Métodos de selección de variables en R para incrementar el poder discriminatorio del Análisis Envolvente de Datos (DEA) *

Andrés Mauricio Gómez *Universidad de Antioquia*
Andrés Mauricio Villegas *Universidad de Nueva Gales del Sur*
Juan Guillermo Villegas *Universidad de Antioquia*

Resumen: El análisis envolvente de datos (DEA) es una técnica no paramétrica muy utilizada en la medición de la eficiencia relativa. Sin embargo, esta no proporciona pautas claras para la selección de las variables. En este trabajo se presenta un estudio comparativo de DEA combinado con técnicas estadísticas para robustecer sus estimaciones y evitar los criterios ad hoc o juicios particulares en la elección de las variables de entrada y salida que se incluyen en los modelos de DEA. Para ello, se implementó en el lenguaje R la combinación de DEA con Análisis de Componentes Principales, Regresión Lasso y un heurístico de búsqueda iterativa. En particular, se usaron las librerías de R: `stats`, `benchmarking`, `glmnet`, `roi` y `ompr`. El desempeño de DEA se midió en función de la diversificación en la clasificación de las unidades productoras evaluadas (denominadas DMU), garantizando la mayor retención en la varianza total de los datos.

Palabras claves: Análisis envolvente de datos, Selección de variables, Problema de dimensionalidad, Medición del desempeño

Introducción

El análisis envolvente de datos (DEA) es una herramienta popular de los métodos cuantitativos para la toma de decisiones, que mide el desempeño relativo de un conjunto de unidades productoras (DMUs) con múltiples métricas de valoración (entradas y salidas de sus procesos productivos). Sin embargo, el número inicial de variables incluidas en un modelo DEA suele ser muy grande, y la formulación de DEA no proporciona pautas objetivas para la selección de entradas y salidas (Villanueva-Cantillo y Munoz-Marquez, 2021). Por lo cual, los problemas de discriminación entre DMUs eficientes e ineficientes a menudo surgen cuando hay un número relativamente grande de medidas de desempeño (variables) en comparación con el número de DMU; esto puede llevar a que las unidades eficientes se clasifiquen incorrectamente como ineficientes y las unidades ineficientes se clasifiquen erróneamente como eficientes (Charles, Aparicio, y Zhu, 2019).

Así mismo, Liu, Lu, y Lu (2016) en su estudio sobre los frentes de investigación en DEA, identificaron que la selección de variables era una subárea de investigación sólida en la literatura de DEA. Aunque en la metodología de aplicación de DEA se contempla la utilización de reglas ad hoc para la selección de variables de entrada y salida, no es un procedimiento sistemático lo cual deja en manos del tomador de decisiones la elección de las variables. Estas reglas carecen de medidas cuantitativas que indiquen el impacto que tiene la eliminación de variables tanto en el desempeño de las DMU como en la retención de información de los datos originales (Dyson y cols., 2001) (Cook, Tone, y Zhu, 2014).

En consecuencia, existen estudios que se han enfocado en mejorar el poder discriminatorio de DEA, algunos con la intención de aumentar el número de DMUs y conservar el mismo número de variables, usando datos de corte transversal y series de tiempo (Charles y cols., 2019). En cambio, otros han incorporado técnicas estadísticas como el análisis de covarianza parcial (Jenkins y Anderson, 2003), o análisis de componentes principales combinado con DEA (Adler y Golany, 2002), con el objetivo de reducir el número de variables utilizadas pero conservando el mismo número de DMUs.

Metodología

Con el fin de ilustrar el desempeño de estas aproximaciones tanto estadísticas como heurísticas, descritas anteriormente, usaremos un conjunto de datos existente en la literatura, reportados en los trabajos de Wong y Beasley (1990) y Adler y Golany (2002) para presentar sus resultados. En esta ilustración numérica, se comparan siete DMUs (facultades universitarias) con tres variables de insumos (número de docentes, salarios de los docentes, número de empleados

*Autor de contacto: andres.gomez8@udea.edu.co

administrativos) y tres variables de salida (número de estudiantes de pregrado, número de estudiantes de posgrado, y cantidad de artículos publicados), como puede verse en la Tabla 1.

