

UNIVERSIDAD MIGUEL HERNÁNDEZ
FACULTAD DE CIENCIAS SOCIALES Y JURÍDICAS DE
ELCHE

GRADO EN ESTADÍSTICA EMPRESARIAL



TRABAJO DE FIN DE GRADO

**Análisis de la Eficiencia y la Productividad medida en los
Institutos de Secundaria de las Comunidades Autónomas**

Curso Académico 2023-2024

Alumna: Beatriz Perpiñán Yáñez

Tutora: Lidia Ortiz Hernarejos

RESUMEN

En este Trabajo Fin de Grado se aplica una técnica de análisis de eficiencia y productividad a los institutos de secundaria de las comunidades autónomas españolas. Esta técnica, denominada Análisis Envolvente de Datos, es una metodología no paramétrica cuyo principal objetivo es el análisis de fronteras de producción y la medición de la eficiencia productiva. El análisis de eficiencia y productividad se realiza con los datos de PISA 2018 sobre las diferentes comunidades autónomas de España.

Para el análisis, se ha empleado el modelo CCR (Charnes, Cooper y Rhodes) orientado a los outputs. Las variables de los inputs incluyen el número de profesores por estudiante, el nivel socioeconómico del estudiante y la infraestructura escolar, mientras que los outputs son las puntuaciones promedio en lectura, matemáticas y ciencias.

El modelo fue implementado utilizando el lenguaje de programación R y el paquete 'lpSolveAPI'. Los resultados muestran diferencias significativas en la eficiencia y productividad de los institutos de secundaria entre las comunidades autónomas. Se identificaron las regiones más eficientes y aquellas con mayor potencial de mejora, proporcionando recomendaciones valiosas para la optimización de los recursos educativos.

Las conclusiones subrayan la necesidad de una evaluación continua de la eficiencia y productividad en los institutos de secundaria, sugiriendo áreas específicas de mejora para maximizar el rendimiento educativo con los recursos disponibles.

Palabras clave: Eficiencia educativa, Productividad, Análisis Envolvente de Datos (DEA), PISA 2018, Institutos de secundaria, Centros educativos, Comunidades Autónomas, España, R, Frontera de referencia.

AGRADECIMIENTOS

En primer lugar, a mis padres, cuyo apoyo incondicional y amor han sido fundamentales a lo largo de toda mi experiencia académica. Gracias por creer en mí y por estar siempre a mi lado, tanto en los momentos de éxito como en los de dificultad.

A mi prometido, gracias por tu paciencia infinita, por tu apoyo constante y por ser mi fuente de motivación y alegría. No podría haber deseado un compañero mejor en este viaje.

A mi hermano y la preciosa familia que ha formado, gracias por celebrar conmigo cada logro y por consolarme en cada fracaso. Vuestra presencia en mi vida es un gran regalo.

Finalmente, quiero agradecer a todos aquellos que, de una manera u otra, han contribuido a mi formación. En especial a mi tutora, Lidia, cuyo conocimiento y constante apoyo han sido esenciales para la realización de este trabajo. Gracias por creer en mí y por guiarme en este proceso con tanta dedicación y profesionalismo.

Con todo mi cariño,

Beatriz Perpiñán Yáñez

ÍNDICE

1. INTRODUCCIÓN	4
1.1. CONTEXTO Y JUSTIFICACIÓN DEL ESTUDIO	4
1.2. OBJETIVOS DEL ESTUDIO	4
1.3. ESTRUCTURA	5
2. ESTADO DE LA CUESTIÓN Y MARCO TEÓRICO	6
2.1. LAS EVALUACIONES PISA	6
2.1.1. CARACTERÍSTICAS PRINCIPALES DE PISA	7
2.1.2. OBJETIVOS DE PISA	8
2.1.3. IMPACTO DE PISA	8
2.2. EL ANÁLISIS ENVOLVENTE DE DATOS	9
2.2.1. EL DEA Y SUS DISTINTAS APLICACIONES	9
2.2.2. EL DEA Y SU APLICACIÓN EN EL ÁMBITO EDUCATIVO	11
3. METODOLOGÍA	14
3.1. METODOLOGÍA DEA	14
3.1.1. PRINCIPIOS BÁSICOS DEL DEA	14
3.1.2. DESCRIPCIÓN DEL MODELO CCR	14
3.1.3. ÍNDICE DE MALMQUIST	16
3.1.4. IMPLEMENTACIÓN DEL MODELO CCR	19
3.2. DATOS Y VARIABLES	20
3.2.1. CONJUNTO DE DATOS	20
3.2.2. PREPROCESAMIENTO DE LA BASE DE DATOS	21
3.2.3. DESCRIPCIÓN DE LOS DATOS	25
4. ANÁLISIS DE RESULTADOS	27
4.1. APLICACIÓN DEL DEA A LOS DATOS DE PISA 2018	27
4.1.1. IMPLEMENTACIÓN DE LA METODOLOGÍA DEA EN R	27
4.1.2. IMPLEMENTACIÓN DE LA METODOLOGÍA DEA CON UNA FRONTERA DE REFERENCIA EN R	30
4.2. COMPARACIÓN DE LA EFICIENCIA DE LAS COMUNIDADES AUTÓNOMAS	36
4.3. IDENTIFICACIÓN DE MEJORES PRÁCTICAS E INTERPRETACIÓN DE LOS RESULTADOS	37
5. CONCLUSIONES Y TRABAJO FUTURO	38
5.1. CONCLUSIONES	38
5.2. RECOMENDACIONES PARA MEJORAR LA EFICIENCIA EDUCATIVA EN ESPAÑA	39
5.3. PROPUESTAS PARA FUTURAS INVESTIGACIONES	40
6. REFERENCIAS	42
7. ANEXOS	45

1. INTRODUCCIÓN

1.1. CONTEXTO Y JUSTIFICACIÓN DEL ESTUDIO

En un mundo globalizado donde el conocimiento y la educación son fundamentales para el desarrollo socioeconómico, evaluar y mejorar la eficiencia de los sistemas educativos se ha convertido en una prioridad para los gobiernos y las instituciones educativas. En este contexto, el Análisis Envolvente de Datos (Data Envelopment Analysis en inglés, DEA) se presenta como una herramienta poderosa para medir la eficiencia relativa de unidades de decisión que utilizan múltiples insumos para producir múltiples productos (Charnes et al., 1978).

Este Trabajo de Fin de Grado se centra en aplicar la metodología DEA a los datos del Programa para la Evaluación Internacional de Alumnos (PISA, 2018), con el objetivo de comparar la eficiencia de las diferentes comunidades autónomas españolas en el ámbito educativo.

1.2. OBJETIVOS DEL ESTUDIO

El objetivo principal de este trabajo es evaluar la eficiencia de los centros educativos en las comunidades autónomas españolas utilizando la metodología DEA aplicada a los datos de PISA 2018.

Específicamente, este estudio busca evaluar la eficiencia relativa de las comunidades autónomas, así como realizar una comparativa de productividad entre ellas, identificando las mejores prácticas y proporcionando recomendaciones específicas para aquellas comunidades autónomas que se encuentran por debajo de la frontera de eficiencia, con el objetivo de mejorar sus resultados educativos.

Se espera que este análisis permita identificar las comunidades autónomas que utilizan sus recursos educativos de manera más eficiente y aquellas que tienen un mayor potencial de mejora. Al comparar los resultados de las comunidades autónomas, se podrán extraer conclusiones valiosas y formular recomendaciones específicas para optimizar el uso de recursos educativos en España.

1.3. ESTRUCTURA

Este trabajo se estructurará de la siguiente manera: Primero, se realizará una revisión de la literatura más relevante y su aplicación en el ámbito educativo. A continuación, se describirá la metodología DEA y las variables seleccionadas para el análisis. Luego, se implementará en R el modelo DEA CCR y algunas variantes para aplicarlo a los datos de PISA 2018, seguido de una discusión sobre las implicaciones de estos resultados. Finalmente, se ofrecerán conclusiones y recomendaciones para futuras investigaciones y políticas educativas.



2. ESTADO DE LA CUESTIÓN Y MARCO TEÓRICO

El análisis de la eficiencia del sistema educativo es crucial para mejorar la calidad y equidad de la educación en cualquier país.

En España, la diversidad regional en términos de recursos y resultados educativos hace que sea particularmente importante evaluar la eficiencia de los centros educativos en diferentes comunidades autónomas. Las evaluaciones PISA proporcionan datos comparativos sobre el rendimiento de estudiantes en materias básicas, ofreciendo una base sólida para este tipo de análisis.

Este programa contribuye a evaluar de forma sistemática lo que los jóvenes saben y son capaces de hacer al finalizar su Educación Secundaria Obligatoria (ESO) en más de 80 países del mundo. España ha participado, desde su primera edición en 2000 (PISA, 2024).

2.1. LAS EVALUACIONES PISA

A finales de la década de 1990, la Organización para la Cooperación y el Desarrollo Económico (OCDE) reconoció la necesidad de contar con una herramienta que permitiera evaluar de manera comparativa los sistemas educativos de sus países miembros y otros países asociados. Hasta ese momento, no existía una evaluación internacional estandarizada que midiera de manera integral las habilidades de los estudiantes cercanos al final de su educación obligatoria. La organización buscaba proporcionar datos relevantes para ayudar a los países a mejorar sus políticas educativas y prácticas escolares.

El Programa para la Evaluación Internacional de Alumnos (PISA, por sus siglas en inglés) es una iniciativa global coordinada por la OCDE que evalúa el rendimiento académico de estudiantes de secundaria en materias como lectura, matemáticas y ciencias. Se llevó a cabo por primera vez en el año 2000 y desde entonces se realiza cada tres años (PISA, 2024).

A continuación, desde el apartado 2.1.1. hasta el 2.1.3. se destacan las principales características y objetivos de PISA cuyos datos han sido obtenidos a través de la página web de la OECD (PISA, 2024).

2.1.1. CARACTERÍSTICAS PRINCIPALES DE PISA

Las características más relevantes de las evaluaciones PISA son las siguientes:

1. **Periodicidad:** la evaluación se realiza cada tres años desde el año 2000. Cada ciclo de evaluación se centra principalmente en una de las tres áreas principales (lectura, matemáticas o ciencias), aunque las tres áreas se evalúan en cada ciclo.
2. **Edad de los Participantes:** la evaluación está dirigida a estudiantes de 15 años, independientemente del grado escolar en el que se encuentren. Esta edad se elige porque es el final de la educación obligatoria en muchos países.
3. **Ámbitos de Evaluación:**
 - a. **Lectura:** Evalúa la capacidad de los estudiantes para comprender, usar y reflexionar sobre textos escritos.
 - b. **Matemáticas:** Mide la capacidad de los estudiantes para formular, emplear e interpretar las matemáticas en una variedad de contextos.
 - c. **Ciencias:** Examina la habilidad de los estudiantes para explicar fenómenos científicamente, diseñar e interpretar investigaciones científicas, y comprender las características de la ciencia como una forma de conocimiento.
4. **Enfoque en la Aplicabilidad:** PISA se centra en la evaluación de competencias prácticas y la capacidad de los estudiantes para aplicar sus conocimientos y habilidades a problemas del mundo real, en lugar de simplemente medir la memorización de contenidos curriculares.
5. **Cuestionarios de Contexto:** además de las pruebas cognitivas, se recopila información a través de cuestionarios que recogen datos sobre el entorno educativo, las condiciones socioeconómicas, las actitudes hacia el aprendizaje, y las prácticas pedagógicas, proporcionando un contexto más amplio para interpretar los resultados académicos.

