  
XXVII Escuela de Perfeccionamiento en Investigación Operativa realizado en 2016 en la  
Pontificia Universidad Católica Argentina. Buenos Aires, Argentina



Esta obra está bajo una [Licencia Creative Commons Atribución-NoComercial-CompartirIgual  
4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)



## ANÁLISIS COMPARATIVO DE TRES MÉTODOS DE SELECCIÓN DE VARIABLES EN DEA

JOSEFINA RACAGNI – HERNÁN PABLO GUEVEL – SOFÍA CORTABERRIA  
Facultad de Ciencias Económicas - Universidad Nacional de Córdoba  
*jracagni@gmail.com - heguevel@eco.unc.edu.ar - scortaberrria@gmail.com*

**Palabras Claves:** Selección de variables - DEA - Empresa de Servicios Financieros

### RESUMEN EXTENDIDO

#### 1. OBJETIVOS DEL TRABAJO

El objetivo del presente trabajo es comparar diferentes métodos para la selección de variables basados en el Análisis Envolvente de Datos (DEA) partiendo de un conjunto más amplio de indicadores, para reducir dimensiones.

#### 2. METODOLOGÍA Y APLICACIÓN

Los métodos de Análisis Envolvente de Datos (DEA) buscan determinar el desempeño de un conjunto de unidades homogéneas que, a partir de un mismo conjunto de “*Inputs*”, genera un mismo conjunto de “*Outputs*”, ordenándolas en términos de su “eficiencia relativa”. Estos métodos determinan una frontera de mejores prácticas, sobre la que se ubican las unidades eficientes, para lo que se considera necesario incluir las variables que contribuyan a describir el problema abarcando todos los aspectos relevantes; excluyendo aquellas que estén evaluando características ya representadas a través de otra variable y analizando adecuadamente las variables seleccionadas como *inputs* o *outputs* (Roy y Bouyssou, 1993).

Basándonos en estudios previos realizados sobre una empresa de servicios financieros<sup>1</sup>, donde se analizaba la eficiencia de un subconjunto de sus sucursales a través de los modelos DEA clásicos y comparando esta medida con sus resultados para el *Balance Scorecard* (Cortaberría y Racagni, 2014), propusimos a la organización trabajar con un conjunto diferente de variables, debido a que los indicadores empleados previamente estaban expresados como “porcentaje de cumplimiento de los objetivos del BSC” y surgían de una elaboración interna de compleja interpretación al momento de efectuar recomendaciones.

Consecuentemente la empresa proporcionó una nueva base de datos con información de los indicadores que considera adecuados para evaluar la eficiencia de las unidades bajo estudio. A los fines del presente trabajo, nos concentramos en los datos del semestre cerrado en Octubre de 2011, para el que contamos con 17 indicadores para 51 sucursales de la organización.

Tal como en las etapas previas de nuestra investigación y con el fin de clasificar las variables como “*inputs*” o “*outputs*”, consideramos a un indicador como “*input*” cuando, manteniendo constantes los valores de todos los demás indicadores, la eficiencia de las DMUs disminuye, en caso de aumentar el valor del indicador considerado; y considerar a un indicador como “*output*” si, al aumentar el valor de tal indicador, manteniéndose constantes los valores de los restantes, la eficiencia de la DMU aumenta siguiendo a Alberto et al. (2000).

Teniendo en cuenta lo indicado en el párrafo anterior las variables

<sup>1</sup> Por razones de confidencialidad no divulgaremos el nombre de la misma ni los de sus sucursales.



candidatas a ser seleccionadas, se presentan en la Tabla 1:

Tabla 1: Variables candidatas para la selección

| Variable                       | Código             | Un. De Medida      | Clasif. | Variable                          | Código         | Un. De Medida      | Clasif. | Variable  | Código           | Un. De Medida       | Clasif. |
|--------------------------------|--------------------|--------------------|---------|-----------------------------------|----------------|--------------------|---------|---|------------------|---------------------|---------|
| Resúmenes Asignados            | ResAsig            | Unidades           | Input   | Ingresos Netos de Costos Directos | IngNet CtosDir | Miles de \$        | Output  | Altas de Servicios en Celular   | AltSNC           | Nro de altas        | Output  |
| Promedio de Empleados          | PromEmp            | Personas           | Input   | Aperturas                         | Apert          | Nro de solicitudes | Output  | Alta de Productos   | AltProd          | Nro de altas        | Output  |
| Gastos de Personal             | GtosPers           | Miles de \$        | Input   | Adherentes                        | Adh            | Nro de solicitudes | Output  | Porcentaje de Recupero Promedio   | %Recup Prom      | % del total en mora | Output  |
| Gastos Ejecutados sin Personal | GtosEjecS/personal | Miles de \$        | Input   | Facturación Comercios Adheridos   | FactCA         | Miles de \$        | Output  | Satisfacción Promedio Clientes  | SatProm Clientes | Ptos (0 a 11)*      | Output  |
| Aperturas Rechazadas           | ApRech             | Nro de solicitudes | Input   | Alta Débitos Automáticos          | AltDeb         | Nro de altas       | Output  | Satisfacción Promedio Comercios   | SatProm CA       | Ptos (0 a 11)*      | Output  |
| Adherentes Rechazados          | AdhRech            | Nro de solicitudes | Input   | Alta de Suscripciones Revista     | AltRev         | Nro de altas       | Output  | * estas variables se miden en base al puntaje obtenido en una encuesta de satisfacción. |                  |                     |         |

