

# DIRECCIÓN DE DESARROLLO DIGITAL

Unidad de Científicos de Datos



**DNP** Departamento  
Nacional  
de Planeación



Unidad de Científicos de Datos  
[ucd@dnpp.gov.co](mailto:ucd@dnpp.gov.co)

# Metodología para el cálculo del índice de eficiencia para la Rama Judicial en Colombia

## Tabla de contenido

1. Introducción.....	2
2. Análisis de envoltura de datos .....	3
2.1 Origen.....	3
2.2 Ventajas .....	3
2.3 Desventajas .....	4
2.4 ¿Por qué calcular la eficiencia con DEA?.....	4
3. Descripción de la metodología DEA.....	5
3.1 DEA orientado a las entradas (VRS).....	7
3.2 DEA orientado a las salidas (VRS).....	9
4. Entradas y salidas para calcular la eficiencia judicial .....	9
5. Tratamiento de los datos .....	10
5.1 Selección de las variables .....	10
5.2 Relevancia de las variables .....	11
5.3 Rendimientos a escala.....	11
5.4 Datos atípicos .....	12
5.5 Tratamiento para datos atípicos.....	14
6. DEA aplicado.....	15
6.1 DEA en el lenguaje R.....	16
6.1.1 Supereficiencia.....	19
6.1.2 Gráfico de la frontera eficiente.....	21
7. Conclusiones .....	21
8. Anexos .....	22
9. Referencias .....	29

## 1. Introducción

El retraso de los procesos de los jueces tiene consecuencias directas sobre el ingreso y crecimiento de una nación. Estas consecuencias surgen de la realización de menos contratos, lo cual desestimula una mayor división del trabajo terminando en la disminución del rendimiento de la economía. La eficiencia es apenas una dimensión del rendimiento judicial, el cual se ve afectado además por la independencia, la accesibilidad, la responsabilidad y la efectividad (Voigt, 2016). Sin embargo, la dimensión de la eficiencia es bastante compleja, pero puede abordarse como un problema económico.

Según Farrel (1957), la eficiencia económica o total se cumple siempre que sus dos componentes también se cumplan:

- Eficiencia técnica: mejor uso posible para unos recursos dados.
- Eficiencia de asignación: los recursos se usan donde sean de mayor provecho para la sociedad.

Existen otras definiciones de eficiencia, como las coincidentes entre Koopmans y Pareto. Sumado a eso, la eficiencia puede verse afectada por valores tales como la verdad o la justicia, por el trade-off entre calidad y velocidad, y por todas aquellas variables que componen los insumos que le permiten a una corte resolver un determinado número de casos. El problema no es fácil en el sentido de que no haya técnicas que permitan calcular un índice, sino en el sentido de que la complejidad del sistema judicial en un determinado país hace que el problema se vuelva muy específico en ese nivel.

Antes de seguir con la metodología escogida para el cálculo del índice de eficiencia, es necesario hacer recuento de las otras alternativas econométricas. En primer lugar, dado que es posible establecer ecuaciones de oferta y demanda para un juzgado, por ejemplo, el problema puede abordarse mediante la estimación por regresión de ecuaciones simultáneas, algo muy común en las ciencias económicas. Aun así, puede presentarse tal sesgo que ocasione signos contradictorios en los parámetros (Murrel, 2001). En segundo lugar, cuando la muestra se basa en datos panel (mismas unidades observadas a lo largo de varios períodos de tiempo) es posible capturar la heterogeneidad no observada utilizando una variable categórica (Dimitrova, 2015). La anterior es una aproximación válida que ha aparecido recientemente en la literatura. En último lugar, se encuentra el uso de variables instrumentales, el cual también ha aparecido de manera reciente en la literatura (Dimitrova, 2012 & 2015).

## 2. Análisis de envoltura de datos

Antes de proceder a explicar en detalle cómo funciona la metodología del DEA (por sus siglas en inglés), es necesario especificar el porqué de su escogencia, su origen y sus virtudes.

### 2.1 Origen

Entre las medidas que pueden ser calculadas relativas a una tecnología eficiente, representadas por una función frontera, se encuentran:

- Programación matemática (DEA).
- Métodos de econometría (fronteras estocásticas).

William W. Cooper, el padre de las ciencias administrativas, un campo interdisciplinario entre economía, ingeniería, matemática y estadística recibió la inspiración para crear el DEA cuando leyó un artículo de la Royal Statistical Society perteneciente al economista Michael James Farrel. Este artículo aborda el problema de la metodología para medición de la productividad, y aunque, Farrel incluyó en sus ejemplos solo una entrada y una salida, tenía claro la posibilidad de extensiones con múltiples variables como entradas y salidas. Esta metodología no tenía en cuenta la asignación de pesos o ponderaciones a priori, lo cual significaría una ventaja en una medición más imparcial de la eficiencia.

Cooper abordó el problema desde el punto de vista de la teoría económica, optando así por la programación lineal o matemática. De esta manera, Cooper, junto con el profesor Abraham Charnes (matemático) y el becario Edwardo Rhodes (hoy profesor emérito), publicaron un artículo en 1978 titulado: Measuring the efficiency of decision making units. Dicho artículo lidió con el problema planteado por Farrel de las múltiples entradas y salidas, pero además amplió las posibilidades en las aplicaciones del análisis de envoltura de datos, lo que a futuro se vería en sectores como la banca, la salud, los deportes, la educación y el de este presente trabajo, el judicial. El trabajo de Cooper, Charnes y Rhodes se acuñaría como el modelo CCR de análisis de envoltura de datos.

### 2.2 Ventajas

- No requiere información sobre las preferencias, los precios, o sobre prioridades al lidiar con múltiples entradas y salidas. Esto permite que no haya juicios de valor por parte del investigador en cuanto a la importancia relativa de las variables.
- La función de producción es flexible/no paramétrica. En ese sentido no tendría que suponerse por ejemplo una forma para la frontera que siga una función Cobb-Douglas o Leontief (supuestos fuertes).

- Sigue la eficiencia de Farrel, la cual es menos fuerte que la de Koopmans/Pareto.
- Rápida computación.
- Reducción de sesgo en la estimación de la eficiencia.
- Provee verdaderas unidades modelo a seguir o referencia para las unidades ineficientes (Bogetoft & Otto, 2001).
- No requiere una hipótesis de relación funcional entre las entradas y las salidas (Hernández, 2016).
- Múltiples entradas y salidas a diferencia del análisis estocástico de frontera (Rapposelli & Nissi, 2012).

### 2.3 Desventajas

- Al ser una metodología determinista, considera todo ruido o valores atípicos como ineficiencia. Un método estocástico, por el contrario, puede capturar esas observaciones distintas a ineficiencia. Hoy día se han hecho avances en modelos DEA de redes dinámicas con el propósito de capturar dichas observaciones no deseadas.
- El modelo de Chanes, Cooper y Rhodes (CCR) de DEA solo contempló calcular la eficiencia técnica y de escala, pero con los distintos softwares que existen hoy es posible calcular la de asignación, aditiva, supereficiencia y el índice de productividad de Malmquist.
- No explica el comportamiento de las unidades tomadoras de decisiones con respecto a un óptimo teórico. Este problema persiste puesto que, en la rama judicial específicamente, no hay un marco teórico que especifique los valores óptimos del retraso en los juzgados, el número de juzgados, el presupuesto, o el tamaño de los juzgados.
- Al ser una metodología no paramétrica, las pruebas de hipótesis presentan mayor dificultad al momento de quererse aplicar. Sin embargo, el paquete rDEA del lenguaje estadístico R, permite hacer una prueba sobre los rendimientos a escala que sigue la frontera de posibilidades de producción.

### 2.4 ¿Por qué calcular la eficiencia con DEA?

El análisis de envoltura de datos se remonta a la definición de eficiencia más citada al tratar problemas de carácter económico (producción). Además, sobre este problema en específico ha sido también la metodología más utilizada en la literatura internacional para evaluar la eficiencia en la rama judicial. Se ha empleado en países como Brasil, Estados Unidos, Noruega, Italia, Egipto y demás, lo cual refleja que a

pesar de sus limitaciones proporciona resultados satisfactorios. Para cada uno de estos países el problema ha sido tratado de manera diferente, incluyendo diferentes entradas y salidas, diferentes rendimientos a escala (sujetos a factores exógenos como la justicia y la verdad), y diferentes etapas y combinaciones con otras metodologías. Lo anterior sucede dependiendo de la complejidad del sistema judicial en cada país, de los insumos y de los incentivos a los que están expuestos los jueces.

