FORMATO DE EVALUACIÓN
PRODUCTOS DE INVESTIGACIÓN
ESAP 2025

En el marco del Proyecto ESAP/CINEP 2025, CINEP se ha comprometido a evaluar productos resultado de evaluación con el fin de sugerir ajustes a los documentos de cara a la presentación en revistas académicas o publicación en cualquier formato. En ese marco, le pedimos hacer una evaluación constructiva orientada a la cualificación del texto.

Título	Estadísticas, probabilidades y software aplicados a la administración pública
Autor/a	Heder Eduardo Gómez Sanabria
Disciplina	Administración Pública/ Estadística/ Ciencia de Datos
Tema	Estadística y probabilidad aplicada
Evaluador/a	CINEP/PPP
Tipo de documento (artículo para revista, libro de investigación, manual, compilación etc.).	Libro tipo manual para estudiantes y funcionarios/as

1. Originalidad y aporte:

El objetivo del libro es servir como una guía de estadística y probabilidad para la administración pública en la era del Big Data. Aunque se trata de una propuesta interesante, especialmente por su intención de ofrecer un manual contextualizado al caso colombiano mediante el uso de fuentes como el Departamento Administrativo Nacional de Estadística, la obra no logra cumplir su objetivo.

Los contenidos aún no logran consolidarse como una guía pedagógica dirigida ni a lectores especializados ni a públicos en formación inicial. Tampoco cumple su objetivo de enfocarse en aplicaciones propias de la era del Big Data, pues las herramientas de software y métodos propuestos son tradicionales. Especialmente los software se presentan de manera desconectada de la línea argumentativa y no son aptos para el trabajo con grandes cantidades de datos, por ejemplo, Excel tiene un límite de 1 millón de filas, para trabajar con Big Data se esperaría la incorporación de software con mayor capacidad como R o Python.

2. Claridad

El objetivo del libro se encuentra bien especificado. Se considera que es relevante la construcción de un libro guía pensada para la aplicación actual de la estadística a la administración pública, su valor agregado es pensarlo para el caso Colombiano con ejemplos de bases de datos del país.

Sin embargo, existe una desconexión entre el objetivo y los medios enseñados. Explicar Macros en Visual Basic o usar calculadoras de lógica proposicional no responde al problema de la gestión pública moderna, que requiere agilidad y manejo de grandes volúmenes de datos.

Como se explicará más adelante, los capítulos se presentan de manera desconectada, no guardan un orden lógico por lo que no resulta una guía útil para la aplicación de software en el análisis estadístico, toda vez que dedica docenas de páginas a capturas de pantalla estáticas de interfaces de software (Excel, SPSS, Tableau), lo que puede confundir a un lector.

Finalmente, como se señaló previamente, los softwares utilizados no corresponden a las herramientas más actualizadas para el análisis estadístico ni resultan óptimos para el manejo de grandes volúmenes de información, lo que limita la coherencia del texto con su propósito declarado de situar el análisis en la era del Big Data.

3. Marco teórico y revisión de literatura

Se identifica la necesidad de realizar una revisión exhaustiva del componente conceptual, teniendo en cuenta que el texto presenta algunas imprecisiones teóricas relevantes —por ejemplo, la clasificación de la estadística descriptiva como paramétrica y no paramétrica, así como ciertas interpretaciones poco precisas de la Ley de los Grandes Números y del Teorema del Límite Central— observaciones que se desarrollan con mayor detalle en los comentarios incluidos en el manuscrito.

De igual forma, considerando que el libro propone ubicarse en el contexto de la era del Big Data y en su aplicación a la Administración Pública, sería pertinente fortalecer la revisión de literatura, incorporando referencias centrales tanto del ámbito colombiano como de la econometría y la ciencia de datos actual.

Se sugiere incorporar, entre otras, las siguientes obras:

Como uno de los libros fundamentales para la evaluación de impacto en Colombia:

- Bernal Salazar, R., & Peña, X. (2011). Guía práctica para la evaluación de impacto. Bogotá, Colombia: Universidad de los Andes.

En cuanto a la profundización requerida para el apartado de regresión se sugiere:

- Wooldridge, J. M. (2015). Introducción a la econometría: un enfoque moderno. Cengage Learning.

En coherencia con la sugerencia de actualizar los conceptos de inferencia, muestreo, representatividad y sesgos, sería pertinente incluir literatura contemporánea en inferencia causal:

- Angrist, J. D., & Pischke, J.-S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton, NJ: Princeton University Press.

- Cunningham, S. (2021). Causal Inference: The Mixtape. London: Yale University Press.

Por último, sobre literatura más reciente en ciencia de datos y estadística aplicada a la administración pública se sugiere:

- Wickham, H., & Grolemund, G. (2016). R for Data Science: Import, tidy, transform, visualize, and model data (2ª ed.). O'Reilly Media.

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. New York, NY: Springer.

- Bruce, P., Bruce, A., & Gedeck, P. (2020). Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python (2.ª ed.). O'Reilly Media.

- Scheller, D. S. (2024). Elementary Statistics for Public Administration: An Applied Perspective. Cambridge: Cambridge University Press.

