

Análisis Estadístico Multivariado

Problem Set 0

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Resuelva los siguientes ejercicios en R y en Stata.

Ejercicio 1

Abra el archivo Hogar_t403_0.dta, el cual contiene la base de microdatos correspondiente al relevamiento realizado en Argentina en el marco de la Encuesta Permanente de Hogares durante el cuarto trimestre de 2003.

- (a) ¿Qué tipo de variable identifica la vivienda?
- (b) ¿Cuántos hogares fueron relevados en total?
- (c) Genere una nueva base de datos (Hogar_t403_1.dta) que reúna únicamente los códigos para distinguir viviendas, hogares, año, semestre, region y las variables que reflejan la representatividad del hogar, la cantidad de miembros en el hogar, el monto total del ingreso familiar, y el monto de ingreso per cápita familiar.
- (d) Ayudándose por el archivo Hogar_t403_2.raw obtenga el monto promedio de ingreso total familiar de los hogares correspondientes al aglomerado Gran Resistencia, que componen la muestra.

Ejercicio 2

A partir de la base obtenida en el inciso (d) del ejercicio anterior, construya la matriz $X_{n \times p}$ (con $(n > p)$), cuyas columnas estarán dadas por las variables que representan la cantidad de miembros del hogar, el monto total del ingreso familiar, y el monto de ingreso per cápita familiar, y sus filas por las primeras 600 observaciones.

- (a) Obtener las matrices X' , $X'X$, XX' , $(X'X)^{-1}$. Antes de hacer cuentas considere de qué tamaño deben ser las matrices resultantes.

(b) Verifique que, dada la matriz,

$$B = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

vale la siguiente igualdad:

$$(X'X + B)' = (X'X)' + B'.$$

(c) Obtener la traza y el determinante de la matriz $X'X$.

(d) Sea $\lambda = 1/1000$. Verifique que

$$|\lambda X'X| = \lambda^n |X'X|.$$

Ejercicio 3

Dadas las matrices

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 23 & 2 \\ 22 & 15 \\ 3 & 7 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 3 \\ 2 & 13 \end{bmatrix}, \quad D = \begin{bmatrix} 34 & 25 & 2 \\ 3 & 14 & 32 \end{bmatrix},$$

(a) Verificar que

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$$

(b) Sea $A_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ una partición de la matriz A . Verifique que

$$(A_{11} + C)^{-1} = C^{-1}(A_{11}^{-1} + C^{-1})^{-1}A_{11}^{-1}$$

Ejercicio 4

Dada la matriz

$$B = \begin{bmatrix} 1 & 1 & 2 & 3 & 2 & 5 & 5 & 2 \\ 1 & 2 & 1 & 2 & 1 & 3 & 3 & 6 \\ 1 & 2 & 3 & 2 & 3 & 1 & 5 & 10 \\ 2 & 4 & 4 & 57 & 5 & 7 & 9 & 1 \\ 2 & 5 & 5 & 5 & 5 & 4 & 4 & 2 \\ 3 & 6 & 5 & 5 & 5 & 6 & 5 & 6 \\ 1 & 7 & 4 & 5 & 6 & 7 & 7 & 8 \\ 3 & 8 & 5 & 6 & 3 & 1 & 1 & 8 \end{bmatrix}$$

obtenga el determinante y la inversa de la matriz, considerando sus particiones:

$$B_{11} = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 1 & 2 & 1 & 2 \\ 1 & 2 & 3 & 2 \\ 2 & 4 & 4 & 57 \end{bmatrix}, \quad B_{12} = \begin{bmatrix} 2 & 5 & 5 & 2 \\ 1 & 3 & 3 & 6 \\ 3 & 1 & 5 & 10 \\ 5 & 7 & 9 & 1 \end{bmatrix}, \quad B_{12} = \begin{bmatrix} 2 & 5 & 5 & 5 \\ 3 & 6 & 5 & 5 \\ 1 & 7 & 4 & 5 \\ 3 & 8 & 5 & 6 \end{bmatrix}, \quad B_{22} = \begin{bmatrix} 5 & 4 & 4 & 2 \\ 5 & 6 & 5 & 6 \\ 6 & 7 & 7 & 8 \\ 3 & 1 & 1 & 8 \end{bmatrix}$$

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.

