



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

**ANÁLISIS DE EFICIENCIA DE BIBLIOTECAS PÚBLICAS MEDIANTE
ANÁLISIS ENVOLVENTE DE DATOS (DEA): UN ESTUDIO EN
CASTILLA Y LEÓN**

Alumno:

Brenda Virginia Moreno Gómez

Tutor:

Jesús Alberto Tapia García

**ANÁLISIS DE EFICIENCIA DE BIBLIOTECAS PÚBLICAS
MEDIANTE ANÁLISIS ENVOLVENTE DE DATOS (DEA): UN
ESTUDIO EN CASTILLA Y LEÓN**

Brenda Virginia Moreno Gómez

*Dedicado a mis padres, con
todo mi amor: gracias por su
apoyo incondicional y la firme
convicción de que puedo
lograr cualquier cosa que me
proponga. Este logro es
también suyo*

Agradecimientos

Este logro no habría sido posible sin el apoyo incondicional y la guía constante de mis padres, a quienes dedico mi más profundo agradecimiento. A mis amigos y a mi pareja, gracias por creer en mí y acompañarme en cada paso del camino. A la universidad, a mis profesores y a todos los que contribuyeron a mi formación académica y personal, gracias por brindarme las herramientas y el conocimiento necesarios para alcanzar mis metas.

Gracias

Resumen

Este Trabajo de Fin de Grado evalúa la eficiencia de las bibliotecas públicas de Castilla y León durante el año 2020 mediante la metodología DEA (Análisis Envolvente de Datos). Se ha seleccionado dicho año por su carácter excepcional debido a la pandemia de COVID-19, lo que permite analizar el desempeño de las bibliotecas en un contexto de fuertes restricciones. Los datos utilizados provienen del Anuario Estadístico de Castilla y León.

Tras un análisis exploratorio, se optó por un modelo con dos inputs (gastos corrientes y número de bibliotecas) y un output (número de visitantes). Aunque inicialmente se consideró incluir más variables, la limitada muestra de nueve provincias restringió las posibilidades del modelo. Se aplicaron los modelos CCR y BCC para distinguir entre eficiencia técnica pura y eficiencia de escala.

Los resultados revelan notables diferencias entre provincias. Solo Burgos y Soria alcanzan una eficiencia plena (1.00), mientras que Segovia presenta la menor eficiencia (0.39), lo que indica que podría reducir sus recursos en un 61% sin afectar el número de visitantes. Seis provincias muestran rendimientos crecientes a escala, lo que sugiere que podrían beneficiarse de una expansión. Por el contrario, Valladolid presenta rendimientos decrecientes, posiblemente debido a un tamaño excesivo. Burgos se posiciona como referente para todas las provincias ineficientes.

Las recomendaciones principales son: fomentar la expansión en provincias con rendimientos crecientes siguiendo el modelo de Burgos, reconsiderar la escala operativa de Valladolid, y realizar una revisión profunda de la gestión en Segovia.

Palabras clave: Análisis Envolvente de Datos, Benchmark, Cargas factoriales, Eficiencia de escala, Eficiencia técnica, Eficiencia técnica global, Eficiencia técnica pura, Frontera de eficiencia, Holguras, Modelo BCC, Modelo CCR, Multicolinealidad, Orientación input, Orientación output, Programación lineal, Rendimientos a escala, Rendimientos constantes a escala, Rendimientos crecientes a escala, Rendimientos decrecientes a escala, Unidad de Decisión, Análisis de Componentes Principales.

Abstract

This Final Degree evaluates the efficiency of public libraries in Castilla y León during the year 2020 using the DEA (Data Envelopment Analysis) methodology. The year was chosen due to its exceptional nature caused by the COVID-19 pandemic, offering a unique opportunity to assess library performance under significant restrictions. The data used comes from the Statistical Yearbook of Castilla y León.

Following an exploratory analysis, a model was selected with two inputs (current expenditures and number of libraries) and one output (number of visitors). Although more variables were initially considered, the limited sample of nine provinces constrained the model's complexity. Both CCR and BCC models were applied to distinguish between pure technical efficiency and scale efficiency.

The results reveal significant differences among provinces. Only Burgos and Soria achieved full efficiency (1.00), while Segovia showed the lowest efficiency (0.39), suggesting it could reduce its resources by 61% without affecting visitor numbers. Six provinces exhibit increasing returns to scale, indicating potential benefits from expansion. In contrast, Valladolid shows decreasing returns, possibly due to excessive size. Burgos serves as a benchmark for all inefficient provinces.

The main recommendations are: provinces with increasing returns should consider expanding following Burgos's model; Valladolid should reassess its operational scale; and Segovia requires a thorough review of its management practices.

Keywords : Data Envelopment Analysis, Benchmark, Factor loadings, Scale efficiency, Technical efficiency, Overall technical efficiency, Pure technical efficiency, Efficiency frontier, Slacks, BCC model, CCR model, Multicollinearity, Input orientation, Output orientation, Linear programming, Returns to scale, Constant returns to scale, Increasing returns to scale, Decreasing returns to scale, Decision Making Unit, Principal Component Analysis.

Índice general

RESUMEN	5
ABSTRACT	6
ÍNDICE DE FIGURAS.....	9
ÍNDICE DE TABLAS.....	10
CAPÍTULO 1. INTRODUCCIÓN	11
1.1. Objetivos del estudio.....	12
CAPÍTULO 2. FUNDAMENTOS TEÓRICOS DEL ANÁLISIS ENVOLVENTE DE DATOS (DEA)	14
2.1. Eficiencia técnica, asignativa y económica	15
2.2. Evolución de los métodos de análisis de eficiencia	17
2.3. Modelo CCR: rendimientos constantes a escala	19
2.3.1. Formulación del problema primal.....	19
2.3.2. Transformación hacia programación lineal.....	20
2.3.3. Interpretación	21
2.3.4. Supuestos del modelo CCR	21
2.4. Modelo BCC: rendimientos variables a escala.....	22
2.4.1. Formulación matemática del modelo.....	22
2.4.2. Descomposición de la eficiencia técnica	23
2.4.3. Eficiencia de escala.....	23
2.5. Orientación input vs output.....	24
2.5.1. Elección de Inputs y Outputs en el Análisis Envolvente de Datos	24
2.5.1. Orientación input	26
2.5.2. Orientación output	27
2.5.3. Selección de orientación para el presente estudio	27

2.6. Holguras y análisis de mejoras potenciales	27
2.7. Benchmarking y unidades de referencia	29
CAPÍTULO 3. ANÁLISIS Y DISEÑO	30
3.1. Fuentes de datos	31
3.2. Análisis exploratorio de las variables	32
3.2.1. Análisis de correlaciones	32
3.2.2. Análisis de Componentes Principales (PCA).....	33
3.3. Restricciones estadísticas en la selección de variables	34
3.4. Criterios para la selección de variables	35
3.5. Especificación del modelo final	36
3.6. Herramientas de software utilizadas	38
CAPÍTULO 4. RESULTADOS DEL ANÁLISIS DE EFICIENCIA	40
4.1. Resultados de eficiencia técnica.....	41
4.2. Análisis de rendimientos de escala	42
4.3. Benchmarks y unidades de referencia.....	42
4.4. Análisis de holguras	43
4.5. Implicaciones prácticas y recomendaciones	44
CAPÍTULO 5. CONCLUSIONES	46
GLOSARIO	48
BIBLIOGRAFÍA.....	52
APÉNDICE A.....	54

Índice de figuras

Ilustración 1. Representación gráfica de la eficiencia técnica según Koopmans	15
Ilustración 2. Eficiencia asignativa: minimización del coste para un nivel dado de producción	16
Ilustración 3. Resultados de eficiencia técnica para las bibliotecas públicas provinciales de Castilla y León	41
Ilustración 4. Benchmarks para las provincias ineficientes (modelo CCR).....	42
Ilustración 5. Frecuencia de referencias para las provincias eficientes.....	43
Ilustración 6. Holguras en el modelo CCR orientado a input	43
Ilustración 7. Holguras en el modelo BCC orientado a input	44

Índice de tablas

Tabla 1. Principales correlaciones entre variables bibliotecarias	32
Tabla 2. Resultados del Análisis de Componentes Principales	33
Tabla 3. Cargas de las variables en los tres primeros componentes principales.....	33
Tabla 4. Combinaciones válidas según el número de inputs y outputs seleccionados	35

Capítulo 1. Introducción

Anualmente, el Servicio de Bibliotecas y Biblioteca de Castilla y León remite una recopilación detallada de datos estadísticos al *Anuario Estadístico de Castilla y León*, lo que permite obtener una descripción precisa del estado y evolución del sistema bibliotecario en la comunidad. Esta información incluye indicadores clave como el número de visitantes, usuarios registrados, fondos bibliográficos disponibles, dotación económica, préstamos realizados y materiales incorporados a través del Depósito Legal.

El anuario recoge información desde el año 2011 hasta 2020. No obstante, este estudio se centrará específicamente en el año 2020, un periodo marcado por el impacto de la pandemia de la COVID-19, que provocó el cierre temporal de las bibliotecas y su posterior reapertura bajo estrictas medidas. Este contexto único permite evaluar la resiliencia del sistema y su capacidad de adaptación, evidenciada en la rápida digitalización de servicios y desarrollo de nuevas modalidades de atención.

1.1. Objetivos del estudio

El objetivo general de este Trabajo de Fin de Grado es medir la eficiencia técnica de los sistemas bibliotecarios de Castilla y León durante el año 2020, mediante la aplicación de la metodología de Análisis Envolvente de Datos (DEA), con el fin de identificar patrones de eficiencia, establecer benchmarks y proponer recomendaciones específicas para la optimización de recursos y mejora de servicios.

Para llegar a este objetivo general, este trabajo se estructura principalmente en los siguientes puntos:

Desarrollar un marco conceptual y metodológico para técnicas de Análisis Envolvente de Datos en el contexto de las bibliotecas públicas, adaptado a las particularidades del sistema bibliotecario de Castilla y León.

Análisis multivariante de datos bibliotecario: Implementar técnicas estadísticas avanzadas (análisis de correlaciones, componentes principales) para identificar las variables más representativas del desempeño bibliotecario, reduciendo la dimensionalidad del conjunto de datos disponible en el Anuario Estadístico de Castilla y León sin perder información esencial.

Especificación óptima del modelo de eficiencia Determinar la combinación más apropiada de inputs (recursos: presupuestos, infraestructura, personal) y outputs (servicios: visitantes, préstamos, actividades culturales) que permita un análisis DEA estadísticamente válido y metodológicamente robusto.

Aplicar los diferentes modelos DEA (CCR y BCC) para evaluar la eficiencia técnica global, la eficiencia técnica pura y la eficiencia de escala de cada provincia.

Identificación de mejores prácticas y unidades de referencia para las provincias menos eficientes, señalando las unidades de referencia que pueden servir de modelo de mejores prácticas.

Analizar las causas de las ineficiencias detectadas, diferenciando entre factores controlables y no controlables por los gestores bibliotecarios.

El trabajo se estructura en cinco capítulos. En primer lugar, se presentan los fundamentos teóricos del análisis DEA, seguidos de una descripción del contexto bibliotecario de Castilla y León. Posteriormente, se realiza un análisis exploratorio de los datos, destacando las variables más relevantes para el estudio. A continuación, se aplican y comparan los modelos DEA seleccionados, y finalmente, se extraen conclusiones orientadas a optimizar la gestión bibliotecaria en cada una de las nueve provincias de la comunidad.

