

Metaanálisis de metaheurísticas empleando R. *

Daniela Alcázar N. *Universidad de Antioquia*

Aura María Molina *Universidad de Antioquia*

Alejandro Arenas-Vasco *Universidad de Antioquia*

Juan G. Villegas *Analítica e Investigación para la toma de Decisiones, Departamento de Ingeniería Industrial, Facultad de Ingeniería, Universidad de Antioquia*

La divergencia en los resultados de estudios análogos es un problema en la investigación de operaciones, debido a que los investigadores suelen centrarse en el desarrollo de soluciones algorítmicas complejas, dejando de lado la evaluación del rendimiento de las existentes. El objetivo de esta investigación es replicar el metaanálisis realizado por [Turkeš, Sörensen and Hvattum \(2021\)](#) utilizando los datos públicos disponibles de [Turkeš et al. \(2020\)](#) y el lenguaje de programación R. En este metaanálisis, se analizó el impacto de agregar el componente adaptativo (ALNS) a la metaheurística Large Neighborhood Search (LNS). Para ello, se ejecutó un modelo de efectos aleatorios en R, utilizando funciones de las librerías meta, metafor y dmetar para obtener el efecto del componente adaptativo y el diagrama de bosque, en conjunto con ggplot2 para la presentación descriptiva de la información fuente.

Keywords: metaanálisis, modelo de efectos, metaheurísticas

Introducción

El metaanálisis es una técnica estadística que se emplea para reunir y combinar los resultados de diversos estudios. El término hace referencia a una revisión sistemática de la literatura que utiliza técnicas estadísticas para integrar y resumir los hallazgos de varios estudios independientes. En contraste con las revisiones tradicionales, que dependen de la subjetividad del revisor y carecen de transparencia en los métodos de búsqueda, las revisiones sistemáticas reducen el sesgo, ofreciendo resultados más confiables. El propósito de esta revisión es abordar una misma pregunta científica y lograr una estimación más precisa del impacto de una intervención o tratamiento, que se define como la diferencia en los resultados entre grupos experimentales y de control en relación con una variable de interés. Esta metodología no solo facilita la síntesis de la evidencia científica, sino que también posibilita la identificación de patrones comunes en los resultados de la investigación y la evaluación de la magnitud de los efectos. El metaanálisis se emplea para combinar los resultados en un único estimador puntual, bajo el enfoque de que no se debe concentrar la investigación en la significancia estadística de los estudios individuales, sino en la magnitud del efecto o su tamaño ([Crowther, Lim and Crowther, 2010](#)).

En general, un metaanálisis se realiza de acuerdo con los siguientes pasos:

1. Identificación de los estudios: búsqueda bibliográfica.
2. Evaluación de calidad de los estudios: calidad metodológica.
3. Extracción de datos: tamaños de efectos, estadísticas significativas y características de los estudios.

* Autor de contacto: daniela.alcazar@udea.edu.co.

4. Análisis de los datos, empleando métodos estadísticos efectos fijos y efectos aleatorios.
5. Interpretación de resultados, teniendo en cuenta las limitaciones del estudio.

Herramientas estadísticas libres como el lenguaje de programación R ([R Core Team, 2019](#)) son de mucha utilidad en la realización de revisiones de literatura basadas en metaanálisis. Esto se debe a que el metaanálisis abarca desde estadísticas descriptivas de los estudios analizados hasta pruebas de hipótesis y meta-regresiones, entre otras técnicas estadísticas ([Harrer et al., 2021](#); [Balduzzi, Rucker and Schwarzer, 2019](#)).

Por otro lado, ha habido un aumento en la calidad y la frecuencia de las revisiones de literatura en el ámbito de la investigación de operaciones. Según [Schryen and Sperling \(2023\)](#), la mayoría de estas revisiones tienden a tener un alcance limitado y adoptan enfoques narrativos o selectivos, mientras que las revisiones integrativas, sistemáticas y los metaanálisis son menos comunes.