Tabla 1: Datos DEA – Facultades universitarias

	Entradas (Inputs)			Salidas (Outputs)		
DMU	**I1** Empleados Docentes	**I2** Salarios Docentes	**I3** Empleados Administrativos	**O1** Estudiantes de pregrado	**O2** Estudiantes de posgrado	**O3** Publicaciones (Artículos)
DMU1	12	400	20	60	35	17
DMU2	19	750	70	139	41	40
DMU3	42	1500	70	225	68	75
DMU4	15	600	100	90	12	17
DMU5	45	2000	250	253	145	130
DMU6	19	730	50	132	45	45
DMU7	41	2350	600	305	159	97

En este trabajo se realiza la combinación de DEA con: (a) el análisis de componentes principales (PCA); (b) la regularización de variables con Regresión Lasso y (c) una aproximación heurística, basada en la eliminación iterativa de variables, tanto de entrada como de salida. Esto con el propósito de mejorar el poder discriminatorio de los modelos de DEA.

(a). PCA – DEA

Como un número excesivo de variables en un modelo DEA provoca en los resultados una gran cantidad de unidades eficientes, es preferible mantener baja la relación entre el número de entradas y salidas con respecto al número de DMU, y para tal fin Adler y Golany (2001) usaron el Análisis de Componentes Principales (PCA) con el propósito de agregar entradas o salidas, con una pérdida mínima de información, desarrollando una nueva formulación del modelo DEA, denominada PCA–DEA.

Se implementó en R (R Core Team, 2021) el modelo aditivo de DEA combinado con PCA presentado por Adler y Golany (2002) (véase la Ecuación 1), con la ayuda de la función `prcomp::stats` (R Core Team, 2021) para el análisis de componentes principales, así como de las librerías `ROI` (Theußl, Schwendinger, y Hornik, 2020) y `ompr` (Schumacher, 2023) encargadas de la infraestructura de optimización en R. Esta formulación para la estimación de eficiencia de la DMU_a ($z = 1, 2, 3, ..., n$; n = número de DMUs), agrega las variables originales y las variables transformadas por las componentes principales para las entradas ($X = \{X_o, X_{Lx}\}$) y salidas ($X = \{Y_o, Y_{Ly}\}$), donde X_o (Y_o) representa las entradas (salidas) de las variables originales, mientras que X_{Lx} (Y_{Ly}) las variables de entrada (salida) transformadas por PCA. Así mismo, L_x (L_y) representa la matriz de ponderaciones para las entradas (salidas) obtenidos con PCA, con la cual, $X_{pc} = L_x X_{Lx}$ y $X_{pc} = L_y Y_{Ly}$.

$$
\begin{aligned}
\underset{s_o, \sigma_o, s^+_{pc}, s^-_{pc}, \sigma^+_{pc}, \sigma^-_{pc}}{\text{Minimizar}} \quad & -e^T s_o - e^T L'^{-1}(s^+_{pc} - s^-_{pc}) - e^T \sigma_o - e^T L'^{-1}(\sigma^+_{pc} - \sigma^-_{pc}) \\
\text{Sujeto a:} \quad & Y_o \lambda - s_o = Y^a_o \\
& Y_{pc} \lambda - (s^-_{pc} - s^+_{pc}) = Y^a_{pc} \\
& -X_o \lambda - \sigma_o = -X^a_o \\
& -X_{pc} \lambda - (\sigma^-_{pc} - \sigma^+_{pc}) = -X^a_{pc} \\
& L_y^{-1}(s^-_{pc} - s^+_{pc}) \geq 0 \\
& L_x^{-1}(\sigma^-_{pc} - \sigma^+_{pc}) \geq 0 \\
& \lambda, s_o, \sigma_o, s^+_{pc}, s^-_{pc} \sigma^+_{pc}, \sigma^-_{pc} \geq 0
\end{aligned}
\tag{1}
$$

Aquí, las holguras para determinar las ineficiencias están dadas para las entradas (σ_o) y salidas (s_o) originales, como para las variables transformadas ($\sigma^+_{pc}, \sigma^-_{pc}, s^+_{pc}, s^-_{pc}$), como variables auxiliares no negativas. Con el propósito de traducir los resultados del modelo adictivo, el cual arroja resultados entre $[0, \infty)$ (donde una holgura igual a cero significa ser eficiente), en un intervalo $[0, 1]$, se usa la propuesta por Adler y Volta (2019) para la comparación con otros modelos.

