

Análisis multivariado: Análisis factorial, PCA, Regresión múltiple

MSc. Javier Linkolk López Gonzales, Ing.

IDEUV

16 de mayo de 2021

Tabla de Contenido

1 Preliminares

- Media
- Varianza
- Desviación estándar
- Coeficiente de correlación
- Coeficiente de determinación

2 Análisis multivariado

Tabla de Contenido

1 Preliminares

- Media
- Varianza
- Desviación estándar
- Coeficiente de correlación
- Coeficiente de determinación

2 Análisis multivariado

Media

Las medidas de tendencia central son las que dan un valor representativo a todas las observaciones.

- Siendo la media una medida tan sensible a los datos, es importante tener sumo cuidado con su uso, puesto que, puede dar una imagen distorsionada de los datos.
- Cuando la distribución de los datos es “normal”

$$X \sim N(\mu, \sigma^2),$$

entonces la mejor medida de localización del centro es la media.

Media

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Varianza

Se define a la **varianza**, como la medida obtenida a partir de la suma de los cuadrados de las desviaciones de las observaciones, respecto a su media, dividida por el número de observaciones. Matemáticamente, se expresa de la siguiente forma:

$$s^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n - 1}$$

ó

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n - 1}$$

Desviación estándar

Posee la misma unidad de medida de los datos y es la medida que es utilizada como síntesis de la dispersión o variabilidad. Mide la concentración de los valores en relación a la media aritmética. Matemáticamente, se expresa de la siguiente forma:

$$s = \sqrt{\frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n - 1}}$$

Importante...

En otros términos, cuanto menor sea la desviación estándar, más representativa es la media aritmética; pues, en este caso, la baja dispersión indica que la mayoría de las medidas están próximas de la media y, por tanto, esta representa bien el conjunto de datos.

Coeficiente de correlación

El coeficiente de correlación oscila entre -1 y 1 . La relación puede ser positiva o negativa, directa o inversa. Ejemplo: La correlación entre A y B es de -0.8 .

Coeficiente de determinación

El coeficiente de determinación oscila entre 0 y 1.

Tabla de Contenido

1 Preliminares

- Media
- Varianza
- Desviación estándar
- Coeficiente de correlación
- Coeficiente de determinación

2 Análisis multivariado

ANÁLISIS FACTORIAL

- Es una técnica que se utiliza para hacer reducción de variables en FACTORES.
- Esta técnica va a extraer la máxima variación de las variables.

EXTRACCIÓN DE FACTORES

- PCA
- AF Común
- Método de máxima verosimilitud

¿CUÁNDO USAR ANÁLISIS FACTORIAL?

- Utilizamos AF para analizar la estructura de los datos, evaluando las correlaciones entre las variables.
- El análisis de factores resume los datos en pocas dimensiones, condensando el mayor número de variables en un reducido conjunto de **FACTORES**.

CARGA DE FACTOR

- Es una especie de índice que muestra la importancia de alguna colección de variables.
- Son valores que representan la fuerza de la relación entre **FACTORES** y variables.
- Como regla general, la carga factorial **0.7** o superior representa que el **FACTOR** extraiga la varianza suficiente de la variable.

ANÁLISIS DE COMPONENTES PRINCIPALES

- Es una técnica de la estadística multivariada.
- Utiliza álgebra lineal para transformar datasets en una forma reducida, que normalmente se conoce como reducción de dimensión.
- Se puede escoger el número de dimensiones (componentes principales) en el resultado transformado.

REGRESIÓN MÚLTIPLE

Abordado en R Studio.

GRACIAS

javierlinkolk@gmail.com

👉 <https://sites.google.com/view/javierlinkolk/home>