2.1.2. OBJETIVOS DE PISA

El objetivo principal de PISA es proporcionar datos comparativos sobre la calidad, equidad y eficiencia de los sistemas educativos de diferentes países, ofreciendo una buena perspectiva a nivel regional e internacional sobre el rendimiento educativo y los factores que lo afectan.

Estas evaluaciones permiten a los países medir el rendimiento de sus estudiantes y compararlo con el de otros países, identificando puntos fuertes y áreas de mejora en sus sistemas educativos. Es por ello, que los resultados sirven como base para el desarrollo y la implementación de políticas educativas basadas en evidencias, ayudando a los formuladores de políticas a entender qué prácticas y políticas educativas son más efectivas.

Más aún, PISA busca identificar desigualdades en el acceso y la calidad de la educación, promoviendo políticas que aseguren que todos los estudiantes, independientemente de su origen socioeconómico, tengan la oportunidad de desarrollar sus capacidades al máximo.

2.1.3. IMPACTO DE PISA

Desde su inicio, PISA ha tenido un impacto significativo en el debate sobre la educación a nivel mundial. Sus resultados son ampliamente discutidos en medios de comunicación, entre educadores, investigadores y responsables de políticas educativas. Los informes de PISA han llevado a muchos países a reevaluar sus sistemas educativos y adoptar reformas para mejorar la calidad y equidad de la educación.

Al proporcionar datos comparativos sobre el rendimiento de los estudiantes y los factores que lo influyen, PISA ayuda a los países a identificar mejores prácticas y políticas educativas efectivas, con el objetivo final de preparar mejor a los estudiantes para los desafíos del futuro.

En España, los resultados de PISA han sido utilizados para comparar el rendimiento educativo entre las distintas comunidades autónomas. Esto es especialmente relevante dado el sistema educativo descentralizado del país, donde cada comunidad autónoma tiene competencias en materia de educación. Los datos de PISA permiten identificar diferencias regionales en el rendimiento educativo y analizar los factores que contribuyen a estas variaciones.

2.2. EL ANÁLISIS ENVOLVENTE DE DATOS

El concepto de eficiencia se remonta a los trabajos de economistas como Farrell (1957), quien propuso una medida de eficiencia basada en la relación entre inputs y outputs. Este enfoque inicial sentó las bases para el desarrollo de métodos más sofisticados para evaluar la eficiencia.

La metodología DEA fue formalmente introducida en 1978 por Abraham Charnes, William W. Cooper y Edward Rhodes en su artículo seminal "Measuring the Efficiency of Decision Making Units" publicado en *European Journal of Operational Research* (Charnes et al., 1978). Este trabajo marcó el nacimiento del DEA como una herramienta de análisis de eficiencia.

El DEA es una técnica no paramétrica utilizada en la investigación operativa y la economía para medir la eficiencia relativa de unidades de decisión, como empresas, hospitales, escuelas, entre otros, que utilizan múltiples insumos para producir múltiples productos o servicios.

2.2.1. EL DEA Y SUS DISTINTAS APLICACIONES

El Análisis Envolvente de Datos ha sido ampliamente utilizado en diversos sectores para evaluar la eficiencia de distintas unidades de decisión.

A continuación se describen algunos estudios generales y sus aplicaciones en campos como la Salud, la banca, el transporte y la industria, demostrando la versatilidad y utilidad del DEA:

- ❖ *"The Measurement of Efficiency and Productivity of Health Care Delivery"*: Se utiliza el DEA para la evaluación de la eficiencia de hospitales y centros de salud. Los inputs incluyen personal médico, camas hospitalarias y equipos médicos, mientras que los outputs son los servicios prestados y los resultados de salud de los pacientes. El estudio concluye que el DEA es una herramienta efectiva para identificar ineficiencias y mejorar la gestión hospitalaria (Hollingsworth, 2008).
- ❖ *"A Comparative Study of Efficiency in European Banking"*: Se aplica el DEA para comparar la eficiencia de los bancos en varios países europeos. Se analizan inputs como los activos y el número de empleados, y outputs como los ingresos por intereses y comisiones. Los resultados revelan diferencias significativas en la eficiencia bancaria entre los países y sugieren áreas de mejora en la gestión de recursos y operaciones bancarias (Casu & Molyneux, 2003).
- ❖ *"Comparing the Efficiency of Public Transportation Subunits Using Data Envelopment Analysis"*: utiliza DEA para evaluar la eficiencia de diferentes subunidades del transporte público en una ciudad estadounidense. Los inputs incluyen el número de vehículos y empleados, y los outputs son el número de pasajeros transportados y los kilómetros recorridos. El análisis identifica las subunidades más eficientes y proporciona recomendaciones para mejorar el rendimiento del sistema de transporte público (Barnum et al., 2007).
- ❖ *"A Survey of DEA Applications"*: Se realiza una revisión exhaustiva de las aplicaciones del DEA en la industria manufacturera. Se analizan estudios que evalúan la eficiencia de plantas de producción, líneas de ensamblaje y empresas manufactureras. Los inputs incluyen materias primas, mano de obra y energía, mientras que los outputs son productos terminados y calidad del producto. El estudio demuestra cómo el DEA ayuda a identificar ineficiencias y mejorar procesos industriales (Liu et al., 2013).

Estos estudios ilustran cómo el Análisis Envolvente de Datos se aplica a múltiples sectores, proporcionando una herramienta valiosa para evaluar y mejorar la eficiencia operativa en diversos contextos. Desde la educación y la salud hasta el sector agroalimentario y la banca, el DEA permite identificar ineficiencias y sugerir recomendaciones prácticas para optimizar el uso de recursos.

2.2.2. EL DEA Y SU APLICACIÓN EN EL ÁMBITO EDUCATIVO

Centrándonos en el contexto educativo, este método implica comparar los recursos educativos (como los recursos del centro, la ratio de alumnos por profesor, y el nivel socioeconómico de los estudiantes) con los resultados académicos (puntuaciones en lectura, matemáticas y ciencias). El DEA construye una frontera de eficiencia a partir de las mejores prácticas observadas, permitiendo identificar las comunidades autónomas que son más eficientes y aquellas que tienen margen para mejorar.

Existen varios estudios que han aplicado diferentes métodos, incluyendo el Análisis Envolvente de Datos, para evaluar la eficiencia de los sistemas educativos en España. A continuación, se presentan algunos de los estudios más relevantes en esta área:

- ❖ *“Non-Parametric Approaches to Education and Health Efficiency in OECD Countries”*: Este estudio examina la eficiencia en los sectores de educación y salud en una muestra de países de la OCDE mediante dos metodologías no paramétricas: FDH y DEA. Se destaca la importancia del gasto público en estos sectores y se identifican implicaciones significativas para la eficiencia del sector público. Los resultados muestran que algunos países, como Japón, Corea y Suecia, destacan en eficiencia en ambos sectores, aunque por diferentes razones (Afonso & St. Aubyn, 2005).
- ❖ *“Technical Efficiency, Managerial Efficiency and Objective Setting in the Educational System: An International Comparison”*: Este estudio evalúa la eficiencia técnica y gerencial de los centros de formación profesional en varios países, incluyendo España, mediante DEA. Los inputs analizados son los recursos financieros y humanos, mientras que los outputs son las tasas de empleo de los graduados y las tasas de graduación (Giménez et al., 2007).
- ❖ *“Costs and Efficiency of Higher Education Institutions in England: A DEA Analysis”*: Este estudio analiza el aumento significativo en el número de estudiantes en el sector de educación superior del Reino Unido en los últimos 15 años. Utilizando el Análisis Envolvente de Datos, se evalúa la estructura de costos y el rendimiento de varias instituciones de educación superior en Inglaterra. Los resultados del DEA sugieren que se pueden lograr ahorros

sustanciales de costos y aumentos en la producción si se destinan a incrementar el número de estudiantes en cada IES, maximizando así la eficiencia operativa y de escala. Además, se utiliza el índice de Malmquist para evaluar el cambio de productividad en los IES del Reino Unido, revelando que la productividad ha disminuido en la mayoría de las IES durante el período de estudio (Thanassoulis et al., 2011).

- ❖ “*¿Ha mejorado la productividad docente e investigadora de las universidades públicas españolas desde la aprobación de la LOU?: Evidencia a partir del bootstrap*”: Este estudio analiza la productividad de las universidades públicas españolas desde la implementación de la LOU (Ley Orgánica de Universidades 6/2001), utilizando el Índice de Productividad de Malmquist y la técnica de remuestreo bootstrap. Los resultados muestran una mejora del 13,4% en la productividad universitaria global, con un aumento significativo en la productividad investigadora en comparación con la docente. El progreso tecnológico también se identifica como un factor clave en estos cambios. Los resultados de este estudio proporcionan información crucial para futuras decisiones de mejoras de la productividad en investigación y enseñanza (Fernandez-Santos & Martínez-Campillo, 2015).
- ❖ “*Comparing school ownership performance using a pseudo-panel database: A Malmquist-type index approach*”: se centra en una nueva forma de utilizar el índice de Malmquist para analizar las diferencias de rendimiento entre dos grupos de unidades de decisión (DMUs) a lo largo del tiempo, utilizando una base de datos pseudo-panel. Los autores extienden el índice de un periodo de Camanho y Dyson (2006), creando un nuevo índice de Malmquist para pseudo-panel (PPMI). El nuevo índice permite evaluar comparativamente el rendimiento de grupos bajo diferentes condiciones y épocas, y cómo esto puede ser utilizado para identificar prácticas efectivas y áreas de mejora en el contexto educativo (Aparicio et al., 2017).
- ❖ “*A note on measuring group performance over time with pseudo-panels*”: se propone una extensión del CDMI para adaptarlo a bases de datos pseudo-panel, donde las mismas unidades no se observan en todos los periodos, lo cual es común en estudios de educación con datos como PISA. El PPMI se descompone para evaluar cambios en la eficiencia técnica y cambios tecnológicos entre los

grupos comparados. También se introduce un componente de divergencia (DC) que indica cuánto se desvía el PPMI de su interpretación directa en términos de cambio de productividad agregado (Aparicio & Santin, 2018).

Estos estudios ofrecen una gran variedad de enfoques y metodologías para evaluar la eficiencia del sistema educativo en España. A través del uso del DEA y otras técnicas de análisis, estos trabajos proporcionan información valiosa sobre cómo se utilizan los recursos educativos y qué factores contribuyen a una mayor eficiencia en diferentes contextos.



3. METODOLOGÍA

A continuación, se describirá la metodología utilizada en el estudio y las variables seleccionadas para el análisis.

3.1. METODOLOGÍA DEA

El Análisis Envolvente de Datos (DEA) es una técnica no paramétrica utilizada para evaluar la eficiencia de diferentes unidades de decisión (DMUs, por sus siglas en inglés) que convierten múltiples insumos (inputs) en múltiples productos (outputs).

