

Análisis de Datos (CO5316)
Enero-Marzo 2020
E1: 25 %, Duración: n horas.
+ Tarea 4 %

Recuerde que el examen es individual.

En el archivo “Elecciones.csv” tienen los porcentajes de votación, por estado, para los candidatos republicanos y demócratas en las últimas diez elecciones presidenciales en los Estados Unidos. Tomando tres años aleatorios, tienen los siguientes ejercicios:

1. Análisis de Componentes Principales:

Realice un análisis de componentes principales sobre los porcentajes de votación por partido. Está debe incluir con cuantos componentes se debe trabajar luego del análisis y que expresan cada uno de esos componentes.

2. Análisis de Conglomerados:

Realice un análisis de conglomerados utilizando dos métodos jerárquicos + el método de k-medias para agrupar los 50 estados + el Distrito de Columbia según los porcentajes de votación. Expresé el número de conglomerados que escogería, y que entidades estarían en cada uno de esos conglomerados.

3. Análisis de Correlación Canónica:

Realice un análisis de correlación canónica de las variables que se refieren a los porcentajes de votación por el partido republicano contra las variables que se refieren a los porcentajes de votación por el partido demócrata. Escriba cuales son los modelos lineales que obtendría, y cuales son las correlaciones entre las nuevas variables y las variables originales.

4. Análisis de Correspondencias:

Para cada uno de los años cuente el número de estados para el que la diferencia entre los porcentajes caen en las siguientes categorías:

- > 10 % a favor de republicanos.
- Entre 2 % y 10 % a favor de republicanos.
- < 2 % de diferencia entre los dos partidos.
- Entre 2 % y 10 % a favor de demócratas.
- > 10 % a favor de demócratas.

Cree una tabla de contingencias bajo con esta clasificación y la variable Región (REG), y realice un análisis de correspondencias para relacionar, la diferencia en porcentajes con la región de los Estados Unidos, ¿qué puede decir acerca de estas dos clasificaciones?

5. Análisis Discriminante:

En el archivo de datos pueden encontrar una variable “Tend” que expresa la tendencia usual del estado, “REP” de republicano, “DEM” de demócrata y “BGS” si suele ser un estado peleado. Realice un análisis discriminante, y diga cuántas y cuales funciones discriminantes utilizaría para dividir los estados + el Distrito de Columbia. Realice también la división, y diga cómo quedaría dicha división. ¿Qué porcentaje de los datos son reclasificados?

6. Análisis Factorial:

Realice un análisis factorial sobre los datos. Consiga el número de factores que serían suficientes para explicar los datos. Escriba los modelos de cada uno de los factores, y los puntajes factoriales que se obtendrían.

7. Escalamiento Multidimensional:

Encuentre una distancia entre los estados a partir de los datos. Utilizando escalamiento multidimensional, realizar un mapa de los estados + el Distrito de Columbia. ¿Qué agrupaciones podría realizar?

De estos siete ejercicios, **solo deben realizar cinco de ellos**. Cada uno tendrá un valor de 5 puntos. Adicionalmente, deberán escoger un ejercicio, bien sea el de Análisis Discriminante, Análisis Factorial o Análisis de Correlación Canónica; para que valga por los 4 puntos de la tarea. Debe dejar claramente expresado cuales son los cinco problemas que resuelve por el parcial, y cuál es el que resuelve por la tarea.

Deberán trabajar con los datos de las siguientes elecciones:

Alimi	"1980"	"1992"	"2004"	"2012"
Andrés	"1984"	"1988"	"2004"	"2008"
Cristian	"1992"	"1996"	"2008"	"2016"
Joseph	"1980"	"1988"	"1996"	"2004"
Laura	"1988"	"1996"	"2004"	"2016"
Leonardo	"1984"	"2000"	"2008"	"2012"
Luis	"1984"	"1992"	"2000"	"2016"
Orlando	"1980"	"1988"	"2000"	"2012"
Rafael	"1980"	"1996"	"2012"	"2016"

Deben entregar un archivo en **.pdf** con las respuestas a las preguntas que decidieron responder.