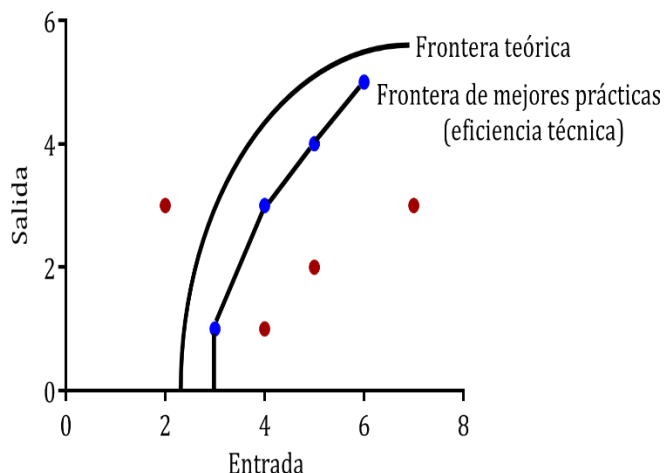
### **3. Descripción de la metodología DEA**

La presente explicación del modelo se basa en el manual del Software DEAP, escrito por el economista Timothy Coelli, afiliado al Centro para la Eficiencia y el Análisis de la Productividad (CEPA). También se basa en el capítulo 2 del libro: Quantitative Models for Performance Evaluation and Benchmarking de Joe Zhu, profesor de análisis de operaciones de la Foisie Business School, Worcester Polytechnic Institute.

Lo que hace el DEA es medir la eficiencia relativa de las unidades tomadoras de decisiones (DMUs) basado en el cálculo de una frontera de producción con variables de entrada y de salida. Gráficamente, es tan simple como observar que las unidades sobre las que pase la frontera se identifican como eficientes, mientras que las que se ubiquen por encima o por debajo de la frontera son ineficientes, ya sea porque gastan más insumos de los que deberían, o porque producen menos resultados. DEA es descriptivo, no analítico; es posible saber las diferencias entre las eficiencias, pero no las razones de dichas diferencias.

La forma de la función de frontera depende de los rendimientos a escala que sigan las entradas y salidas. Por ejemplo, en la rama judicial para que se presentasen rendimientos a escala constantes, equivaldría a decir que, ante un aumento o disminución del número de jueces, conllevará necesariamente a un aumento o disminución en igual proporción al número de casos resueltos por dichos jueces. Esto anterior puede llegar a cumplirse para una determinada unidad tomadora de decisiones, pero es poco probable suponerlo para todas. Lo más probable es que los rendimientos sean crecientes o decrecientes, con algunos casos constantes; en este caso la mejor decisión es considerarlos variables a escala (VRS): los niveles de producción óptimos varían con el tamaño de los DMUs incluidos en la muestra.

## Frontera de posibilidades de producción



Fuente: elaboración propia.

Considérese un conjunto de observaciones para cada  $DMU(n)$ , cada observación  $DMU_j$  ( $j = 1, \dots, n$ ) utiliza  $m$  entradas  $x_{ij}$  ( $i = 1, \dots, m$ ) para producir  $s$  salidas  $y_{rj}$  ( $r = 1, \dots, s$ ). La frontera de mejores prácticas es determinada por las  $n$  observaciones. Esta frontera se estima mediante una aproximación lineal por partes, puesto que la forma empírica de la frontera es muy difícil de estimar.

Existen 2 propiedades fundamentales que aseguran la construcción lineal de la frontera:

### 1. Convexidad en economía

$$\sum_{j=1}^n \lambda_j x_{ij} \quad (i = 1, \dots, m)$$

$$\sum_{j=1}^n \lambda_j y_{rj} \quad (r = 1, \dots, s)$$

donde  $\lambda_j$  ( $j = 1, \dots, n$ ) son escalares no negativos tales que  $\sum_{j=1}^n \lambda_j = 1$ . En un espacio vectorial, un conjunto convexo conformado por subconjuntos convexos no es más que una suma ponderada (Newman, 1987).

### 2. Ineficiencia

El mismo  $y_{rj}$  puede obtenerse con  $\hat{x}_{ij}$ , donde  $\hat{x}_{ij} \geq x_{ij}$ . Las mismas salidas empleando más entradas.

El mismo  $x_{ij}$  puede usarse para obtener  $\hat{y}_{rj}$ , donde  $\hat{y}_{rj} \leq y_{rj}$ . Las mismas entradas para producir menos salidas.

En ese sentido, para  $x_i$  ( $i = 1, \dots, m$ ) y para  $y_i$  ( $r = 1, \dots, s$ ) se tiene que:

$$\sum_{j=1}^n \lambda_j x_{ij} \leq x_i \quad i = 1, \dots, m$$

$$\sum_{j=1}^n \lambda_j y_{rj} \geq y_r \quad r = 1, \dots, s$$

$$\sum_{j=1}^n \lambda_j = 1$$

### 3.1 DEA orientado a las entradas (VRS)

Se minimizan las entradas manteniendo constantes las salidas:

$$\theta^* = \min \theta$$

sujeto a:

$$\sum_{j=1}^n \lambda_j x_{ij} \leq \theta x_{i0} \quad i = 1, \dots, m$$

$$\sum_{j=1}^n \lambda_j y_{rj} \geq y_{r0} \quad r = 1, \dots, s$$

$$\sum_{j=1}^n \lambda_j = 1 \quad (VRS)$$

$$\lambda_j \geq 0 \quad j = 1, \dots, n$$

donde  $DMU_0$  representa a uno de los  $n$   $DMUs$  bajo evaluación.  $x_{i0}$  y  $y_{r0}$  son la  $i$ -ésima entrada y  $r$ -ésima salida para  $DMU_0$ , respectivamente.

Suponiendo que se está resolviendo el problema para algún  $DMU_5$ , ya sea a través de las condiciones de Karush-Kuhn-Tucker o algún software de problemas lineales, se obtiene por ejemplo los siguientes resultados:

$$\theta^* = 0.5 \quad \lambda_2^* = 1 \quad \lambda_j^* = 0 \quad (j \neq 2)$$

Lo anterior se interpreta de la siguiente manera:  $DMU_2$  es el benchmark de  $DMU_5$ ,  $DMU_5$  debería reducir sus insumos a las mismas cantidades empleadas por  $DMU_2$ .

Es posible que existan diferentes soluciones que satisfagan todas las restricciones del problema de minimización, a estas diferentes soluciones se les denomina holguras/slacks, que pueden ser tanto de entrada como de salida.



$$S_i^- = \theta^* x_{i0} - \sum_{j=1}^n \lambda_j x_{ij} \quad i = 1, \dots, m$$

(input slack)

$$S_r^+ = \sum_{j=1}^n \lambda_j y_{rj} - y_{r0} \quad r = 1, \dots, s$$

(output slack)

Dado que puede haber diferentes soluciones con holguras diferentes a cero, se plantea el siguiente modelo de programación lineal.

$$\max \sum_{i=1}^m S_i^- + \sum_{r=1}^s S_r^+$$

sujeto a:

$$\sum_{j=1}^n \lambda_j x_{ij} + S_i^- = \theta^* x_{i0} \quad i = 1, \dots, m$$

$$\sum_{j=1}^n \lambda_j y_{rj} - S_r^+ = y_{r0} \quad r = 1, \dots, s$$

$$\sum_{j=1}^n \lambda_j = 1 \quad (VRS)$$

$$\lambda_j \geq 0 \quad j = 1, \dots, n$$

$DMU_0$  es eficiente sii  $\theta^* = 1$  y  $S_i^- = S_r^+ = 0$  para todo  $i$  y  $r$  (a esto se le denomina eficiencia fuerte o en el sentido de Koopmans/Pareto).  $DMU_0$  es débilmente eficiente si  $\theta^* = 1$  y  $S_i^- \neq 0$  y  $S_r^+ \neq 0$  para algunos  $i$  y  $r$ .

