

Reseña

Libro: ANALISIS EXPLORATORIO DE DATOS: NUEVAS TECNICAS ESTADISTICAS

Xiomara Núñez Estupiñán¹
Universidad Nacional de Colombia. Bogotá, Colombia.

Feixa, M., Salafranca, L., Guàrdia, J., Ferrer M &
Turbany, J. (1992) *Análisis exploratorio de datos:
Nuevas técnicas estadísticas.*
Barcelona: PPPU S.A.

El análisis exploratorio de datos es un conjunto de técnicas estadísticas que permite un entendimiento básico de los datos y de las relaciones existentes entre las variables analizadas. El desarrollo inicial de este tipo de técnicas se inició con el trabajo de John Tukey en 1977 *Exploratory Data Analysis*, en el que expuso los principios fundamentales. El análisis exploratorio de datos (A.E.D) proporciona métodos sistemáticos sencillos para organizar y preparar los datos, detectar fallos en la recogida de los mismos, tratamiento de missing o datos perdidos, identificación de outliers (datos atípicos) y comprobación de los supuestos de normalidad, linealidad y homocedasticidad en las técnicas multivariantes.

Para la realización de un adecuado A.E.D se sugieren varios pasos: a) preparación de los datos para hacerlos accequibles a cualquier técnica estadística: examen gráfico de la naturaleza de las variables individuales a analizar y análisis descriptivo numérico para cuantificar algunos aspectos gráficos de los datos; mediante estas representaciones visuales fácilmente se descubren las posibles estructuras de los datos y su comportamiento, b) examen gráfico de las relaciones entre las variables analizadas, análisis descriptivo numérico para cuantificar interrelación, c) identificar posibles casos atípicos (*outliers*) y evaluar el impacto potencial que puedan ejercer en análisis estadísticos posteriores y d) evaluar, si fuera necesario, el impacto potencial que pueden tener los datos ausentes (*missing*) sobre la representatividad de los datos analizados.

En este texto, los autores exponen un análisis detallado de cada sistema representacional que permite detectar de forma rápida, el conjunto de índices descriptivos necesarios para la comprensión de la estructura de los datos a analizar desde una perspectiva exploratoria. El libro se divide en 6 grandes módulos, donde se aborda detalladamente los pasos a seguir y la aproximación estadística para el análisis requerido.

1. Organización, reducción y representación de datos: Reducción del número de condiciones previas para la aplicación de técnicas específicas. La configuración y organización de los datos determinará la selección del modelo que se utilizará para intentar ajustar los datos. Esto implica por lo tanto la selección del método de entrada y codificación de los datos como también la de un paquete estadístico que permita procesarlos de forma correcta. Este módulo a su vez se divide en 4 apartados que corresponden a los índices descriptivos generados a partir de la perspectiva del A.E.D (localización, dispersión, forma y gráficos)

2. Transformación de las variables: transformación de la escala de medida, en lugar de la utilización de las puntuaciones directas. En este módulo se exponen las transformaciones de potencia, transformaciones lineales, transformaciones no lineales y transformaciones monótonas no lineales,

¹ Laboratorio de Psicometría. Universidad Nacional de Colombia. E-mail: xnuneze@unal.edu.co

transformaciones para promover simetría, transformaciones para conseguir dispersión estable, transformaciones comparadas, y transformaciones de las variables tratadas mediante intervalos.

3. **Línea resistente:** También denominada línea de Tukey. Se refiere a un suavizador lineal resistente, línea recta obtenida de la relación de las medianas cruzadas del primer y último tercio de los casos a lo largo de los valores de X .

4. **Técnica de suavizado:** Descubrir cuál es el patrón que mejor se ajusta a los datos sin que sea necesariamente lineal frente a una forma especial de relación en la secuencia de los datos, donde una de las variables se encuentra en un orden determinado. Dentro de los procedimientos de suavizado se pueden encontrar los procedimientos básicos y los procedimientos sofisticados que se encuentran claramente especificados en esta sección.

5. **Ajuste de medianas:** En este apartado se puede observar la serie de técnicas resistentes y robustas para examinar relaciones entre dos o más variables cualitativas y una cuantitativa.

6. **Introducción a la estimación robusta:** Explicación de la forma como se intenta extraer la máxima información de los datos de la muestra. Según el análisis exploratorio de datos se observan dos características fundamentales en los estimadores. La primera, cada vez que sea posible se calculan varios estimadores lo que permite evaluar la constancia. La segunda, se evita realizar hipótesis respecto a la función de distribución de las variables aleatorias estudiadas.

En este texto los autores muestran claramente el análisis exploratorio de datos como herramienta de análisis hecha en profundidad que puede mostrar no sólo cuantiosa información acerca de los datos objeto de análisis, sino que vista como nueva “filosofía” en la aplicación de los métodos de análisis de datos, permite extraer cuanta información sea posible para generar nuevas hipótesis sobre las observaciones de las que se dispone.