Capítulo 2. Fundamentos teóricos del análisis envolvente de datos (DEA)

2.1. Eficiencia técnica, asignativa y económica

Entendemos por eficiencia la capacidad de transformar recursos disponibles en resultados valorables. La literatura económica distingue principalmente tres tipos de eficiencia: técnica, asignativa y económica (también denominada eficiencia global)

Eficiencia técnica

La eficiencia técnica es la capacidad de una unidad para obtener el máximo output posible a partir de un conjunto dado de Inputs (Orientación Output) o bien utilizar el mínimo de recursos para obtener un nivel determinado de output (Orientación Input).

Koopmans (1951; p. 60) proporcionó una definición de lo que llamamos eficiencia técnica, dado un vector de inputs $x \in R_+^m$ y un vector de outputs $y \in R_+^s$, la eficiencia técnica implica que no existe ningún otro vector (x', y') /

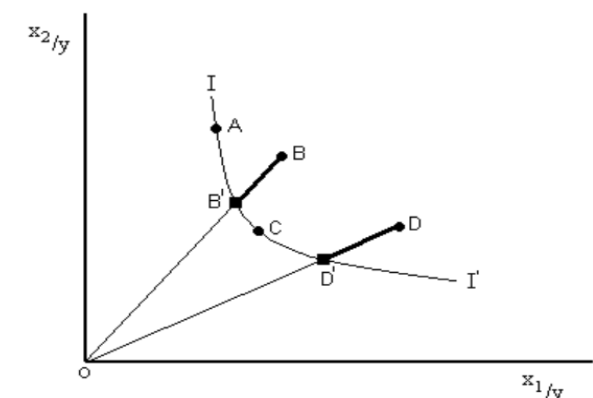
$$x' \geq x$$

$$y' \geq y$$

y al menos una de estas desigualdades sea estricta.

Es decir, es eficiente si no existe otro vector de inputs que utilice igual o menos inputs u otro vector que produzca más outputs.

Ilustración 1. Representación gráfica de la eficiencia técnica según Koopmans



La figura representa gráficamente el concepto de eficiencia técnica según la definición de Koopmans (1951). Los puntos A, B, C y D se encuentran sobre la frontera de posibilidades de producción (FPP), lo que indica que son técnicamente eficientes ya que no es posible mejorar en un input sin empeorar en otro. En cambio, los puntos como I o I', situados por debajo de la frontera, representan situaciones ineficientes, ya que existen combinaciones alternativas (como B o D) que permiten obtener más output con los mismos o incluso menos inputs.

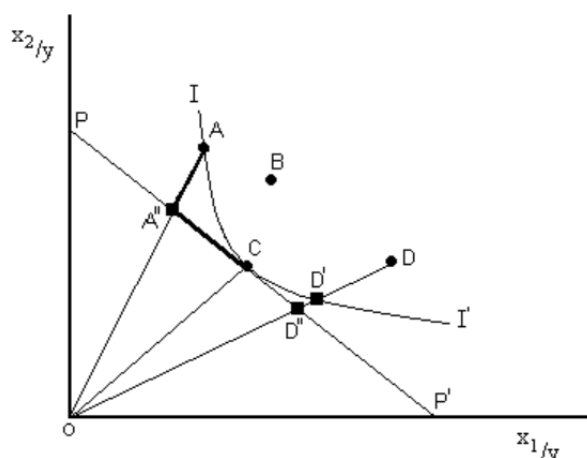
Este gráfico ilustra visualmente que una unidad de decisión (como una biblioteca) es eficiente si se encuentra sobre la frontera de producción, y es ineficiente si se encuentra por debajo de ella, ya que podría mejorar su desempeño sin necesidad de más recursos.

Eficiencia asignativa

La eficiencia asignativa, denominada también como eficiencia de precios, se refiere a la capacidad de una unidad productiva para utilizar los inputs en proporciones óptimas desde el punto de vista de los costes, ingresos o beneficios.

Matemáticamente, para un vector de precios de inputs $w \in R_+^m$ y una función de producción $f(x)$, la eficiencia asignativa implica que se minimiza el coste wx sujeto a la restricción $f(x) \geq y$.

Ilustración 2. Eficiencia asignativa: minimización del coste para un nivel dado de producción



En la ilustración, los ejes (x_1/y , x_2/y) representan la cantidad relativa de dos inputs necesarios para producir una unidad de output. Las curvas (isocuantas) muestran combinaciones de inputs que

producen lo mismo. Las líneas rectas (isocostes) representan combinaciones con el mismo coste. La eficiencia se alcanza donde ambas se **tocan** (punto **C**), ya que es la opción más económica. Otros puntos como **A''** o **D''** son ineficientes porque podrían lograr lo mismo con menor coste.

Eficiencia económica

La eficiencia económica o global combina la eficiencia técnica y la asignativa. Una unidad productiva es económicamente eficiente cuando produce una cantidad determinada de output al mínimo coste posible, lo que requiere ser tanto técnica como asignativa eficiente.

La relación entre estos tres tipos de eficiencia puede expresarse como:

$$\text{Eficiencia económica} = \text{Eficiencia técnica} \times \text{Eficiencia asignativa}$$

Esta descomposición, propuesta originalmente por Farrell (1957), permite identificar las fuentes de ineficiencia global, distinguiendo entre la incapacidad para operar en la frontera de producción (ineficiencia técnica) y la utilización de una combinación subóptima de inputs dados sus precios relativos (ineficiencia asignativa).

2.2. Evolución de los métodos de análisis de eficiencia

El desarrollo de técnicas para la medición de la eficiencia ha seguido dos trayectorias metodológicas principales: los enfoques paramétricos y los enfoques no paramétricos. Cada uno de estos enfoques presenta características, ventajas y limitaciones específicas que condicionan su idoneidad para diferentes contextos de aplicación.

- Los **métodos paramétricos** se caracterizan por especificar una forma funcional particular para la relación entre inputs y outputs, cuyos parámetros se estiman mediante técnicas econométricas convencionales. El método paramétrico más ampliamente utilizado es el Análisis de Fronteras Estocásticas (SFA), desarrollado simultáneamente por Aigner, Lovell y Schmidt (1977) y Meeusen y van den Broeck (1977).

El SFA se basa en la especificación de una función de producción estocástica de la forma:

$$y_i = f(x_i; B) \times \exp(v_i - u_i)$$

donde y_i representa el output de la unidad i , x_i el vector de inputs, β el vector de parámetros a estimar, v_i el término de error aleatorio (distribuido normalmente), y u_i el término de ineficiencia técnica (con distribución semi-normal, exponencial o gamma truncada).

Las **ventajas** de los métodos paramétricos incluyen la posibilidad de realizar contrastes de hipótesis estadísticas, construir intervalos de confianza para las estimaciones de eficiencia, y calcular elasticidades de producción. Sin embargo, presentan **limitaciones** importantes: requieren supuestos restrictivos sobre la forma funcional y la distribución de los errores, pueden ser sensibles a la especificación del modelo, y resultan complejos de implementar en presencia de múltiples outputs.

- Los **métodos no paramétricos** construyen la frontera de eficiencia directamente a partir de los datos observados, mediante técnicas de programación matemática, sin imponer restricciones sobre la forma funcional de la tecnología de producción.

El método no paramétrico más utilizado es el Análisis Envolvente de Datos (DEA), desarrollado por Charnes, Cooper y Rhodes (1978) como extensión de los trabajos de Farrell (1957). El DEA construye una frontera de eficiencia lineal por tramos que envuelve todas las observaciones, identificando las unidades eficientes como aquellas que definen la frontera.

Las **ventajas** de los métodos no paramétricos incluyen su flexibilidad (no requieren supuestos sobre formas funcionales), su capacidad para manejar múltiples inputs y outputs simultáneamente, y su robustez ante especificaciones incorrectas del modelo.

Entre sus **limitaciones** se encuentran la ausencia de un componente estocástico (todos los desvíos de la frontera se atribuyen a ineficiencia), la sensibilidad a observaciones atípicas, y la imposibilidad de realizar inferencia estadística formal.

Selección metodológica para el contexto bibliotecario

En el ámbito específico de la evaluación de bibliotecas públicas, los métodos no paramétricos, y en particular el DEA, presentan ventajas significativas:

- **Capacidad para múltiples outputs:** Las bibliotecas producen simultáneamente diversos servicios (préstamos, consultas, actividades culturales, servicios digitales) que son difíciles de agregar en una medida única.
- **Ausencia de forma funcional conocida:** No existe consenso teórico sobre la función de producción apropiada para caracterizar la tecnología bibliotecaria.
- **Dificultades de valoración:** Muchos inputs (tiempo voluntario) y outputs (beneficio social de la lectura) son difíciles de valorar monetariamente.
- **Tamaño muestral reducido:** Con frecuencia, los estudios bibliotecarios trabajan con muestras pequeñas donde los métodos no paramétricos pueden ser más apropiados.

2.3. Modelo CCR: rendimientos constantes a escala

El desafío fundamental que abordaron Charnes, Cooper y Rhodes en 1978 consistía en desarrollar un marco matemático capaz de medir comparativamente el desempeño de organizaciones que utilizan diversos recursos para generar múltiples servicios simultáneamente.

2.3.1. Formulación del problema primal

Para una biblioteca provincial específica (denominada DMU), el objetivo consiste en determinar qué combinación de pesos maximiza su ratio de eficiencia. Matemáticamente, esto se expresa como un problema de optimización fraccional:

$$\text{Max } h_0 = \frac{\sum_{r=1}^S u_r y_{r0}}{\sum_{i=1}^m v_i x_{i0}}$$

Sujeto a:

$$\frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1 ; j = 1, \dots, n \text{ donde } u_r, v_i \geq 0$$

$$\forall r \ r = 1 \dots s ; \forall i \ i = 1 \dots m$$

Cada DMU_j utiliza m inputs para producir s outputs.

x_{ij} es la cantidad del input i utilizado por la DMU_j , donde $i=1,2,\dots,m$ $i=1,2,\dots,m$.

y_{rj} es la cantidad del output r producido por la DMU_j , donde $r=1,2,\dots,s$

v_i es el peso asignado al input i

u_r es el peso asignado al input r

Esta formulación busca los pesos más favorables para la DMU bajo evaluación, sujeto a la restricción de que ninguna DMU (incluyendo la propia DMU_0) obtenga una puntuación de eficiencia superior a la unidad con esos mismos pesos.

2.3.2. Transformación hacia programación lineal

La formulación fraccional presenta dificultades computacionales que se resuelven mediante la transformación de Charnes-Cooper, que convierte el problema en uno de programación lineal equivalente:

Modelo transformado (orientación input):

$$\min \vartheta$$

Sujeto a:

$$\sum_{j=1}^m \lambda_j x_{ij} \leq \vartheta x_{i0} ; i = 1..m$$

$$\sum_{j=1}^m \lambda_j y_{rj} \leq y_{r0} ; r = 1..s$$

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

Donde:

- θ representa la puntuación de eficiencia de la DMU₀
- λ_j son las variables de intensidad que definen la combinación convexa de DMUs eficientes

2.3.3. Interpretación

Una DMU se considera eficiente si y si solo si $\theta^*=1$ y todas las holguras son cero.

La **solución óptima θ^*** proporciona la puntuación de eficiencia de la DMU₀:

Si $\theta^* = 1$: la DMU₀ es eficiente en el modelo CCR

Si $\theta^* < 1$: la DMU₀ es ineficiente, y $(1-\theta^*)$ representa el porcentaje de reducción proporcional posible en todos los inputs manteniendo los outputs constantes

El valor de θ^* indica qué proporción de los inputs podría reducirse sin afectar la producción actual.