- Weber, J. G. (2024). Statistics for Public Policy: A practical guide to being mostly right (or at least respectably wrong). University of Chicago Press.

- Llaudet, E., & Imai, K. (2023). Data Analysis for Social Science: A Friendly and Practical Introduction. Princeton, NJ: Princeton University Press.

4. Coherencia conceptual y argumentativa

El libro presenta contradicciones internas y errores conceptuales como la clasificación errónea de estadista descriptiva como paramétrica, la falta de delimitación entre muestra y muestra representativa, la confusión en la definición entre muestreo conglomerado y estratificado y problemas en la explicación de la regresión, entre otros aspectos señalados en los comentarios hechos en el manuscrito.

El documento no logra establecer una línea argumentativa clara, dando saltos tanto en la aplicación de los software como de los conceptos. Salta de temas básicos a tutoriales de interfaz de software y luego a conceptos matemáticos avanzados sin una transición pedagógica.

Por su parte, la inclusión del Capítulo 6.3 sobre "Tablas de Verdad" y lógica proposicional es inusual en un texto enfocado en estadística aplicada para administración pública. Aunque la lógica es fundamental, la forma en que está presentada se siente como un anexo desconectado del flujo del análisis de datos.

En línea con lo anterior el capítulo 2 no permite al lector familiarizarse con ningún software y en el resto del libro se dan saltos entre los software, lo que no resulta útil para una persona que se aproxime al material buscando una guía de conceptos y aplicación. Resultaría conveniente estructurar la exposición desde los conceptos básicos hacia los avanzados, utilizando un único software como soporte pedagógico. A este respecto, se recomienda revisar la secuencia y claridad metodológica del libro “Data Analysis for Social Science: A Friendly and Practical Introduction”.

Por último, La estructura del libro debe mejorar en su exposición pedagógica. Como se encuentra redactado expone Inferencia (Intervalos de Confianza, Cap. 3) antes de Probabilidad (Cap. 6) o Distribuciones (Cap. 7), lo que no tiene sentido lógico. Así mismo, como se señaló previamente la incorporación del apartado de tablas de verdad, aunque es importante epistemológicamente, no se conecta con el objetivo del libro.

5. Organización y estructura

El Capítulo 2 presenta un listado de softwares que funciona más como un inventario descriptivo que como una guía aplicada. Los Capítulos 6 y 7 (Fundamentos teóricos) están al final, cuando deberían ser el inicio. Finalmente, algunos apartados dentro de los capítulos combinan temáticas afectando la claridad expositiva y la secuencia lógica del material.

6. Calidad de la escritura

Hay errores ortográficos, se requiere revisión editorial. También hay errores tipográficos en las fórmulas. Se sugiere escribir el texto en LaTeX, pues es el estándar en este tipo de producciones académicas. Es importante revisar la notación, por ejemplo, usa (:) para probabilidad condicional en lugar de (|), siendo este último el estándar. Hay varios diagramas que son ilegibles o cuentan con mala resolución.

7. Rigor académico y ética

El documento requiere ajustes en el rigor matemático. Se identifica, por ejemplo, la referencia a un “algoritmo de Sturges Sterling”, inexistente en la literatura, la denominación correcta es la Regla de Sturges, formulada como k=1+3,322log10 (N). Además, en el texto se le añade una raíz cuadrada a esta regla sin explicar el motivo por el cual se hace.

Aunque se hace uso de referencias, la falta de homogeneidad en la forma de referenciar, la introducción de conceptos inexistentes, el uso inadecuado de algunas fórmulas y la imprecisión en definiciones estadísticas comprometen la solidez académica del texto.

8. Conclusiones

El libro, en su estado actual, no logra responder a los objetivos planteados. Si bien su objetivo es claro, la ejecución falla al entregar herramientas desconectadas y una estructura desordenada. Por lo tanto, en su estado actual, más que aclarar conceptos puede confundir dadas las contradicciones listadas y señaladas en el documento.

12. Consideraciones editoriales

El libro no es apto para ser publicado en su estado actual, requiere un replanteamiento profundo tanto en estructura como en su precisión y profundidad conceptual.

El principal reto es lograr una línea argumentativa que permita un acercamiento pedagógico a las herramientas estadísticas presentadas.

Se sugiere corregir la dispersión de softwares, pues no resulta pedagógicamente viable que los lectores tengan que familiarizarse con más de 5 interfaces, por lo anterior se sugiere usar para la línea argumentativa solo un software que permita trabajar con Big Data como R o Python. En caso contrario se sugiere trabajar con Excel, pero teniendo en cuenta los limitantes que tiene este programa para trabajar con Big Data dado su límite de filas.

De manera adicional se considera irrelevante el uso de calculadoras de tablas de verdad. Aunque la lógica cumple un rol fundamental en las demostraciones matemáticas, para el caso del libro se presenta de manera desconectada de su propósito y consume espacio que se podría usar profundizando temas centrales, como por ejemplo explicar a mayor profundidad la regresión lineal o añadir evaluación de impacto de políticas públicas, este último tema es un eje central para la administración pública.

De manera adicional se debe revisar y corregir las fórmulas matemáticas, estandarizar la notación y ajustar las definiciones conceptuales.