3.1.1. PRINCIPIOS BÁSICOS DEL DEA

Según (Charnes et al., 1978) se establecen los siguientes principios básicos del DEA:

- ❖ Enfoque no paramétrico: A diferencia de otros métodos de análisis de eficiencia que requieren supuestos específicos sobre la forma funcional de la producción, el DEA no impone ninguna forma predeterminada a la relación entre insumos y productos.
- ❖ Frontera de eficiencia: El DEA construye una frontera de eficiencia a partir de las observaciones de las unidades de decisión. Esta frontera representa las mejores prácticas dentro del grupo analizado. Las unidades que se encuentran en esta frontera se consideran eficientes, mientras que las que se encuentran por debajo de la frontera se consideran ineficientes.
- ❖ Unidades de decisión (DMUs): Las unidades que se están evaluando se conocen como unidades de decisión (Decision Making Units, DMUs). Cada DMU utiliza un conjunto de insumos para producir un conjunto de productos.

3.1.2. DESCRIPCIÓN DEL MODELO CCR

Dentro del DEA, uno de los modelos más utilizados es el modelo CCR, desarrollado por Charnes, Cooper y Rhodes en 1978. Este modelo, asume que una proporción constante de inputs producirá una proporción constante de outputs, sin importar el tamaño de la unidad de decisión.

Este modelo se puede formular tanto de manera orientada a los inputs como a los outputs, dependiendo de si el objetivo es minimizar los insumos para un nivel dado de productos o maximizar los productos para un nivel dado de insumos.

En nuestro caso, nos centraremos en maximizar los outputs manteniendo los niveles de inputs constantes. La eficiencia técnica (ϕ_0) de una DMU₀ (centro de secundaria que se está evaluando) se define como el factor por el cual se pueden aumentar los outputs sin cambiar los niveles de inputs.

La formulación matemática del modelo orientado a los outputs es la siguiente:

$$\begin{aligned}
 & \text{Max } \phi_0 \\
 \text{s. a. } & \sum_{j=1}^n \lambda_j x_{ij} \leq x_{i0}, \quad i = 1, \dots, m \\
 & \sum_{j=1}^n \lambda_j y_{rj} \geq \phi_0 y_{r0}, \quad r = 1, \dots, s \\
 & \lambda_j \geq 0, \quad j = 1, \dots, n \quad (1)
 \end{aligned}$$

donde:

- ϕ es la eficiencia técnica de la DMU o (valor mayor o igual a 1);
- x_{ij} es la cantidad de insumo i usado por la DMU j .
- y_{rj} es la cantidad de output r producido por la DMU j .
- λ_j es la ponderación asignada a la DMU j .
- n es el número de DMUs.
- m es el número de insumos.
- s es el número de outputs.

Además, se define

$$D(X_j^A, Y_j^A) = \frac{1}{\phi_j^A}, j = 1, \dots, n_A \quad (2)$$

como la distancia de la DMU $_j$ (centro de secundaria j) de la comunidad autónoma A a su frontera. Siendo n_A el número de colegios de la comunidad autónoma A .

3.1.3. ÍNDICE DE MALMQUIST

El Índice de Malmquist fue propuesto por Caves, Christensen y Diewert en 1982 (Caves et al., 1982) y es una medida de productividad que se utiliza para evaluar cambios en la eficiencia y la productividad de unidades de toma de decisiones (DMUs) a lo largo del tiempo. Se basa en el concepto de fronteras de producción y compara la eficiencia de una DMU en dos períodos diferentes utilizando un enfoque no paramétrico.

Se ha convertido en una herramienta popular en la evaluación de la productividad debido a su capacidad para descomponer los cambios en la productividad en componentes de cambio tecnológico y cambio en la eficiencia técnica.

A continuación, se muestra una recopilación de información obtenida de algunos artículos que detallan en profundidad los siguientes conceptos a tratar (Aparicio et al., 2017; Aparicio & Santin, 2018).

El índice de Malmquist emplea programas de Análisis Envolvente de Datos para calcular las distancias necesarias para su cálculo. Según Färe, Grosskopf y Lovell (1994), el índice de productividad de Malmquist orientado a los outputs para dos períodos bajo una tecnología de rendimientos constantes se formula de la siguiente manera:

$$M(x^{t+1}, y^{t+1}, x^t, y^t) = \left[\left(\frac{D^t(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)} \right) \cdot \left(\frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^{t+1}(x^t, y^t)} \right) \right]^{1/2} \quad (3)$$

Si el índice es mayor que uno, indica que se ha producido una mejora en la productividad del periodo t al periodo t+1, mientras que un valor menor que uno indica una disminución en productividad.

Así mismo, también demostraron que este índice puede descomponerse en cambio de eficiencia (EC) y cambio técnico (TC):

$$\begin{aligned}
 & M(x^{t+1}, y^{t+1}, x^t, y^t) \\
 &= \underbrace{\frac{D^{t+1}(x^{t+1}, y^{t+1})}{D^t(x^t, y^t)}}_{EC(x^{t+1}, y^{t+1}, x^t, y^t)} \cdot \underbrace{\left[\left(\frac{D^t(x^{t+1}, y^{t+1})}{D^{t+1}(x^{t+1}, y^{t+1})} \right) \cdot \left(\frac{D^t(x^t, y^t)}{D^{t+1}(x^t, y^t)} \right) \right]^{1/2}}_{TC(x^{t+1}, y^{t+1}, x^t, y^t)} \quad (4)
 \end{aligned}$$

Donde:

- ❖ EC: representa los cambios en la eficiencia técnica, que captura las mejoras ($EC > 1$) o mermas ($EC < 1$) en la eficiencia en el período $t+1$ con respecto al período t si $EC > 1$ ($EC < 1$). Si $EC = 1$, significa que no hay cambios en la eficiencia técnica.
- ❖ TC: representa el cambio tecnológico en el período $t+1$ con respecto al período t , en el que $TC > 1$ significa progreso tecnológico.

Sin embargo, este índice no permitía explorar las posibles disparidades de rendimiento entre diferentes grupos de unidades cuyas circunstancias varían con el tiempo. Para superar parte de este problema, Camanho y Dyson (2006) propusieron originalmente un índice de tipo Malmquist con el fin de lograr un indicador promedio del rendimiento relativo de dos o más grupos de DMUs (dentro del mismo período t) cuando las directrices organizacionales y de gestión difieren: el CDMI (Índice de Malmquist de Camanho y Dyson).

El objetivo del nuevo índice era comparar el rendimiento de dos o más grupos de DMUs, y fue el resultado de reemplazar en el índice original los superíndices de tiempo (t y $t+1$) por indicadores de grupo (grupo G y grupo P). También, se utilizaron funciones de distancia de Shephard para evaluar el rendimiento relativo con respecto a las fronteras de producción de cada grupo.

El CDMI proporciona una herramienta para evaluar las diferencias de rendimiento entre estos grupos en un único período temporal:

$$\text{CDMI}_t^{\text{GP}} = \left[\frac{\left(\prod_{j=1}^N D^G(x_j^G, y_j^G) \right)^{1/N}}{\left(\prod_{i=1}^M D^G(x_i^P, y_i^P) \right)^{1/M}} \cdot \frac{\left(\prod_{i=j}^N D^P(x_j^G, y_j^G) \right)^{1/N}}{\left(\prod_{i=1}^M D^P(x_i^P, y_i^P) \right)^{1/M}} \right]^{1/2} \quad (5)$$

Este índice puede interpretarse como una medida general para la comparación del rendimiento relativo entre las unidades de los dos grupos G y P, y su descomposición en EG y TG se puede expresar del siguiente modo:

$$\text{CDMI}_t^{\text{GP}} = \underbrace{\frac{\left(\prod_{j=1}^N D^G(x_j^G, y_j^G) \right)^{1/N}}{\left(\prod_{i=1}^M D^P(x_i^P, y_i^P) \right)^{1/M}}}_{EG_t^{\text{GP}}} \times \underbrace{\left[\frac{\left(\prod_{j=1}^N D^P(x_j^G, y_j^G) \right)^{1/N}}{\left(\prod_{i=1}^M D^P(x_i^P, y_i^P) \right)^{1/M}} \cdot \frac{\left(\prod_{i=1}^M D^P(x_i^P, y_i^P) \right)^{1/M}}{\left(\prod_{j=1}^N D^G(x_j^G, y_j^G) \right)^{1/N}} \right]^{1/2}}_{TG_t^{\text{GP}}} \quad (6)$$

A menudo, en campos como la educación, es común extraer olas aleatorias de muestras representativas de centros educativos, con el fin de analizar su rendimiento a lo largo del tiempo. Esta selección de DMUs aleatoria, conocida como base de datos pseudo-panel, puede ser utilizada para comparar el rendimiento de estos grupos representativos de DMUs a lo largo del tiempo utilizando un índice tipo Malmquist de Pseudo-Panel (PPMI).

El índice PPMI es una combinación de dos índices tipo Malmquist y mide el cambio en la brecha de rendimiento relativo entre DMUs en períodos consecutivos. Sin embargo, su interpretación puede ser complicada debido a un factor denominado componente de divergencia (DC), que debe ser igual a uno para una correcta interpretación.

Dada la dificultad de interpretar el DC, se define un nuevo PPMI que fija un grupo de referencia como tecnología base.

Supongamos que hemos observado las DMUs de un grupo de referencia R (Aparicio et al., 2017; Aparicio & Santin, 2018). Entonces, el CDMI de grupo base para comparar el rendimiento de A y B en un período de tiempo s se define de la siguiente manera:

$$\text{CDMI}_s^{\text{AB}}(R^s) = \frac{\left(\prod_{j=1}^{n^{A^s}} D^{R^s}(x_j^{A^s}, y_j^{A^s}) \right)^{1/n^{A^s}}}{\left(\prod_{i=1}^{n^{B^s}} D^{R^s}(x_i^{B^s}, y_i^{B^s}) \right)^{1/n^{B^s}}} \quad (7)$$

Dicho índice, tiene la siguiente descomposición:

$$\text{CDMI}_s^{\text{AB}}(R^s) = \underbrace{\frac{\left(\prod_{j=1}^{n^{A^s}} D^{A^s}(x_j^{A^s}, y_j^{A^s}) \right)^{1/n^{A^s}}}{\left(\prod_{i=1}^{n^{B^s}} D^{B^s}(x_i^{B^s}, y_i^{B^s}) \right)^{1/n^{B^s}}}}_{EG_s^{\text{AB}}} \cdot \underbrace{\left[\frac{\left(\prod_{i=j}^{n^{A^s}} D^{R^s}(x_j^{A^s}, y_j^{A^s}) \right)^{1/n^{A^s}}}{\left(\prod_{j=1}^{n^{A^s}} D^{A^s}(x_j^{A^s}, y_j^{A^s}) \right)^{1/n^{A^s}}} \cdot \frac{\left(\prod_{i=1}^{n^{B^s}} D^{B^s}(x_i^{B^s}, y_i^{B^s}) \right)^{1/n^{B^s}}}{\left(\prod_{i=1}^{n^{B^s}} D^{R^s}(x_i^{B^s}, y_i^{B^s}) \right)^{1/n^{B^s}}} \right]}_{TG_s^{\text{AB}}(R^s)} \quad (8)$$

3.1.4. IMPLEMENTACIÓN DEL MODELO CCR

Para implementar el modelo correctamente, seguiremos los siguientes pasos:

1. Selección de DMUs: las DMUs en este estudio son los centros educativos de las comunidades autónomas de España.
2. Definición de Inputs y Outputs:
 - a. Inputs: Recursos del centro educativo, ratio alumno/profesor y nivel socioeconómico y cultural
 - b. Outputs: Puntuación promedio en lectura, en matemáticas, y en ciencias.