Esta formulación matemática establece los cimientos teóricos que permiten evaluar objetivamente la eficiencia relativa de las bibliotecas públicas provinciales de Castilla y León, proporcionando tanto medidas cuantitativas de desempeño como identificación específica de mejores prácticas a emular.

2.3.4. Supuestos del modelo CCR

El modelo CCR se basa en varios supuestos fundamentales:

- **Rendimientos constantes a escala:** La tecnología de producción exhibe rendimientos constantes, de manera que escalamientos proporcionales de inputs generan escalamientos proporcionales de outputs.

Libre disposición: Es siempre posible producir menos outputs con los mismos inputs, o producir los mismos outputs con más inputs.

Convexidad: El conjunto de posibilidades de producción es convexo.

Homogeneidad: Todas las DMUs emplean la misma tecnología de producción y persiguen objetivos similares.

2.4. Modelo BCC: rendimientos variables a escala

El modelo BCC, desarrollado por Banker, Charnes y Cooper (1984), constituye una generalización del modelo CCR que permite la existencia de rendimientos variables a escala (Variable Returns to Scale, VRS). Esta extensión reconoce que la eficiencia puede variar según la escala de operación de las unidades evaluadas.

2.4.1. Formulación matemática del modelo

El modelo BCC mantiene la estructura básica del CCR pero añade una restricción adicional de convexidad que modifica fundamentalmente la forma de la frontera de eficiencia. Formalmente, la versión primal del modelo BCC orientado a input se plantea como:

Modelo BCC (orientación input):

$$\min \theta$$

Sujeto a:

$$\sum_{j=1}^n \lambda_j x_{ij} \leq \theta x_{oj} ; i = 1..m$$

$$\sum_{i=1}^n \lambda_i x_{ij} \leq y_{oj} ; j = 1..s$$

$$\sum_{i=1}^n \lambda_i \leq 1$$

$$\lambda_j \geq 0 ; i = 1, \dots, n$$

Restricción adicional

$$\sum_{i=1}^n \lambda_i \leq 1$$

Es una restricción de convexidad que garantiza que las DMUs tengan un tamaño similar, impidiendo que unidades pequeñas se comparen con versiones escaladas de unidades grandes, o viceversa.

2.4.2. Descomposición de la eficiencia técnica

Una de las principales contribuciones del modelo BCC es que permite descomponer la eficiencia técnica global (medida por el modelo CCR) en dos componentes:

Eficiencia técnica pura: Medida por el modelo BCC, refleja exclusivamente la eficiencia operativa de la DMU sin considerar su escala de operación.

Eficiencia de escala: Indica si la DMU opera a una escala óptima (rendimientos constantes) o subóptima (rendimientos crecientes o decrecientes).

Matemáticamente, esta descomposición se expresa como:

Eficiencia técnica global (CCR) = Eficiencia técnica pura (BCC) × Eficiencia de escala

Esta descomposición resulta particularmente valiosa en el contexto bibliotecario, donde coexisten unidades de diferentes tamaños (desde pequeñas bibliotecas municipales hasta grandes bibliotecas provinciales) que pueden enfrentar diferentes condiciones operativas según su escala.

2.4.3. Eficiencia de escala

La eficiencia de escala (SE) cuantifica qué tan cercana está una DMU a su tamaño óptimo de operación. Se calcula como el cociente entre la eficiencia CCR y la eficiencia BCC:

$$SE = \theta_{CCR} / \theta_{BCC}$$

Donde:

θ_{CCR} es la puntuación de eficiencia según el modelo CCR

θ_{BCC} es la puntuación de eficiencia según el modelo BCC

Basándose en la eficiencia de escala, las DMUs pueden clasificarse en tres categorías:

1. **Rendimientos constantes a escala (CRS):** Eficiencia de escala = 1

La DMU opera en su escala óptima

Incrementos proporcionales en inputs generan incrementos proporcionales en outputs.

2. **Rendimientos crecientes a escala (IRS):** Eficiencia de escala < 1

La DMU es demasiado pequeña para su entorno operativo

Incrementos en inputs generan incrementos más que proporcionales en outputs

La DMU se beneficiaría de expandir sus operaciones.

3. **Rendimientos decrecientes a escala (DRS):** Eficiencia de escala < 1

La DMU es demasiado grande para operar eficientemente

Incrementos en inputs generan incrementos menos que proporcionales en outputs

La DMU se beneficiaría de reducir su escala operativa

Identificación del tipo de rendimientos a escala

Para determinar si una DMU ineficiente en escala presenta rendimientos crecientes o decrecientes, se resuelve un modelo DEA adicional con la restricción $\sum_{j=1}^n \lambda_j \leq 1$ (rendimientos no crecientes a escala):

Si la eficiencia obtenida es igual a la del modelo BCC: la DMU presenta rendimientos decrecientes a escala

Si la eficiencia obtenida es diferente: la DMU presenta rendimientos crecientes a escala

2.5. Orientación input vs output

2.5.1. Elección de Inputs y Outputs en el Análisis Envolvente de Datos

La correcta selección de inputs y outputs es uno de los aspectos más cruciales en la aplicación del Análisis Envolvente de Datos (DEA), ya que determina la validez, interpretabilidad y utilidad de los

resultados obtenidos. Esta elección debe basarse en fundamentos teóricos sólidos, criterios estadísticos y consideraciones prácticas del contexto de estudio.

Fundamento teórico

En DEA, los inputs representan los recursos utilizados por las unidades de decisión (DMUs), mientras que los outputs reflejan los resultados o productos generados. La eficiencia técnica se evalúa comparando cómo cada DMU transforma sus inputs en outputs, en relación con las mejores prácticas observadas en el conjunto.

Según Charnes, Cooper y Rhodes (1978), los inputs y outputs deben cumplir las siguientes condiciones:

- **Relevancia económica o funcional:** Deben reflejar aspectos significativos del proceso productivo.
- **Controlabilidad:** Los inputs deben ser controlables por la gestión de la DMU.
- **Homogeneidad:** Las variables deben ser comparables entre DMUs.
- **No negatividad:** DEA requiere que todas las variables sean no negativas.
- **Variabilidad:** Deben existir diferencias entre DMUs para que el modelo pueda discriminar entre eficientes e ineficientes.

Criterios teóricos

- Los inputs deben representar recursos clave (económicos, humanos, físicos).
- Los outputs deben reflejar el impacto o resultado del servicio prestado.

Criterios estadísticos

- Evitar multicolinealidad entre variables.
- Mantener una proporción adecuada entre el número de DMUs y el número total de variables (regla de Banker y Morey: $n \geq \max\{\frac{m+s}{3}, 3\}$).

Criterios prácticos

- Disponibilidad y fiabilidad de los datos.
- Interpretabilidad de los resultados por los gestores.
- Sensibilidad al contexto específico (por ejemplo, impacto de la pandemia en bibliotecas).
- Implicaciones de una mala elección

Una selección inadecuada puede llevar a:

- Resultados triviales (todas las DMUs eficientes).
- Falta de discriminación.
- Interpretaciones erróneas.
- Recomendaciones poco útiles para la gestión.

Por ello, la elección debe ser cuidadosamente justificada y, cuando sea posible, respaldada por análisis exploratorios como correlaciones, PCA o revisión bibliográfica.

2.5.2. Orientación input

La orientación input busca minimizar los inputs empleados manteniendo al menos el mismo nivel de outputs. Esta orientación resulta apropiada cuando:

- Los gestores tienen mayor control sobre los recursos que sobre los resultados
- Existe presión presupuestaria para reducir costes operativos
- El objetivo principal consiste en identificar recursos excedentes o mal utilizados.
- Los outputs están determinados por la demanda externa y resultan difíciles de modificar a corto plazo.

En el contexto bibliotecario de 2020, la orientación input resulta especialmente pertinente dado que las bibliotecas enfrentaron simultáneamente restricciones presupuestarias y una demanda reducida debido a las limitaciones sanitarias.

2.5.3. Orientación output

La orientación output busca maximizar los outputs producidos sin incrementar los inputs empleados. Esta orientación es apropiada cuando:

- Los recursos están predeterminados o resultan difíciles de modificar
- El objetivo principal consiste en expandir los servicios prestados
- Existe capacidad instalada infrautilizada
- La mejora del servicio constituye una prioridad política

En el contexto de bibliotecas públicas, una orientación input puede resultar más apropiada si el objetivo principal es identificar posibilidades de optimización de recursos en un contexto de restricción presupuestaria. Por otro lado, una orientación output podría ser más adecuada si el interés está en identificar el potencial de expansión de servicios manteniendo los recursos existentes, especialmente en el ámbito digital donde las bibliotecas tuvieron que innovar rápidamente.

2.5.4. Selección de orientación para el presente estudio

Para este análisis se adopta la orientación input por las siguientes razones contextuales:

- **Restricciones presupuestarias:** Las administraciones públicas españolas enfrentaron presiones fiscales significativas durante 2020
- **Demanda reducida:** Las restricciones sanitarias limitaron la capacidad de las bibliotecas para expandir sus servicios
- **Control gerencial:** Los gestores bibliotecarios tienen mayor capacidad de influir sobre los recursos empleados que sobre la demanda ciudadana
- **Coherencia temporal:** La orientación input permite identificar eficiencias operativas relevantes para el período post-pandemia

2.6. Holguras y análisis de mejoras potenciales

El análisis DEA proporciona información no solo sobre las puntuaciones de eficiencia, sino también sobre las mejoras específicas necesarias para que una DMU ineficiente alcance la frontera eficiente.

Concepto de holguras

Las holguras (slacks) representan las mejoras adicionales posibles en inputs u outputs después de aplicar la mejora radial indicada por la puntuación de eficiencia. Formalmente, las holguras se definen como:

- **Holguras de Inputs (s_i^-):** Indican cuánto se podrían reducir los inputs sin afectar la producción actual.
- **Holguras de Outputs (s_r^+):** Indican cuánto se podrían aumentar los outputs sin necesidad de más inputs.

Identificación de holguras

Para identificar las holguras, se resuelve un problema de programación lineal secundario tras obtener la puntuación de eficiencia óptima θ^* :

$$\max_{s^+, s^-} \sum_{i=1}^m s_i^- + \sum_{r=1}^s s_r^+$$

Sujeto a las restricciones:

$$\sum_{j=1}^n \lambda_j x_{ij} + s_i^- \leq \theta^* x_{io} ; i = 1..m$$

$$\sum_{j=1}^n \lambda_j y_{rj} + s_r^+ \leq y_{ro} ; r = 1..s$$

$$\lambda_j, s_i^-, s_r^+ \geq 0 ; \forall j, i, r$$

Eficiencia en sentido de Pareto-Koopmans

Una DMU se considera completamente eficiente (eficiente en sentido de Pareto-Koopmans) si y solo si $\theta^* = 1$ y todas las holguras son nulas. Esta condición garantiza que no es posible mejorar ningún input u output sin empeorar algún otro.

2.7. Benchmarking y unidades de referencia

Identificación de benchmarks

Los valores $\lambda_j^* > 0$ en la solución óptima identifican las DMUs que sirven como referencias (benchmarks) para una DMU ineficiente. Estas unidades de referencia definen la proyección de la DMU ineficiente sobre la frontera eficiente.

El conjunto de referencia para una DMU ineficiente puede interpretarse como una "DMU virtual" que combina las mejores prácticas observadas en las DMUs eficientes, proporcionando objetivos específicos y alcanzables para la mejora.