Combinando ambos problemas se conforma un DEA en 2 etapas:

$$\min \theta - \varepsilon \left( \sum_{i=1}^m S_i^- + \sum_{r=1}^s S_r^+ \right)$$

sujeto a:

$$\sum_{j=1}^n \lambda_j x_{ij} + S_i^- = \theta x_{i0} \quad i = 1, \dots, m$$

$$\sum_{j=1}^n \lambda_j y_{rj} - S_r^+ = y_{r0} \quad r = 1, \dots, s$$

$$\sum_{j=1}^n \lambda_j = 1 \quad (VRS)$$

$$\lambda_j \geq 0 \quad j = 1, \dots, n$$

El  $\varepsilon$  no arquimediano permite darle prioridad a minimización (comparación infinitesimal). Así, en una primera etapa se minimiza  $\theta$  para hallar  $\theta^*$ , y posteriormente se optimizan las holguras.

### 3.2 DEA orientado a las salidas (VRS)

$$\max \phi + \varepsilon \left( \sum_{i=1}^m S_i^- + \sum_{r=1}^s S_r^+ \right)$$

sujeto a:

$$\sum_{j=1}^n \lambda_j x_{ij} + S_i^- = x_{i0} \quad i = 1, \dots, m$$

$$\sum_{j=1}^n \lambda_j y_{rj} - S_r^+ = \theta y_{r0} \quad r = 1, \dots, s$$

$$\sum_{j=1}^n \lambda_j = 1 \quad (VRS)$$

$$\lambda_j \geq 0 \quad j = 1, \dots, n$$

Esta es la orientación que será considerada para el caso de la Rama Judicial. El objetivo será entonces maximizar los egresos manteniendo constantes los insumos del modelo. Lo contrario no se contempla por el momento pues llevaría a decisiones polémicas o muy difíciles de cumplir. Además, la orientación hacia las salidas es la más común que se encuentra en la literatura para este problema. A este modelo en específico se le conoce como BCC por Banker, Charnes y Cooper (1984).

## 4. Entradas y salidas para calcular la eficiencia judicial

El número de casos resueltos se toma normalmente como output de las cortes. Si los jueces son calificados por el número de casos resueltos, tenderán a atender los menos complejos primero.

Como entradas debería definitivamente incluirse el número de jueces, número de empleados, habitaciones, computadores... Todo ello a nivel de juzgado. Un estudio de

Gillespie (1976) hace la pregunta si estas variables deberían ser fijas o sustituibles entre sí (como una función de producción Leontief). Esto depende del sistema judicial, puede suceder que, por ejemplo, además del juez haya personas capacitadas para escuchar un caso (haciéndolos sustitutos). Lo anterior es importante ya que la eficiencia relativa de la metodología DEA es sensible al número de entradas que se incluyan, puesto que estas hacen parte de la construcción de la frontera. El tiempo de retraso podría ser una variable de entrada importante porque este representa el verdadero rendimiento de un juzgado. Según Dimitrova (2012), la tasa de congestión es una variable de entrada más apropiada que los casos entrantes, puesto que los jueces no tienen control sobre la demanda. Esta carga de congestión se calcula como la proporción de carga de casos por juez sobre número de casos (siendo la carga de casos = resueltos + nuevos + pendientes).

## 5. Tratamiento de los datos

Previo a aplicar el modelo, es necesario saber tres cosas fundamentales: qué otras variables pueden llegar a tenerse en cuenta, los rendimientos a escala que siguen los datos, y si hay presencia de datos atípicos, puesto que la presencia de 'outliers' puede afectar la forma de la frontera.

### 5.1 Selección de las variables

Variables adicionales se consideraron de acuerdo con su pertinencia y a la posibilidad de su obtención.

#### 1. Entradas adicionales

- Carga de trabajo: Esta variable es la combinación entre la demanda de trabajo y el inventario de los jueces. No se considera conveniente agregar la demanda de manera individual como una entrada del modelo puesto que los jueces no la controlan. Una alternativa sería considerarla como una variable no discrecional (no es controlable, pero afecta la eficiencia de todas formas). Sin embargo, estando dentro de la carga de trabajo junto con el inventario, se resuelve dicho problema.
- Número de personal administrativo.
- Número de computadores.
- Gastos del juzgado o presupuesto.

La disponibilidad de estas 3 últimas está sujeta al envío de la información.

#### 2. Salidas adicionales

- Sentencias.
- Casos resueltos por tipo de proceso.

La disponibilidad de estas también está sujeta al envío de la información.

## 5.2 Relevancia de las variables

Si se dispone de más variables podría reducirse la dimensionalidad de los datos con un análisis de componentes principales, por el momento no se realiza disponiendo solo de 3. Además, se sabe que las variables seleccionadas por el momento son muy importantes de acuerdo con la literatura consultada.

## 5.3 Rendimientos a escala

El paquete rDEA del lenguaje R contiene una función 'rts.test' que realiza una prueba de hipótesis sobre los rendimientos a escala. Dicha prueba de hipótesis está basada en un método Bootstrap propuesto por el matemático Léopold Simar y el economista Paul Wilson. Este test, inicialmente pensado para estimar intervalos de confianza para las diferentes unidades, se ha adaptado para estimar los rendimientos a escala.

```
> rts.test(inputs1,outputs1,model="output",H0="constant",bw="cv", B=100,alpha = 0.05)
[1] 0.28
```

Como el p-valor es mayor a un valor de 0.05, es posible afirmar que con un 95% de confianza, se rechaza la hipótesis nula de que los rendimientos son constantes. Se concluye que los datos siguen rendimientos variables: crecientes, decrecientes y constantes en distintos vecindarios.

Con el software DEAP es posible ver el rendimiento específico de cada circuito para el caso de ejemplo:

**Tabla de rendimientos**

Circuito	Rendimientos
Bogotá	-
Leticia	irs
Mocoa	irs
Pamplona	drs
Pasto	drs
Zipaquirá	irs
Medellín	irs
Tunja	irs
Villavicencio	drs
San Andrés	drs
Popayán	-
Cartagena	irs
Santa Rosa de Viterbo	drs
Bucaramanga	irs
Florencia	drs

Cali	irs
Barrancabermeja	-
Valledupar	drs
Sincelejo	-
Buga	irs
Santa Marta	-
San Gil	irs
Barranquilla	-
Turbo	irs
Pereira	drs
Arauca	irs
Montería	irs
Armenia	irs
Ibagué	irs
Buenaventura	irs
Riohacha	irs
Quibdó	irs
Cúcuta	irs
Neiva	irs
Yopal	drs
Facatativá	-
Manizales	irs
Girardot	irs

*Fuente: elaboración propia*

Donde irs y drs representan rendimientos crecientes y decrecientes a escala, respectivamente. Los rendimientos crecientes quieren decir que, ante un aumento en el número de insumos, los egresos aumentarán en mayor proporción a dicho aumento. Los decrecientes significan que, ante un aumento en el número de insumos, los egresos aumentarán en menor proporción (aumentar el número de jueces u otros insumos no necesariamente conlleva a un mayor número de egresos deseados). Por último, los rendimientos constantes, los cuales ya fueron explicados, están representados por un guion.

## 5.4 Datos atípicos

Los datos atípicos en estadística alteran la media y sesgan los resultados en general. Para este modelo, el cual es de optimización, los datos atípicos de extremo superior son los que afectan la construcción de la frontera. En otras palabras, aquellas unidades por lo general eficientes cuyas variables no sigan la tendencia general de los datos, se consideran atípicas. Por otro lado, dichas variables podrían tener valores muy altos comparadas a las demás y estar cumpliendo con un papel de apalancamiento

(predictores extremos). Los datos atípicos de extremo inferior no se tienen en cuenta puesto que no contribuyen en la construcción de la frontera, simplemente son unidades muy ineficientes.