Análisis Estadístico Multivariado | UTDT

Problem Set 2

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Ejercicio 1

El archivo *ine_dta* es una base de datos que contiene los gastos promedio, en euros, de los hogares españoles, por grandes rubros y comunidad autónoma, correspondientes a los relevamientos de la encuesta de presupuestos familiares del año 2005 realizada por el Instituto Nacional de Estadísticas de España (INE). La descripción de las variables y de la base puede consultarse en las etiquetas asociadas a cada una.

- (a) Realice un análisis descriptivo de los datos.
- (b) Realice un análisis de componentes principales.
- (c) Efectúe la selección de los componentes principales de acuerdo con los siguientes modos:
 - i. Búsqueda del “codo”.
 - ii. Búsqueda por umbral de varianza a explicar, considerando un 80%.
 - iii. Búsqueda por tope mínimo al valor de los eigenvalores, considerando la varianza media.
- (d) Interprete los componentes seleccionados. Para ello, puede emplear el archivo *renta.csv*.

Aclaración: el archivo *mmta_csv* contiene información referida al ingreso promedio por hogar, por comunidad; relevado por el INE para el año 2005.

Ejercicio 2

En relación al ejercicio anterior, realice el *biplot* asociado a las dos dimensiones principales, considerando un parámetro $c = 0.5$.

Ejercicio 3

Teniendo en cuenta la misma base de datos del primer punto, efectúe los *biplots* correspondientes a las dos dimensiones principales y a la primera y tercera, considerando en ambos casos, un parámetro $c = 0$. Comente brevemente el resultado en relación a lo obtenido mediante el análisis de componentes principales efectuado en el primer ejercicio.

Ejercicio 4

La base *wb.dta* contiene información relevada por el Banco Mundial en 77 países durante el año 2007 (los detalles de las variables relevadas se pueden observar en las etiquetas asignadas a las mismas). A partir de estos datos, se pide:

- (a) Realice un análisis normado de componentes principales a partir de los datos.
- (b) ¿Cuántos componentes sugiere extraer?
- (c) ¿Cuál es el porcentaje de variabilidad total explicado por las componentes seleccionadas?
- (d) ¿Qué interpretación sugiere de las componentes, con arreglo a las correlaciones existentes con las variables originales? Para este punto puede ayudarse con la estructura de análisis presente en el libro *P2.xls*.
- (e) Clasifique a los países de acuerdo con las dos componentes principales.
- (f) Realice un biplot considerando las dos dimensiones principales de análisis para un parámetro $c = 0.5$. ¿Qué puede decir sobre la posición de los Estados Unidos y China, y su influencia en el análisis, a la luz de los resultados obtenidos?

Ejercicio 5

La base *hspendusa2009.dta* contiene información relevada durante los años 2008-2009 por el Instituto de Estadísticas de los Estados Unidos, correspondiente a los gastos medios de los hogares por capítulos y por área metropolitana (los detalles de las variables se pueden observar en el archivo *usa.xls*). Efectúe un análisis de componentes principales.

Análisis Estadístico Multivariado | UTDT

Problem Set 3

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Ejercicio 1

Utilizando la base de datos *firmas.dta*, la cual contiene información financiera referida a dos tipos de firmas: 12 firmas consideradas de buena performance, y 12 firmas no tan buenas en ese sentido; y dispone de dos características relevada: EBITASS (gcias. despues de impuestos e intereses sobre activos; y ROTC (retorno sobre el capital), se pide:

- (a) Realizar un análisis de clusters empleando el algoritmo de las k -medias. Determinar el número de grupos a partir de un test F de reducción de variabilidad.
- (b) Realizar un análisis de conglomerados empleando métodos jerárquicos. Trabajar con distintas medidas de distancias entre clusters, comentando los pros y las contras de cada una de ellas.

Ejercicio 2

Utilice la base de datos *iris* para realizar un análisis de conglomerados. La base contiene datos de especies de flores de iris con registros de las medidas del sépalo y pétalo de cada una. Originalmente los datos ya están agrupados por especie, pero en este ejercicio se propone utilizar *k-means clustering* para aplicar el método y ver cómo se compara con las categorías verdaderas.

Nota: la base *iris* ya está precargada en R, porque podemos acceder a ella directamente.