Frecuencia de referencias

La frecuencia con que una DMU eficiente aparece como referencia para otras DMUs ineficientes constituye un indicador de la robustez de sus prácticas y de su relevancia como modelo a emular. DMUs con alta frecuencia de referencias representan patrones de gestión especialmente valiosos para la transferencia de conocimiento.

Capítulo 3. Análisis y Diseño

3.1. Fuentes de datos

Los datos empleados en este estudio proceden del Anuario Estadístico de Castilla y León, publicación oficial que recopila anualmente información detallada sobre el funcionamiento de los servicios públicos en la comunidad autónoma. Específicamente, se han utilizado los datos correspondientes al ejercicio 2020, extraídos de los epígrafes 13.8.1 al 13.8.5 del citado anuario, que recogen información sobre bibliotecas públicas provinciales.

Variables disponibles

El conjunto inicial de variables consideradas incluye:

Variables de recursos (inputs potenciales):

- Bibliotecas públicas 2020
- Gastos en bienes corrientes y servicios 2020
- Fondo bibliográfico 2020
- Otras inversiones 2020
- Número de bibliotecas
- Fondo de libros en bibliotecas municipales 2020
- Fondo de publicaciones periódicas en bibliotecas municipales 2020
- Fondo audiovisual y electrónico en bibliotecas municipales 2020

Variables de servicios (outputs potenciales):

- Visitantes 2020
- Adultos inscritos 2020
- Infantiles inscritos 2020
- Nuevos usuarios adultos 2020

- Nuevos usuarios infantiles 2020
- Actividades culturales 2020
- Usuarios inscritos en bibliotecas municipales 2020
- Libros prestados en bibliotecas municipales 2020

Unidades de análisis

Las unidades de decisión (DMUs) del estudio son las nueve provincias que conforman la comunidad autónoma de Castilla y León: Ávila, Burgos, León, Palencia, Salamanca, Segovia, Soria, Valladolid y Zamora.

3.2. Análisis exploratorio de las variables

Antes de proceder a la selección final de variables para el modelo DEA, se realizó un análisis exploratorio de los datos disponibles mediante dos técnicas complementarias: análisis de correlaciones y análisis de componentes principales (PCA). Estos análisis permiten identificar relaciones entre variables y su importancia relativa, facilitando una selección informada de inputs y outputs.

3.2.1. Análisis de correlaciones

El análisis de correlaciones reveló fuertes relaciones entre diversas variables del sistema bibliotecario de Castilla y León. Las correlaciones más significativas (superiores a 0.9) se presentan en la Tabla 1:

Tabla 1. Principales correlaciones entre variables bibliotecarias

Variable1	Variable2	Correlación
Número de bibliotecas	Bibliotecas públicas 2020	0.9955159
Infantiles inscritos 2020	Adultos inscritos 2020	0.9722656
Libros prestados en bibliotecas municipales 2020	Visitantes 2020	0.9655051
Fondo de Libros en bibliotecas municipales 2020	Bibliotecas públicas 2020	0.9605948
Otras inversiones 2020	Fondo bibliográfico 2020	0.9452689
Total Usuarios Inscritos en bibliotecas municipales 2020	Nuevos usuarios adultos 2020	0.9383806
Nuevos usuarios infantiles 2020	Nuevos usuarios adultos 2020	0.9306882
Libros prestados en bibliotecas municipales 2020	Nuevos usuarios adultos 2020	0.9224162
Total Usuarios Inscritos en bibliotecas municipales 2020	Nuevos usuarios infantiles 2020	0.9184482

Estas correlaciones elevadas indican posibles redundancias entre variables, lo que sugiere la necesidad de seleccionar cuidadosamente aquellas que aporten información diferenciada al modelo.

3.2.2. Análisis de Componentes Principales (PCA)

Para complementar el análisis de correlaciones, se realizó un análisis de componentes principales que permitió identificar las dimensiones subyacentes en los datos. Los resultados principales se resumen en la Tabla 2:

Tabla 2. Resultados del Análisis de Componentes Principales

Componente	Desviación estándar	Proporción de σ^2	Proporción acumulada
PC1	32.905	0.6369	0.6369
PC2	15.751	0.1459	0.7829
PC3	129.749	0.09903	0.88188
PC4	0.91619	0.04938	0.93125
PC5	0.74958	0.03305	0.96430
PC6	0.62038	0.02264	0.98694
PC7	0.38640	0.00878	0.99573
PC8	0.26955	0.00427	100.000
PC9	1,78E-13	0.000e+00	1,00E+03

El primer componente principal (PC1) explica el 63.69% de la varianza total, lo que indica una fuerte estructura subyacente en los datos. Los tres primeros componentes explican conjuntamente el 88.19% de la varianza, proporcionando una representación adecuada de la información contenida en el conjunto original de variables.

Las cargas (loadings) de las variables en los tres primeros componentes principales se presentan en la Tabla 3:

Tabla 3. Cargas de las variables en los tres primeros componentes principales

Componente	Desviación estándar	Proporción de varianza	Proporción acumulada
Bibliotecas públicas 2020	0.26659057	0.22359354	-0.21479056
Visitantes 2020	0.28192029	-0.09972412	0.19834686
Adultos inscritos 2020	0.27666626	0.06733935	0.09397381
Infantiles inscritos 2020	0.28519960	0.15586107	0.01528859
Nuevos usuarios adultos 2020	0.28858387	-0.11557625	-0.11238602
Nuevos usuarios infantiles 2020	0.28203563	-0.07631511	-0.16897422
Gastos en bienes corrientes y servicios 2020	0.18960713	-0.15332708	0.27211851

Fondo bibliográfico 2020	0.22551895	-0.33274178	-0.03266222
Actividades culturales 2020	0.01067682	0.19448815	0.67570967
Otras inversiones 2020	0.16792294	-0.40195239	-0.02193260
Número de bibliotecas	0.26435394	0.20560819	-0.25867138
Fondo de Libros en bibliotecas municipales 2020	0.26863746	0.21847061	-0.18163841
Fondo de P. Periódicas en bibliotecas municipales 2020	0.22730832	0.32706153	0.20267268
Fondo Audiovisuales y Electrónicos en bibliotecas municipales 2020	0.24752624	0.08023267	0.35051120
Nuevos Usuarios Inscritos en bibliotecas municipales 2020	0.05547431	0.51869123	-0.19178530
Total Usuarios Inscritos en bibliotecas municipales 2020	0.26689819	-0.26772662	-0.13350650
Libros prestados en bibliotecas municipales 2020	0.29038153	-0.11960544	0.11684443

En el PC1, las variables con mayores cargas absolutas son:

- Libros prestados en bibliotecas municipales 2020 (0.29038153)
- Nuevos usuarios adultos 2020 (0.28858387)
- Infantiles inscritos 2020 (0.28519960)
- Visitantes 2020 (0.28192029)

Estas variables tienen un peso significativo en la dimensión principal que captura la mayor parte de la variabilidad del sistema bibliotecario. Por otro lado, en el PC2, destacan variables como "Nuevos Usuarios Inscritos en bibliotecas municipales 2020" y "Otras inversiones 2020", mientras que en el PC3 sobresale claramente "Actividades culturales 2020".

3.3. Restricciones estadísticas en la selección de variables

La construcción de un modelo DEA efectivo requiere un equilibrio entre el número de unidades evaluadas (DMUs) y la cantidad de variables incluidas (inputs y outputs). Este equilibrio es fundamental para garantizar la robustez estadística del modelo y su capacidad discriminadora entre unidades eficientes e ineficientes.

Para determinar el número máximo de variables que pueden incluirse en un modelo DEA con validez estadística, se aplica la siguiente regla generalmente aceptada en la literatura especializada:

$$nDMUs \geq \max\{m \cdot s, 3 \cdot (m+s)\}$$

Donde:

- DMU's = número de unidades evaluadas (9 provincias en nuestro caso)
- m = número de inputs
- s = número de outputs

En el contexto del presente estudio, con 9 DMUs correspondientes a las provincias de Castilla y León, las posibles combinaciones de inputs y outputs quedan limitadas según se muestra en la Tabla 4:

Tabla 4. Combinaciones válidas según el número de inputs y outputs seleccionados

Inputs (m)	Outputs (s)	Total ($m+s$)	$3(m+s)$	$m \times s$	$\text{Max}(m \times s, 3 \times (m+s))$	¿Válido?
2	1	3	9	2	9	Sí
1	2	3	9	2	9	Sí
2	2	4	12	4	12	No
3	1	4	12	3	12	No
1	3	4	12	3	12	No

Por tanto, con 9 DMUs, las únicas combinaciones estadísticamente válidas son:

1. Dos inputs y un output (2-1)
2. Un input y dos outputs (1-2)

Esta restricción ha condicionado necesariamente la selección de variables para nuestro modelo DEA.

3.4. Criterios para la selección de variables

Criterios teóricos

- **Relevancia conceptual:** Las variables seleccionadas deben representar adecuadamente los recursos empleados (inputs) y los servicios prestados (outputs) por el sistema bibliotecario.
- **Relación causal:** Debe existir una relación lógica de causalidad entre los inputs y outputs seleccionados, de manera que incrementos en los recursos conduzcan plausiblemente a incrementos en los servicios.
- **Controlabilidad:** Los inputs deben corresponder a recursos efectivamente controlables por los gestores bibliotecarios.

- **Comparabilidad:** Las variables deben ser comparables entre provincias, sin sesgos sistemáticos relacionados con diferencias en criterios de medición o contabilización.

Criterios estadísticos

- **Variabilidad:** Las variables deben presentar variabilidad suficiente entre DMUs para permitir la discriminación entre unidades eficientes e ineficientes.
- **Independencia:** Se prefieren variables con correlaciones moderadas para evitar redundancias informativas.
- **Disponibilidad:** Deben estar disponibles datos completos y fiables para todas las DMUs.
- **Poder discriminatorio:** La combinación de variables debe permitir que el modelo identifique un número razonable de DMUs eficientes (ni todas ni muy pocas).

Criterios contextuales

- **Relevancia COVID-19:** Las variables seleccionadas deben ser sensibles a las transformaciones operativas impuestas por la pandemia.
- **Significado práctico:** Los resultados deben ser interpretables y útiles para la toma de decisiones de gestión bibliotecaria.
- **Precedentes en la literatura:** Se considera la experiencia previa en estudios similares de eficiencia bibliotecaria.

3.5. Especificación del modelo final

Variables seleccionadas

La elección de inputs y outputs constituye una etapa crítica en la construcción de modelos DEA, ya que determina la validez y relevancia de los resultados obtenidos. En este estudio, la selección se ha basado en una combinación de criterios teóricos, estadísticos y contextuales, con especial atención al impacto de la pandemia de COVID-19 en el funcionamiento de las bibliotecas públicas.

Inputs seleccionados:

- **Gastos en bienes corrientes y servicios (2020):** Representa los recursos económicos operativos empleados por cada provincia. Incluye costes de personal, suministros, mantenimiento y adaptación a protocolos sanitarios durante la pandemia.
- **Número de bibliotecas:** Representa la infraestructura física disponible, determinando la capacidad de servicio y la cobertura territorial de cada provincia

Output seleccionado:

- **Visitantes (2020):** Representa los recursos presupuestarios operativos empleados anualmente por cada provincia en el funcionamiento de sus bibliotecas públicas. Su elección responde a su relevancia como medida de impacto social y a su sensibilidad frente a las restricciones sanitarias impuestas en 2020.

La combinación de estas variables cumple con los requisitos de validez estadística ($\text{número de DMUs} \geq \max\{m \cdot s, 3(m+s)\}$), evita problemas de multicolinealidad y permite una interpretación clara de los resultados. Además, se alinea con estudios previos en el ámbito de la eficiencia bibliotecaria.