3. Normalización y limpieza de los datos: asegurarse de que los datos estén en unidades comparables y normalizar si es necesario.
4. Programación y aplicación del modelo CCR en Rstudio para calcular las eficiencias técnicas de cada comunidad autónoma: resolver los modelos de programación lineal para cada DMU, obteniendo los valores de ϕ (orientación al output).
5. Aplicación del modelo CCR fijando una frontera de referencia: marcar una frontera aleatoria que se comportará como la comunidad autónoma de referencia y resolver los modelos de programación lineal para cada DMU, obteniendo los valores de ϕ .
6. Por último, cálculo del índice CDMI (con frontera de referencia) y su descomposición para comparar dos a dos, la eficiencia y productividad de todas las CCAA.

3.2. DATOS Y VARIABLES

Para llevar a cabo el Análisis Envoltente de Datos, es crucial describir detalladamente las variables involucradas en el estudio.

Esta sección proporciona una descripción de los datos, las variables utilizadas como insumos y productos en el análisis, así como un resumen estadístico de los datos recopilados para las comunidades autónomas españolas.

3.2.1. CONJUNTO DE DATOS

Trabajaremos con los datos recogidos en PISA 2018 (PISA, 2018) para realizar nuestro estudio, centrándonos únicamente en los datos de España.

En nuestra base de datos se recoge información de 1047 centros de las diferentes comunidades autónomas de España durante el año 2018.

Contamos con un total de 10 variables, cuyos nombres en la base de datos son los siguientes:

- ❖ *SCHOOLID*: variable categórica que representa el Id de la escuela
- ❖ *Region*: variable categórica que representa la Comunidad Autónoma a la que pertenece la escuela
- ❖ *SCHLTYPE*: variable categórica que representa el tipo de escuela (Privada, Concertada, Pública)
- ❖ *PVSCIE_m*: variable numérica que representa la nota/resultados en la asignatura de Ciencias (en una escala sobre 700 puntos)
- ❖ *PVMATH_m*: variable numérica que representa la nota/resultados en la asignatura de Matemáticas (en una escala sobre 700 puntos)
- ❖ *PVREAD_m*: variable numérica que representa la nota/resultados en la asignatura de Lectura (en una escala sobre 700 puntos)
- ❖ *EDUSHORT2*: variable numérica que representa el inverso del índice de escasez de material educativo, que incluye datos sobre recursos similares (libros de texto, biblioteca, material de laboratorio y equipos informáticos).
- ❖ *TSRATIO*: variable numérica que nos indica el ratio de profesores por cada cien alumnos.
- ❖ *OBS*: variable numérica que representa el número de alumnos/observaciones en la escuela.
- ❖ *ESCS_m2*: variable numérica que representa la situación económica, social y cultural de los alumnos del centro

3.2.2. PREPROCESAMIENTO DE LA BASE DE DATOS

Inicialmente contamos con el siguiente número de centros de secundaria por comunidad autónoma:

- Andalucía: 52 escuelas
- Aragón: 52 escuelas
- Asturias: 54 escuelas
- Islas Baleares: 53 escuelas

- Islas Canarias: 53 escuelas
- Cantabria: 56 escuelas
- Castilla y León: 58 escuelas
- Castilla-La Mancha: 53 escuelas
- Cataluña: 50 escuelas
- Extremadura: 54 escuelas
- Galicia: 57 escuelas
- La Rioja: 44 escuelas
- Madrid: 142 escuelas
- Murcia: 52 escuelas
- Navarra: 49 escuelas
- País Vasco: 116 escuelas
- Comunidad Valenciana: 52 escuelas

Puesto que hemos decidido extraer del estudio los centros con menos de 10 estudiantes y también aquellos que tengan valores faltantes, nos quedaremos con 999 registros.

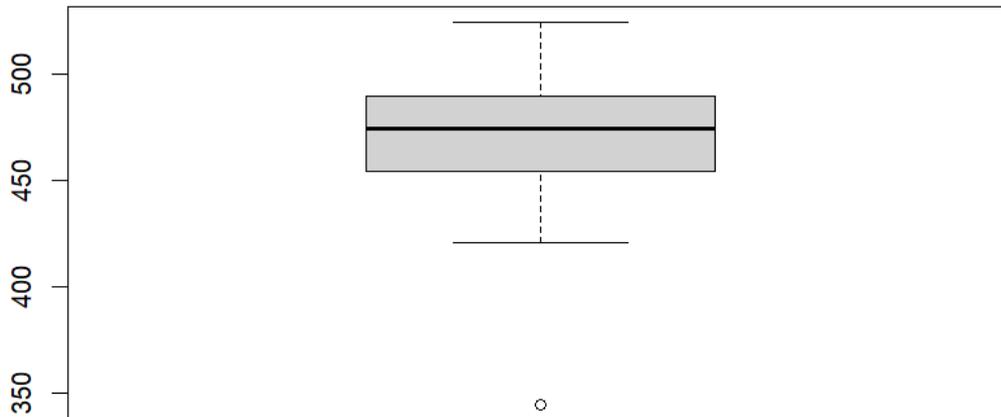
Posteriormente, se procede a eliminar los valores atípicos. Una buena estrategia para la identificación de valores atípicos, consiste en normalizar la variable de interés de la forma tradicional, lo que permite obtener una nueva variable que se interpreta como el número de unidades (positivas o negativas, dependiendo del signo) en que una observación se encuentre alejada de la media de la serie.

Una vez tenemos la variable normalizada, se consideran como valores atípicos las observaciones que se encuentren a más de 3 desviaciones estándar del promedio.

El primer paso para eliminar los outliers será separar nuestros datos por comunidades autónomas. Posteriormente se realizará una visualización de cada una de las variables mediante un gráfico de cajas, que nos mostrará de forma rápida si existen valores atípicos. Por último, procederemos a la eliminación de dichos valores en caso de ser necesario.

Por ejemplo, veamos el caso de la variable $PVSCIE_m$ (notas en ciencias) para Andalucía:

FIGURA 1: Boxplot de la puntuación en ciencias



Fuente: Elaboración propia

Se observa un valor atípico y se extrae de la base de datos.

Si realizamos la depuración de datos para todas las comunidades, finalmente nos quedamos con los siguientes registros.

TABLA 1: Número de centros por CA

CCAA	Centros de secundaria
1. Andalucía	48
2. Aragón	49
3. Asturias	51
4. Islas Baleares	46
5. Islas Canarias	50
6. Cantabria	50
7. Castilla y León	55
8. Castilla-La Mancha	51
9. Cataluña	46

CCAA	Centros de secundaria
10. Extremadura	52
11. Galicia	54
12. La Rioja	41
13. Madrid	125
14. Murcia	51
15. Navarra	46
16. País Vasco	106
17. Comunidad Valenciana	49

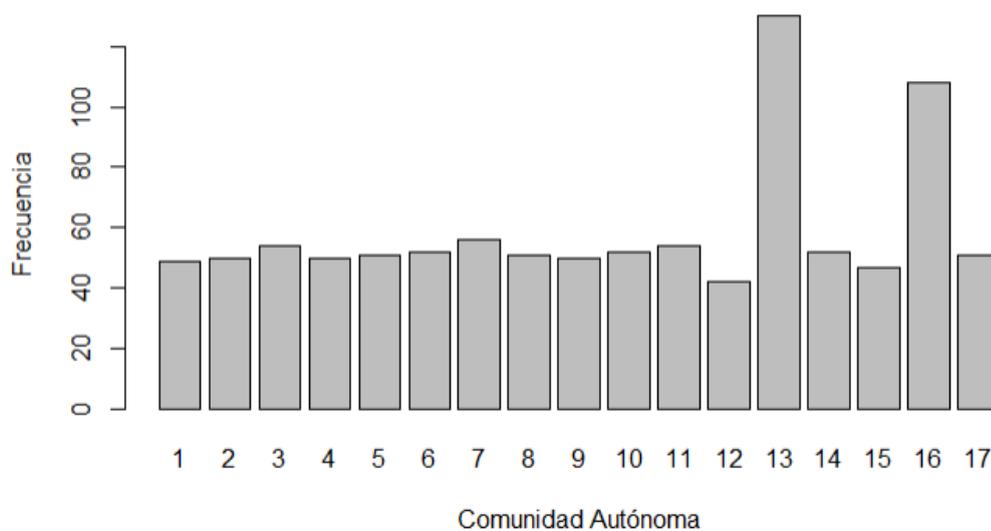
FUENTE: Elaboración propia

Lo que supone un total de **971 centros**.

Podemos observar el siguiente gráfico que nos permite visualizar mejor las diferencias entre el número de centros en cada comunidad autónoma:

FIGURA 2: Número de centros por CA

Centros por Comunidad Autónoma



FUENTE: Elaboración propia

Vemos que Madrid y País Vasco son las comunidades con más centros educativos, casi el doble que las otras comunidades.

3.2.3. DESCRIPCIÓN DE LOS DATOS

Uno de los primeros pasos para poder evaluar la eficiencia de los centros educativos mediante el DEA es decidir las variables que tomaremos como insumos y productos:

❖ **Inputs:**

- Recursos del centro educativo: *EDUSHORT2*
- Ratio profesor/alumno: *TSRATIO*
- Índice Socioeconómico y Cultural de los alumnos: *ESCS_m2*

❖ **Outputs:**

- Resultados en la asignatura de Ciencias: *PVSCIE_m*
- Resultados en la asignatura de Matemáticas: *PVMATH_m*
- Resultados en la asignatura de Lectura: *PVREAD_m*

A continuación, se presenta un resumen estadístico de las variables utilizadas en el análisis para las comunidades autónomas españolas. Este resumen incluye medidas como el promedio, mediana, la desviación típica, el valor mínimo y el valor máximo.

TABLA 2: Resumen descriptivo de los datos

	Media	Mediana	Máximo	Mínimo	Desviación típica
Ciencias	487,1	488,6	576,8	344,6	33,96
Matemáticas	487,0	488,3	592,2	369,2	35,99
Lectura	478,7	479,9	598,2	355,1	38,66
Recursos	3,84	3,86	5,38	1,00	1,06

	Media	Mediana	Máximo	Mínimo	Desviación típica
Ratio	11,21	9,14	100	1,94	12,01
Nivel socioeconómico	2,82	2,79	4,35	1,14	0,49
Nº alumnos	33,64	36,00	43,00	10,00	7,32

FUENTE: Elaboración propia

El análisis estadístico preliminar de las variables proporciona una visión general de la variabilidad y distribución de los insumos y de los productos. El ratio de profesores por estudiante muestra una variación considerable entre las comunidades autónomas, lo que puede influir significativamente en la eficiencia educativa. De manera similar, las puntuaciones en lectura, matemáticas y ciencias reflejan las diferencias en el rendimiento académico entre las comunidades (ver desviación típica en TABLA 2). También se han extraído dichos descriptivos por comunidad autónoma, que se pueden consultar en la sección 1 de los anexos.

