

# Análisis exploratorio de datos

El **análisis exploratorio de datos** definido por John W. Tukey (*E.D.A.: Exploratory data analysis*) es, básicamente, el tratamiento estadístico al que se someten las muestras recogidas durante un proceso de investigación en cualquier campo científico. Para mayor rapidez y precisión, todo el proceso suele realizarse por medios informáticos, con aplicaciones específicas para el tratamiento estadístico. Los *E.D.A.*, no necesariamente, se llevan a cabo con una base de datos al uso, ni con una hoja de cálculo convencional; no obstante el programa SPSS y R (lenguaje de programación) son las aplicaciones más utilizadas, aunque no las únicas.

Por ejemplo, en el campo de la Arqueología el análisis técnico de una pieza puede ser simultáneo a la introducción de los datos, bien porque las fichas estén directamente informatizadas o, bien, porque se usen formularios en papel cuyos datos sean fáciles de introducir en el ordenador o computadora. Es posible, incluso, usar en la propia excavación, una serie de PDAs conectados en red inalámbrica instalada en el yacimiento arqueológico, que envíen numerosos datos de campo a una base de datos central que luego se usarán con fines diversos, entre ellos éste. Los pasos seguidos en el *E. D. A.* son básicamente dos:

- **Medición y descripción** de los datos tecnológicos —tipológicos— y dimensiones, por medio de la Estadística descriptiva. Aquí tenemos, por un lado, las medidas de tendencia central (promedios que, en una sola cifra, resumen todos los valores de una muestra: media, mediana y moda son las más habituales) y, por otro, las medidas de dispersión (que calculan hasta qué punto la muestra se agrupa o no en torno a esos promedios). Dentro de este apartado, se ha de procurar, además, calibrar la confianza de las muestras a través de tres estadímetros básicos: la desviación estándar de la muestra, la curtosis y la asimetría.

*La siguiente es una tabla de ejemplo:*

Estadísticos descriptivos	Longitud	Anchura	Grosor	Peso
Número de Mediciones	383	383	383	383
Error estándar	1%	0,7%	0,7%	8,6%
Medición máxima	142 mm	127 mm	94 mm	1025 g
Medición mínima	29 mm	27 mm	12 mm	16 g
Recorrido	115 mm	98 mm	82 mm	1009 g
Moda	82 mm	60 mm	38 mm	236 g
Mediana	75 mm	61 mm	39 mm	219 g
Rango intercuartil	25 mm	18 mm	17 mm	207 g
Media aritmética	77 mm	62 mm	39 mm	247 g
Desviación estándar	19 mm	14 mm	13 mm	167 g
Coefficiente de variación	25 %	23 %	33 %	68 %
Varianza	376,84	198,67	170,96	27 838,44
Simetría	0,53	0,53	0,48	1,32
Curtosis	0,47	0,83	0,43	2,44

- **Comparación** de los caracteres de una muestra, o de varias muestras diferentes por medio de la Estadística inferencial. Las pruebas más frecuentemente utilizadas comienzan por las más sencillas comparaciones visuales —a través de gráficas como la campana de Gauss, nubes de dispersión o diagramas de caja y arbotantes—, pasando por las socorridas tablas de contingencia (incluido la prueba del  $\chi^2$ ), y por los típicos Análisis de Varianza (que no es más que una confrontación muy precisa de los promedios de varias muestras), hasta llegar a los más complejos análisis multivariantes de conglomerados.



# Fuentes y contribuyentes del artículo

**Análisis exploratorio de datos** Fuente: <http://es.wikipedia.org/w/index.php?oldid=57871212> Contribuyentes: Filemon, Fjherrero, Juan Mayordomo, Locutus Borg, Oblongo, Rosarinagazo, Sabbut, 4 ediciones anónimas

## Licencia

---

Creative Commons Attribution-Share Alike 3.0 Unported  
[//creativecommons.org/licenses/by-sa/3.0/](https://creativecommons.org/licenses/by-sa/3.0/)

---