En virtud de la estrecha relación entre el número de variables y el número de unidades a evaluar, que incide generalmente sobre el poder de discriminación entre unidades eficientes e ineficientes, y sin perjuicio de la existencia de otros criterios, Cooper, Seiford y Tone. (2007) proponen evaluar los grados de libertad del modelo de la siguiente manera:

$$n \geq \max \{m \times s; 3 (m + s)\} \quad (1)$$

Donde:

**n** es el número de DMUs a analizar, en este caso el mínimo de unidades necesarias para que el modelo DEA pueda discriminar correctamente, **m** es el número de *inputs*, y **s**, es el número de *outputs*.

En el presente caso observamos que, si se quisiera trabajar con todas las variables candidatas a ser incluidas en el análisis, deberíamos contar con, al menos 66 DMUs, por lo que resulta necesario reducir dimensiones, para aumentar el poder de discriminación del modelo DEA. A tal fin, nos hemos concentrado en los métodos propuestos por Lins y Moreira (1999), Soares de Mello et al. (2004) y Wagner y Shimshak (2007).

**Método de Lins y Moreira (1999):** este método parte de un par inicial de variables (1 input y 1 output, con la mayor correlación absoluta) e incorpora de a una variable repitiendo los siguientes pasos: 1) calcula la medida de eficiencia y obtiene el ordenamiento de las DMUs con este par inicial de variables de acuerdo al modelo DEA que se desee o corresponda aplicar; 2) calcula la correlación entre el ranking DEA obtenido en el paso anterior y todas las variables candidatas restantes; 3) analiza los coeficientes de correlación y selecciona las próximas variables candidatas a ingresar al modelo, una entrada (la que verifique la mayor correlación positiva con el ordenamiento) y una salida (la que verifique la mayor correlación negativa con el ranking); 4) incorpora alternativamente al modelo estas variables y calcula la eficiencia promedio de los nuevos ordenamientos obtenidos (este valor indicará el grado de ajuste a la frontera de eficiencia que se obtiene de agregar las nuevas variables); 5) se incorporará al subconjunto preseleccionado la variable que genere la mayor eficiencia media. El proceso se detiene cuando no se logra incrementar significativamente este índice o bien cuando se alcanza un número ideal de variables.



**Método de Soares de Mello et al. (2004):** también parte de un par inicial de variables, para el que calcula la medida de eficiencia, e incorpora paso a paso nuevas variables evaluando a cada una de las candidatas en base a un índice que pondera la bondad de ajuste a la frontera y el poder de discriminación del modelo, agregándolos en una única medida “de compromiso”, de acuerdo a la siguiente fórmula de cálculo:

$$S = \omega S_{EF} + (1-\omega) S_{DIS} \quad (2)$$

Donde:

$S_{EF}$ , corresponde al valor normalizado de la eficiencia media para el ranking DEA resultante de agregar cada una de las variables candidatas, asumiendo un valor de 10 para la que aporte la mayor eficiencia media y 0 para la menor.

y  $S_{DIS}$  representa el valor normalizado del poder de discriminación del modelo, medido a través del número de unidades eficientes, que asume el valor 10 para la variable candidata con menos unidades eficientes y 0 para la que incluye el mayor número de ellas.

El ponderador  $\omega$  es un número entre cero y uno, que en caso de que no se desee que alguno de los criterios predomine por sobre el otro deberá asumir el valor 0,5. En este caso, dado que nos interesaba más el poder de discriminación del modelo que el ajuste a la frontera, trabajamos con  $\omega=0,40$ .

**Método de Wagner y Shimshak (2007):** parte de calcular el ranking de eficiencia de las DMUs usando todas las variables disponibles y las va eliminando de a una secuencialmente del conjunto de variables teniendo en cuenta el promedio de cambio en las eficiencias para las DMUs. Este procedimiento se detiene cuando en el conjunto de variables sólo resta un *input* y un *output*, de acuerdo a la propuesta de los autores, o cuando se alcanza un número ideal de variables a criterio de algún experto o del analista.