(b). Lasso – DEA

Lasso (*Least Absolute Shrinkage and Selection Operator - Lasso*) pretende seleccionar un modelo de menor dimensión, llevando a cero el efecto estimado de algunas variables mediante la incorporación de un coeficiente de regularización o penalización al problema original, usualmente por mínimos cuadrados (Tibshirani, 1996). Esta regularización ayuda a minimizar el error de predicción del modelo, haciéndolo más parsimonioso y fácil de explicar. Este método es popular cuando la dimensión del problema es mayor que el tamaño de la muestra (Chen, Tsionas, y Zelenyuk, 2021).

Para esta aproximación se usó la librería `Benchmarking` (Bogetoft y Otto, 2020) para el cálculo de eficiencia de los modelos, y de manera conjunta `glmnet` (Friedman, Tibshirani, y Hastie, 2010) para la selección de variables. Aquí, las variables independientes eran las métricas de valoración (entradas y salidas del proceso) y la variable dependiente era la estimación de eficiencia, e iterativamente tras la regularización Lasso, se depuraba una a una las variables, creando diferentes modelos parsimoniosos que permitieran identificar las DMUs con mejor desempeño en el grupo evaluado.

(c). ISA – DEA

Finalmente, se realizó la comparación con una propuesta heurística para la selección de características, tanto de entrada como de salida, basada en la eliminación iterativa de variables (Iterative Search Algorithm, ISA-DEA). La metodología propuesta elige en cada iteración la configuración de variables donde se maximiza una métrica de impureza (índice Gini o entropía) de las eficiencias de las DMU, proporcionando así un conjunto de modelos seleccionados por índices estadísticos de dispersión (Gómez Ardila, Villegas Ramírez, y Villegas Ramírez, 2021).

Resultados

Como punto de partida, al aplicar DEA con todas las variables para las siete unidades productoras, se encuentra que la DMU_4 es la única ineficiente (véase la Tabla 2). A continuación, se presentan los resultados de las técnicas PCA-DEA, Lasso-DEA e ISA-DEA, las cuales buscan mejor la clasificación de eficiencia en DEA.

Tabla 2: Puntajes de eficiencia con DEA con todas las variables

Puntajes de eficiencia (DMU eficiente= 1, DMU ineficiente $\in$ (0,1))							
DMU	DMU1	DMU2	DMU3	DMU4	DMU5	DMU6	DMU7
Puntaje	1	1	1	0.820	1	1	1

Se aplico la técnica de PCA-DEA, con la formulación de la Ecuación 1 para todas las posibles configuraciones de componentes principales, con el objetivo de abarcar todo el espectro en la reducción de dimensionalidad y su impacto en la estimación de eficiencia. Para los nueve posibles modelos evaluados, se presentan en la Tabla 3 las métricas de impureza (Gini para la eficiencia, la ineficiencia y el ponderado) (Gómez Ardila y cols., 2021), la cantidad y la proporción de DMUs eficientes, las varianzas retenidas para las entradas y salidas (métricas propias de PCA), y la varianza retenida por el total de variables (usando la Covarianza Parcial de Jenkins y Anderson (2003))).

Tabla 3: Métricas de impureza y varianza retenida para los modelos con PCA-DEA

Modelo	Componentes Principales		Gini	Gini	Gini	DMUs	% DMUs	Varianza Retenida		
	Entradas (CPi)	Salidas (CPo)	Eficiencia	Ineficiencia	Ponderado	Eficientes	Eficientes	Entradas	Salidas	Total
PCA_DEA_1	3	3	0.0412	0.7500	0.3956	6	85.7 %	1.0000	1.0000	1.0000
PCA_DEA_2	3	2	0.0398	0.7500	0.3949	6	85.7 %	1.0000	0.9755	1.0000
PCA_DEA_3	3	1	0.0363	0.5546	0.2955	3	42.9 %	1.0000	0.9456	1.0000
PCA_DEA_4	2	3	0.1036	0.5663	0.3349	4	57.1 %	0.9995	1.0000	1.0000
PCA_DEA_5	2	2	0.1010	0.5638	0.3324	4	57.1 %	0.9995	0.9755	0.9728
PCA_DEA_6	2	1	0.0673	0.2255	0.1464	1	14.3 %	0.9995	0.9456	0.9122
PCA_DEA_7	1	3	0.1253	0.5431	0.3342	4	57.1 %	0.8679	1.0000	0.9839
PCA_DEA_8	1	2	0.1204	0.4907	0.3056	2	28.6 %	0.8679	0.9755	0.8893
PCA_DEA_9	1	1	0.0803	0.2167	0.1485	1	14.3 %	0.8679	0.9456	0.7882