En particular, [Turkeš, Sörensen and Hvattum \(2021\)](#) llevaron a cabo un metaanálisis que abarcó 25 estudios sobre el efecto del componente de adaptabilidad en la búsqueda adaptativa de vecindarios de gran escala (ALNS – Adaptive large neighborhood search). La ALNS constituye una técnica metaheurística que se utiliza en la resolución de problemas de optimización combinatoria. El objetivo del metaanálisis de [Turkeš, Sörensen and Hvattum \(2021\)](#) fue determinar si la adaptabilidad de la ALNS tiene un efecto significativo en su eficacia. Como resultado, encontraron que la adaptabilidad tiene un efecto positivo significativo en la eficacia de la ALNS, pues para los estudios analizados, se obtuvo en general mejores resultados cuando se incluía este componente, en comparación con aquellas veces donde la búsqueda era no adaptativa. Cabe resaltar que los autores señalan que, si bien la capa adaptativa puede aportar valor (y en un número limitado de estudios lo hace), también agrega complejidad, por lo que su recomendación se limita a situaciones específicas.

Metaanálisis en R con los datos de [Turkeš et al. \(2020\)](#)

A continuación se detalla el procedimiento para reproducir el estudio de [Turkeš, Sörensen and Hvattum \(2021\)](#) utilizando los datos públicos de [Turkeš et al. \(2020\)](#) y las directrices de [Harrer et al. \(2021\)](#), guía práctica proporciona instrucciones detalladas y ejemplos para realizar un metaanálisis con el lenguaje de programación R ([R Core Team, 2019](#)). Se empleó RStudio ([Allaire, 2012](#)) como entorno de desarrollo integrado.

[Turkeš et al. \(2020\)](#) recopilaron datos de 25 estudios para el metaanálisis. De estos, 22 abordaron problemas de minimización, y solo tres buscaban maximizar la función objetivo. Cada estudio definió un conjunto de instancias y un número de repeticiones específico. El promedio de instancias por estudio fue de aproximadamente 150, exceptuando uno que tenía más de 11000 instancias. En cuanto a las repeticiones, 13 estudios resolvieron cada instancia cinco veces, 10 realizaron 10 repeticiones, y los dos restantes llevaron a cabo una y 20 repeticiones, respectivamente.

1. Preparación de los datos.

En esta etapa inicial se utilizan los datos no procesados proporcionados por [Turkeš et al. \(2020\)](#), datos que contienen la información sobre los valores obtenidos en cada repetición de las instancias del problema en cuestión. Se cargan al entorno de **RStudio** ([Allaire, 2012](#)), donde se realiza una exploración de los datos, imputando valores faltantes, eliminando

duplicados y corrigiendo errores de entrada, y se analiza gráficamente la distribución de los datos obtenidos por factores y sus características generales, usando los paquetes **skimr** (Waring et al., 2022) y **ggplot2** (Wickham, 2016). Tras esto, el paquete por excelencia en manipulación de marcos de datos **dplyr** (Wickham et al., 2023) se emplea para así poder sintetizar la información por instancias, siguiendo lo definido por Turkeš, Sörensen and Hvattum (2021) acerca de tomar el promedio de las instancias como valor de la función objetivo para la metaheurística con y sin el componente, para posteriormente hallar el efecto del componente adaptativo en cada una de ellas y el error estándar asociado.

2. Metaanálisis.

Siguiendo la guía práctica de Harrer et al. (2021), se usan los paquetes **meta** (Balduzzi, Rucker and Schwarzer, 2019) y **metafor** (Viechtbauer, 2010) para la creación del modelo de efectos aleatorios y el diagrama de bosque, que representa visualmente las estimaciones obtenidas en el modelo de efectos.

3. Análisis de sensibilidad.

Se evalúa la robustez de los resultados del metaanálisis mediante un análisis de sensibilidad, identificando los estudios altamente influyentes en la estimación del efecto global, utilizando el paquete **dmeter** (Harrer et al., 2019).

4. Evaluación de sesgo de publicación.

Se verifica la presencia de sesgo de publicación en los estudios analizados utilizando el paquete **dmeter** (Harrer et al., 2019).