Tabla 2: Variables seleccionadas en cada método

| Variables \ Metodología | Res Asig | Prom Emp | Gtos Pers | GtosEjecS /personal | ApRech | Adh Rech | IngNet CtosDir | Apert | Adh | FactCA | AltDeb | AltRev | AltSNC | AltProd | %Recup Prom | SatProm Clientes | SatProm CA | DMU EF | EF PROM  |
|-------------------------|----------|----------|-----------|---------------------|--------|----------|----------------|-------|-----|--------|--------|--------|--------|---------|-------------|------------------|------------|--------|----------|
| Lins & Moreira          | SI       | -        | SI        | SI                  | -      | -        | SI             | SI    | SI  | -      | -      | -      | -      | -       | SI          | SI               | SI         | 22     | 0,958747 |
| Soares de Mello, et al  | SI       | SI       | SI        | SI                  | SI     | -        | SI             | SI    | -   | SI     | -      | -      | SI     | -       | -           | -                | -          | 12     | 0,887741 |
| Wagner & Shimshak       | SI       | -        | SI        | SI                  | SI     | -        | -              | SI    | SI  | SI     | -      | SI     | -      | -       | SI          | -                | -          | 25     | 0,933448 |

La base de datos fue procesada con el paquete “*Benchmarking*” del software “R”. En la Tabla 2 comparamos las variables que son seleccionadas por cada método, el número de unidades eficientes resultantes de calcular el ranking DEA y el valor de la eficiencia promedio. En los tres casos detuvimos el proceso de selección al llegar a un total de 9 variables, a fin de que los conjuntos seleccionados fueran comparables.

Podemos notar que el método sugerido por Soares de Melo et al (2004) arroja el menor número de DMUs eficientes y una menor eficiencia promedio. Esto resulta razonable dado el valor escogido para el ponderador de impacto del criterio “ajuste a la frontera”, y el consiguiente ponderador para el “poder de discriminación”. El método propuesto Wagner y Shimshak (op cit), muestra una mayor eficiencia promedio, y es el que mayor número de unidades eficientes registra al detener el proceso de selección. Finalmente, el método propuesto por Lins y Moreira (op cit) es el que muestra un mejor ajuste a la frontera de



eficiencia, arrojando un número intermedio de unidades eficientes.

### 3. CONCLUSIONES

De la comparación de los resultados obtenidos, y del conjunto de variables seleccionadas por cada método, concluimos que 4 de las 9 variables son incluidas por todos ellos. Además consideramos que la propuesta de Lins y Moreira permite reflejar de mejor manera la relación de causalidad entre los *inputs* y *outputs* y la eficiencia de cada sucursal en la actividad de la Empresa. Por otra parte, este método también logra el mejor ajuste a la frontera, con un número no tan elevado de unidades eficientes.

El conjunto de variables seleccionado constituye el punto de partida de un análisis completo de eficiencia para los semestres finalizados en Octubre de 2011; Abril y Octubre de 2012 y Abril de 2013 para la empresa y nos permite mejorar la relación entre el número de *Inputs-Outputs* y DMUs a considerar.

Sería deseable repetir el análisis para los demás períodos a fin de estudiar si la selección de variables obtenida permanece o no inalterable, o si sería conveniente instrumentar un método de selección diferente, como el propuesto por González Araya et al (2013) que trabaja con series de tiempo.

### 4. REFERENCIAS

- ALBERTO, C. L.; PÉREZ MACKEPRANG, C. O., CARIGNANO, C. E., FUNES, M. (2000): "Evaluación de la Eficiencia en Investigación Científica y Desarrollo Experimental en Países Iberoamericanos". Publicado en Anales del X CLAIO. México D.F., México.
- Cooper, W. W., Seiford L. M., K. Tone. (2007) "Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software" Springer
- CORTABERRIÁ, S. y RACAGNI, J. (2014) "Evaluación de Eficiencia Empleando DEA Aplicada a Sucursales de una Empresa Financiera". Publicado en Anales del XVII CLAIO. Monterrey, México.
- GONZÁLEZ ARAYA, M.; CAMPOS HERNÁNDEZ, G. y ACOSTA ESPEJO, L.G. (2013). "Selección de Variables en Modelos DEA Usando Series de Tiempo". XLV Simposio Brasileiro de Pesquisa Operacional, <http://www.din.uem.br/sbpo/sbpo2013/pdf/arg0384.pdf>. (consulta: marzo de 2016).
- LINS, M.P.E. & MOREIRA, M.C.B. (1999). "Método I-O Stepwise para Seleção de Variáveis em Modelos de Análise Envolvória de Dados". Pesquisa Operacional, 19(1), 39-50.
- ROY, B., BOUYSSOU, D. (1993): "Aide Multicritère À La Décision: Méthodes Et Cas". Economica – Paris.
- SOARES DE MELLO, J. C. C. B., GOMES, E. G., ANGULO MEZA, L., & LINS, M. P.E. (2004). Selección de variables para el incremento del poder de discriminación de los modelos DEA. Revista de la 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(1), 57-67.