



Análisis de Componentes Principales

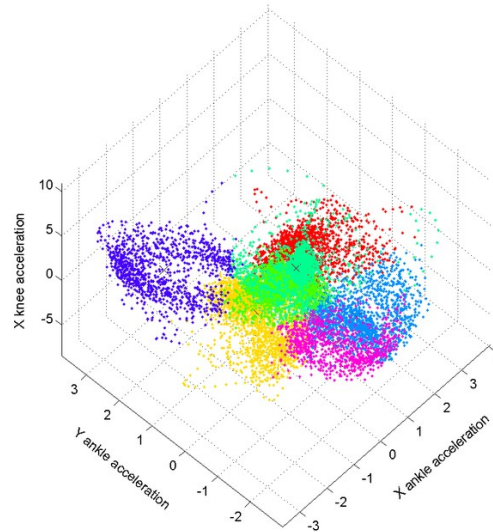
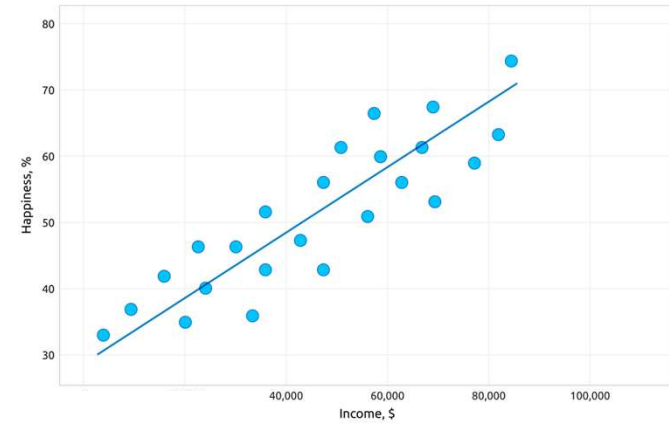
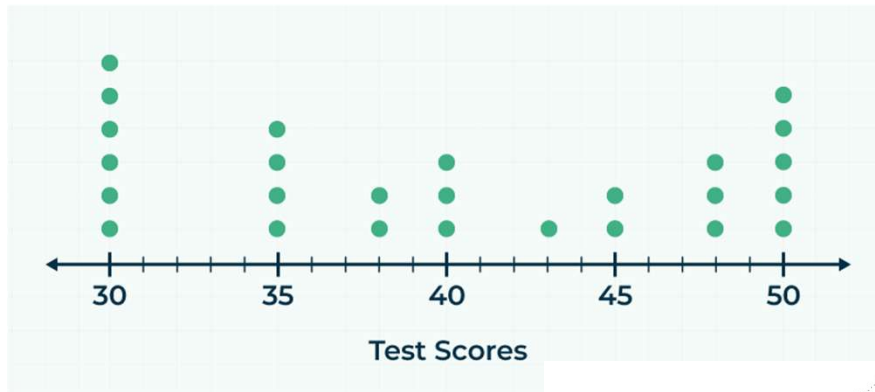
Dra. Mariana-Edith Miranda-Varela

1-sept-2025

Introducción

- Recolección de información
 - Mayor número de variables
 - Se complica visualizar su relación
 - Algunas variables presentan una fuerte correlación
- Reducir el número de variables

Visualización de relaciones



Componentes principales

- Estudiar las relaciones que existe entre p variables correlacionadas
- Transformar el conjunto original (tamaño p) en otro conjunto de nuevas variables incorrelacionadas (tamaño m)

$$m < p$$

- Nuevas variables son combinaciones lineales de las anteriores
- Se construyen según el orden de importancia en cuanto a variabilidad total

Calculo de componentes principales

- Entrada: (x_1, x_2, \dots, x_p)
- Salida (y_1, y_2, \dots, y_m)
- Cada y_i es una combinación lineal de las x_i originales
$$y_j = a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jp}x_p = a'_j x$$
donde a'_j es un vector de constantes y x es el vector de entrada
- Ortogonalidad de la transformación

Función prcomp

- Realiza el análisis de componentes principales (PCA)
 - x – datos a analizar
 - center – centrar las variables restando su media (TRUE* o FALSE)
 - scale. – escalar las variables dividiendo por su desviación estándar (TRUE o FALSE*), BD con unidades diferentes
 - tol – umbral de tolerancia para decidir cuantos CP conservar (NULL*)
 - rank – número de CP que se van a conservar (NULL*)

Función prcomp

- Salida
 - sdev
 - Desviación estándar de los CP, es decir, miden cuanto varían los datos a lo largo de cada componente principal.
 - Valores más altos indican que ese CP explica una mayor parte de la variabilidad en los datos
 - rotation
 - Matriz que contiene los coeficientes de cada variable original en cada CP
Describe cómo se combinan las VO para formar los CP
Columna CP, filas VO
 - center and scale
 - center = True -> medias
 - scale = True -> desviaciones estándar
- Antes de que fueran transformadas

Función biplot

- Visualiza el resultado de PCA
- Genera un gráfico que indica tanto las observaciones como las VO en el espacio de los CP
- ¿Qué muestra?
 - Observaciones (filas) proyectadas en los 2 primeros PC
 - Cercanas entre sí – similares en términos de VO
 - Flechas (VO, columnas)
 - Contribución – flechas largas
 - Dirección y longitud – relación entre VO y PC
 - Direcciones similares – variables correlacionadas
 - Direcciones opuestas – variables con correlación negativa



Actividad Grupal



Análisis de residuos generados en hogares

- Los Residuos Sólidos Urbanos (RSU) corresponden a la generación de residuos de casas habitación.
- Tres R



Dataset

Variable	Tipo	Valores
id_hogar	Discreta	[1, 200]
día	Discreta	[1, 30]
residuos_plasticos	Continua	(0, 2]
residuos_organicos	Continua	(0, 3]
residuos_metal	Continua	(0, 1.5]
residuos_papel_carton	Continua	(0, 2.5]
residuos_total	Continua	Suma de residuos
tipo_vivienda	Categórica	Apartamento, Casa, Cabaña
compromiso	Categórica	Alto, Medio, Bajo
miembros	Discreta	[1, 5]

Previo al planteamiento

1. EDA
2. Tratamiento de nulos
3. Tratamiento de outliers

Problema 1

¿El tipo de vivienda influye en la cantidad de residuos plásticos generados?

Modelos:

i. ANOVA

$$H_0: \mu_{cabaña} = \mu_{apartamento} = \mu_{casa}$$

H_1 : Al menos una media es diferente

ii. Test no paramétrico de Kruskal-Wallis

Problema 2

¿El número de miembros de la vivienda permite predecir la cantidad de residuos_total?

Modelos:

- i. Regresión lineal simple
- ii. Modelo de Regresión Poisson



Avisos



Actividad Grupal

Fecha de entrega
15 de septiembre 23:59

Última sesión

Semana 15

15 al 19 de septiembre

「 muchas gracias. 」