5. Interpretación y publicación de los resultados.

Por último, se presentan los resultados obtenidos durante el metaanálisis usando R Markdown (Xie, Dervieux and Riederer, 2020) para crear el informe, que facilita la combinación de texto plano, código y salidas de R (R Core Team, 2019) en un solo documento.

Resultados y discusión

Turkeš, Sörensen and Hvattum (2021) proponen que el efecto del componente adaptativo en la metaheurística de la ALNS para cada estudio como la mejora promedio. Utilizan la Ecuación 1 si el objetivo es maximizar y la Ecuación 2, si es minimizar. Donde A_k representa la mejora promedio atribuida al componente adaptativo de la metaheurística en el estudio k , N_k es el número de instancias, \mathcal{I} es el conjunto de instancias, $\bar{f}_k(x_{r_k}(I))$ es el valor promedio de la función objetivo para la instancia I con el componente adaptativo (con un factor de adaptatividad r definido para el estudio k), y $\bar{f}_k(x_0(I))$ es el valor promedio de la función objetivo sin la capa adaptativa de la metaheurística.

$$A_k = \frac{1}{N_k} \sum_{I \in \mathcal{I}} \frac{\bar{f}_k(x_{r_k}(I)) - \bar{f}_k(x_0(I))}{\bar{f}_k(x_0(I))} \quad (1)$$

$$A_k = \frac{1}{N_k} \sum_{I \in \mathcal{I}} \frac{\bar{f}_k(x_0(I)) - \bar{f}_k(x_{r_k}(I))}{\bar{f}_k(x_0(I))} \quad (2)$$

El error estándar del efecto se calcula mediante la Ecuación 3, donde σ_k^2 representa la varianza del efecto entre las instancias del estudio y N_k es el número de instancias.

$$SE_k = \frac{\sigma_k^2}{N_k} \quad (3)$$

De los datos de [Turkeš et al. \(2020\)](#) se excluyeron tres estudios con valores atípicos para el error estándar (error estándar muy grande). Se aplicó un modelo de efectos aleatorios para abordar la heterogeneidad mencionada, mediante el cual se obtuvo un promedio de efecto del componente adaptativo para la ALNS del 0.136%, con un intervalo al 95% de confianza (−0.09;0.36). El resumen del modelo, representado en la Figura 1, incluye el autor principal y el año de publicación del artículo, el efecto promedio, el error estándar, el intervalo de confianza y el peso asignado a cada estudio. El diagrama de bosque visualiza estos datos, destacando el efecto, el peso (a través del tamaño), y el intervalo de confianza (la línea horizontal alrededor del efecto), con un índice de heterogeneidad del 94.4% y una varianza del efecto real de 0.015.