Entre las múltiples formas que existen para su detección en modelos DEA, se ha escogido el método de Nube de Datos, consultado en Bogetoft y Otto (2011):

Sea  $X = (x^1, \dots, x^k)$  y  $Y = (y^1, \dots, y^k)$  matrices  $k \times m$  y  $k \times n$  de entradas y salidas respectivamente. La matriz combinada  $[XY]$  contiene entonces a todas las observaciones. Las diferentes filas pueden verse como una nube de puntos en el espacio  $\mathbb{R}_+^m \times \mathbb{R}_+^n$ , donde cada punto representa a un  $DMU$ . El volumen de esta nube es proporcional al determinante de la matriz combinada  $[XY]'[XY]$ .

$$\text{Volumen de datos} \approx D(X, Y)$$

Si se retira al  $DMU$  de la primera fila, siendo este un valor extremo superior, el volumen de la nube puede disminuir en gran medida. Si este es el caso, entonces es posible concluir que dicho  $DMU$  es un outlier. El proceso puede repetirse para más de un  $DMU$  a la vez.

Sea  $D^{(i)}$  el determinante después de quitar al  $DMU_i$  y  $R^i$  la relación entre el nuevo volumen de datos y el viejo.  $R^i = \frac{D^{(i)}}{D}$ ; donde  $R^i$  es adimensional (no depende de las unidades de  $X$  ni  $Y$ ). Si el  $DMU$  no es un outlier,  $R^i$  tendrá un valor cercano a 1; caso contrario en donde sí lo es, se observará un valor cercano a cero. Para nuestro caso de ejemplo de circuitos judiciales, al eliminar el dato extremo 'Bogotá' el resultado es el siguiente:

```
> x1 <- with(juzgcirc, cbind(juzgcirc$Jueces))
> y1 <- with(juzgcirc, cbind(juzgcirc$Carga))
> w1 <- with(juzgcirc, cbind(juzgcirc$Egresos))
> x1y1w1 <- cbind(x1,y1,w1)
> D1 <- det(t(x1y1w1)%*%x1y1w1)
> i1 <- c(1) #número de DMUs a quitarse, sobre los que se sospeche sean atípicos
> x1y1w1i1 = x1y1w1[-i1,]
> D1i <- det( t(x1y1w1i1) %*% x1y1w1i1 )
> R1i <- D1i/D1
> R1i
[1] 0.9452631
```

Se concluye que Bogotá no es un dato atípico y sigue la tendencia general de los datos. Si se quiere hallar valores pequeños de  $R$  para múltiples  $DMUs$  que se sospechen sean atípicos, es conveniente utilizar el siguiente método:

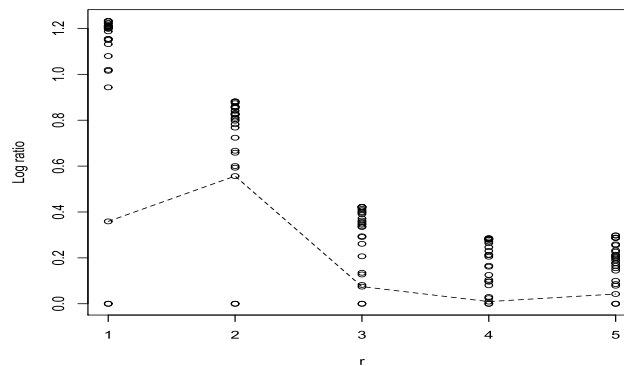
Si existe un grupo de  $s$  outliers y se buscan borrando grupos de  $1, \dots, r$   $DMUs$ , entonces para  $r < s$  no se espera encontrar un  $R$  pequeño puesto que todavía hay outliers en los datos restantes. En cambio, para  $s < r$  se tiene un  $R$  en donde todos los

outliers se borran, y este  $R$  sería presuntamente muy pequeño. Al examinar los valores de  $R$ , se busca el primer valor aislado que sea pequeño, si este existe, se encontró a un grupo de outliers.

Un valor aislado pequeño es un valor aislado mínimo. Fijándolo en una escala  $\frac{R_{min}}{R_{min}} = 1$  debería estar aislado de otros valores de  $\frac{R^{(r)}}{R_{min}} = 1$ ; o cero debería estar aislado de otros valores de  $\log\left(\frac{R^{(r)}}{R_{min}}\right)$ . Así, al graficar el par ordenado  $(r, (\log\frac{R^{(r)}}{R_{min}}))$  aquel  $r$  con puntos bajos da un indicio sobre el número de outliers.

```
> atipico1 <- outlier.ap(inputs1,outputs1,NDEL=5)
> outlier.ap.plot(atipico1$ratio)
```

Par ordenado  $(r, (\log\frac{R^{(r)}}{R_{min}}))$



*Fuente: elaboración propia.*

En el caso de ejemplo se han eliminado 5 *DMUs* del extremo superior, se observa que la distancia punteada entre los casos de eliminar 2 *DMUs* y 0 es mayor que para los otros casos. Con la prueba de Nube de Datos se pudo comprobar que dichos *DMUs* actúan de palanca, más no de outliers.

## 5.5 Tratamiento para datos atípicos

Para el caso de ejemplo y los encontrados en los anexos, no se detectaron outliers, sin embargo, en caso de encontrarse lo más conveniente es retirarlos del análisis. La presencia de un outlier puede afectar en gran medida los resultados de los puntajes de eficiencia técnica.

Ante su detección, se prosigue a sacarlos de la muestra y calcular las verdaderas eficiencias estimadas de los datos restantes. Un tratamiento formal si se desea calcular la eficiencia de los datos atípicos puede consultarse en Gholam (2017).

## 6. DEA aplicado

A continuación, se muestra el modelo implementado en el lenguaje de programación R. Adicionalmente, se utilizaron distintos softwares para comprobar que la metodología se está ejecutando de manera correcta; entre esos se utilizó el programa escrito en Fortran llamado DEAP de autoría de Timothy Coelli.

El modelo se realizará para el caso de juzgados entre un número determinado de circuitos, contemplando un tipo de proceso denominado ‘reparación directa’ y a todas las especialidades para el año 2016. Otros ejemplos se incluirán como anexos.

### Base de datos

CIRCUITO	Ingresos	Egresos	Inventario	Jueces	Carga
Arauca	360	70	716	2	1076
Armenia	240	81	430	6	670
Barrancabermeja	68	33	102	2	170
Barranquilla	617	270	1218	15	1835
Bogotá	2926	1468	7632	65	10558
Bucaramanga	434	225	553	15	987
Buenaventura	107	33	221	3	328
Buga	352	122	592	3	944
Cali	978	389	2462	21	3440
Cartagena	554	410	1699	15	2253
Cúcuta	317	102	1055	10	1372
Facatativá	47	8	82	3	129
Florencia	600	187	1151	4	1751
Girardot	43	11	144	3	187
Ibagué	534	173	1251	12	1785
Leticia	4	4	10	1	14
Manizales	259	55	627	8	886
Medellín	1247	705	3409	36	4656
Mocoa	493	275	407	2	900
Montería	309	121	672	7	981
Neiva	406	110	1079	9	1485
Pamplona	30	24	75	1	105
Pasto	485	776	1056	9	1541
Pereira	238	114	559	7	797
Popayán	790	397	2281	10	3071
Quibdó	371	64	593	4	964
Riohacha	148	43	443	3	591
San Andrés	34	13	101	1	135
San Gil	162	51	139	3	301
Santa Rosa de	323	129	210	7	533



Viterbo					
Santa Marta	267	181	774	8	1041
Sincelejo	243	153	549	9	792
Tunja	180	159	327	15	507
Turbo	102	38	219	2	321
Valledupar	631	272	1235	8	1866
Villavicencio	427	378	894	9	1321
Yopal	124	22	319	2	443
Zipaquirá	28	23	32	3	60

*Fuente: datos de la Rama Judicial.*

De esta manera las variables de entrada son *Jueces* y *Carga*, mientras que la de salida es *Egresos*. El dato sobre el número de jueces para Santa Rosa de Viterbo no estaba disponible, por lo que se decidió promediar el número de jueces de aquellos circuitos con un número similar de ingresos y egresos para estimar dicho dato faltante (se promedió específicamente a Buga, Cúcuta y Montería).

## 6.1 DEA en el lenguaje R

En el lenguaje R se han desarrollado varios paquetes para análisis de envoltura de datos; para este trabajo se utilizó el siguiente:

### Benchmarking

Este paquete contiene las funciones que calculan el DEA con sus diferentes variantes, grafican la frontera de producción o las isocuantas dado el caso, y, por último, permite calcular también la super eficiencia.

