

Análisis Estadístico Multivariado | UTDT

Problem Set 1

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Ejercicio 1

1. En el archivo *eurosec.dta* es una base de datos que contiene los porcentajes de empleo de los distintos sectores económicos para un grupo de países europeos. Los sectores son: S1 (Agricultura), S2 (Minería), S3 (Industria), S4 (Energía), S5 (Construcción), S6 (Servicios Industriales), S7 (Finanzas), S8 (Servicios), S9 (Transporte y Telecomunicaciones).
 - (a) Obtenga la media, varianza y el coeficiente de variación asociados a cada una de las variables.
 - (b) Obtenga la matriz de varianzas y covarianzas.
 - (c) Obtenga las medidas globales de variabilidad.
 - (d) Obtenga la matriz de correlaciones.
 - (e) Regrese el porcentaje de empleo en el sector Minería, respecto de las variables restantes. Obtenga la varianza de los residuos y el coeficiente de determinación. ¿Es posible obtener las estimaciones a partir de la matriz de varianzas y covarianzas?
 - (f) Obtenga el coeficiente de correlación parcial entre los porcentajes de empleo en el sector agrícola respecto del sector minero. ¿Cómo efectuaría el cómputo a partir de la matriz de varianzas y covarianzas?
 - (g) Obtenga los autovalores asociados a la matriz de correlaciones. ¿Existe alguna relación entre el número de variables y los mismos?
 - (h) Obtenga el coeficiente de dependencia efectiva.
 - (i) Comente brevemente la información que brinda la matriz de precisión.

Ejercicio 2

2. En el archivo *individual_t410.dta* encontrará el corte por personas de la Encuesta Permanente de Hogares correspondiente al cuarto trimestre del año 2010. Se propone la construcción de una base para analizar la estructura de la muestra ocupada mayor de 15 años de edad, por aglomerado y por rama de actividad, de acuerdo con la clasificación CAES-Mercosur, considerando las grandes ramas: Agricultura, Ganadería, Caza y Pesca; Minería; Industria; Energía; Construcción; Comercio; Correo y Telecomunicaciones; Ss. Financieros; Otros Ss.; y Administración Pública.
- (a) Analice la variabilidad de la proporción de ocupados por grandes de ramas de actividad.
 - (b) Analice la estructura de correlaciones entre las proporciones de ocupados de las ramas de actividad consideradas.
 - (c) Regrese la proporción de ocupados de la rama servicios, respecto de las proporciones observadas en las ramas restantes. Obtenga la varianza de los residuos y el coeficiente de determinación. ¿De qué otra forma hubiera podido obtener estas estimaciones?
 - (d) Obtenga el coeficiente de correlación parcial entre los porcentajes de ocupados en el sector comercial respecto del sector Otros Ss.
 - (e) Obtenga los autovalores asociados a la matriz de correlaciones.
 - (f) Proponga una métrica que resuma la dependencia entre las proporciones de ocupados entre los distintos sectores.

Ejercicio 3

El archivo *records.dta* contiene información sobre records obtenidos por atletas de diferentes nacionalidades en varias especialidades. La siguiente tabla muestra la descripción del contenido de cada variable.

m_100	100 metros
m_200	200 metros
m_400	400 metros
m_800	800 metros
m_1000	1000 metros
m_1500	1500 metros
km_5	5 kilómetros
km_10	10 kilómetros
maraton	42 kilómetros

- (a) Realice un análisis descriptivo de los datos. Señale si en el caso de querer efectuar un análisis de componentes principales, recomendaría la estandarización de las variables.
- (b) Realice un análisis de componentes principales.
- (c) Efectue la selección de los componentes principales de acuerdo con los siguientes modos:
 - i. Búsqueda del “codo”.
 - ii. Graficando el porcentaje de varianza explicada por cada componente.
- (d) Grafique las componentes en pares, en función de lo determinado en el punto anterior. Interprete.