Ejercicio 3

El archivo “govt_bond_returns.xlsx” contiene los rendimientos anuales para los inversores (a partir de septiembre de 2010) de bonos del gobierno, enumerados por duración y emisor. Estos retornos están influenciados por las respectivas tasas de inflación y fortalezas de las diversas monedas, así como las percepciones del mercado para estas tasas en el futuro. Realizar un agrupamiento jerárquico (*hierarchical clustering*) de estos retornos a identificar países con situaciones financieras similares. Comente como el dendrograma agrupa países en regiones geográficas similares.

Análisis Estadístico Multivariado | UTDT

Problem Set 4

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Ejercicio 1

Construya un test de normalidad multivariada, y aplíquelo sobre los datos correspondientes al primer ejercicio de la práctica 2. ¿Puede concluir que el vector de medias de los datos es distinto del vector nulo?

Recordemos del ejercicio 1 de la práctica 2 que el archivo *ine_dta* es una base de datos que contiene los gastos promedio, en euros, de los hogares españoles, por grandes rubros y comunidad autónoma, correspondientes a los relevamientos de la encuesta de presupuestos familiares del año 2005 realizada por el Instituto Nacional de Estadísticas de España (INE).

```
ine <- haven::read_dta("C:/Users/fiona/Dropbox/Materias MAECO/Análisis Estadístico Multivariado/Practica
mardia(ine[,2:ncol(ine)])
```

```
## $mv.test
##           Test Statistic p-value Result
## 1      Skewness    302.0658  0.2459   YES
## 2      Kurtosis     0.2887  0.7728   YES
## 3  MV Normality      <NA>    <NA>   YES
##
```

```
## $uv.shapiro
##           W      p-value UV.Normality
## alybnh    0.7665 5e-04    No
## vestcal    0.9284 0.1819   Yes
## vivagelo    0.9725 0.8428   Yes
## mobymant    0.982  0.9682   Yes
## salud      0.9499 0.4232   Yes
## transp     0.9292 0.1881   Yes
## comu       0.9715 0.8249   Yes
## ocio       0.9851 0.9872   Yes
## educ       0.9489 0.4083   Yes
## esparc     0.9473 0.3837   Yes
## otros      0.9468 0.3765   Yes
```

```
k <- 11
N <- 18
(6*((k+1)*(N+1)-6))*302.0658/(k+1)*(N+1)*(N+3)
```

```
## [1] 13378192
```

```
0.2887*sqrt(8*k*(k+2)/N)+k*(k+2)*(N-1)/(N+1) #para la igualdad con STATA
```

```
## [1] 130.2489
```

```
qchisq((1-0.2459),(11*12*13)/6)
```

```
## [1] 302.0691
```

Ahora realizamos un test de significatividad sobre el vector de medias:

```
ICSNP::HotellingsT2(ine[,2:ncol(ine)])
```

```
##  
## Hotelling's one sample T2-test  
##  
## data: ine[, 2:ncol(ine)]  
## T.2 = 70.918, df1 = 11, df2 = 7, p-value = 4.258e-06  
## alternative hypothesis: true location is not equal to c(0,0,0,0,0,0,0,0,0,0,0)
```

Rechazo la hipotesis nula de que el vector de medias sea el vector nulo.

Ejercicio 2

Generar una muestra de 100 observaciones de una población normal multivariada teniendo en cuenta los siguientes vectores de parámetros:

$$\mu = \begin{bmatrix} 2 \\ 0.45 \\ 0.23 \\ 0.54 \\ 0.12 \\ 0.63 \\ 0.66 \\ 0.32 \end{bmatrix}, \quad \sigma = \begin{bmatrix} 0.67 \\ 0.89 \\ 0.56 \\ 0.90 \\ 0.56 \\ 0.34 \\ 0.76 \\ 0.13 \end{bmatrix}$$

- Efectúe un test para verificar la normalidad de la base de datos construída.
- ¿Existe evidencia empírica suficiente que permita concluir que el vector de medias es estadísticamente distinto del vector nulo?
- Efectúe un test para contrastar la hipótesis acerca de la igualdad de las medias de las variables que componen la matriz de datos.
- Contraste la esfericidad de la distribución de los datos.
- ¿Se verifica estadísticamente la diagonalidad de la matriz de varianzas y covarianzas en los datos?