Justificación teórica

- Este modelo captura la esencia de la función de producción bibliotecaria: la transformación de recursos presupuestarios e infraestructurales en servicios utilizados por la ciudadanía. Los gastos corrientes representan los recursos de funcionamiento (personal, suministros, servicios, mantenimiento), mientras que el número de bibliotecas representa la capacidad instalada y la accesibilidad geográfica del servicio.

Justificación empírica

El análisis de correlaciones muestra que:

- Los dos inputs seleccionados presentan una correlación moderada ($r = 0.734$), evitando problemas severos de multicolinealidad

- Ambos inputs muestran correlaciones significativas con el output ($r = 0.612$ y $r = 0.821$ respectivamente)
- La combinación permite discriminación adecuada entre provincias

Justificación contextual COVID-19

En el contexto de 2020, esta especificación resulta especialmente pertinente:

- **Gastos corrientes:** Incluyen los costes de adaptación a protocolos sanitarios, implementación de servicios digitales, y reorganización operativa necesaria durante la pandemia.
- **Número de bibliotecas:** Determinó la capacidad para mantener servicios distribuidos geográficamente respetando limitaciones de aforo y distanciamiento social.
- **Visitantes:** Refleja la capacidad de cada provincia para atraer y servir a los ciudadanos pese a las restricciones sanitarias y los cambios en hábitos de consumo cultural.

Normalización de variables

Para evitar problemas relacionados con diferentes escalas de medición, todas las variables fueron normalizadas dividiendo cada observación por el valor máximo de la variable correspondiente. Esta transformación preserva las relaciones de eficiencia mientras facilita la interpretación de los resultados.

3.6. Herramientas de software utilizadas

Para el desarrollo del presente estudio se ha utilizado el lenguaje de programación R (R Core Team, 2023), un entorno de software libre ampliamente empleado en análisis estadístico y científico. La elección de R se debe a su flexibilidad, potencia para el tratamiento de datos y disponibilidad de paquetes especializados para el análisis de eficiencia.

Las principales librerías utilizadas fueron:

readxl: para la importación de datos desde archivos Excel.

dplyr y **tidyr**: para la manipulación y transformación de datos.

stats: para análisis estadístico básico, incluyendo el análisis de componentes principales (PCA).

Benchmarking: paquete específico para la implementación de modelos DEA (Data Envelopment Analysis), tanto CCR como BCC, con orientación input y output.

ggplot2 y **gridExtra**: para la generación de gráficos y visualizaciones.

corrplot: para la representación gráfica de matrices de correlación.

writexl: para la exportación de resultados a archivos Excel.

Entre las funciones más relevantes utilizadas destacan:

dea(): función del paquete Benchmarking para calcular la eficiencia técnica bajo distintos supuestos de rendimientos a escala y orientaciones.

prcomp(): función base de R para realizar el análisis de componentes principales.

cor(): para calcular matrices de correlación entre variables.

scale(): para normalizar los datos antes del análisis multivariante.

cbind() y **as.data.frame()**: para combinar y estructurar los datos en tablas finales.

Capítulo 4. Resultados del análisis de eficiencia

4.1. Resultados de eficiencia técnica

Tras aplicar el análisis DEA a las bibliotecas públicas de Castilla y León utilizando el Modelo seleccionado, que considera como inputs "Gastos en bienes corrientes y servicios 2020" y "Número de bibliotecas", y como output "Visitantes 2020", se obtuvieron los siguientes resultados de eficiencia técnica, mostrados en la Ilustración 3:

Ilustración 3. Resultados de eficiencia técnica para las bibliotecas públicas provinciales de Castilla y León

Provincia	CCR (RCE)	BCC (RVE)	Eficiencia de Escala	Tipo de Rendimientos
Ávila	0.5451	0.5777	0.9435	Crecientes
Burgos	1.0000	1.0000	1.0000	Constantes
León	0.5292	0.6479	0.8168	Crecientes
Palencia	0.6057	0.8704	0.6959	Crecientes
Salamanca	0.7300	0.8238	0.8861	Crecientes
Segovia	0.3886	0.7443	0.5221	Crecientes
Soria	1.0000	1.0000	1.0000	Constantes
Valladolid	0.9153	1.0000	0.9153	Decrecientes
Zamora	0.5693	0.7634	0.7457	Crecientes

Nota: RCE = Rendimientos Constantes a Escala; RVE = Rendimientos Variables a Escala

Los resultados muestran que sólo dos provincias (Burgos y Soria) alcanzan la eficiencia técnica global (CCR = 1), lo que significa que operan en la escala óptima y utilizan eficientemente sus recursos. Valladolid, aunque no alcanza la eficiencia global, sí logra la eficiencia técnica pura (BCC = 1), lo que indica que su ineficiencia se debe exclusivamente a factores de escala y no a la gestión de sus recursos.

La provincia con menor eficiencia técnica global es Segovia (0.3886), seguida de León (0.5292) y Ávila (0.5451), lo que revela un amplio margen de mejora en la gestión de sus recursos bibliotecarios.

4.2. Análisis de rendimientos de escala

El análisis de los rendimientos a escala proporciona información valiosa sobre la escala operativa óptima de las bibliotecas provinciales. Como se observa en la Ilustración 3, la mayoría de las provincias (seis de nueve) presentan rendimientos crecientes a escala, lo que indica que un aumento proporcional en sus inputs generaría un incremento más que proporcional en su output. Esto sugiere que estas provincias podrían beneficiarse de una expansión en sus operaciones.

Burgos y Soria operan con rendimientos constantes a escala, lo que significa que trabajan en su escala óptima. Por otro lado, Valladolid presenta rendimientos decrecientes a escala, lo que indica que un incremento en sus inputs generaría un aumento menos que proporcional en sus outputs, sugiriendo que podría beneficiarse de una reducción en su escala operativa.

La **eficiencia de escala** (calculada como el cociente entre la eficiencia CCR y la eficiencia BCC) revela que Segovia (0.5221) y Palencia (0.6959) son las provincias con mayores problemas relacionados con la escala de operación, mientras que Ávila (0.9435) y Valladolid (0.9153) están operando más cerca de su escala óptima, aunque todavía con margen de mejora.

4.3. Benchmarks y unidades de referencia

Una de las ventajas del análisis DEA es que permite identificar unidades de referencia (benchmarks) para cada DMU ineficiente. La Ilustración 4 muestra los benchmarks para cada provincia ineficiente, junto con sus correspondientes valores lambda, que indican la importancia relativa de cada referencia.

Ilustración 4. Benchmarks para las provincias ineficientes (modelo CCR)

Provincia ineficiente	Benchmarks y valores lambda
Ávila	Burgos (0.09), Soria (0.98)
León	Burgos (0.21), Soria (1.13)
Palencia	Burgos (0.43)
Salamanca	Burgos (0.47), Soria (0.96)
Segovia	Burgos (0.26)
Valladolid	Burgos (0.97), Soria (0.70)
Zamora	Burgos (0.22), Soria (0.36)

La Ilustración 5 complementa este análisis mostrando la frecuencia con que cada provincia eficiente

sirve como referencia para otras, lo que puede interpretarse como un indicador de su solidez como modelo de mejores prácticas.

Ilustración 5. Frecuencia de referencias para las provincias eficientes

Provincia	Eficiencia	CCR	Veces referencia	Porcentaje
Burgos	1.0000		8	100%
Soria	1.0000		6	75%

Burgos emerge como el benchmark más importante, sirviendo como referencia para todas las provincias ineficientes (100%), seguido por Soria, que es referencia para el 75% de ellas. Esto sugiere que el modelo de gestión bibliotecaria de Burgos representa un caso ejemplar de buenas prácticas que podría servir como guía para la mejora del sistema bibliotecario en otras provincias.

4.4. Análisis de holguras

El análisis de holguras permite identificar los excesos en inputs o déficits en outputs que persisten incluso después de la reducción radial de inputs (en modelos orientados a input). La Ilustración 6 muestra las holguras detectadas en el modelo CCR.

Ilustración 6. Holguras en el modelo CCR orientado a input

Provincia	Holgura en "Gastos"	Holgura en "Número de bibliotecas"	Holgura en "Visitantes"
Ávila	0.0000	0.0000	0.0000
Burgos	0.0000	0.0000	0.0000
León	0.0000	0.0000	0.0000
Palencia	0.0087	0.0000	0.0000
Salamanca	0.0000	0.0000	0.0000
Segovia	0.0934	0.0000	0.0000
Soria	0.0000	0.0000	0.0000
Valladolid	0.0000	0.0000	0.0000
Zamora	0.0000	0.0000	0.0000

Las holguras en el modelo CCR son mínimas, indicando que la ineficiencia detectada es

principalmente de naturaleza radial. Solo Palencia y Segovia presentan pequeñas holguras en "Gastos en bienes corrientes y servicios", lo que sugiere un exceso específico en este input que debería reducirse adicionalmente a la reducción radial para alcanzar la frontera eficiente.

En el modelo BCC, sin embargo, se observan holguras más significativas (Ilustración 7), especialmente en "Gastos en bienes corrientes y servicios" para Palencia, Segovia y Zamora, y en "Número de bibliotecas" para Ávila, León y Salamanca.

Ilustración 7. Holguras en el modelo BCC orientado a input

Provincia	Holgura en "Gastos"	Holgura en "Número de bibliotecas"	Holgura en "Visitantes"
Ávila	0.0000	0.0163	0.0000
Burgos	0.0000	0.0000	0.0000
León	0.0000	0.1302	0.0000
Palencia	0.3118	0.0000	0.0000
Salamanca	0.0000	0.0580	0.0000
Segovia	0.5642	0.0000	0.0000
Soria	0.0000	0.0000	0.0000
Valladolid	0.0000	0.0000	0.0000
Zamora	0.1722	0.0000	0.0000

4.5. Implicaciones prácticas y recomendaciones

A partir del análisis realizado, se pueden extraer las siguientes implicaciones prácticas y recomendaciones para mejorar la eficiencia del sistema bibliotecario provincial de Castilla y León:

- **Para provincias con rendimientos crecientes a escala** (Ávila, León, Palencia, Salamanca, Segovia y Zamora), se recomienda una expansión de sus operaciones bibliotecarias, ya que esto generaría incrementos más que proporcionales en el número de visitantes. Esta expansión debería realizarse siguiendo el modelo de las provincias eficientes (Burgos y Soria).
- **Para Valladolid**, que presenta rendimientos decrecientes a escala, sería recomendable reevaluar su dimensionamiento, posiblemente reduciendo la escala de operaciones para alcanzar una mayor eficiencia global.
- **Segovia**, como la provincia con menor eficiencia técnica (0.3886), necesita una revisión profunda de su modelo de gestión bibliotecaria. Específicamente, debería reducir sus gastos

en bienes corrientes y servicios (donde presenta una holgura significativa de 0.5642 en el modelo BCC) mientras mantiene o incluso incrementa el número de visitantes.

- **León y Ávila**, también con bajos niveles de eficiencia, podrían beneficiarse de una reorganización de su red de bibliotecas, ya que presentan holguras en el "Número de bibliotecas" en el modelo BCC.
- **Burgos** emerge como el principal referente de mejores prácticas para el sistema bibliotecario provincial. Se recomienda un estudio detallado de su modelo de gestión para identificar las estrategias específicas que contribuyen a su alta eficiencia, tanto en la utilización de recursos como en su escala de operación.

Capítulo 5. Conclusiones

El análisis DEA ha permitido evaluar la eficiencia de las bibliotecas públicas provinciales de Castilla y León, identificando tanto las unidades eficientes como aquellas con margen de mejora, así como los factores específicos que contribuyen a la ineficiencia.