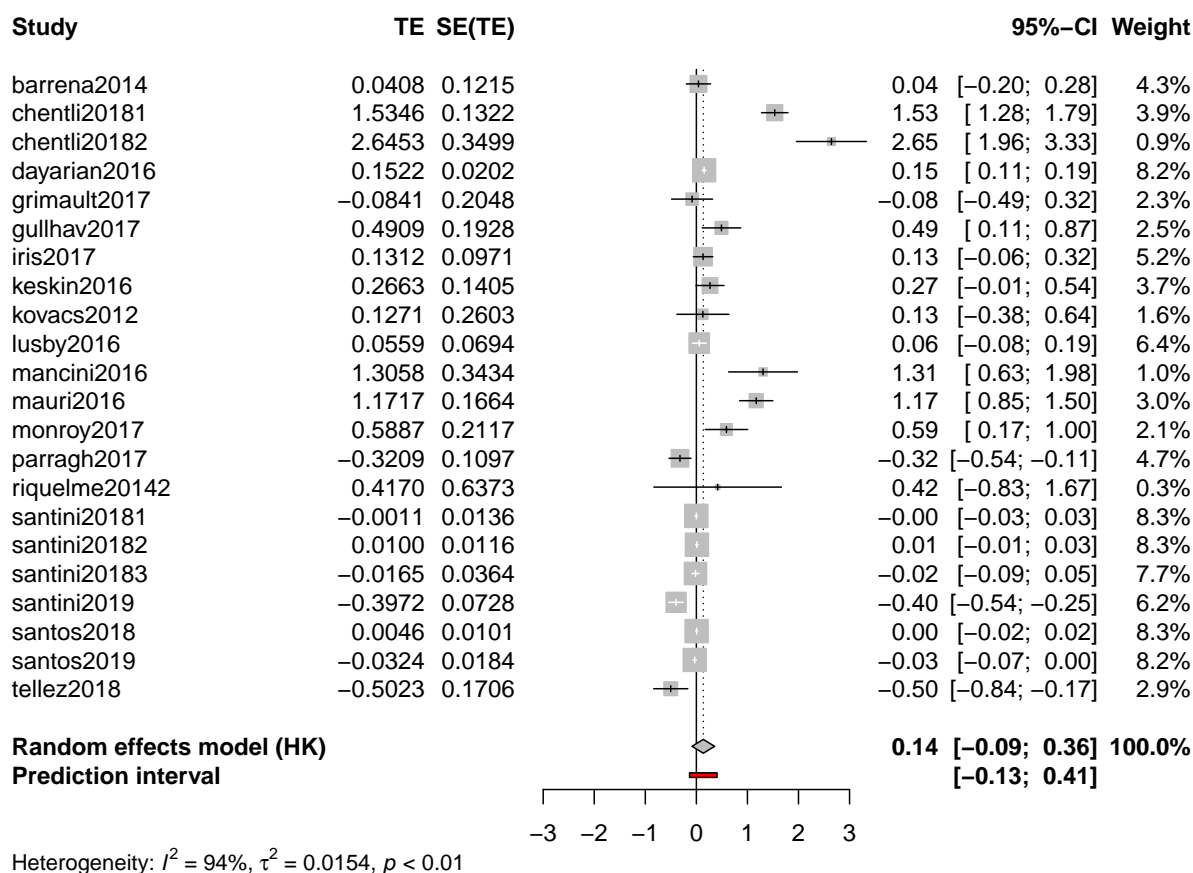


Figure 1: Diagrama de bosque

Según [Valentine, Pigott and Rothstein \(2010\)](#), se considera que al menos 10 estudios son necesarios para llevar a cabo un metaanálisis confiable. No obstante, esta cifra puede variar según diversos factores, como las características requeridas y los recursos disponibles. En este contexto, los 25 estudios utilizados por [Turkeš, Sørensen and Hvattum \(2021\)](#) podrían considerarse ‘pocos’, especialmente al tener en cuenta la complejidad del ALNS. Es importante señalar que la limitación en la obtención de datos, ya que no todos los estudios que emplean ALNS reportan resultados con y sin el componente de adaptabilidad, llevó a la exclusión de aproximadamente 40 estudios de la lista preliminar de elegibles.

Aunque no es posible identificar directamente la presencia de sesgo de manera estadística, existen herramientas que permiten examinar ciertas propiedades de los datos que podrían indicar su presencia. Algunos de estos métodos se enfocan en analizar los efectos de los estudios pequeños y parten de la suposición de que los estudios pequeños tienen más probabilidad de ser víctimas del sesgo de publicación, pues, al no implicar un gran compromiso de recursos y tiempo, si los resultados no son tan favorables suelen quedar archivados (Harrer et al., 2021). Un estudio pequeño se traduce en un error estándar alto (pocas instancias o mucha variación); a pesar de que esto se abordó mediante la eliminación de atípicos, en el diagrama de embudo de la Figura 2 se puede apreciar un posible sesgo de publicación en los datos incluidos para el metaanálisis, ya que los estudios más pequeños (o con error estándar más grande) son los que tienen resultados más positivos.

