



Profundiza más

Recurso de Profundización

Reducción de datos por numerosidad

Consiste en reemplazar el conjunto original por una representación más pequeña del mismo. Entre las técnicas más conocidas están el muestreo (sampling) y la agregación de datos utilizando cubos de información.

Muestreo

Un muestreo se puede realizar mediante un muestreo aleatorio simple, seleccionando un N% del total de los datos, donde cada registro tiene la misma probabilidad de ser elegido. También se puede llevar a cabo un muestreo estratificado, en el cual la probabilidad de seleccionar un registro dependerá de alguna característica específica.

Agregación de datos

Consiste en reducir los datos agregándolos según un atributo. Por ejemplo, si se dispone de la información de ventas diarias de una compañía del año pasado y se necesita realizar un análisis para entender el comportamiento de las ventas mensuales, se pueden agregar los datos a nivel mensual en lugar de mantenerlos desglosados por día.

Análisis de Correspondencias

El análisis de correspondencias es una técnica de análisis multivariado utilizada para analizar y visualizar las relaciones entre variables categóricas. Su objetivo principal es representar en un espacio gráfico las correspondencias entre filas y columnas de una tabla de contingencia. El proceso del análisis de correspondencias se puede resumir en los siguientes pasos:

- **Crear la tabla de contingencia:** Se elabora una tabla que muestra la frecuencia de cada combinación de categorías de dos o más variables.
- **Cálculo de correspondencias:** Se calculan las distancias y asociaciones entre las categorías, utilizando medidas como el chi-cuadrado.
- **Representación gráfica:** Los resultados se proyectan en un gráfico bidimensional, donde las categorías cercanas indican una relación fuerte.

El siguiente ejemplo muestra gráficamente un análisis de correspondencias aplicado a los datos de la Encuesta Nacional de Salud y Nutrición (ENSANUT) 2012. Se puede observar una gran similitud entre las personas con "peso bajo" e "hipotensión", así como entre aquellos con niveles de cintura "normales" y "peso adecuado", y las personas con "obesidad" y niveles de cintura "no normales".



Profundiza más