Los resultados revelan una heterogeneidad significativa en los niveles de eficiencia entre provincias, con valores que oscilan entre 0.3886 (Segovia) y 1.0000 (Burgos y Soria), lo que indica un potencial considerable para la mejora global del sistema bibliotecario regional.

La mayoría de las provincias operan con rendimientos crecientes a escala, lo que sugiere que políticas orientadas a la expansión de servicios bibliotecarios podrían tener un impacto positivo y más que proporcional en el número de visitantes, maximizando así el retorno social de la inversión en estos servicios culturales.

El análisis de benchmarks ha identificado a Burgos como el principal referente de mejores prácticas, seguido por Soria. La transferencia de conocimiento y la adaptación de estrategias de gestión desde estas provincias hacia las menos eficientes podría constituir un eje fundamental para la mejora global del sistema bibliotecario de Castilla y León.

Finalmente, el análisis de holguras ha permitido identificar aspectos específicos de mejora para cada provincia, principalmente relacionados con la gestión del gasto corriente y la organización de la red de bibliotecas, que complementan las medidas generales de mejora de eficiencia

Glosario

Análisis Envolvente de Datos (DEA): Metodología no paramétrica de programación lineal que construye una frontera de eficiencia empírica para evaluar el desempeño relativo de unidades organizacionales homogéneas (DMUs). Desarrollada originalmente por Charnes, Cooper y Rhodes (1978).

Benchmark: Unidad de decisión que exhibe mejores prácticas y sirve como referencia para unidades menos eficientes. En DEA, los benchmarks son las DMUs que conforman la frontera de eficiencia y definen los objetivos de mejora para las unidades ineficientes.

Cargas factoriales (loadings): Coeficientes que indican la correlación entre cada variable original y cada componente principal en el análisis PCA. Valores altos (en valor absoluto) indican que la variable contribuye significativamente a ese componente. También llamadas saturaciones factoriales.

Eficiencia de escala: Medida que indica si una DMU opera a su tamaño óptimo. Se calcula como el cociente entre la eficiencia técnica global (CCR) y la eficiencia técnica pura (BCC). Valores menores a 1 indican escala subóptima.

Eficiencia técnica: Capacidad de una unidad para obtener el máximo output posible con un conjunto dado de inputs (orientación output) o emplear el mínimo de inputs para un nivel determinado de outputs (orientación input).

Eficiencia técnica global: Medida de eficiencia obtenida mediante el modelo CCR, que asume rendimientos constantes a escala. Combina tanto eficiencia técnica pura como eficiencia de escala.

Eficiencia técnica pura: Medida de eficiencia obtenida mediante el modelo BCC, que considera rendimientos variables a escala. Refleja la eficiencia operativa independientemente de los efectos de escala.

Frontera de eficiencia: Línea o superficie que delimita el conjunto de posibilidades de producción, definida por las DMUs más eficientes. Las unidades sobre la frontera son eficientes; las ubicadas por debajo son ineficientes.

Holguras (slacks): Mejoras adicionales posibles en inputs u outputs después de aplicar la mejora radial indicada por la puntuación de eficiencia. Indican ajustes específicos necesarios más allá de cambios proporcionales. En orientación input, las holguras de input (s^-) representan reducciones adicionales posibles, mientras que las holguras de output (s^+) representan incrementos adicionales factibles.

Modelo BCC: Modelo DEA que permite rendimientos variables a escala, desarrollado por Banker, Charnes y Cooper (1984). Incorpora una restricción de convexidad ($\sum \lambda_j = 1$) que permite identificar diferentes tipos de rendimientos a escala.

Modelo CCR: Modelo DEA básico que asume rendimientos constantes a escala, desarrollado por Charnes, Cooper y Rhodes (1978). No incluye restricciones de convexidad en las variables de intensidad.

Multicolinealidad: Fenómeno estadístico que ocurre cuando dos o más variables explicativas en un modelo están altamente correlacionadas entre sí. En el contexto DEA, la multicolinealidad entre inputs o outputs puede reducir la capacidad discriminatoria del modelo y generar inestabilidad en las

estimaciones de eficiencia.

Orientación input: Perspectiva del análisis DEA que busca minimizar los inputs empleados manteniendo al menos el mismo nivel de outputs. Apropia cuando los gestores tienen mayor control sobre recursos que sobre resultados.

Orientación output: Perspectiva del análisis DEA que busca maximizar los outputs producidos sin incrementar los inputs empleados. Apropia cuando los recursos están predeterminados y el objetivo es expandir servicios.

Programación lineal: Técnica de optimización matemática que busca maximizar o minimizar una función objetivo lineal sujeta a restricciones lineales. Base computacional de los modelos DEA.

Rendimientos a escala: Relación entre cambios proporcionales en inputs y outputs. Pueden ser constantes (CRS), crecientes (IRS) o decrecientes (DRS), indicando diferentes comportamientos de la productividad según el tamaño operativo.

Rendimientos constantes a escala (CRS): Situación donde incrementos proporcionales en inputs generan incrementos proporcionales en outputs. Caracteriza el funcionamiento en escala óptima.

Rendimientos crecientes a escala (IRS): Situación donde incrementos proporcionales en inputs generan incrementos más que proporcionales en outputs. Indica que la DMU es demasiado pequeña y se beneficiaría de expandir operaciones.

Rendimientos decrecientes a escala (DRS): Situación donde incrementos proporcionales en inputs generan incrementos menos que proporcionales en outputs. Indica que la DMU es demasiado grande y sufre deseconomías de escala.

Unidad de Decisión (DMU): Entidad organizacional evaluada en el análisis DEA. Debe ser homogénea en términos de objetivos, tecnología y entorno operativo. En este estudio, las DMUs son las provincias de Castilla y León.

Análisis de Componentes Principales (PCA): Técnica estadística de reducción dimensional que transforma un conjunto de variables posiblemente correlacionadas en un conjunto menor de variables no correlacionadas llamadas componentes principales. Cada componente es una combinación lineal de las variables originales que captura la máxima varianza posible

Bibliografía

0 Datos estadísticos | Red de Bibliotecas | Junta de Castilla y León

0 Junta de Castilla y León. (2021). *Anuario de Estadísticas Culturales 2020*. Consejería de Cultura y Turismo.

0 Instituto Nacional de Estadística. (2021). *Estadística de Bibliotecas 2020*.

0 https://www.uv.es/vcoll/libros/2006_evaluacion_eficiencia_DEA.pdf

0 (2007). The Measurement of Efficiency. In: *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Studies in Productivity and Efficiency*, vol 4. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-35231-2_2

0 Camanho, A. S., & D’Inverno, G. (2023). *Data Envelopment Analysis: A Review and Synthesis*. En *Advanced Mathematical Methods for Economic Efficiency Analysis* (pp. 33–54). SpringerLink

0 <https://www.mimp.gob.pe/files/omep/archivos/sesion2.pdf>

0 <https://qualityanalytics.net/analisis-parametricos-vs-no-parametricos-como-y-cuando-utilizarlos/>

0 <https://dialnet.unirioja.es/descarga/articulo/7286075.pdf>

0 https://www.academia.edu/39481471/Measuring_the_efficiency_of_decision_making_units

0 Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429-444.

0 Banker, R.D., Charnes, A., & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078-1092.

0 Cooper, W.W., Seiford, L.M., & Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software* (2nd ed.). Springer Science & Business Media.

0 Vitaliano (1998): *Assessing public library efficiency using DEA*

0 <https://rua.ua.es/dspace/bitstream/10045/19658/6/Materiales.Teor%C3%ADa.Bloque.III.pdf>

0 Pearson, K. (1896). Mathematical contributions to the theory of evolution III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London*, 187, 253-318.

0 Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572.

0 R Core Team (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>

Apéndice A

Código R

```
# =====  
# ANÁLISIS DE EFICIENCIA DE BIBLIOTECAS PÚBLICAS MEDIANTE DEA  
# CASO DE ESTUDIO: CASTILLA Y LEÓN (2020)  
# =====  
  
# Carga de librerías necesarias  
library(readxl)  # Para leer archivos Excel  
library(tidyr)   # Para manipulación de datos  
library(dplyr)   # Para manipulación de datos  
library(stats)   # Para análisis estadístico  
library(Benchmarking) # Para análisis DEA  
library(ggplot2) # Para visualizaciones  
library(writexl) # Para exportar resultados  
library(gridExtra) # Para combinar gráficos  
library(corrplot) # Para visualizar correlaciones  
  
# =====  
# PASO 1: CARGA Y PREPARACIÓN DE DATOS  
# =====  
  
# Ruta del archivo (ajustar según sea necesario)  
filename = "/Users/conchitagomezmiguel/Downloads/Bibliotecas.xls"  
  
# EPIGRAFE 1  
EP1 <- as.data.frame(read_excel(filename, sheet = "13.8.1.a.", skip = 11,  
                                col_names = FALSE, n_max = 9,  
                                .name_repair = ~ substr(.x, start = 1, stop = 40)))  
  
ep1_colnames <- c("Provincia",  
                  "Bibliotecas publicas 2020",  
                  "Visitantes 2020",  
                  "Adultos inscritos 2020",  
                  "Infantiles inscritos 2020",  
                  "Total de inscritos 2020",  
                  "",  
                  "Nuevos usuarios adultos 2020",
```

```

"Nuevos usuarios infantiles 2020",
"Total de Nuevos usuarios 2020")
colnames(EP1) <- ep1_colnames

EP1 = EP1[,-c(6,7,10)]

# EPIGRAFE 2
EP2 <- as.data.frame(read_excel(filename, sheet = "13.8.1.b.", skip = 11,
                                col_names = FALSE, n_max = 9,
                                .name_repair = ~ substr(.x, start = 1, stop = 40)))

ep2_colnames <- c("Provincia",
                  "Gastos en bienes corrientes y servicios 2020",
                  "Fondo bibliográfico 2020",
                  "Actividades culturales 2020",
                  "Otras inversiones 2020",
                  "Total Consejería de Cultura y Turismo 2020")

colnames(EP2) <- ep2_colnames

# EPIGRAFE 3
EP3 <- as.data.frame(read_excel(filename, sheet = "13.8.2.a.", skip = 12,
                                col_names = FALSE, n_max = 9,
                                .name_repair = ~ substr(.x, start = 1, stop = 40)))

ep3_colnames <- c("Provincia",
                  "Fondo bibliográfico de Libros en bibliotecas provinciales en 2020",
                  "Fondo bibliográfico de P. Periódicas",
                  "",
                  "Fondo audiovisual de Vídeos en bibliotecas provinciales en 2020",
                  "Fondo audiovisual de Discos y Casetes en bibliotecas provinciales 2020",
                  "Fondo audiovisual de Documentos electrónicos en bibliotecas provinciales 2020")

colnames(EP3) <- ep3_colnames
EP3 = EP3[,-4]

# EPIGRAFE 4
EP4 <- as.data.frame(read_excel(filename, sheet = "13.8.2.b.", skip = 11,
                                col_names = FALSE, n_max = 9,
                                .name_repair = ~ substr(.x, start = 1, stop = 40)))

ep4_colnames <- c("Provincia",
                  "Nº de Usuarios infantiles en bibliotecas provinciales 2020",
                  "Nº de Usuarios adultos en bibliotecas provinciales 2020",
                  "Nº de Usuarios totales en bibliotecas provinciales 2020",

