

Análisis de datos

Bajo la denominación "análisis de datos" se engloba en estadística a un conjunto de métodos descriptivos multidimensionales. Para aplicar estos métodos, se necesita que la información esté organizada de la manera siguiente: "n" individuos estadísticos (entidades espaciales, parejas, firmas,...) descritos por "p" variables. Estos métodos permiten resumir la información contenida en las tablas de datos en dimensiones importantes (tabla de n líneas por p columnas). Se pueden distinguir dos "familias" de métodos:

- Los análisis factoriales: consisten en transformar la tabla de datos inicial en una nueva tabla que contiene la misma información, pero bajo forma *jerarquizada*. Está compuesta de *ejes factoriales*. El primer eje factorial corresponde a la combinación lineal de variables iniciales, que diferencia al máximo a los individuos entre ellos. Este es de *varianza máxima*. Los ejes factoriales son independientes unos de otros y están clasificados en función de su varianza. En general, para extraer lo esencial de la información contenida en la tabla inicial, es suficiente un pequeño número de ejes factoriales (tres o cuatro). La interpretación de estos ejes factoriales permite poner en evidencia la forma de las interrelaciones entre las variables estudiadas, y las semejanzas y diferencias entre los individuos con respecto a esas variables. Los dos métodos más comúnmente utilizados son el análisis en *componentes principales* (adaptado para datos heterogéneos que combinan variables expresadas en escalas de medida diferentes, o incluso para variables expresadas en porcentajes), y el análisis de *correspondencias* (adaptado para tablas de contingencia o variables cualitativas).

- Las clasificaciones: permiten elaborar tipologías y agrupar individuos por clases en función de sus semejanzas con respecto al conjunto de las variables. Un criterio empleado a menudo desde el punto de vista técnico es el de buscar la clasificación que minimiza la varianza intraclase (variabilidad entre los individuos de una misma clase), y maximiza la varianza interclase (variabilidad entre las clases). Los métodos más clásicos son la clasificación *jerárquica ascendente*, y la clasificación por *nubes dinámicas*.

Ver también: Variables cuantitativas

Bibliographie