Revisando dichos descriptivos, podemos extraer las siguientes conclusiones:

- ❖ La media más alta en las calificaciones de **Ciencias** la tienen los centros de Galicia, mientras que la más baja la tienen en Andalucía.
- ❖ La media más alta en las calificaciones de **Matemáticas** la tienen en Castilla y León, siendo Andalucía la peor.
- ❖ La media más alta en las calificaciones de **Lectura** la tiene Castilla y León, mientras que La Rioja tiene la peor.
- ❖ La media más alta en **Recursos** la tiene Cantabria, siendo Castilla-La Mancha la peor.
- ❖ La media más alta en el **Ratio** de profesores por alumno la tienen en Islas Baleares y la peor en Andalucía.
- ❖ La media más alta en **Nivel Socioeconómico** de los alumnos se localiza en Madrid, mientras que en Canarias se localiza el nivel más bajo.

4. ANÁLISIS DE RESULTADOS

A continuación, se presentarán los resultados del análisis DEA aplicado a los datos de PISA 2018, seguido de una discusión sobre las implicaciones de estos resultados.

4.1. APLICACIÓN DEL DEA A LOS DATOS DE PISA 2018

En esta sección, se aplica el Análisis Envolvente de Datos a los datos de PISA 2018 para evaluar la eficiencia de los centros educativos en las comunidades autónomas españolas. Utilizando el modelo CCR orientado al output, se midió la eficiencia relativa de cada comunidad autónoma en términos de maximización de resultados educativos dados ciertos niveles de recursos.

4.1.1. IMPLEMENTACIÓN DE LA METODOLOGÍA DEA EN R

Para calcular las eficiencias por comunidad autónoma, se utilizó el lenguaje de programación R (R Core Team, 2023) y 'lpSolveAPI'.

El paquete 'lpSolveAPI' (Berkelaar & others, 2020) proporciona una interfaz para resolver problemas de programación lineal y entera utilizando el solver lp_solve (Berkelaar & Eikland, 2004).

Se empezó programando la función 'func_dea_CCR' (ver anexo A18), en la que se aplica el método DEA CCR orientado al output (ver (1) sección 3.1.3) para una DMU concreta, cuyos parámetros de entrada son:

- ❖ DMU: indica, mediante un número entero, la posición de la DMU que se va a analizar
- ❖ datos_entrada: indica, mediante una matriz, los valores de los distintos inputs por filas
- ❖ datos_salidas: indica, mediante una matriz, los valores de los distintos outputs por filas

La función será capaz de reconocer el número de inputs, outputs y el número de DMUs. Posteriormente, construye el modelo lineal y lo resuelve para obtener las soluciones para la eficiencia técnica (ϕ) de una DMU que maximizan el problema planteado.

Posteriormente, se programó la función 'efic_CCR' (ver anexo A19) que recorrerá todas las DMUs, calculará sus eficiencias mediante la función 'func_dea_CCR' y almacenará dichas eficiencias en un vector.

Los parámetros de entrada para esta función son:

- ❖ n: indica, mediante un número entero, el total de DMUs que se desean analizar
- ❖ datos_entrada: indica, mediante una matriz, los valores de los distintos inputs por filas
- ❖ datos_salidas: indica, mediante una matriz, los valores de los distintos outputs por filas

Como modo de ejemplo, se mostrará el cálculo de eficiencias para los centros educativos (DMUs) de Andalucía:

```
n = 47 # DMUs
m = 3 # num. entradas (Xi)
s = 3 # num. salidas (Yi)

# Matriz de inputs

EDUSHORT2 <- Andalucia$EDUSHORT2
TSRATIO <- Andalucia$TSRATIO
ESCS_m2 <- Andalucia$ESCS_m2

datos_entradas = matrix(c(EDUSHORT2, TSRATIO, ESCS_m2),
nrow=m, ncol=n, byrow=TRUE)

# Matriz de outputs

PVSCIE_m <- Andalucia$PVSCIE_m
PVMATH_m <- Andalucia$PVMATH_m
PVREAD_m <- Andalucia$PVREAD_m
```

```

datos_salidas      =      matrix(c(PVSCIE_m,PVMATH_m,PVREAD_m),
nrow=s,ncol=n,byrow=TRUE)

Ef_Andalucia=efic_CCR(n,datos_entradas,datos_salidas)

```

Se obtuvo un vector de eficiencias similar a éste:

	Eficiencia
DMU1	1.153266
DMU2	1.092786
DMU3	1.043412
DMU4	1.128717
DMU5	1.125141
DMU6	1.000000
DMU7	1.101973
DMU8	1.153424
DMU9	1.025487
DMU10	1.000000
DMU11	1.205962
DMU12	1.085391
DMU13	1.116826

A continuación, mediante el uso del paquete ‘psych’ (Revelle, 2023) se calculó su media geométrica para obtener la media de eficiencia de Andalucía: 1,1174.

Los resultados del análisis DEA se presentan a continuación, mostrando las eficiencias medias calculadas para cada comunidad autónoma. El valor de las eficiencias varía de entre $[1, +\infty)$, donde un valor de 1 indica que la comunidad autónoma es eficiente.

TABLA 3: Eficiencias por comunidad autónoma

	Eficiencia media
Andalucía	1,117
Aragón	1,118
Asturias	1,118
Islas Baleares	1,157
Islas Canarias	1,138
Cantabria	1,081
Castilla y León	1,178
Castilla-La Mancha	1,121

	Eficiencia media
Cataluña	1,102
Extremadura	1,183
Galicia	1,111
La Rioja	1,214
Madrid	1,254
Murcia	1,167
Navarra	1,078
País Vasco	1,325
Comunidad Valenciana	1,146

Fuente: Elaboración Propia

Como se puede observar, las regiones más eficientes son Navarra y Cantabria, por tener una eficiencia más próxima a 1. Sin embargo, las más alejadas en eficiencia serán País Vasco y Madrid.

4.1.2. IMPLEMENTACIÓN DE LA METODOLOGÍA DEA CON UNA FRONTERA DE REFERENCIA EN R

Para poder realizar comparaciones entre comunidades autónomas, es necesario tener una frontera de referencia para evaluar cómo de “buena” es la comunidad autónoma con respecto a la comunidad de referencia.

Se generó una comunidad autónoma de referencia de forma aleatoria del siguiente modo:

1. La dimensión de dicha comunidad se fijó como la media de las dimensiones del resto de comunidades: 47 centros educativos
2. Se procedió a extraer una muestra aleatoria de la base de datos de tamaño 47 sin reemplazamiento con el comando ‘sample’
3. Se estudiaron los outliers de cada variable, pero no se encontró ninguno

4. Tras realizar el análisis de eficiencia con la función ‘efic_CCR’ y hallar su eficiencia media, se obtuvo un resultado de: 1,1772

El objetivo será resolver el siguiente problema, llamado modelo CCR con frontera de referencia, siendo A la comunidad autónoma que se desea analizar y R la comunidad de referencia:

$$\begin{aligned}
 & \text{Max } \phi_0^A \\
 & \text{s. a. } \sum_{j=1}^n \lambda_{j0} x_{ij}^R \leq x_{i0}^A, \quad i = 1, \dots, m \\
 & \sum_{j=1}^n \lambda_{j0} y_{ij}^R \geq \phi_0 y_{r0}^A, \quad r = 1, \dots, s \\
 & \lambda_{j0} \geq 0, \quad j = 1, \dots, n_A \quad (9)
 \end{aligned}$$

Para proceder a realizar el análisis DEA fijando la frontera de referencia seleccionada, se realizaron una serie de modificaciones en las funciones iniciales quedando definidas las nuevas funciones ‘func_dea_CCR_referencia’ y ‘efic_CCR_referencia’ (ver anexos A20 y A21).

Los cambios consistieron, fundamentalmente, en fijar los valores de los inputs y outputs de la frontera de referencia en el lado izquierdo de las restricciones y manteniendo en el lado derecho los datos referentes a la DMU que se desea analizar. En los parámetros de entrada, por tanto, se añadieron las matrices con la información referente a los inputs y outputs de la frontera de referencia.

A continuación, se mostrará el cálculo de eficiencias para los centros educativos de Andalucía:

```

Ef_Andalucia_referencia =
efic_CCR_referencia(n, datos_entradas, datos_salidas, datos_entradas_referencia, datos_salidas_referencia)

```

Se calcularon los valores “distancia”

$$D^R(X_j^A, Y_j^A) = \frac{1}{\phi_j^A}, j=1, \dots, n_A \quad (10)$$

para cada centro educativo de Andalucía, que son los valores inversos de las eficiencias ϕ halladas con la función ‘efic_CCR_referencia’. A continuación, se realizó la media geométrica de todos esos valores:

```
D_Andalucia_referencia=1/Ef_Andalucia_referencia
Media_D_Andalucia_referencia=geometric.mean(D_Andalucia_referencia)
```

El mismo procedimiento se realizó para todas las comunidades autónomas.

Posteriormente, se calculó el índice CDMI (ver (7) sección 3.1.3), que se encarga de medir la brecha de productividad entre A y B, donde ambas comunidades se miden con respecto a la de referencia (R):

$$CDMI^{AB}(R) = \frac{\left(\prod_{j=1}^{n^A} D^R(X_j^A, Y_j^A) \right)^{1/n^A}}{\left(\prod_{k=1}^{n^B} D^R(X_k^B, Y_k^B) \right)^{1/n^B}}$$

Si el CDMI es mayor que 1, significa que la comunidad autónoma A se ha comportado mejor que la comunidad autónoma B, en términos de productividad.