Los resultados de la eficiencia técnica para el ejemplo fueron los siguientes:

```
> dea1 <- dea(inputs1,outputs1,RTS="vrs",ORIENTATION = "out",SLACK = TRUE, DUAL = TRUE)
```

**Tabla de eficiencias**

Ranking	Circuito	Eficiencia técnica
1	Bogotá	1
1	Leticia	1
1	Mocoa	1
1	Pamplona	1
1	Pasto	1
2	Zipaquirá	0,843849
3	Medellín	0,694542
4	Tunja	0,627852
5	Villavicencio	0,568613

6	San Andrés	0,541667
7	Popayán	0,503579
8	Cartagena	0,493594
9	Santa Rosa de Viterbo	0,484254
10	Bucaramanga	0,453706
11	Florencia	0,447216
12	Cali	0,422029
13	Barrancabermeja	0,398222
14	Valledupar	0,386129
15	Sincelejo	0,38507
16	Buga	0,364582
17	Santa Marta	0,345937
18	San Gil	0,342058
19	Barranquilla	0,338107
20	Turbo	0,293094
21	Pereira	0,285101
22	Arauca	0,254545
23	Montería	0,245494
24	Armenia	0,241322
25	Ibagué	0,217685
26	Buenaventura	0,202768
27	Riohacha	0,174787
28	Quibdó	0,164849
29	Cúcuta	0,147706
30	Neiva	0,14712
31	Yopal	0,137262
32	Facatativá	0,128741
33	Manizales	0,123636
34	Girardot	0,120267

*Fuente: elaboración propia.*

Sabiendo esto, las recomendaciones que deberían seguirse son las siguientes:

>  $\lambda$

#### Tabla de unidades modelo a seguir y recomendaciones

Circuito	Bogotá	Leticia	Mocoa	Pamplona	Pasto	Eficiencia	Aumento recomendado en porcentaje
Bogotá	1	0	0	0	0	1	0
Leticia	0	1	0	0	0	1	0

Mocoa	0	0	1	0	0	1	0
Pamplona	0	0	0	1	0	1	0
Pasto	0	0	0	0	1	1	0
Zipaquirá	0	0,969876	0	0	0,030124	0,843849	15,61509
Medellín	0,345459	0	0	0	0,654541	0,694542	30,5458
Tunja	0	0,677145	0	0	0,322855	0,627852	37,21477
Villavicencio	0	0,144073	0	0	0,855927	0,568613	43,13869
San Andrés	0	0	0	1	0	0,541667	45,83333
Popayán	0,017857	0	0	0	0,982143	0,503579	49,64211
Cartagena	0,078962	0	0	0	0,921038	0,493594	50,64057
Santa Rosa de Viterbo	0	0,660118	0	0	0,339882	0,484254	51,57458
Bucaramanga	0	0,362803	0	0	0,637197	0,453706	54,62943
Florencia	0	0	0,714286	0	0,285714	0,447216	55,27844
Cali	0,210602	0	0	0	0,789398	0,422029	57,79706
Barrancabermeja	0	0,897839	0	0	0,102161	0,398222	60,17781
Valledupar	0	0	0,142857	0	0,857143	0,386129	61,38714
Sincelejo	0	0,490504	0	0	0,509496	0,38507	61,49303
Buga	0	0,059881	0,788707	0	0,151412	0,364582	63,54176
Santa Marta	0	0,327439	0	0	0,672561	0,345937	65,40631
San Gil	0	0,81205	0	0	0,18795	0,342058	65,79421
Barranquilla	0,032605	0	0	0	0,967395	0,338107	66,18926
Turbo	0	0,728826	0,167056	0	0,104118	0,293094	70,69062
Pereira	0	0,48723	0	0	0,51277	0,285101	71,48992
Arauca	0	0	1	0	0	0,254545	74,54545
Montería	0	0,366732	0	0	0,633268	0,245494	75,45055
Armenia	0	0,570399	0	0	0,429601	0,241322	75,86783
Ibagué	0,02706	0	0	0	0,97294	0,217685	78,23148
Buenaventura	0	0,794368	0	0	0,205632	0,202768	79,72324
Riohacha	0	0,504226	0,280885	0	0,214889	0,174787	82,52135
Quibdó	0	0,149973	0,542888	0	0,307139	0,164849	83,51509
Cúcuta	0	0,110675	0	0	0,889325	0,147706	85,22936
Neiva	0	0,036673	0	0	0,963327	0,14712	85,28799
Yopal	0	0,575256	0,342564	0	0,082179	0,137262	86,2738
Facatativá	0	0,924689	0	0	0,075311	0,128741	87,12587
Manizales	0	0,428946	0	0	0,571054	0,123636	87,63639
Girardot	0	0,886706	0	0	0,113294	0,120267	87,97328

*Fuente: elaboración propia.*

Las unidades eficientes, las cuales son Bogotá, Leticia, Mocoa, Pamplona y Pasto no tienen recomendación por el momento al ser las unidades de referencia o modelo a seguir para las demás. Zipaquirá, por ejemplo, tiene una eficiencia técnica o relativa de 0.84 y su unidad de referencia fue en mayor medida Leticia (el escalar no negativo 'lambda' de Leticia para su problema de optimización fue de 0.969 sobre 1, mientras que el de Pasto fue de 0.03). Por lo que la recomendación que debería seguir Zipaquirá es aumentar sus egresos en aproximadamente 15.6% manteniendo constantes su número de jueces y carga de trabajo para ser tan eficiente como Leticia. Nótese que, entre las recomendaciones, no se aconseja en ningún caso ser tan eficiente como Bogotá.

Cabe mencionar que, si se estuviese evaluando la orientación hacia las entradas, la recomendación sería diferente: habría que disminuir los insumos en aproximadamente el porcentaje  $\frac{1-eficiencia}{eficiencia} * 100$ .

El paquete permite calcular las holguras, es decir, saber quién es eficiente en el sentido fuerte o débil, y también calcula los pesos de las variables para hacer un pequeño análisis de sensibilidad. Sin embargo, se excluye lo primero por practicidad y lo segundo ante carencia de un mayor número de variables de salida.

De todas formas, suponiendo que en el momento se disponen de más variables de salida, entre mayor sea el peso de estas variables, mayor será su contribución a la eficiencia, por lo que la recomendación irá enfocada en aumentar esa variable en específico. Los pesos de las variables de entrada no tiene sentido interpretarlos si la orientación del modelo es hacia las salidas, como es el del ejemplo, pero no hay que confundir interpretación con importancia.

### 6.1.1 Supereficiencia

La supereficiencia elimina la restricción de convexidad en donde los escalares no negativos tienen que sumar uno, de esta manera es posible encontrar eficiencias mayores a uno. Esto se emplea con dos propósitos: en primer lugar, es otra alternativa para detectar datos atípicos si las supereficiencias tienen valores extremadamente altos. En segundo lugar, dado que no se hace distinción entre las unidades eficientes, es posible hacer un verdadero ranking de eficiencia (y de esta manera se evita que las unidades eficientes no tengan ningún incentivo para mejorar). El ranking de eficiencias sigue el siguiente orden:

```
> superdea1 <- sdea(inputs1,outputs1,RTS = "vrs",ORIENTATION = "out")
```

**Tabla del ranking de circuitos basado en supereficiencias**

Ranking	Circuito	Supereficiencia
1	Leticia	No solución/Hipereficiente
2	Mocoa	2,330508

3	Pamplona	2,228703
4	Pasto	2,000149
5	Bogotá	1,891753
6	Zipaquirá	0,843849
7	Medellín	0,694542
8	Tunja	0,627852
9	Villavicencio	0,568613
10	San Andrés	0,541667
11	Popayán	0,503579
12	Cartagena	0,493594
13	Santa Rosa de Viterbo	0,484254
14	Bucaramanga	0,453706
15	Florencia	0,447216
16	Cali	0,422029
17	Barrancabermeja	0,398222
18	Valledupar	0,386129
19	Sincelejo	0,38507
20	Buga	0,364582
21	Santa Marta	0,345937
22	San Gil	0,342058
23	Barranquilla	0,338107
24	Turbo	0,293094
25	Pereira	0,285101
26	Arauca	0,254545
27	Montería	0,245494
28	Armenia	0,241322
29	Ibagué	0,217685
30	Buenaventura	0,202768
31	Riohacha	0,174787
32	Quibdó	0,164849
33	Cúcuta	0,147706
34	Neiva	0,14712
35	Yopal	0,137262
36	Facatativá	0,128741
37	Manizales	0,123636
38	Girardot	0,120267