Resulta claro que al utilizar menos componentes principales, la varianza retenida disminuye, y se evidencia en la Tabla 3 como esto tiene un impacto en la reducción de DMUs estimadas como eficientes, pasando de un 85.7 % a 14.3 % de DMUs eficientes entre el PCA_DEA1 y el PCA_DEA9 , modelos con la máxima y mínima retención de varianza respectivamente.

Por otra parte, tras emplear la regresión Lasso en la búsqueda de aquellas variables que mejor explican la eficiencia, se evidencia en la Tabla 4 el resultado de la penalización Lasso sobre todas las variables disponibles (entradas y salidas), y al igual que en el ejercicio anterior, se calcular las métricas de impureza y la covarianza parcial para cada uno de los modelos, con el propósito de unificar los resultados de las diferentes técnicas presentes en este trabajo.

Tabla 4: Métricas de impureza y varianza retenida para los modelos con Lasso-DEA

Modelo	λ	Varibles						Gini	Gini	Gini	DMUs	% DMUs	Varianza Retenida		
		I1	I2	I3	O1	O2	O3	Eficiencia	Ineficiencia	Ponderado	Eficientes	Eficientes	Entradas	Salidas	Total
LASSO_DEA$_1$	0	✓	✓	✓	✓	✓	✓	0.0198	0.7500	0.3849	6	85.7 %	1.0000	1.0000	1.0000
LASSO_DEA$_2$	0.0012	✓	✓	✓	✓	✓		0.0198	0.7500	0.3849	6	85.7 %	1.0000	0.9602	1.0000
LASSO_DEA$_3$	0.0021	✓	✓		✓	✓	✓	0.0334	0.6274	0.3304	5	71.4 %	0.9958	1.0000	1.0000
LASSO_DEA$_4$	0.0037		✓	✓	✓	✓	✓	0.0334	0.6368	0.3351	5	71.4 %	0.9983	1.0000	1.0000
LASSO_DEA$_5$	0.0095			✓	✓	✓	✓	0.2400	0.4161	0.3281	2	28.6 %	0.6955	1.0000	0.9920
LASSO_DEA$_6$	0.0115			✓	✓		✓	0.2426	0.3898	0.3162	1	14.3 %	0.6955	0.9631	0.9879
LASSO_DEA$_7$	0.0429			✓	✓			0.2629	0.3773	0.3201	1	14.3 %	0.6955	0.8913	0.9435
LASSO_DEA$_8$	0.0518	✓		✓	✓			0.0435	0.5863	0.3149	4	57.1 %	0.9992	0.8913	0.9649
LASSO_DEA$_9$	0.0754	✓			✓			0.0719	0.3742	0.2230	1	14.3 %	0.7515	0.8913	0.9222

En la Tabla 4 se resumen los modelos para cada λ de penalización para el cual se dio una cambio en la configuración de variables, es decir, la inclusión u omisión de alguna variable. Por ejemplo, el LASSO_DEA$_1$ tiene un $\lambda = 0$ para el cual se marca que se usaron todas las variables. Entre las nueve configuraciones de variables, se encuentran que los índices Gini para la eficiencia más altos son 0.2400, 0.2426 y 0.2629, y corresponden al LASSO_DEA$_5$, LASSO_DEA$_6$ y LASSO_DEA$_7$, respectivamente, en los cuales se logra obtener una reduccción en el porcentaje de DMUs eficientes por debajo al 28.5 %.

En cuanto a los resultados del algoritmo heurístico (ISA-DEA), tras realizar la búsqueda iterativa de modelos orientada por la maximización del "Gini ponderado" ($GiniPonderado = \alpha(GiniEficiencia) + (1 - \alpha)(GiniIneficiencia)$) propuesto por Gómez Ardila y cols. (2021) para diferentes ponderaciones con valores de λ entre [0, 1], se identifican siete modelos con métricas únicas después de omitir los resultados para aquellos λ donde no se presentó un cambio en los resultados.

