**Universidad del Valle - Escuela de Estadística**

**Curso: Estadística Aplicada 3**

**Profesor: Andrés Felipe Ochoa Muñoz**

**Periodo 2018-2. Grupo 02**

**Laboratorio 1: Análisis de Componentes Principales**

En este laboratorio realice un ACP usando la librerías del software R: ade4 o FactoMineR.

Estudie las relaciones en la nube de individuos y variables, analizando un número de ejes que usted considere adecuado. Utilice los indicadores como las contribuciones y cosenos cuadrados. Además, identifique si es posible realizar un índice usando las variables de la parte activa y relacione el índice con variables suplementarias. Si usted quiere realice: transforme del índice en una escala de 0 a 100, como la propuesta en (<http://www.bdigital.unal.edu.co/3029/1/832213.2010.pdf>). A continuación, se presentan los siguientes casos:

**Caso 1. Base de datos decathlon:**

Una base de datos con 41 filas y 13 columnas: los primeros diez columnas corresponde a la actuación de los atletas para los 10 eventos del decatlón. Las columnas 11 y 12 corresponden, respectivamente, al ranking y la puntuación obtenida. La última columna es una variable categórica que corresponde al evento deportivo (2004 Juego Olímpico ó Decastar)

El objetivo de este estudio es conocercuáles son las relaciones entre las competencias de unos deportistas olímpicos de varios países del mundo.

Trabaje las variables “Rank” “Points” y “Competition” como variables suplementarias

**Unidades de medida**

100m (s) ;Long.jump(m); Shot.put (kg); high.jump (m) ; 400m(s); 110m.hurdle (s) ;

Discus (m); Pole.vault (m); Javeline (m);1500m (s)

**Caso 2. Base de datos Consumo de carne**

Se dispone de la estimación del consumo promedio de proteínas, en gramos, por persona y por día, en Europa (datos del año 1981). Se considera el consumo para:

• x1: Carnes de animales rumiantes (RUMI),  
• x2: Carnes de cerdo y aves de corral (AVES),  
• x3: Huevos (HUEV),  
• x4: Leche (LECH),  
• x5: Carne de pescado (PESC),  
• x6: Cereales (CERE),  
• x7: Comidas almidonadas (ALMI),  
• x8: Semillas oleaginosas y semillas secas (LEGU),  
• x9: Frutas y verduras frescas (VERD).

El objetivo de estudio es analizar las dinámicas asociadas con los hábitos alimenticios de los países. Para la interpretación de los resultados, se le aconseja consultar un mapa de Europa.

**Caso 3. Base de datos Importaciones**

La base de datos contiene las importaciones hechas por los países suramericanos, éstas provenientes de Estados Unidos, entre 1991 y 2010.  
Efectúe el A.C.P. de esta tabla. ¿Cuántos ejes es suficiente considerar? En el plano principal, se pueden trazar líneas entre los puntos que corresponden a años consecutivos; estas líneas se llaman trayectorias. ¿Qué puede interpretar de estas trayectorias?

**Caso 4. Base de datos migración en aeropuertos Colombianos**

En la página migración del gobierno se reporta cuantas personas entran y salen de los aeropuertos. Esta base de datos consta de información desde enero del 2016 hasta julio del 2017. El objetivo de este estudio es analizar los flujos identificando en que periodos salen o entran más colombianos. Si usted considera pertinente utilice elementos suplementarios para mejorar el ACP.

http://migracioncolombia.gov.co/index.php/es/component/content/article?id=718

**Caso 5. Base de datos PIB.**

El archivo *PIB\_2015.xlxs* contiene la información del PIB desglosado por sectores económicos para cada uno de los departamentos además del distrito capital Bogotá, con esta base de datos realice un análisis de Componentes Principales considerando como individuos los *Departamentos* y como variables los 10 *Sectores Económicos.* Identifique diferentes opciones para corregir el problema de datos atípicos en el ACP.

**Caso 6. Calidad del aire**

Se dispone de una muestra de 41 ciudades de USA en las que se midieron diferentes variables relacionadas con la contaminación atmosférica.

* Contenido en SO2
* Temperatura anual en grados F.
* Número de empresas mayores de 20 trabajadores.
* Población (en miles de habitantes).
* Velocidad media del viento.
* Precipitación anual media.
* Días lluviosos al año.

Es de interés investigar la relación entre la concentración en SO2 y el resto de variables, utilice las herramientas que dispone el análisis de componentes principales y conforme un análisis con sus resultados.

**Caso 7. Aplicación en leche**

Se cuenta con la base de datos leche que se encuentra en la librer´ıa de FactoMineR, est´a tiene 85 leches descritas por 6  
variables las cu´ales son: densidad, contenido de grasa, prote´ına, case´ına, seco, rendimiento

**El conjunto de datos se encuentra en el paquete FactoMineR (data(milk))**

**Caso 8. Análisis morfometrico en tortugas**

En un estudio realizado por Jolicoeur & Mossimann (1), se tomaron 84 especiménes de tortugas (de la especie Chrysemys Picta Marginada) el 2 de agosto de 1956 a 35*millas* del sur de Montreal, Canadá. Luego seleccionaron una muestra representativa de 24 machos y hembras y a cada caparazón le midieron la Longitud, el Ancho y el Altura. El objetivo del estudio era poder modelar el crecimiento del caparazón de las tortugas identificando aquellos factores que describían la variación en el tamaño y la forma del caparazón

**Aspectos a tener en cuenta**

Conforme un informe en Latex de máximo 10 páginas donde muestre los resultados encontrados en el ACP. Piense en otras metodologías que puedan mejorar o complementar el ACP, como los modelos regresión o algún método de clasificación. Recuerde que aleatoriamente se seleccionaran los grupos en cada caso, igualmente se seleccionarán 4 grupos para que realicen una exposición con los resultados.