*Fuente: elaboración propia.*

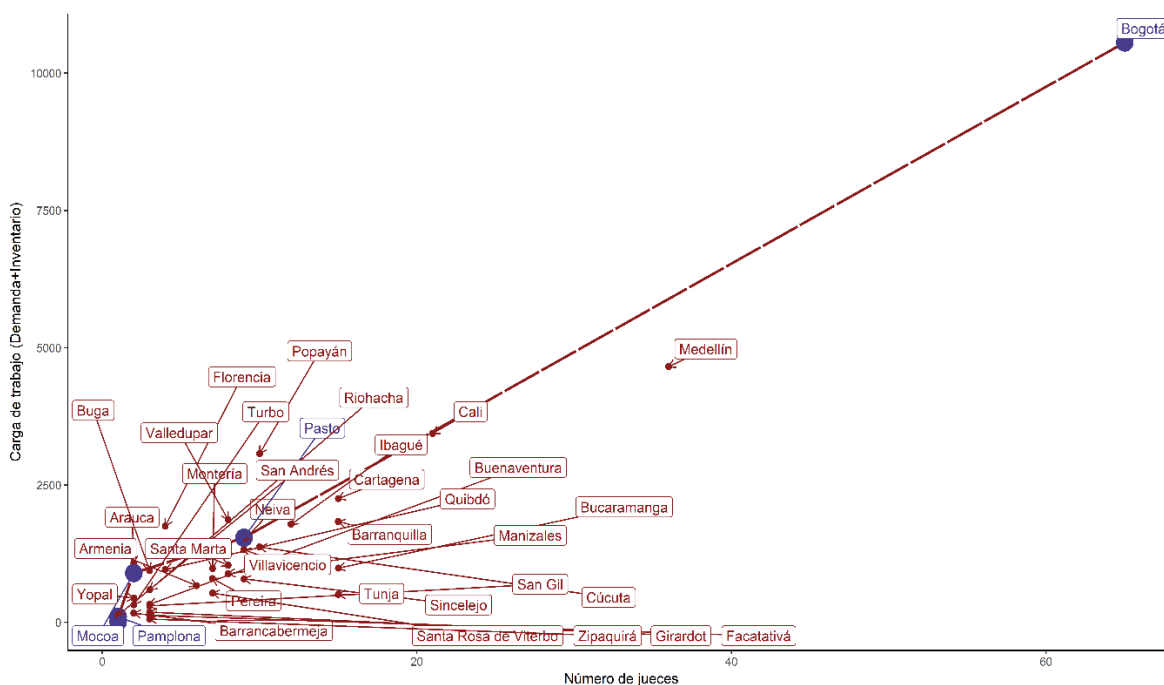
Leticia aparece en primer lugar puesto que no se encontró circuito de referencia para su problema de optimización, pero se identifica como el más eficiente entre los

eficientes (a esto se le denomina hipereficiencia). Mientras que a Bogotá se le podría catalogar como el peor entre los mejores.

### 6.1.2 Gráfico de la frontera eficiente

Finalmente, se muestra el gráfico de la frontera de posibilidades de producción en donde con las variables ‘carga de trabajo’ y ‘número de jueces’ se produce la variable ‘casos resueltos’ (al ser un gráfico de dos dimensiones esta última no aparece visible). Se construyó utilizando la función ‘dea.plot.frontier’ del paquete ‘Benchmarking’ para saber la forma de la frontera, y se personalizó con el paquete ‘ggplot’.

#### Frontera de producción para proceso de reparación directa entre circuitos



*Fuente: elaboración propia.*

## 7. Conclusiones

La metodología empleada para el cálculo de la eficiencia ha proporcionado resultados satisfactorios. Asimismo, las pruebas estadísticas sobre los rendimientos y datos atípicos han permitido que dichos resultados reflejen una aproximación a los verdaderos índices de eficiencia. La metodología puede mejorarse incluyendo más variables tanto de entrada como de salida, para incluso contemplar la orientación hacia las entradas. El modelo planteado en este trabajo es uno de los más básicos y

más utilizados en la literatura, y se tuvieron todas las consideraciones necesarias para evitar errores comunes.

También, como en trabajos recientes y experimentales, a futuro puede combinarse con metodologías de Minería de Datos e Inteligencia Artificial, una vez exista una base teórica sólida para emplear dichos métodos. En este trabajo se emplearon test de estadística no paramétrica adicionales al modelo básico para su validación, consultados en el libro guía de Joe Zhu y el artículo de Simar y Wilson.

Por último, cabe mencionar que el caso de Bogotá en el ejemplo explicado no necesariamente quiera decir que sea eficiente. Puede que por ser un dato extremo esté ayudando a la construcción de la frontera, y es bastante posible que su eficiencia se vea afectada con la inclusión de más variables. Adicionalmente, la eficiencia de Bogotá puede haberse dado ante la falta de unidades de comparación. La eficiencia también puede verse afectada por factores exógenos que no se puedan percibir con los datos, y que deban tomarse en cuenta antes de hacer conclusiones definitivas. Los modelos son solo aproximaciones de la realidad, y, por ende, deben seguirse solo como recomendación.

## 8. Anexos

Código de programación para el caso de ejemplo y para los casos de primera y segunda instancia.

```
#####
## ANÁLISIS DE ENVOLTURA DE DATOS ##
#####

library(readxl)

#####
# JUZGADOS proceso reparación directa entre CIRCUITOS #
#####
juzgcirc <- read_excel("C:/Users/CamiloAndrés/Desktop/DNP/Proyectos/Distribución de la oferta
judicial/DEA en R/Repo/Data-Envelopment-Analysis/Bases_de_Rama_Judicial-DEA.xlsx",sheet = 1)

inputs1 <- data.frame(c(juzgcirc[5],juzgcirc[6])) #vector de entradas
inputs1 <- as.matrix(inputs1) #matriz necesaria para la función dea de Benchmarking
outputs1 <- data.frame(juzgcirc[3]) #vector de salidas
outputs1 <- as.matrix(outputs1)
library(rDEA) #paquete con test de rendimientos a escala
# test de rendimientos a escala
rendimientos1 <- rts.test(inputs1,outputs1,model="output",H0="constant",bw="cv", B=100,alpha =
0.05)
rendimientos1$pvalue #si es mayor a 0.05, los rendimientos son variables
detach(package:rDEA)

# detección de DATOS ATÍPICOS con nube de datos
```

```

x1 <- with(juzgcirc, cbind(juzgcirc$Jueces))
y1 <- with(juzgcirc, cbind(juzgcirc$Carga))
w1 <- with(juzgcirc, cbind(juzgcirc$Egresos))
x1y1w1 <- cbind(x1,y1,w1)
D1 <- det(t(x1y1w1)%*%x1y1w1)
i1 <- c(1) #número de DMUs a quitarse, que se sospeche que sean atípicos: En este caso, Bogotá
x1y1w1i1 = x1y1w1[-i1,]
D1i <- det( t(x1y1w1i1) %*% x1y1w1i1 )
R1i <- D1i/D1
R1i #valores pequeños de Ri me indican presencia de datos atípicos
library(Benchmarking) #paquete con funciones dea
# GRÁFICAMENTE, entre más DMUs se quiten, más tarda la gráfica en generarse
atipico1 <- outlier.ap(inputs1,outputs1,NDEL=5) #proceso lento
# para que se sospeche de un dato atípico, la distancia entre el punto ubicado en cero y la línea punteada
tiene que ser muy grande
outlier.ap.plot(atipico1$ratio)

dea1 <- dea(inputs1,outputs1,RTS="vrs",ORIENTATION = "out",SLACK = TRUE, DUAL = TRUE)
#rendimientos variables, orientado a salidas y 2 etapas
# EFICIENCIA técnica
eftcircj <- 1/dea1$eff
eftcircj
eft_circj <- data.frame(Circuito=juzgcirc$CIRCUITO,Eficiencia_técnica=eftcircj)
eft_circj_ordenada <- eft_circj[order(-eft_circj$Eficiencia_técnica),] # ordenar por eficiencia
eft_circj_ordenada
# unidades de REFERENCIA
unidades1 <- data.frame(dea1$lambda)
# RECOMENDACIÓN
recomen1 <- data.frame(Circuito=juzgcirc$CIRCUITO,Bogotá=unidades1$L5,
  Leticia=unidades1$L16,
  Mocoa=unidades1$L19,
  Pamplona=unidades1$L22,
  Pasto=unidades1$L23,
  Eficiencia_técnica=eftcircj,
  "Aumento_recomendado_en_porcentaje"=(1-eftcircj)*100)
recomen1_ordenada <- recomen1[order(-eft_circj$Eficiencia_técnica),]
recomen1_ordenada
# HOLGURAS (eficiencia fuerte o débil)
dea1$sx
dea1$sy
# PESOS (sensibilidad/impacto sobre la eficiencia)
dea1$ux
dea1$vy

# SUPEREFICIENCIA (Para detectar valores atípico/Diferenciar unidades frontera)
superdea1 <- sdea(inputs1,outputs1,RTS = "vrs",ORIENTATION = "out")
1/superdea1$eff
superref1 <- data.frame(Circuito=juzgcirc$CIRCUITO,Super_eficiencia=1/superdea1$eff)
superref1

superdea1$lambda

