

## Guía 1

Con el objetivo de obtener índices útiles para la gestión hospitalaria basados en técnicas exploratorias se recogió información del Hospital San Pedro correspondiente a los ingresos hospitalarios del período 2018-2019. Se estudiaron las variables habitualmente monitorizadas por el Servicio de Salud, del Sistema Nacional de Salud Colombiana:

NI: número de ingresos,

MO: mortalidad,

RE: número de reingresos,

NE: número de consultas externas,

ICM: índice case-mix, Un índice case-mix superior o inferior a 1 indicará mayor o menor complejidad de los enfermos atendidos en el servicio con respecto al patrón de referencia.

ES: número de camas

IF: índice funcional. Índice global para evaluar los servicios hospitalarios

Realizar el análisis de componentes principales para el conjunto de datos.

### Actividad

#### Ejercicio 1

Sea A la matriz de varianzas y covarianzas:

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 5 \end{pmatrix}$$

Correspondiente al vector aleatorio  $X = (X_1, X_2, X_3)'$  de media 0

- Hallar los autovalores y autovectores de la matriz de varianzas y covarianzas
- Escribir la expresión de las componentes principales  $Y = (Y_1, Y_2, Y_3)'$  e indique que proporción de la variabilidad explica cada una de ellas.
- Hallar las primeras dos componentes principales correspondientes a la observación  $X = (2, 2, 1)$

#### Ejercicio 2

Sea S la matriz de varianzas y covarianzas poblacionales:

$$S = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 4 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

Correspondiente al vector aleatorio  $X = (X_1, X_2, X_3)'$  donde:

$X_1$ : puntuación media obtenida en las asignaturas de econometría

X2: puntuación media obtenida en las asignaturas de derecho

X3: puntuación media obtenida en asignaturas libres

Para alumnos de la carrera de economía

a. Calcule los autovalores de la matriz S.

b. Interprete la segunda componente principal sabiendo que el autovector correspondiente:  $e_1 = (0,5744; -0,5744; 0,5744)$

¿Cómo se debería interpretar si un estudiante tuviera una puntuación en esta componente principal muy inferior a la de sus compañeros?

c. ¿Cuántas componentes principales serán necesarias para explicar al menos el 80% de la variabilidad total del conjunto?

### EJERCICIO 3

El conjunto de datos de clientes.csv mide los consumos en clientes de un supermercado.

id_cliente:	Identificador del cliente.
alimentos_frescos	Consumo en productos de la categoría alimentos frescos.(\$total dólares)
leche	Consumo en productos lácteos. (\$total dólares)
productos_enlatados	consumo en productos de categoría enlatados .(\$total dólares)
alimentos_mar	Consumo en productos de mar. (\$total dólares)
papel_detergentes	Consumo en productos de la categoría papeles y detergentes. (\$total dólares)
alimentos_gourmet	Consumo en productos gourmet .(\$total dólares)

Realizar el análisis de componentes principales sobre el conjunto de datos estandarizado.

- ¿Cuál es la variabilidad explicada por cada una de las componentes principales?
- ¿Según el criterio de káiser cuántas componentes elegiría?
- ¿Cuál es la variabilidad máxima que explican las 2 primeras componentes?
- ¿Cuál es la fórmula de la primera componente principal?
- ¿Interprete las dos primeras componentes principales, cuando toman valores muy altos y cuando muy bajos?
- Realice un biplot y de una explicación de este mismo.