```

```
"Servicio de Préstamo infantil en bibliotecas provinciales 2020",
"Servicio de Préstamo adultos en bibliotecas provinciales 2020",
"Servicio de Préstamo total en bibliotecas provinciales 2020")
```

```
colnames(EP4) <- ep4_colnames
EP4 <- EP4[,-c(4,7)]
```

```
# EPIGRAFE 5
```

```
EP5 <- as.data.frame(read_excel(filename, sheet = "13.8.2.c.", skip = 10,
                             col_names = FALSE, n_max = 9,
                             .name_repair = ~ substr(.x, start = 1, stop = 40)))
```

```
ep5_colnames <- c("Provincia",
                  "Fondo bibliográfico en bibliotecas provinciales en 2011",
                  "Fondo bibliográfico en bibliotecas provinciales en 2012",
                  "Fondo bibliográfico en bibliotecas provinciales en 2013",
                  "Fondo bibliográfico en bibliotecas provinciales en 2014",
                  "Fondo bibliográfico en bibliotecas provinciales en 2015",
                  "Fondo bibliográfico en bibliotecas provinciales en 2016",
                  "Fondo bibliográfico en bibliotecas provinciales en 2017",
                  "Fondo bibliográfico en bibliotecas provinciales en 2018",
                  "Fondo bibliográfico en bibliotecas provinciales en 2019",
                  "Fondo bibliográfico en bibliotecas provinciales en 2020")
```

```
colnames(EP5) <- ep5_colnames
```

```
# EPIGRAFE 6
```

```
EP6 <- as.data.frame(read_excel(filename, sheet = "13.8.2.d.", skip = 11,
                             col_names = FALSE, n_max = 9,
                             .name_repair = ~ substr(.x, start = 1, stop = 40)))
```

```
ep6_colnames <- c("Provincia",
                  "Fondo bibliográfico y audiovisual infantil en bibliotecas provinciales 2017",
                  "Fondo bibliográfico y audiovisual adultos en bibliotecas provinciales 2017",
                  "Fondo bibliográfico y audiovisual total en bibliotecas provinciales 2017",
                  "",
                  "Fondo bibliográfico y audiovisual infantil en bibliotecas provinciales 2018",
                  "Fondo bibliográfico y audiovisual adultos en bibliotecas provinciales 2018",
                  "Fondo bibliográfico y audiovisual total en bibliotecas provinciales 2018",
                  "",
                  "Fondo bibliográfico y audiovisual infantil en en bibliotecas provinciales 2019",
                  "Fondo bibliográfico y audiovisual adultos en bibliotecas provinciales 2019",
                  "Fondo bibliográfico y audiovisual total en bibliotecas provinciales 2019",
                  "",
                  "Fondo bibliográfico y audiovisual infantil en bibliotecas provinciales 2020",
```


"Fondo bibliográfico y audiovisual adultos en bibliotecas provinciales 2020",

"Fondo bibliográfico y audiovisual total en bibliotecas provinciales 2020")

```
colnames(EP6) <- ep6_colnames
```

```
# EPIGRAFE 7
```

```
EP7 <- as.data.frame(read_excel(filename, sheet = "13.8.3.a.", skip = 11,  
  col_names = FALSE, n_max = 9,  
  .name_repair = ~ substr(.x, start = 1, stop = 40)))
```

```
ep7_colnames <- c("Provincia",  
  "Número de bibliotecas",  
  "Fondo de Libros en bibliotecas municipales 2020",  
  "Fondo de P. Periódicas en bibliotecas municipales 2020",  
  "Fondo Audiovisuales y Electrónicos en bibliotecas municipales 2020",  
  "Nuevos Usuarios Inscritos en bibliotecas municipales 2020",  
  "Total Usuarios Inscritos en bibliotecas municipales 2020")
```

```
colnames(EP7) <- ep7_colnames
```

```
# EPIGRAFE 8
```

```
EP8 <- as.data.frame(read_excel(filename, sheet = "13.8.3.b.", skip = 9,  
  col_names = TRUE, n_max = 9,  
  .name_repair = ~ substr(.x, start = 1, stop = 40)))
```

```
ep8_colnames <- c("Provincia",  
  "Fondo bibliográfico en bibliotecas municipales 2011",  
  "Fondo bibliográfico en bibliotecas municipales 2012",  
  "Fondo bibliográfico en bibliotecas municipales 2013",  
  "Fondo bibliográfico en bibliotecas municipales 2014",  
  "Fondo bibliográfico en bibliotecas municipales 2015",  
  "Fondo bibliográfico en bibliotecas municipales 2016",  
  "Fondo bibliográfico en bibliotecas municipales 2017",  
  "Fondo bibliográfico en bibliotecas municipales 2018",  
  "Fondo bibliográfico en bibliotecas municipales 2019",  
  "Fondo bibliográfico en bibliotecas municipales 2020")
```

```
colnames(EP8) <- ep8_colnames
```

```
# EPIGRAFE 9
```

```
EP9 <- as.data.frame(read_excel(filename, sheet = "13.8.3.c.", skip = 10,  
  col_names = FALSE, n_max = 9,  
  .name_repair = ~ substr(.x, start = 1, stop = 40)))
```

```
ep9_colnames <- c("Provincia",  
  "Libros prestados en bibliotecas municipales 2011",  
  "Libros prestados en bibliotecas municipales 2012",  
  "Libros prestados en bibliotecas municipales 2013",  
  "Libros prestados en bibliotecas municipales 2014",  
  "Libros prestados en bibliotecas municipales 2015",
```

```
"Libros prestados en bibliotecas municipales 2016",
"Libros prestados en bibliotecas municipales 2017",
"Libros prestados en bibliotecas municipales 2018",
"Libros prestados en bibliotecas municipales 2019",
"Libros prestados en bibliotecas municipales 2020")
```

```
colnames(EP9) <- ep9_colnames
```

```
# EPIGRAFE 10
```

```
EP10 <- as.data.frame(read_excel(filename, sheet = "13.8.3.d.", skip = 11,
                                col_names = FALSE, n_max = 9,
                                .name_repair = ~ substr(.x, start = 1, stop = 40)))
```

```
EP10 = EP10[,-c(4,7)]
```

```
ep10_colnames <- c("Provincia",
                  "Préstamo infantil de libros en bibliotecas municipales 2020",
                  "Préstamo infantil de Audiovisuales y electrónicos en bibliotecas municipales 2020",
                  "Préstamo adultos de libros en bibliotecas municipales 2020",
                  "Préstamo adultos de Audiovisuales y electrónicos en bibliotecas municipales 2020",
                  "Préstamo total de libros en bibliotecas municipales 2020",
                  "Préstamo total de Audiovisuales y electrónicos en bibliotecas municipales 2020")
```

```
colnames(EP10) <- ep10_colnames
```

```
EP10 <- EP10[,-c(6,7)]
```

```
# EPIGRAFE 16
```

```
EP16 <- as.data.frame(read_excel(filename, sheet = "13.8.5.", skip = 10,
                                col_names = TRUE, n_max = 12,
                                .name_repair = ~ substr(.x, start = 1, stop = 40)))
```

```
ep16_colnames <- c("Provincia",
                  "Ávila",
                  "Burgos",
                  "León",
                  "Palencia",
                  "Salamanca",
                  "Segovia",
                  "Soria",
                  "Valladolid",
                  "Zamora",
                  "Castilla & León")
```

```
colnames(EP16) <- ep16_colnames
```

```
EP16 = EP16[,-ncol(EP16)]
```

```
EP16 = t(EP16)
```

```
colnames(EP16) = EP16[1,]  
EP16 = as.data.frame(EP16[-1,])
```

```
# Convertir a numérico
```

```
library(data.table)
```

```
setDT(EP16)
```

```
EP16 = EP16[, names(EP16) := lapply(.SD, as.numeric)]
```

```
EP16 <- as.data.frame(EP16)
```

```
# =====
```

```
# PASO 2: UNIFICACIÓN DE DATOS Y PREPARACIÓN PARA ANÁLISIS
```

```
# =====
```

```
# Combinar todos los epígrafes en un único conjunto de datos
```

```
data = cbind(EP1, EP2, EP3, EP4, EP4, EP5, EP6, EP7, EP8, EP9, EP10, EP10)
```

```
# Eliminar columnas de provincia duplicadas
```

```
data <- data[, !grepl("Provincia", names(data))]
```

```
# Convertir todas las columnas a numérico
```

```
setDT(data)
```

```
data = as.data.frame(data[, names(data) := lapply(.SD, as.numeric)])
```

```
# Asignar nombres de provincias como nombres de fila
```

```
rownames(data) = c("Ávila", "Burgos", "León", "Palencia", "Salamanca",  
                  "Segovia", "Soria", "Valladolid", "Zamora")
```

```
# =====
```

```
# PASO 3: ANÁLISIS EXPLORATORIO Y SELECCIÓN DE VARIABLES
```

```
# =====
```

```
# 3.1 Selección de variables potencialmente relevantes
```

```
datos_relevantes <- data %>%
```

```
select(
```

```
  'Bibliotecas publicas 2020',
```

```
  'Visitantes 2020',
```

```
  'Adultos inscritos 2020',
```

```
  'Infantiles inscritos 2020',
```

```
  'Nuevos usuarios adultos 2020',
```

```
  'Nuevos usuarios infantiles 2020',
```

```
  'Gastos en bienes corrientes y servicios 2020',
```

```
  'Fondo bibliográfico 2020',
```

```
  'Actividades culturales 2020',
```

```
  'Otras inversiones 2020',
```

```
  'Número de bibliotecas',
```

```
'Fondo de Libros en bibliotecas municipales 2020',  
'Fondo de P. Periódicas en bibliotecas municipales 2020',  
'Fondo Audiovisuales y Electrónicos en bibliotecas municipales 2020',  
'Nuevos Usuarios Inscritos en bibliotecas municipales 2020',  
'Total Usuarios Inscritos en bibliotecas municipales 2020',  
'Libros prestados en bibliotecas municipales 2020'
```

```
)
```

3.2 Matriz de correlación

1. Crear la matriz de correlación

```
correlacion <- cor(datos_relevantes, use = "complete.obs")
```

2. Convertir a data.frame largo (tipo lista)

```
library(tidyr)
```

```
library(dplyr)
```

```
cor_larga <- as.data.frame(as.table(correlacion)) %>%
```

```
  rename(Variable1 = Var1, Variable2 = Var2, Correlacion = Freq) %>%
```

```
  filter(Variable1 != Variable2) %>% # quitar correlaciones de una variable consigo misma
```

```
  mutate(AbsCor = abs(Correlacion)) %>% # añadir valor absoluto
```

```
  arrange(desc(AbsCor)) # ordenar por fuerza de correlación
```

3. Eliminar duplicados (por ejemplo, A~B y B~A)

```
cor_limpia <- cor_larga %>%
```

```
  rowwise() %>%
```

```
  mutate(Pareja = paste(sort(c(Variable1, Variable2)), collapse = "_")) %>%
```

```
  distinct(Pareja, keep_all = TRUE) %>%
```

```
  select(-Pareja)
```

4. Ver las correlaciones más importantes (por ejemplo, top 10)

```
View(head(cor_limpia, 10))
```

3.3 Análisis de Componentes Principales

Escalar los datos

```
datos_escalados <- scale(datos_relevantes)
```

Realizar el PCA

```
pca <- prcomp(datos_escalados, scale = FALSE) # scale = FALSE porque ya escalamos los datos
```