```



```
#####
# JUZGADOS proceso reparación directa entre DISTRITOS #
#####
juzgdist <- read_excel("C:/Users/CamiloAndrés/Desktop/DNP/Proyectos/Distribución de la oferta
judicial/DEA en R/Repo/Data-Envelopment-Analysis/Bases_de_Rama_Judicial-DEA.xlsx",sheet = 4)

inputs2 <- data.frame(c(juzgdist[5],juzgdist[6])) #vector de entradas
inputs2 <- as.matrix(inputs2) #matriz necesaria para la función dea de Benchmarking
outputs2 <- data.frame(juzgdist[3]) #vector de salidas
outputs2 <- as.matrix(outputs2)

detach(package:Benchmarking)
library(rDEA) #paquete con test de rendimientos a escala
# test de rendimientos a escala
rendimientos2 <- rts.test(inputs2,outputs2,model="output",H0="constant",bw="cv", B=100,alpha =
0.05)
rendimientos2$pvalue #si es mayor a 0.05, los rendimientos son variables
detach(package:rDEA)
library(Benchmarking)

# detección de DATOS ATÍPICOS con nube de datos
x2 <- with(juzgdist, cbind(juzgdist$Jueces))
y2 <- with(juzgdist, cbind(juzgdist$Carga))
w2 <- with(juzgdist, cbind(juzgdist$Egresos))
x2y2w2 <- cbind(x2,y2,w2)
D2 <- det(t(x2y2w2)%*%x2y2w2)
i2 <- c(1) #número de DMUs a quitarse, que se sospeche que sean atípicos
x2y2w2i2 = x2y2w2[-i2,]
D2i <- det( t(x2y2w2i2) %*% x2y2w2i2 )
R2i <- D2i/D2
R2i #valores pequeños de Ri me indican presencia de datos atípicos

# GRÁFICAMENTE, entre más DMUs se quiten, más tarda la gráfica en generarse
atipico2 <- outlier.ap(inputs2,outputs2,NDEL=5)
# para que se sospeche de un atípico, la distancia entre el punto ubicado en cero y la línea punteada tiene
que ser muy grande
outlier.ap.plot(atipico2$ratio)

dea2 <- dea(inputs2,outputs2,RTS="vrs",ORIENTATION = "out",SLACK = TRUE, DUAL = TRUE)
#rendimientos variables, orientado a salidas y 2 etapas

# EFICIENCIA técnica
eftdistj <- 1/dea2$eff
eftdistj
eft_distj <- data.frame(Circuito=juzgdist$DISTRITO,Eficiencia_técnica=eftdistj)
eft_distj
eft_distj_ordenada <- eft_distj[order(-eft_distj$Eficiencia_técnica),] #ordenar por eficiencia
eft_distj_ordenada
# unidades de REFERENCIA
dea2$lambda
# HOLGURAS (eficiencia fuerte o débil)
dea2$sx
```

```

dea2$sy

# SUPEREFICIENCIA (para detectar valores atípico/diferenciar unidades frontera)
superdea2 <- sdea(inputs2,outputs2,RTS = "vrs",ORIENTATION = "out")
1/superdea2$eff
superef2 <- data.frame(Distrito=juzgdist$DISTRITO,Super_eficiencia=1/superdea2$eff)
superef2

superdea2$lambda

#####
# TRIBUNALES proceso reparación directa entre DISTRITOS #
#####
tribdist <- read_excel("C:/Users/CamiloAndrés/Desktop/DNP/Proyectos/Distribución de la oferta
judicial/DEA en R/Repo/Data-Envelopment-Analysis/Bases_de_Rama_Judicial-DEA.xlsx",sheet = 6)

inputs3 <- data.frame(c(tribdist[5],tribdist[6])) #vector de entradas
inputs3 <- as.matrix(inputs3) #matriz necesaria para la función dea de Benchmarking
outputs3 <- data.frame(tribdist[3]) #vector de salidas
outputs3 <- as.matrix(outputs3)

detach(package:Benchmarking)
library(rDEA) #paquete con test de rendimientos a escala
# test de rendimientos a escala
rendimientos3 <- rts.test(inputs3,outputs3,model="output",H0="constant",bw="cv", B=100,alpha =
0.05)
rendimientos3$pvalue #si es mayor a 0.05, los rendimientos son variables
detach(package:rDEA)
library(Benchmarking)

# detección de DATOS ATÍPICOS con nube de datos
x3 <- with(tribdist, cbind(tribdist$Jueces))
y3 <- with(tribdist, cbind(tribdist$Carga))
w3 <- with(tribdist, cbind(tribdist$Egresos))
x3y3w3 <- cbind(x3,y3,w3)
D3 <- det(t(x3y3w3)%*%x3y3w3)
i3 <- c(1) # DMUs a quitarse, que se sospeche que sean atípicos
x3y3w3i3 = x3y3w3[-i3,]
D3i <- det( t(x3y3w3i3) %*% x3y3w3i3 )
R3i <- D3i/D3
R3i # Valores pequeños de Ri me indican presencia de datos atípicos

# GRÁFICAMENTE, entre más DMUs se quiten, más tarda la gráfica en generarse
atipico3 <- outlier.ap(inputs3,outputs3,NDEL=5)
# para que se sospeche de un dato atípico, la distancia entre el punto ubicado en cero y la línea punteada
tiene que ser muy grande
outlier.ap.plot(atipico3$ratio)

dea3 <- dea(inputs3,outputs3,RTS="vrs",ORIENTATION = "out",SLACK = TRUE, DUAL = TRUE)
#rendimientos variables, orientado a salidas y 2 etapas

```

```

# EFICIENCIA técnica
eftdistt <- 1/dea3$eff
eftdistt
eft_distt <- data.frame(Distrito=tribdist$DISTRITO,Eficiencia_técnica=eftdistt)
eft_distt
eft_distt_ordenada <- eft_distt[order(-eft_distt$Eficiencia_técnica),] #ordenar por eficiencia
eft_distt_ordenada
# unidades de REFERENCIA
unidades3 <- data.frame(dea3$lambda)
# RECOMENDACIÓN
recomen3 <- data.frame(Distrito=tribdist$DISTRITO,Arauca=unidades3$L2,
  Cundinamarca=unidades3$L13,
  San_Andrés=unidades3$L22,
  Pamplona=unidades3$L22,
  Eficiencia_técnica=eftdistt,
  "Aumento_recomendado_en_porcentaje"=(1-eftdistt)*100)
recomen3
# holguras (evidencia eficiencia fuerte o débil)
dea3$sx
dea3$sy
# SUPEREFICIENCIA (para detectar valores atípico/diferenciar unidades frontera)
superdea3 <- sdea(inputs3,outputs3,RTS = "vrs",ORIENTATION = "out")
1/superdea3$eff
superef3 <- data.frame(Distrito=tribdist$DISTRITO,Super_eficiencia=1/superdea3$eff)
superef3

superdea3$lambda

# COMPARACIÓN entre primera y segunda instancia
comp <-
data.frame(Distrito=tribdist$DISTRITO,eficiencia_juzgados=eftdistj,eficiencia_tribunales=eftdistt)
comp

#####
## EXPORTAR eficiencias ##
#####
library(xlsx)
write.xlsx(eft_circj_ordenada,"C:/Users/CamiloAndrés/Desktop/basejc_ordenada.xlsx")
write.xlsx(eft_distj_ordenada,"C:/Users/CamiloAndrés/Desktop/basejd_ordenada.xlsx")
write.xlsx(eft_distt_ordenada,"C:/Users/CamiloAndrés/Desktop/basetd_ordenada.xlsx")

#####
####
## GRÁFICOS DE FRONTERAS DE PRODUCCIÓN (2 inputs para producir 1 output) ##
#####
####
library(ggplot2)
library(ggrepel)

#####
# Gráfico para JUZGADOS entre CIRCUITOS #
#####