Tabla 5: Métricas de impureza y varianza retenida para los modelos con ISA-DEA

Modelo	α	Variables						Gini	Gini	Gini	DMUs	% DMUs	Varianza Retenida		
		I1	I2	I3	O1	O2	O3	Eficiencia	Ineficiencia	Ponderado	Eficientes	Eficientes	Entradas	Salidas	Total
ISA_DEA$_1$	0	✓		✓	✓	✓	✓	0.0202	0.7500	0.7500	6	85.7 %	0.9992	1.0000	1.0000
ISA_DEA$_2$	0.02	✓		✓	✓	✓		0.0202	0.7500	0.7354	6	85.7 %	0.9992	0.9602	0.9954
ISA_DEA$_3$	0.46	✓		✓		✓		0.1636	0.5062	0.3486	3	42.9 %	0.9992	0.8998	0.9903
ISA_DEA$_4$	0.49			✓		✓	✓	0.2661	0.4108	0.3399	2	28.6 %	0.6955	0.9565	0.9660
ISA_DEA$_5$	0.62			✓			✓	0.2662	0.3470	0.2969	1	14.3 %	0.6955	0.8957	0.9553
ISA_DEA$_6$	0.68			✓		✓		0.3201	0.2339	0.2925	1	14.3 %	0.6955	0.8998	0.9129
ISA_DEA$_7$	0.93	✓				✓		0.1945	0.3240	0.2036	1	14.3 %	0.7515	0.8998	0.9359

Como se ha mencionado y expuesto en todo el documento, la importancia de la selección de variables es crucial en DEA, sobre todo bajo criterios objetivos, por ello para las tres metodologías en la Tabla 6 se presentan solo aquellos modelos donde la métrica de impureza es máxima y los que son aptos bajo la mirada de las reglas empíricas. Esta regla busca que la relación entre la cantidad de variables y unidades de evaluación esté en el rango $[\frac{1}{3}\ ,\ \frac{1}{2}\ ]$.

Tabla 6: Modelos sugeridos con PCA-DEA, LASSO-DEA e ISA-DEA

Modelo	Gini Eficiencia	Gini Ineficiencia	Gini Ponderado	DMUs Eficientes	% DMUs Eficientes	Varianza Retenida
PCA_DEA$_7$	0.1253	0.5431	0.3342	4	0.571	0.9839
PCA_DEA$_7$	0.1204	0.4907	0.3056	2	0.286	0.8893
LASSO_DEA$_7$	0.2629	0.3773	0.3201	1	0.143	0.9435
LASSO_DEA$_8$	0.0435	0.5863	0.3149	4	0.571	0.9646
ISA_DEA$_3$	0.1636	0.5262	0.3486	3	0.429	0.9903
ISA_DEA$_4$	0.2661	0.4108	0.3399	2	0.286	0.966
ISA_DEA$_6$	0.3201	0.2339	0.2925	1	0.143	0.9129

Como resultado general, se evidencia que los modelos sugeridos por la metodología propuesta (ISA-DEA) presentan las mejores valoraciones, considerando las métricas trasversales anteriormente mencionadas, e independientemente de las particularidades metodológicas de cada técnica, ya sea la sustitución de los datos por nuevas variables no correlacionadas o la identificación de unas cuantas variables con la mayor relación ante una variable independiente

(en este caso la estimación de eficiencia con DEA). El ISA_DEA3 e ISA_DEA4 con valores de α de 0.46 y 0.49 respectivamente, obtienen los valores más altos para el Gini Ponderado y de varianza explicada (véase la Tabla 6), con los cuales se logra disminuir considerablemente el porcentaje de DMUs clasificadas como eficientes, usando la mitad de las variables iniciales.

En definitiva, el ISA_DEA3 es el de mejor rendimiento, ya que reduce del 85.7 % al 42.9 % el número de DMUs clasificadas como eficientes, con tan solo la selección del 50 % de las variables iniciales. Este modelo usa dos insumos (I1: Empleados Docentes, e I3: Empleados Administrativos) y un producto (O2: Estudiantes de posgrado) en la estimación de eficiencia de las siete facultades universitarias (DMUs), variables con las cuales se explica el 99.03 % de la varianza total de los datos.