Resumen del PCA

```
summary_pca <- summary(pca)
```

```
print(summary_pca)
```

```
View(print(summary_pca))
```

```

# Visualizar las cargas factoriales de los primeros componentes
print(pca$rotation[, 1:3])

# Scree plot
plot(pca, type = "l", main = "Scree Plot")
abline(h = 1, col = "red", lty = 2) # Criterio de Kaiser

# Biplot
biplot(pca, scale = 0, cex = c(0.8, 0.8),
       xlabs = rownames(datos_relevantes),
       main = "Biplot: PCA para variables bibliotecarias")

# 3.4 Selección final de variables para el modelo DEA
# Basado en el análisis PCA y consideraciones teóricas

# Variables seleccionadas para el modelo DEA
# INPUTS

X <- data %>% select("Gastos en bienes corrientes y servicios 2020", "Número de bibliotecas")
Y <- data %>% select("Visitantes 2020")

# Normalización (para evitar problemas de escala)
X_norm <- apply(X, 2, function(x) x / max(x))
Y_norm <- apply(Y, 2, function(x) x / max(x))

X <- as.matrix(X_norm)
Y <- as.matrix(Y_norm)
# =====
# PASO 4: ANÁLISIS DEA
# =====

# 4.1 Modelo CCR (Rendimientos constantes a escala)
# Orientación input
ccr_input <- dea(X, Y, RTS = "crs", ORIENTATION = "in", SLACK = TRUE)
eficiencia_ccr_in <- ccr_input$eff
names(eficiencia_ccr_in) <- rownames(data)

# Orientación output (opcional)
ccr_output <- dea(X, Y, RTS = "crs", ORIENTATION = "out", SLACK = TRUE)
eficiencia_ccr_out <- ccr_output$eff
names(eficiencia_ccr_out) <- rownames(data)

# 4.2 Modelo BCC (Rendimientos variables a escala)
# Orientación input

```

```

bcc_input <- dea(X, Y, RTS = "vrs", ORIENTATION = "in",SLACK = TRUE)
eficiencia_bcc_in <- bcc_input$eff
names(eficiencia_bcc_in) <- rownames(data)

# Orientación output (opcional)
bcc_output <- dea(X, Y, RTS = "vrs", ORIENTATION = "out",SLACK = TRUE)
eficiencia_bcc_out <- bcc_output$eff
names(eficiencia_bcc_out) <- rownames(data)

# 4.3 Cálculo de la eficiencia de escala
eficiencia_escalas <- eficiencia_ccr_in / eficiencia_bcc_in
names(eficiencia_escalas) <- rownames(data)

# 4.4 Determinación del tipo de rendimientos a escala
# Modelo con rendimientos no crecientes a escala
nirs <- dea(X, Y, RTS = "irs2", ORIENTATION = "in",SLACK = TRUE)

eficiencia_nirs <- nirs$eff

# Identificación del tipo de rendimientos
tipo_rendimientos <- character(length(rownames(data)))
names(tipo_rendimientos) <- rownames(data)

for (i in 1:length(rownames(data))) {
  if (abs(eficiencia_ccr_in[i] - eficiencia_bcc_in[i]) < 1e-6) {
    # Si CCR = BCC, rendimientos constantes
    tipo_rendimientos[i] <- "Constantes"
  } else if (abs(nirs$eff[i] - eficiencia_bcc_in[i]) < 1e-6) {
    # Si NIRS = BCC, rendimientos decrecientes
    tipo_rendimientos[i] <- "Decrecientes"
  } else {
    # Si NIRS != BCC, rendimientos crecientes
    tipo_rendimientos[i] <- "Crecientes"
  }
}

# 4.5 Extracción y preparación de holguras
# Extraer holguras del modelo CCR orientado a input
sx_ccr_in <- ccr_input$sx
sy_ccr_in <- ccr_input$sy

# Extraer holguras del modelo BCC orientado a input
sx_bcc_in <- bcc_input$sx
sy_bcc_in <- bcc_input$sy

```

```

# Extraer holguras del modelo CCR orientado a output
sx_ccr_out <- ccr_output$sx
sy_ccr_out <- ccr_output$sy

# Extraer holguras del modelo BCC orientado a output
sx_bcc_out <- bcc_output$sx
sy_bcc_out <- bcc_output$sy

# Nombrar las columnas de holguras
input_names <- colnames(X)
output_names <- colnames(Y)

colnames(sx_ccr_in) <- paste0("Holgura_Input_CCR_IN_", input_names)
colnames(sy_ccr_in) <- paste0("Holgura_Output_CCR_IN_", output_names)
colnames(sx_bcc_in) <- paste0("Holgura_Input_BCC_IN_", input_names)
colnames(sy_bcc_in) <- paste0("Holgura_Output_BCC_IN_", output_names)
colnames(sx_ccr_out) <- paste0("Holgura_Input_CCR_OUT_", input_names)
colnames(sy_ccr_out) <- paste0("Holgura_Output_CCR_OUT_", output_names)
colnames(sx_bcc_out) <- paste0("Holgura_Input_BCC_OUT_", input_names)
colnames(sy_bcc_out) <- paste0("Holgura_Output_BCC_OUT_", output_names)

# Convertir matrices de holguras a data frames
sx_ccr_in_df <- as.data.frame(sx_ccr_in)
sy_ccr_in_df <- as.data.frame(sy_ccr_in)
sx_bcc_in_df <- as.data.frame(sx_bcc_in)
sy_bcc_in_df <- as.data.frame(sy_bcc_in)
sx_ccr_out_df <- as.data.frame(sx_ccr_out)
sy_ccr_out_df <- as.data.frame(sy_ccr_out)
sx_bcc_out_df <- as.data.frame(sx_bcc_out)
sy_bcc_out_df <- as.data.frame(sy_bcc_out)

# Asignar nombres de provincias como nombres de fila
rownames(sx_ccr_in_df) <- rownames(sx_bcc_in_df) <- rownames(data)
rownames(sy_ccr_in_df) <- rownames(sy_bcc_in_df) <- rownames(data)
rownames(sx_ccr_out_df) <- rownames(sx_bcc_out_df) <- rownames(data)
rownames(sy_ccr_out_df) <- rownames(sy_bcc_out_df) <- rownames(data)

# 4.6 Procesamiento de los benchmarks (referencias)
# Extraer valores lambda para el modelo CCR orientado a input
lambda_ccr_in <- ccr_input$lambda

# Asegurarse de que los nombres de filas y columnas sean correctos
rownames(lambda_ccr_in) <- rownames(data)

# Si las columnas tienen prefijos como "L_", los eliminamos
if(any(grepl("^L_", colnames(lambda_ccr_in)))) {

```

```

colnames(lambda_ccr_in) <- gsub("^L_", "", colnames(lambda_ccr_in))
} else {
  # Si no tienen prefijo, asignamos los nombres de provincias
  colnames(lambda_ccr_in) <- rownames(data)
}
# Convertir valores muy pequeños a cero (por redondeo numérico)
lambda_ccr_in[lambda_ccr_in < 1e-6] <- 0

# Hacer lo mismo para el modelo BCC
lambda_bcc_in <- bcc_input$lambda
rownames(lambda_bcc_in) <- rownames(data)
if(any(grepl("^L_", colnames(lambda_bcc_in)))) {
  colnames(lambda_bcc_in) <- gsub("^L_", "", colnames(lambda_bcc_in))
} else {
  colnames(lambda_bcc_in) <- rownames(data)
}
lambda_bcc_in[lambda_bcc_in < 1e-6] <- 0

# Convertir a data frames con nombres descriptivos
lambda_ccr_in_df <- as.data.frame(lambda_ccr_in)
colnames(lambda_ccr_in_df) <- paste0("Ref_CCR_IN_", colnames(lambda_ccr_in_df))

lambda_bcc_in_df <- as.data.frame(lambda_bcc_in)
colnames(lambda_bcc_in_df) <- paste0("Ref_BCC_IN_", colnames(lambda_bcc_in_df))

# Opcional: hacer lo mismo para los modelos orientados a output
lambda_ccr_out <- ccr_output$lambda
lambda_bcc_out <- bcc_output$lambda
# (código similar al anterior para procesar estos lambda)

# 4.7 Compilación de todos los resultados
resultados_dea <- data.frame(
  Provincia = rownames(data),
  CCR_Input = round(eficiencia_ccr_in, 4),
  BCC_Input = round(eficiencia_bcc_in, 4),
  CCR_Output = round(eficiencia_ccr_out, 4),
  BCC_Output = round(eficiencia_bcc_out, 4),
  Eficiencia_Escala = round(eficiencia_escalas, 4),
  Tipo_Rendimientos = tipo_rendimientos
)

# Añadir las holguras al dataframe de resultados
resultados_dea <- cbind(
  resultados_dea,
  round(sx_ccr_in_df, 4),

```



```

round(sy_ccr_in_df, 4),
round(sx_bcc_in_df, 4),
round(sy_bcc_in_df, 4)
)

```

```

# Añadir los valores lambda (benchmarks)

```

```

resultados_dea <- cbind(
  resultados_dea,
  round(lambda_ccr_in_df, 4),
  round(lambda_bcc_in_df, 4)
)

```

```

# Ver resultados completos

```

```

print(resultados_dea)

```

```

# Opcionalmente, guardar los resultados en un archivo Excel

```

```

write_xlsx(resultados_dea, "Resultados_DEA_completos.xlsx")

```

```

# 4.8 Crear resúmenes y tablas específicas para el análisis

```

```

# Resumen de eficiencia y benchmarks principales

```

```

resumen_benchmarks <- data.frame(
  Provincia = resultados_dea$Provincia,
  CCR_Input = resultados_dea$CCR_Input,
  BCC_Input = resultados_dea$BCC_Input,
  Eficiencia_Escala = resultados_dea$Eficiencia_Escala,
  Tipo_Rendimientos = resultados_dea$Tipo_Rendimientos,
  Principales_Referencias = character(nrow(resultados_dea))
)

```

```

# Añadir información sobre benchmarks para cada provincia

```

```

for (i in 1:nrow(resumen_benchmarks)) {
  if (resumen_benchmarks$CCR_Input[i] < 0.9999) {
    # Para provincias ineficientes, identificar sus principales referencias
    # Obtener valores lambda del modelo CCR
    lambda_valores <- as.numeric(lambda_ccr_in[i,])
    # Obtener nombres de provincias con lambda > 0
    referencias <- rownames(data)[lambda_valores > 0.01]
    # Obtener valores lambda para estas referencias
    valores <- lambda_valores[lambda_valores > 0.01]
    # Crear texto con referencias y sus valores lambda
    refs_texto <- paste(referencias, " (", round(valores, 2), ")", sep="", collapse=", ")
    resumen_benchmarks$Principales_Referencias[i] <- refs_texto
  } else {
    # Para provincias eficientes

```

```

resumen_benchmarks$Principales_Referencias[i] <- "Es benchmark (DMU eficiente)"
}
}

# Crear tabla de referencias recibidas (cuántas veces cada DMU es benchmark)
# Contar cuántas veces cada provincia aparece como referencia para otras
referencias_count <- numeric(nrow(data))
names(referencias_count) <- rownames(data)

for (i in 1:nrow(lambda_ccr_in)) {
  for (j in 1:ncol(lambda_ccr_in)) {
    if (lambda_ccr_in[i,j] > 0.01) {
      referencias_count[j] <- referencias_count[j] + 1
    }
  }
}

tabla_referencias <- data.frame(
  Provincia = rownames(data),
  Eficiencia_CCR = resultados_dea$CCR_Input,
  Veces_Referencia = referencias_count,
  Porcentaje = round(referencias_count / (nrow(data)-1) * 100, 1)
)
tabla_referencias <- tabla_referencias[order(-tabla_referencias$Veces_Referencia),]

print(resumen_benchmarks)
print(tabla_referencias)

```