```

```
##### Gráfico con función del paquete para saber forma de la frontera
#inputs1 <- as.matrix(inputs1)
#outputs1 <- as.matrix(outputs1)
#plotef <- dea.plot.frontier(inputs1,outputs1,RTS = "vrs")
#####

tablaf1 <- data.frame(juzgcirc,eficiencia=dea1$eff)
myColors <- c("firebrick4", "slateblue4") #colores de eficiencia e ineficiencia
tablalabelf1 <- data.frame(juzgcirc,eficiencia=dea1$eff)
tablalabelf1$eficiencia[-which(tablalabelf1$eficiencia==1)]=0
tablalabelf1$color=rep("firebrick4",length(tablalabelf1$CIRCUITO))
tablalabelf1$color[which(tablalabelf1$eficiencia==1)]=rep("slateblue4", 5) #número de unidades
eficientes
tablalabelf1$eficiencia = as.factor(tablalabelf1$eficiencia)

tablaff1 <- tablaf1[which(tablaf1$eficiencia==1),]
tablaff1

fronteraJC <- ggplot(data=tablalabelf1,
  aes(x=Jueces,
    y=Carga,
    label=CIRCUITO)
  )+
  geom_line(data=tablaff1,
    aes(x=Jueces,
      y=Carga),
    color="firebrick4",
    cex=1,
    linetype="F1"
  )+
  geom_point(aes(color=eficiencia,
    size=eficiencia)
  )+
  scale_color_manual(values=myColors
  )+
  theme(legend.position = "none",
    rect=element_rect(fill = "transparent"),
    plot.title = element_text(hjust = 0.5),
    # para fondo blanco
    panel.grid=element_blank(),
    panel.background = element_blank(),
    axis.line = element_line(colour = "black")
  )+

  ggtitle("Eficiencia técnica proceso reparación directa entre circuitos (juzgados)"
  )+
  labs(x="Número de jueces",
    y="Carga de trabajo (Demanda+Inventario)"
  )+
  geom_label_repel(aes(label=CIRCUITO,
    color=eficiencia),
```

```

        force=8,
        arrow = arrow(length = unit(0.5, 'picas'))
    )
fronteraJC

# Guardar .png en directorio actual
ggsave("frontera1.png",fronteraJC,dpi = 700, bg= "white")

#####
# Gráfico para TRIBUNALES entre DISTRITOS #
#####

##### Gráfico con función del paquete para saber forma de la frontera
#inputs3 <- as.matrix(inputs3)
#outputs3 <- as.matrix(outputs3)
#plotef3 <- dea.plot.frontier(inputs3,outputs3,RTS = "vrs")
#####

tablaf3 <- data.frame(tribdist,eficiencia=dea3$eff)
tablalabelf3 <- data.frame(tribdist,eficiencia=dea3$eff)
tablalabelf3$eficiencia[-which(tablalabelf3$eficiencia==1)]=0
tablalabelf3$color=rep("firebrick4",length(tablalabelf3$DISTRITO))
tablalabelf3$color[which(tablalabelf3$eficiencia==1)]=rep("slateblue4",3) #número de unidades
eficientes
tablalabelf3$eficiencia = as.factor(tablalabelf3$eficiencia)

tablaff3 <- tablaf3[which(tablaf3$eficiencia==1),]

fronteraTD <- ggplot(data=tablalabelf3,
  aes(x=Jueces,
    y=Carga,
    label=DISTRITO)
  )+
geom_line(data=tablaff3,
  aes(x=Jueces,
    y=Carga),
  color="firebrick4",
  cex=1,
  linetype="F1"
  )+
geom_point(aes(color=eficiencia,
  size=eficiencia)
  )+
scale_color_manual(values=myColors
  )+
theme(legend.position = "none",
  rect=element_rect(fill = "transparent"),
  plot.title = element_text(hjust = 0.5),
  # para fondo blanco
  panel.grid=element_blank()

```

```

panel.background = element_blank(),
axis.line = element_line(colour = "black")
)+
ggtitle("Eficiencia técnica proceso reparación directa entre distritos para tribunales"
)+
labs(x="Número de jueces",
y="Carga de trabajo (Demanda+Inventario)"
)+
geom_label_repel(aes(label=DISTRITO,
color=eficiencia),
force=8,
arrow = arrow(length = unit(0.5, 'picas'))
)
fronteraTD

# Guardar .png en directorio actual
ggsave("frontera2.png",fronteraTD,dpi = 700, bg= "white")

```

## 9. Referencias

Banker R., Charnes A. & Cooper W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 30(9), 1078-1092.

Bogetoft, P. & Otto, L. (2011). *Benchmarking with DEA, SFA, and R*. USA: Stanford University.

Bogetoft, P. & Otto, L. (2015). Benchmark and Frontier Analysis Using DEA and SFA. R Foundation for Statistical Computing, Vienna, Austria. Disponible en: <https://cran.r-project.org/web/packages/Benchmarking/index.html>

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2,429-444.

Coelli, T. (1996). A Guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program. *Centre for Efficiency and Productivity Analysis (CEPA) Working Paper No. 8*.

Dimitrova-Grajzl, V., et al. (2012). Court output, judicial staffing, and the demand for court services: Evidence from Slovenian courts of first instance. *International Review of Law and Economics*, 32(1), 19-29.

Dimitrova-Grajzl, V., et al. (2015). Courts in a transition economy: Case disposition and the quantity-quality tradeoff in Bulgaria. *Economic Systems*, 40(1), 19-38.

- Farrel, M. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society*, 120(3), 253-290.
- Gholam, A. (2017). Impact of Outliers in Data Envelopment Analysis. *Int. J. Industrial Mathematics*, 9(4), 14 p.
- Gillespie, R. (1976). The production of court services: Analysis of scale effects and other factors. *The Journal of Legal Studies*, 5(2), 243-265.
- Hernández, V. (2016). Análisis de la eficiencia relativa en el sistema judicial colombiano como funciones de producción, mediante análisis envolvente de datos (DEA). Tesis de maestría: Universidad Distrital Francisco José de Caldas.
- Murrel, P. (2001). Demand and Supply in Romanian Commercial Courts: Generating Information for Institutional Reform. *IRIS Center*: Universidad de Maryland.
- Newman, P. (1987). *Convexity* en Eatwell, J & Milgrate, M. (2016). *The New Palgrave: A Dictionary of Economics* (Primera ed.). UK: Palgrave Macmillan.
- Rapposelli, A. & Nissi, E. (2012). Analyzing Industrial Accidents in European Countries Using Data Envelopment Analysis. *AIEL Series in Labour Economics*, 6,93-101.
- Simar, L. & Wilson, P. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research*, 139, 115-132.
- Simm, J. & Besstremyannaya, G. (2016). rDEA: Robust Data Envelopment Analysis (DEA) for R. R Foundation for Statistical Computing, Vienna, Austria. Disponible en: <https://cran.r-project.org/web/packages/rDEA/index.html>
- Voigt, S. (2016). Determinants of Judicial Efficiency: a survey. *Eur J Law Econ*, 42, 183-208.
- Zhu, J. (2009). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets* (Segunda ed.). USA: Springer.