Conclusiones

En este trabajo se ilustra el problema de dimensionalidad presente en DEA ante la modelación con muchas variables de entrada y salida, y pocas DMUs, donde es alta la probabilidad de clasificar una DMU como eficiente cuando no lo es. Resulta claro que cada metodología tiene sus bondades y desventajas, como por ejemplo el PCA que aprovecha todas las variables y las sintetiza por medio de combinaciones lineales no correlacionadas, de ahí que este enfoque combinado con DEA resulte una alternativa muy beneficiosa para la solución de la maldición de la dimensionalidad. Sin embargo, con respecto a las otras aproximaciones como Lasso-DEA e ISA-DEA, con PCA-DEA no se puede identificar de forma directa cuáles variables están impactando el modelo. Estas otras aproximaciones indican las variables usadas y omitidas en cada modelo, lo cual le permite al analista de información ver exploratoriamente que variables son más relevantes y contrastarlo con su experticia y conocimiento acerca del sector estudiado.

Los resultados en este estudio permiten determinar, dada una configuración de variables y bajo cualquiera de las tres aproximaciones, el nivel de entropía o grado de desigualdad de los puntajes de eficiencia (índice Gini). Además, se indica la varianza explicada para los modelos elegidos en cada iteración, lo cual resulta relevante para el entendimiento del sacrificio en la eliminación de variables en función de aumentar el poder discriminatorio de DEA.

Referencias

Adler, N., y Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to western europe. *European Journal of Operational Research*, *132*(2), 260–273.

Adler, N., y Golany, B. (2002). Including principal component weights to improve discrimination in data envelopment analysis. *Journal of the Operational Research Society*, *53*(9), 985–991.

Adler, N., y Volta, N. (2019). Ranking methods within data envelopment analysis. En *The palgrave handbook of economic performance analysis* (pp. 189–224). Springer.

Bogetoft, P., y Otto, L. (2020). Benchmarking with dea and sfa [Manual de software informático]. (R package version 0.29)

Charles, V., Aparicio, J., y Zhu, J. (2019). The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis. *European Journal of Operational Research*, *279*(3), 929–940.

Chen, Y., Tsionas, M., y Zelenyuk, V. (2021). Lasso+ dea for small and big wide data. *Omega*, 102419.

Cook, W. D., Tone, K., y Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega*, *44*, 1–4.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., y Shale, E. A. (2001). Pitfalls and protocols in dea. *European Journal of operational research*, *132*(2), 245–259.

Friedman, J., Tibshirani, R., y Hastie, T. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. doi: 10.18637/jss.v033.i01

Gómez Ardila, A., Villegas Ramírez, A., y Villegas Ramírez, J. (2021). Método heurístico de selección de variables para el incremento del poder discriminatorio de dea: caso de aplicación a las eps colombianas. *XXX Simposio Internacional de Estadística 2021 - Evento virtual*, 183–190.

Jenkins, L., y Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research*, *147*(1), 51–61.

Liu, J. S., Lu, L. Y., y Lu, W.-M. (2016). Research fronts in data envelopment analysis. *Omega*, *58*, 33–45.

R Core Team. (2021). R: A language and environment for statistical computing [Manual de software informático]. Vienna, Austria. Descargado de https://www.R-project.org/

Schumacher, D. (2023). ompr: Model and solve mixed integer linear programs [Manual de software informático]. Descargado de https://github.com/dirkschumacher/ompr (R package version 1.0.4.9000)

Theußl, S., Schwendinger, F., y Hornik, K. (2020). Roi: An extensible r optimization infrastructure. *Journal of Statistical Software*, *94*(15), 1–64. Descargado de https://www.jstatsoft.org/index.php/jss/article/view/v094i15 doi: 10.18637/jss.v094.i15

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Villanueva-Cantillo, J., y Munoz-Marquez, M. (2021). Methodology for calculating critical values of relevance measures in variable selection methods in data envelopment analysis. *European Journal of Operational Research*, *290*(2), 657–670.

Wong, Y.-H., y Beasley, J. (1990). Restricting weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*, *41*(9), 829–835.