Para poder realizar una lectura de las tablas que vienen a continuación, será necesario tener en cuenta la siguiente leyenda:

CCAA
1. Andalucía
2. Aragón
3. Asturias
4. Islas Baleares

CCAA
5. Islas Canarias
6. Cantabria
7. Cataluña
8. Comunidad Valenciana
9. Extremadura
10. Galicia
11. Castilla-La Mancha
12. La Rioja
13. Castilla y León
14. Madrid
15. Murcia
16. Navarra
17. País Vasco

La siguiente tabla muestra todos los índices CDMI:

TABLA 4: Índices CDMI entre dos comunidades autónomas

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1,00	1,05	1,08	1,11	1,01	1,12	1,14	1,08	1,10	1,08	0,97	1,01	1,07	1,09	0,99	1,10	1,09
2	-	1,00	1,03	1,05	0,96	1,06	1,09	1,03	1,05	1,03	0,92	0,97	1,02	1,04	0,94	1,05	1,04
3	-	-	1,00	1,03	0,94	1,04	1,06	1,00	1,02	1,00	0,90	0,94	0,99	1,02	0,92	1,02	1,01
4	-	-	-	1,00	0,91	1,01	1,03	0,98	0,99	0,97	0,87	0,92	0,97	0,99	0,89	1,00	0,99
5	-	-	-	-	1,00	1,11	1,13	1,07	1,09	1,07	0,96	1,01	1,06	1,09	0,98	1,09	1,08
6	-	-	-	-	-	1,00	1,02	0,97	0,98	0,96	0,87	0,91	0,96	0,98	0,88	0,99	0,98
7	-	-	-	-	-	-	1,00	0,95	0,96	0,94	0,85	0,89	0,94	0,96	0,87	0,97	0,96
8	-	-	-	-	-	-	-	1,00	1,02	1,00	0,90	0,94	0,99	1,01	0,91	1,02	1,01
9	-	-	-	-	-	-	-	-	1,00	0,98	0,88	0,92	0,97	1,00	0,90	1,00	0,99
10	-	-	-	-	-	-	-	-	-	1,00	0,90	0,94	0,99	1,02	0,92	1,03	1,01
11	-	-	-	-	-	-	-	-	-	-	1,00	1,05	1,10	1,13	1,02	1,14	1,13
12	-	-	-	-	-	-	-	-	-	-	-	1,00	1,05	1,08	0,97	1,09	1,07

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
13	-	-	-	-	-	-	-	-	-	-	-	-	1,00	1,03	0,92	1,03	1,02
14	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00	0,90	1,01	1,00
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00	1,12	1,11
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00	0,99
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00

Fuente: Elaboración propia.

() Las comunidades situadas en las filas, se consideran como comunidad autónoma A, mientras que las columnas se consideran comunidad autónoma B.*

A continuación, se calcula la descomposición del índice anterior en función de la eficiencia técnica y el cambio tecnológico: EG y TG (ver (8) sección 3.1.3).

La siguiente tabla muestra los datos de EG (Eficiencia Global):

TABLA 5: Datos del cambio en la Eficiencia Global entre cada par de comunidades autónomas

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1,00	1,00	1,00	1,04	1,00	0,97	0,99	1,03	1,06	1,06	1,00	1,09	1,05	1,12	1,04	0,97	1,19
2	-	1,00	1,00	1,04	1,00	0,97	0,99	1,03	1,06	1,06	1,00	1,09	1,05	1,12	1,04	0,96	1,18
3	-	-	1,00	1,04	1,00	0,97	0,99	1,03	1,06	1,06	1,00	1,09	1,05	1,12	1,04	0,96	1,19
4	-	-	-	1,00	0,96	0,94	0,95	0,99	1,02	1,02	0,97	1,05	1,02	1,08	1,01	0,93	1,15
5	-	-	-	-	1,00	0,97	0,99	1,03	1,06	1,06	1,01	1,09	1,06	1,12	1,05	0,97	1,19
6	-	-	-	-	-	1,00	1,02	1,06	1,09	1,09	1,03	1,12	1,09	1,16	1,08	0,99	1,22
7	-	-	-	-	-	-	1,00	1,04	1,07	1,07	1,02	1,10	1,07	1,14	1,06	0,98	1,20
8	-	-	-	-	-	-	-	1,00	1,03	1,03	0,98	1,06	1,03	1,09	1,02	0,94	1,16
9	-	-	-	-	-	-	-	-	1,00	1,00	0,95	1,03	1,00	1,06	0,99	0,91	1,12
10	-	-	-	-	-	-	-	-	-	1,00	0,95	1,03	1,00	1,06	0,99	0,91	1,12
11	-	-	-	-	-	-	-	-	-	-	1,00	1,08	1,05	1,12	1,04	0,96	1,18
12	-	-	-	-	-	-	-	-	-	-	-	1,00	0,97	1,03	0,96	0,89	1,09
13	-	-	-	-	-	-	-	-	-	-	-	-	1,00	1,06	0,99	0,92	1,12
14	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00	0,93	0,86	1,06
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00	0,92	1,14
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00	1,23

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1,00	1,00	1,00	1,04	1,00	0,97	0,99	1,03	1,06	1,06	1,00	1,09	1,05	1,12	1,04	0,97	1,19
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00

Fuente: Elaboración propia.

() Las comunidades situadas en las filas, se consideran como comunidad autónoma A, mientras que las columnas se consideran comunidad autónoma B.*

La siguiente tabla muestra los datos de TG:

TABLA 6: Datos del cambio tecnológico entre cada par de comunidades autónomas

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1,00	1,05	1,08	1,07	1,01	1,15	1,16	1,05	1,04	1,02	0,96	0,93	1,01	0,98	0,94	1,14	0,92
2	-	1,00	1,03	1,02	0,96	1,10	1,10	1,00	0,99	0,97	0,92	0,89	0,97	0,93	0,90	1,09	0,88
3	-	-	1,00	0,99	0,94	1,07	1,07	0,98	0,96	0,94	0,89	0,87	0,94	0,91	0,88	1,06	0,85
4	-	-	-	1,00	0,95	1,08	1,08	0,99	0,97	0,95	0,90	0,88	0,95	0,91	0,89	1,07	0,86
5	-	-	-	-	1,00	1,14	1,15	1,04	1,03	1,01	0,95	0,93	1,00	0,97	0,94	1,13	0,91
6	-	-	-	-	-	1,00	1,01	0,92	0,90	0,88	0,84	0,81	0,88	0,85	0,82	1,00	0,80
7	-	-	-	-	-	-	1,00	0,91	0,90	0,88	0,83	0,81	0,88	0,84	0,82	0,99	0,80
8	-	-	-	-	-	-	-	1,00	0,99	0,97	0,92	0,89	0,96	0,93	0,90	1,09	0,87
9	-	-	-	-	-	-	-	-	1,00	0,98	0,93	0,90	0,98	0,94	0,91	1,10	0,89
10	-	-	-	-	-	-	-	-	-	1,00	0,95	0,92	1,00	0,96	0,93	1,13	0,91
11	-	-	-	-	-	-	-	-	-	-	1,00	0,97	1,05	1,01	0,98	1,19	0,95
12	-	-	-	-	-	-	-	-	-	-	-	1,00	1,09	1,04	1,01	1,22	0,98
13	-	-	-	-	-	-	-	-	-	-	-	-	1,00	0,96	0,93	1,13	0,91
14	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00	0,97	1,17	0,94
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00	1,21	0,97
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00	0,80
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1,00

Fuente: Elaboración propia.

() Las comunidades situadas en las filas, se consideran como comunidad autónoma A, mientras que las columnas se consideran comunidad autónoma B.*

4.2. COMPARACIÓN DE LA EFICIENCIA DE LAS COMUNIDADES AUTÓNOMAS

Esta sección presenta los principales resultados de nuestro análisis para comparar las brechas de productividad y sus cambios en las diferentes comunidades autónomas de España, utilizando el grupo de referencia generado aleatoriamente (R).

Los resultados de la TABLA 4 referidos al índice CDMI, muestran las posibles brechas de productividad entre las diferentes regiones de España. La comunidad A habrá sido más productiva que la comunidad B si el índice es mayor que 1, igual de productivo si el índice es 1 y menos productivo si el índice es menor que 1.

La Comunidad Valenciana, por ejemplo, supera en productividad a las Islas Baleares, Cantabria, Cataluña, Extremadura, Madrid, Navarra y País Vasco, y se comporta igual que Galicia y Asturias. Sin embargo, queda por debajo de comunidades como Andalucía, Aragón, Islas Canarias, Castilla-La Mancha, La Rioja, Castilla y León y Murcia.

Se puede observar una importante brecha positiva de productividad que sitúa a Andalucía en ventaja respecto a prácticamente todas las comunidades, excepto a Castilla-La Mancha y Murcia en las que se obtiene un índice menor que 1.

Sin embargo, Murcia supera a todas excepto a Castilla-La Mancha. Por lo que, en general, Castilla-La Mancha supera en productividad a todas las comunidades autónomas.

El origen de la brecha entre Castilla-La Mancha y el resto de comunidades se explica por el componente tecnológico en la mayoría de los casos, excepto para Castilla-La Mancha, La Rioja, Castilla y León, Madrid, Murcia y País Vasco (ver TABLA 5). En estos casos excepcionales, la brecha queda explicada por la componente del cambio de eficiencia (ver TABLA 6).

Por otro lado, también se identifica un importante déficit de productividad en el que Cantabria y Cataluña van a la cabeza, siendo Cataluña la peor de todas las regiones. La

brecha entre Cataluña y el resto de comunidades queda explicado por el componente tecnológico, excepto para Cantabria y Navarra que lo explica la componente de eficiencia.

4.3. IDENTIFICACIÓN DE MEJORES PRÁCTICAS E INTERPRETACIÓN DE LOS RESULTADOS

A modo de resumen, sería interesante reconocer las *mejores prácticas* para tomar como referentes para posibles mejoras.

En cuanto a nivel general se refiere, Castilla-La Mancha es quien menos brecha en productividad tiene de todas las comunidades autónomas.

Sin embargo, a nivel tecnológico, tendríamos que tomar como referente a La Rioja por ser quien supera al resto de comunidades en este aspecto. Sin embargo, en cuanto a nivel de eficiencia, sería Navarra la comunidad que tomaríamos como ejemplo a seguir.

Por otro lado, también sería interesante reconocer aquellas comunidades que han tenido las *peores prácticas* durante 2018 para tenerlas en cuenta a la hora de mejorar.

Como hemos visto en el apartado anterior, Cataluña es la que peor se comporta en productividad a nivel general. Si nos centramos en el nivel tecnológico, Cataluña, Navarra y Cantabria son las regiones más afectadas, mientras que a nivel de eficiencia, lo son La Rioja y Madrid.

5. CONCLUSIONES Y TRABAJO FUTURO

En este capítulo, se ofrecerán conclusiones y recomendaciones para futuras investigaciones y políticas educativas.

5.1. CONCLUSIONES

El análisis de eficiencia de los centros educativos en las comunidades autónomas españolas, basado en la metodología DEA y utilizando los datos de PISA 2018, ha proporcionado una visión clara sobre el uso de los recursos educativos y la productividad en cada región.

A continuación, se presentan las conclusiones principales basadas en los objetivos fijados:

- ❖ **Castilla-La Mancha** se destaca como la comunidad autónoma con la menor brecha en productividad en comparación con el resto de comunidades autónomas, lo que sugiere un uso eficiente de sus recursos educativos en general.
- ❖ **Navarra** ha demostrado ser la más eficiente en términos de uso de recursos, lo que la posiciona como un modelo a seguir para otras regiones que buscan mejorar su eficiencia.
- ❖ **La Rioja** sobresale en el aspecto tecnológico, superando al resto de las comunidades autónomas. Esto indica que La Rioja ha implementado prácticas tecnológicas avanzadas que han contribuido significativamente a su alta productividad.
- ❖ **La Comunidad Valenciana**, por ejemplo, se destaca por superar en productividad a las Islas Baleares, Cantabria, Cataluña, Extremadura, Madrid, Navarra y el País Vasco. Además, su rendimiento es comparable al de Galicia y Asturias, situándose en un nivel intermedio en cuanto a eficiencia educativa. No obstante, queda por debajo de otras comunidades que han demostrado un rendimiento superior en términos de productividad.