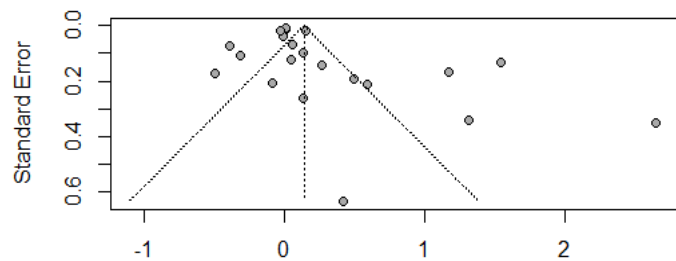


Figure 2: Diagrama de embudo

Conclusiones y trabajo futuro

Como señalan Turkeš, Sörensen and Hvattum (2021), la investigación en metaheurística se ha centrado en desarrollar métodos que resuelvan mejor los problemas de optimización, pero se ha desarrollado poco entendimiento de la razón del desempeño de los métodos y el aporte individual de los componentes. Su trabajo es el primer metaanálisis realizado en este campo de estudio. Las conclusiones de dicho estudio permiten ganar entendimiento relacionado con un camino deseable en la investigación en metaheurísticas, donde se ha reiterado recientemente que la verdadera innovación no se logra a partir de otro método que funcione mejor que sus competidores si no se comprende por qué el método funciona bien.

Por esta razón, como línea de investigación futura se espera utilizar la infraestructura computacional desarrollada en este trabajo para estudiar las metaheurísticas que utilizan post-optimización en la solución problemas de enrutamiento de vehículos con partición de conjuntos, trabajo que actualmente se encuentra en proceso.

References

- Allaire, J. 2012. "RStudio: integrated development environment for R." *Boston, MA* 770(394):165–171.
- Balduzzi, Sara, Gerta Rucker and Guido Schwarzer. 2019. "How to perform a meta-analysis with R: a practical tutorial." *Evidence-Based Mental Health* (22):153–160.
- Crowther, Mark, Wendy Lim and Mark A Crowther. 2010. "Systematic review and meta-analysis methodology." *Blood, The Journal of the American Society of Hematology* 116(17):3140–3146.
- Harrer, Mathias, Pim Cuijpers, Toshi Furukawa and David Daniel Ebert. 2019. *dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'*. R package version 0.1.0.
- Harrer, Mathias, Pim Cuijpers, Toshi Furukawa and David Ebert. 2021. *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall/CRC.
- R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
URL: <https://www.R-project.org/>
- Schryen, Guido and Martina Sperling. 2023. "Literature reviews in operations research: A new taxonomy and a meta review." *Computers & Operations Research* p. 106269.
- Turkeš, Renata, Kenneth Sörensen and Lars Magnus Hvattum. 2021. "Meta-analysis of meta-heuristics: Quantifying the effect of adaptiveness in adaptive large neighborhood search." *European Journal of Operational Research* 292(2):423–442.
- Turkeš, Renata, Kenneth Sörensen, Lars Magnus Hvattum, Eva Barrena, Hayet Chentli, Leandro C Coelho, Iman Dayarian, Axel Grimault, Anders N Gullhav, Çağatay Iris et al. 2020. "Data for a meta-analysis of the adaptive layer in adaptive large neighborhood search." *Data in Brief* 33:106568.
- Valentine, Jeffrey C, Therese D Pigott and Hannah R Rothstein. 2010. "How many studies do you need? A primer on statistical power for meta-analysis." *Journal of Educational and Behavioral Statistics* 35(2):215–247.
- Viechtbauer, Wolfgang. 2010. "Conducting meta-analyses in R with the metafor package." *Journal of Statistical Software* 36(3):1–48.
- Waring, Elin, Michael Quinn, Amelia McNamara, Eduardo Arino de la Rubia, Hao Zhu and Shannon Ellis. 2022. *skimr: Compact and Flexible Summaries of Data*. R package version 2.1.5.
URL: <https://CRAN.R-project.org/package=skimr>
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
URL: <https://ggplot2.tidyverse.org>
- Wickham, Hadley, Romain FranC'ois, Lionel Henry, Kirill MC<ller and Davis Vaughan. 2023. *dplyr: A Grammar of Data Manipulation*. R package version 1.1.3.
URL: <https://CRAN.R-project.org/package=dplyr>
- Xie, Yihui, Christophe Dervieux and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman and Hall/CRC.
URL: <https://bookdown.org/yihui/rmarkdown-cookbook>