

# Estadística Multivariada: Análisis de Correspondencia.

Andrés García

Fecha de entrega: 27 de Mayo de 2020.

**Ejercicio 1.** Investigue y *describa* (20 pts) la formulación de Análisis de Correspondencia mediante Mínimos Cuadrados Ponderados.

**Ejercicio 2.** (10 pts) Considere una tabla de contingencia  $f$  de dimensión  $n \times n$  con elementos diferentes de cero sólo en la diagonal ¿Como espera que sean las frecuencias relativas  $r_i$  y  $c_i$ , para  $i = 1, \dots, n$ ? ¿Que consecuencias tiene esto?

**Ejercicio 3.** (30 pts) Considere los datos dados en el archivo adjunto (FrenchFoodData.csv). El conjunto de datos consiste en el gasto promedio en ciertos alimentos para diferentes tipos de familias en Francia (manual workers D MA, employees D EM, managers D CA). Los datos se tomaron de Lebart, Morineau, y Fénelon (1982). Dado que todas las variables son medidas en la misma unidad (Francos franceses), explique como es que pueden verse estos datos como una tabla de contingencia. Realice el análisis de correspondencia y compare los resultados con los obtenidos mediante PCA en el análisis de Härdle y Simar (2015, Chap. 11).

Nota: no olvide obtener las inercias, determinar si las variables son independientes, e interpretar los resultados adecuadamente.

**Ejercicio 4.** (10 pts) Para el ejemplo de puntuaciones visto en clase:

- ¿Cuáles son las puntuaciones escaladas correspondientes (del 0 al 10)?
- Grafique la proyección de los profesores y las puntuaciones, e interprete los resultados apropiadamente. Para lo último se puede apoyar de la inercia y significancia estadística.

**Ejercicio 5.** (20 pts) Este conjunto de datos fue estudiado por Madsen (1976) para las condiciones de alojamiento en áreas seleccionadas de Copenhague, Dinamarca.

Un total de 1,681 residentes rentando en casas construidas durante el periodo de 1960-1968 fueron seleccionados para contestar una encuesta acerca de su nivel de satisfacción (categorizado como bajo (ls), medio (ms), alto (hs)), el nivel de contacto con otros residentes (bajo (lc), alto (hc)), y su sensación de influencia sobre el manejo de su vivienda (bajo (li), medio (mi), alto (hi)). Las unidades de renta fueron categorizadas como *tower blocks* (tb), apartamentos (ap), *atrium houses* (ah) y casas con terraza (th).

El proposito del estudio es identificar si existe algun asociación entre el grado de contacto, influencia, satisfacción, y el tipo de vivienda.

- Aplique el Análisis de Correspondencia Múltiple al archivo adjunto y explique las asociaciones principales de cada eje.
- Diga si existen asociaciones significativas entre los pares de variables categóricas al nivel de significancia de 0.05.
- Explique y justifique el por qué es posible reducir dimensión directamente sobre la tabla  $(\frac{1}{nQ}\mathbf{B})$  cuando naturalmente uno debería de realizar el análisis sobre los bloques no diagonales de  $\mathbf{B}$ .
- Describa las implicaciones de los resultados obtenidos.

**Ejercicio 6.** (10 pts) Estimar  $\hat{\beta}_1$  y  $\hat{\beta}_2$  para el ejemplo guía visto en la clase de JMCA. Representar el mapa bidimensional y dar la inercia asociada en esta proyección de dos dimensiones. No olvide interpretar los resultados.