Las diferencias observadas muestran la necesidad de adoptar y adaptar las mejores prácticas de las comunidades más eficientes. Por ejemplo, la Comunidad Valenciana podría beneficiarse de analizar y tomar como referente las estrategias implementadas en regiones como La Rioja, que sobresale en términos de avances tecnológicos, o Navarra, que se destaca en eficiencia.

5.2. RECOMENDACIONES PARA MEJORAR LA EFICIENCIA EDUCATIVA EN ESPAÑA

A continuación, se muestran posibles propuestas de mejoras para implementar en los centros educativos españoles:

❖ Adopción de Tecnologías Avanzadas:

- Inversión en infraestructura Tecnológica: Aumentar la inversión en infraestructuras tecnológicas, incluyendo hardware, software y conectividad de alta calidad.
- Formación en TIC: Implementar programas de formación continua para docentes en el uso de tecnologías de la información y la comunicación (TIC) para mejorar la enseñanza y el aprendizaje.

❖ Gestión Eficiente de Recursos:

- Optimización de recursos: Implementar sistemas de gestión que optimicen el uso de recursos humanos, materiales y financieros.
- Evaluación y monitoreo: Desarrollar mecanismos de evaluación y monitoreo continuo para asegurar que los recursos se utilicen de manera eficiente.

❖ Mejora de la Eficiencia Técnica:

- Buenas prácticas en gestión: Adoptar las mejores prácticas de gestión observadas en comunidades autónomas como Navarra, incluyendo la planificación estratégica, la gestión del tiempo y la organización del trabajo.

- Capacitación y desarrollo profesional: Fomentar programas de desarrollo profesional y capacitación para los docentes y el personal administrativo.
- ❖ **Colaboración y Transferencia de Conocimientos:**
 - Redes de colaboración: Crear redes de colaboración entre las comunidades autónomas para compartir conocimientos, experiencias y buenas prácticas.
 - Estudios comparativos: Realizar estudios comparativos periódicos para identificar y evaluar nuevas áreas de mejora.
- ❖ **Políticas Educativas Sostenibles:**
 - Políticas de largo plazo: Diseñar y aplicar políticas educativas sostenibles que tengan en cuenta las necesidades a largo plazo y que fomenten la innovación y la eficiencia.
 - Participación de la comunidad educativa: Involucrar a la comunidad educativa, incluidos padres y estudiantes, en la toma de decisiones para asegurar que las políticas educativas respondan a las necesidades locales.

En resumen, la implementación de estas recomendaciones puede contribuir significativamente a mejorar la eficiencia y la productividad de los centros educativos en España, asegurando un mejor uso de los recursos y optimizando los resultados educativos a nivel nacional.

5.3. PROPUESTAS PARA FUTURAS INVESTIGACIONES

Para futuras investigaciones, es esencial considerar estudios que examinen la evolución de la eficiencia educativa en las comunidades autónomas a lo largo del tiempo.

Una futura línea de investigación sería realizar este estudio haciendo uso del índice PPMI con grupo de referencia (Aparicio et al., 2017; Aparicio & Santin, 2018), aplicado a los datos PISA de varios años para poder valorar cómo evoluciona la productividad, la eficiencia y el cambio tecnológico a lo largo de esos años. Este enfoque permitirá identificar tendencias y cambios en la eficiencia, así como el impacto de diversas políticas y reformas educativas implementadas en diferentes periodos. Además, es importante investigar cómo las tecnologías avanzadas, como la Inteligencia

Artificial y el contenido on-line, están influyendo en la eficiencia y la calidad de la educación. Evaluar la incorporación de estas tecnologías en el aula puede proporcionar una mejor comprensión de su eficacia y cómo pueden optimizarse para mejorar los resultados educativos.

Otra área de interés es la evaluación de la eficiencia en la educación inclusiva. Investigaciones futuras podrían centrarse en cómo los centros educativos están atendiendo a estudiantes con necesidades especiales y qué prácticas pueden ser implementadas para mejorar la eficiencia en este contexto. También sería útil realizar comparaciones de la eficiencia de los sistemas educativos españoles con los de otros países, para identificar prácticas exitosas a nivel internacional que podrían ser adaptadas al contexto español.

Finalmente, se recomienda desarrollar y utilizar nuevos indicadores de calidad educativa que puedan integrarse en los análisis de eficiencia. Estos indicadores podrían incluir aspectos como la satisfacción de los estudiantes y padres, el bienestar emocional de los estudiantes y la capacidad de las escuelas para fomentar habilidades críticas y creativas. Al integrar estos nuevos indicadores, las investigaciones futuras pueden proporcionar una visión más completa y global del rendimiento educativo en España, ofreciendo datos valiosos para la mejora continua del sistema educativo nacional.

6. REFERENCIAS

- Afonso, A., & St. Aubyn, M. (2005). Non-Parametric Approaches to Education and Health Efficiency in OECD Countries. *Journal of Applied Economics*, 8(2), 227-246. [10.1080/15140326.2005.12040626](https://doi.org/10.1080/15140326.2005.12040626)
- Aparicio, J., Crespo-Cebada, E., Pedraja-Chaparro, F., & Santín, D. (2017). Comparing school ownership performance using a pseudo-panel database: A Malmquist-type index approach. *European Journal of Operational Research*, 256(2), 533-542. <https://doi.org/10.1016/j.ejor.2016.06.030>
- Aparicio, J., & Santin, D. (2018). A note on measuring group performance over time with pseudo-panels. *European Journal of Operational Research*, 267(1), 227-235. <https://doi.org/10.1016/j.ejor.2017.11.049>
- Barnum, D. T., McNeil, S., & Hart, J. (2007). Comparing the Efficiency of Public Transportation Subunits Using Data Envelopment Analysis. *Journal of Public Transportation*, 10(2), 1-16. <http://doi.org/10.5038/2375-0901.10.2.1>
- PISA. (2018). Web del Ministerio de Educación. <https://www.educacionfpydeportes.gob.es/inee/evaluaciones-internacionales/pisa/pisa-2018.html>
- Berkelaar, M., & Eikland, K. (2004). *lpsolve: Open Source (Mixed-Integer) Linear Programming System*.
- Berkelaar, M., & others. (2020). *lpSolveAPI: R Interface to 'lp_solve'* (Version 5.5.2.0. R package version 5.5.2.0-17 ed.). CRAN
- Camanho, A. S. , & Dyson, R. G. (2006). Data envelopment analysis and Malmquist indices for measuring group performance. *Journal of Productivity Analysis*, 26 , 35–49 . <https://doi.org/10.1007/s11123-006-0004-8>

- Casu, B., & Molyneux, P. (2003). A comparative study of efficiency in European banking. *Applied Economics*, 35(17), 1865–1876.
<https://doi.org/10.1080/0003684032000158109>
- Caves, D.W., Christensen, L.R., & Diewert, W.E. (1982). The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity. *Econometrica*, 50(6), 1393–1414. <https://doi.org/10.2307/1913388>
- Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*. *European Journal of Operational Research*, 2(6), 429-444.
[https://doi.org/10.1016/0377-2217\(78\)90138-8](https://doi.org/10.1016/0377-2217(78)90138-8)
- Färe, R. , Grosskopf, S. , & Lovell, C. A. K. (1994). Production frontiers . *Cambridge University Press*. <https://doi.org/10.1017/CBO9780511551710>
- Fernandez-Santos, Y., & Martínez-Campillo, A. (2015). *¿Ha mejorado la productividad docente e investigadora de las universidades públicas españolas desde la aprobación de la LOU?: Evidencia a partir del bootstrap*. Ministerio de Educación, Formación Profesional y Deportes.
<https://www.educacionfpydeportes.gob.es/revista-de-educacion/numeros-revista-educacion/numeros-anteriores/2015/367/367-6.html>
- Giménez, V., Prior, D., & Thieme, C. (2007). Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison. *Journal of the Operational Research Society*, 58(8), 996–1007.
<https://doi.org/10.1057/palgrave.jors.2602213>
- Hollingsworth, B. (2008). The Measurement of Efficiency and Productivity of Health Care Delivery. *Health Economics*, 17(10), 1107-1128.
<https://doi.org/10.1002/hec.1391>

Liu, J. S., Lu, L. Y., Lu, W. M., & Lin, B. J. (2013). A survey of DEA applications.

Omega, 41(5), 893-902. <https://doi.org/10.1016/j.omega.2012.11.004>

PISA. (2024). Web Oficial de PISA. Retrieved Abril, 2024, <https://oecd.org/pisa/>

R Core Team. (2023). R: A language and environment for statistical computing. *R*

Foundation for Statistical Computing. <https://www.R-project.org/>

Revelle, W. (2023). *An Introduction to psych: Package Documentation*.

<https://cran.r-project.org/web/packages/psych/index.html>

Thanassoulis, E., Kortelainen, M., Johnes, G., & Johnes, J. (2011). Costs and efficiency

of higher education institutions in England: a DEA analysis. *Journal of the operational research society*, 62(7), 1282-1297.



7. ANEXOS

1. Descriptivos por comunidad autónoma

❖ A1. Descriptivos de Andalucía

	Media	Mediana	Máximo	Mínimo	Desviación típica
Ciencias	472,00	474,10	524,20	420,50	28,13
Matemáticas	467,90	467,50	523,70	416,40	28,18
Lectura	467,40	470,00	528,10	362,00	36,71
Recursos	3,44	3,50	5,38	1,00	1,14
Ratio	8,23	8,47	15,42	4,99	2,16
Nivel socioecon	2,56	2,46	3,46	1,54	0,49
Nº alumnos	34,55	36,00	42,00	17,00	5,62

❖ A2. Descriptivos de Aragón

	Media	Mediana	Máximo	Mínimo	Desviación típica
Ciencias	492,60	494,00	557,70	436,70	29,15
Matemáticas	495,20	496,50	565,70	431,00	29,57
Lectura	488,00	486,80	560,30	417,60	33,26
Recursos	3,86	3,86	5,38	1,00	1,10
Ratio	9,08	9,55	18,21	1,94	3,22
Nivel socioecon	2,87	2,84	3,71	2,15	0,38
Nº alumnos	34,71	36,00	41,00	12,00	5,62

❖ A3. Descriptivos de Asturias

	Media	Mediana	Máximo	Mínimo	Desviación típica
Ciencias	496,1	492,3	496,1	434,3	29,85
Matemáticas	490,7	488,3	574,0	425,0	31,53
Lectura	494,6	494,3	598,2	412,4	33,12
Recursos	3,80	3,86	5,38	1,00	1,07
Ratio	13,06	10,24	97,30	5,30	1,07
Nivel socioecon	2,84	2,76	4,16	1,95	0,52
Nº alumnos	34,94	37,50	42,00	12,00	7,08

❖ A4. Descriptivos de Islas Baleares

	Media	Mediana	Máximo	Mínimo	Desviación típica
Ciencias	481,1	482,6	527,5	382,5	28,84
Matemáticas	481,9	484,2	535,9	372,3	29,88
Lectura	478,2	476,7	533,8	393,8	31,11
Recursos	3,69	3,69	5,38	1,77	1
Ratio	15,27	11,10	100	5,17	18,53
Nivel socioecon	2,76	2,79	3,82	1,14	0,49
Nº alumnos	33,42	36,00	40,00	14,00	6,53

❖ A5. Descriptivos de Islas Canarias

	Media	Mediana	Máximo	Mínimo	Desviación típica
Ciencias	468,8	464,4	551,6	403,1	32,75
Matemáticas	459,6	454,6	539,1	385,5	33,48
Lectura	471,3	468,2	559,0	413,1	35,28
Recursos	3,52	3,53	5,38	1,00	1,11
Ratio	9,18	8,97	30,59	4,64	3,92
Nivel socioecon	2,50	2,45	3,85	1,79	0,5
Nº alumnos	33,86	36,00	42,00	17,00	5,33

❖ A6. Descriptivos de Cantabria

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	497,97	498,89	552,81	429,05	21,88
Matemáticas	501,87	501,25	549,26	428,70	24,26
Lectura	485,95	486,86	538,96	424,08	26,07
Recursos	4,13	4,13	5,38	2,71	0,76
Ratio	9,92	10,32	14,58	5,04	2,76
Nivel socioecon	2,93	2,83	3,87	2,15	0,38
Nº alumnos	34,96	37,00	42,00	16,00	6,52

❖ A7. Descriptivos de Castilla y León

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	501,48	504,51	549,70	440,42	27,88
Matemáticas	502,71	503,57	563,45	435,79	28,75
Lectura	496,44	498,89	555,55	428,42	31,33
Recursos	3,93	4,07	5,38	1,00	1,07
Ratio	11,38	9,43	94,73	4,27	12,02
Nivel socioecon	2,86	2,81	3,91	1,93	0,40
Nº alumnos	32,82	37,00	42,00	12,00	8,81

❖ A8. Descriptivos de Castilla-La Mancha

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	485,29	490,70	524,05	420,15	26,66
Matemáticas	479,60	480,11	530,31	409,72	27,87
Lectura	478,63	484,45	529,62	405,47	30,77
Recursos	2,97	2,75	5,38	1,00	1,20
Ratio	8,87	8,95	15,03	4,93	2,32
Nivel socioecon	2,66	2,55	3,77	1,99	0,48
Nº alumnos	34,67	37,00	41,00	14,00	6,07

❖ A9. Descriptivos de Cataluña

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	489,50	490,49	574,84	381,94	35,20
Matemáticas	491,01	487,01	592,18	402,35	34,03
Lectura	485,53	483,68	585,70	371,02	36,86
Recursos	4,04	3,86	5,38	1,82	1,03
Ratio	10,48	9,62	35,96	5,12	5,11
Nivel socioecon	3,00	2,95	4,14	2,22	0,47
Nº alumnos	34,39	37,00	41,00	13,00	6,64

❖ A10. Descriptivos de Extremadura

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	472,58	472,40	535,44	400,10	32,85
Matemáticas	468,39	465,78	542,34	398,47	31,90
Lectura	463,31	467,26	545,08	385,85	35,65
Recursos	3,84	3,90	5,38	1,00	1,00
Ratio	11,19	11,24	19,93	4,78	3,46
Nivel socioecon	2,52	2,46	3,60	1,64	0,43
Nº alumnos	33,37	35,00	42,00	13,00	6,58

❖ A11. Descriptivos de Galicia

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	510,17	506,86	554,58	461,16	22,46
Matemáticas	497,23	496,56	556,58	447,72	24,57
Lectura	492,40	488,16	556,09	433,74	28,14
Recursos	4,07	4,03	5,38	2,03	1,03
Ratio	10,38	10,01	19,94	5,56	3,56
Nivel socioecon	2,82	2,78	4,09	2,00	0,45
Nº alumnos	33,46	37,50	42,00	12,00	8,56

❖ A12. Descriptivos de La Rioja

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	482,26	488,65	550,06	396,55	36,14
Matemáticas	492,99	500,92	556,33	400,06	38,09
Lectura	461,81	469,96	537,86	353,14	43,83
Recursos	3,97	3,89	5,38	2,24	0,92
Ratio	8,31	7,89	14,11	2,09	2,34
Nivel socioecon	2,71	2,74	3,65	1,67	0,44
Nº alumnos	34,22	37,00	42,00	10,00	7,53

❖ A13. Descriptivos de Madrid

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	494,50	492,30	575,21	393,30	36,99
Matemáticas	495,11	497,15	583,66	392,10	40,16
Lectura	481,56	481,78	579,52	366,47	46,43
Recursos	4,08	4,06	5,38	1,82	0,95
Ratio	8,56	7,61	44,16	2,63	4,92
Nivel socioecon	3,21	3,25	4,35	1,63	0,62
Nº alumnos	34,98	37,00	43,00	12,00	6,21

❖ A14. Descriptivos de Murcia

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	480,43	480,04	558,69	392,78	36,06
Matemáticas	474,71	472,64	547,85	390,70	35,95
Lectura	482,77	482,91	586,66	377,17	38,59
Recursos	3,65	3,65	5,38	1,50	1,05
Ratio	8,25	8,53	11,49	2,62	1,74
Nivel socioecon	2,51	2,48	3,81	1,32	0,50
Nº alumnos	32,31	35,00	42,00	11,00	7,25

❖ A15. Descriptivos de Navarra

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	491,57	490,90	565,72	413,70	34,91
Matemáticas	501,71	499,96	584,09	439,93	34,34
Lectura	470,88	473,40	562,49	383,65	41,76
Recursos	4,08	3,88	5,38	2,19	0,98
Ratio	10,02	10,03	15,14	5,63	3,16
Nivel socioecon	2,89	2,79	3,79	1,64	0,45
Nº alumnos	35,43	37,00	42,00	12,00	6,10

❖ A16. Descriptivos de País Vasco

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	482,55	486,09	558,53	383,51	34,32
Matemáticas	492,05	501,12	580,61	387,51	40,17
Lectura	470,04	472,99	563,77	366,53	39,23
Recursos	3,99	3,96	5,38	1,00	1,03
Ratio	9,88	8,55	31,61	3,06	4,06
Nivel socioecon	2,89	2,94	3,84	1,14	0,52
Nº alumnos	30,07	33,00	42,00	10,00	9,62

❖ A17. Descriptivos de la Comunidad Valenciana

	Media	Mediana	Máximo	Mínimo	Desviación Típica
Ciencias	476,85	476,15	551,07	428,00	26,23
Matemáticas	471,63	468,40	536,54	424,80	24,92
Lectura	471,77	471,74	568,28	403,53	34,67
Recursos	3,78	3,91	5,38	1,00	1,09
Ratio	10,07	9,73	32,19	5,59	4,23
Nivel socioecon	2,69	2,63	3,69	1,72	0,44
Nº alumnos	34,00	36,00	42,00	15,00	6,66

2. Implementación en RStudio

❖ A18. Función para aplicar el método CCR (programación propia en R)

```
func_dea_CCR= function(DMU,datos_entradas,datos_salidas) {
  DMU=DMU
  X =datos_entradas #inputs
  Y =datos_salidas #outputs
  n =dim(X)[2] #nº DMU's
  m =dim(X)[1] #nº entradas
  s =dim(Y)[1] #nº salidas

  #Construimos el modelo para la DMU indicada
  model = make.lp(nrow = 0, ncol = n + 1) #0 restricciones y
n + 1 variables
  set.objfn(model, c(1,rep(0,n))) #la primera corresponde a
theta
  set.bounds(model, lower = "-inf", columns = (1)) #para que
theta sea libre
  lp.control(model, sense = "max")

  #Restricciones inputs
  for (i in 1:m){
    coeficientes = c(0, X[i,])
    add.constraint(model,coeficientes,"<=",X[i,DMU])
  }
}
```

```
#Restricciones outputs
for (j in 1:s){
  #coeficientes = c(-Y[j,DMU],Y[j,DMU], Y[j,])
  coeficientes = c(-Y[j,DMU], Y[j,])
  add.constraint(model,coeficientes,">=",0)
}

#Resolvemos
solve(model)
sol= get.objective(model) #valor óptimo función objetivo
return(sol)
}
```

- ❖ A19. Función para aplicar el método a todas las DMUs y calcular la eficiencia por Comunidad Autónoma (programación propia en R)

```
efic_CCR=function(n,datos_entradas,datos_salidas) {
  eficiencia = matrix(NA , nrow=n, ncol=1, byrow=TRUE)
  colnames(eficiencia)<-c("Eficiencia")

  DMUS = c(1:n)

  for (i in 1:n){
    DMU=i
    DMUS[i]=paste0("DMU",i)

  eficiencia[i,1]=func_dea_CCR(DMU,datos_entradas,datos_salidas
  )
  }

  rownames(eficiencia)=DMUS
  return(eficiencia)
}
```

❖ A20. Función para aplicar el método CCR con frontera de referencia (programación propia en R)

```

func_dea_CCR_referencia=
function(DMU,datos_entradas,datos_salidas,datos_entradas_referencia,datos_salidas_referencia) {
  DMU=DMU
  X =datos_entradas #inputs
  Y =datos_salidas #outputs
  X_ref = datos_entradas_referencia #inputs referencia
  Y_ref = datos_salidas_referencia #outputs referencia
  n =dim(X) [2] #n° DMU's
  m =dim(X) [1] #n° entradas
  s =dim(Y) [1] #n° salidas
  p =dim(X_ref) [2] #n° DMU's referencia

  #Construimos el modelo para la DMU indicada
  model = make.lp(nrow = 0, ncol = p + 1) #0 restricciones y
p + 1 variables
  set.objfn(model, c(1,rep(0,p))) #la primera corresponde a
theta
  set.bounds(model, lower = "-inf", columns = (1)) #para que
theta sea libre
  lp.control(model, sense = "max")

  #Restricciones inputs
  for (i in 1:m){
    coeficientes = c(0, X_ref[i,])
    add.constraint(model,coeficientes,"<=",X[i,DMU])
  }

  #Restricciones outputs
  for (j in 1:s){
    coeficientes = c(-Y[j,DMU], Y_ref[j,])
    add.constraint(model,coeficientes,">=",0)
  }

  #Resolvemos
  solve(model)
  sol= get.objective(model) #valor óptimo función objetivo
  return(sol)
}

```

- ❖ A21. Función para aplicar el método con frontera de referencia a todas las DMUs y calcular la eficiencia por Comunidad Autónoma (programación propia en R)

```
efic_CCR_referencia=function(n,datos_entradas,datos_salidas,datos_entradas_referencia, datos_salidas_referencias) {  
  
  eficiencia = matrix(NA , nrow=n, ncol=1, byrow=TRUE)  
  colnames(eficiencia)<-c("Eficiencia")  
  
  DMUS = c(1:n)  
  
  for (i in 1:n){  
    DMU=i  
    DMUS[i]=paste0("DMU",i)  
  
  eficiencia[i,1]=func_dea_CCR_referencia(DMU,datos_entradas,datos_salidas,datos_entradas_referencia,datos_salidas_referencia)  
  }  
  
  rownames(eficiencia)=DMUS  
  return(eficiencia)  
}
```