Ejercicio 3

Repetir el ejercicio anterior, considerando en este caso la siguiente estructura:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \quad X_1 \sim N(\mu_1, \Sigma_1), \quad X_2 \sim N(\mu_2, \Sigma_2), \quad X_3 \sim N(\mu_3, \Sigma_3).$$

siendo

$$\mu_1 = \begin{bmatrix} 2 \\ 0.45 \\ 0.23 \\ 0.54 \\ 0.12 \\ 0.63 \\ 0.66 \\ 0.32 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0.67 \\ 0.89 \\ 0.56 \\ 0.90 \\ 0.56 \\ 0.34 \\ 0.76 \\ 0.13 \end{bmatrix}, \quad \pi_1 = 40\%$$

$$\mu_2 = \begin{bmatrix} 3 \\ 1.45 \\ 1.23 \\ 1.54 \\ 1.12 \\ 1.63 \\ 1.66 \\ 1.32 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 1.67 \\ 1.89 \\ 1.56 \\ 1.90 \\ 1.56 \\ 1.34 \\ 1.76 \\ 1.13 \end{bmatrix}, \quad \pi_2 = 50\%$$

$$\mu_3 = \begin{bmatrix} 4 \\ 2.45 \\ 2.23 \\ 2.54 \\ 2.12 \\ 2.63 \\ 2.66 \\ 2.32 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 2.67 \\ 2.89 \\ 2.56 \\ 2.90 \\ 2.56 \\ 2.34 \\ 2.76 \\ 2.13 \end{bmatrix}, \quad \pi_3 = 10\%$$

- Comente lo observado en contraste con el ejercicio anterior.
- Efectúe similares contrastes pero considerando cada partición de X por separado.
- Proponga soluciones al problema observado, incluso considerando la eventualidad de que existan valores atípicos en la muestra que conlleven al mismo. Aplica la/s soluciones sufridas y comentar brevemente el resultado encontrado.

Análisis Estadístico Multivariado | UTDT

Problem Set 5

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Ejercicio 1

El archivo *ind.dta* contiene información sobre 17 indicadores económicos:

Variable	Indicador
X1	Agricultural land (% of land area)
X2	Agriculture, value added (% of GDP)
X3	CO2 emissions (metric tons per capita)
X4	Exports of goods and services (% of GDP)
X5	Fertility rate, total (births per woman)
X6	Fixed line and mobile phone subscribers (per 1000 people)
X7	GDP growth (annual)
X8	Immunization, measles (% of children ages 12-23 months)
X9	Imports of goods and services (% of GDP)
X10	Industry, value added (% of GDP)
X11	Inflation, GDP deflator (annual %)
X12	Internet users (per 1000 people)
X13	Life expectancy at birth, total (years)
X14	Merchandise trade (% of GDP)
X15	Mortality rate, infant (per 1000 live births)
X16	Mortality rate, under 5 (per 1000)
X17	Population growth (annual %)

- (a) Realice un test de normalidad multivariada.
- (b) Determine el número de factores a extraer.
- (c) Realice el análisis exploratorio de acuerdo con el modelo factorial.
- (d) Reporte las comunales.

Ejercicio 2

El archivo *ipc2dig.dta* contiene datos sobre variaciones mensuales de precios desagregados en componentes de la canasta básica de alimentos para el período comprendido entre marzo de 2022 y marzo de 2006.

Variable	Descripción
div1	Alimentos para consumir en el hogar

Variable	Descripción
div2	Bebidas e infusiones para consumir en el hogar
div3	Alimentos y bebidas consumidos fuera del hogar
div4	Ropa
div5	Calzado
div6	Accesorios y servicios para la indumentaria
div7	Alquiler de la vivienda
div8	Servicios básicos y combustibles para la vivienda
div9	Reparaciones y gastos comunes de la vivienda
div10	Equipamiento del hogar
div11	Mantenimiento del hogar
div12	Productos medicinales y accesorios terapéuticos
div13	Servicios para la salud
div14	Transporte
div15	Comunicaciones
div16	Turismo
div17	Equipos, conexiones y servicios de audio, televisión y computación
div18	Diarios, revistas y libros
div19	Juguetes y artículos para deporte
div20	Flores, planes y atención de animales domésticos
div21	Otros servicios de esparcimiento
div22	Servicios educativos
div23	Textos y útiles escolares
div24	Cigarrillos y accesorios
div25	Artículos y servicios para el cuidado personal
div26	Servicios diversos

- Realice un análisis factorial y determine el número de factores a considerar en el mismo.
- Estime los factores.
- Analice los residuos del modelo, con arreglo a los supuestos considerados.
- Obtenga las principales medidas estadísticas que permitan inferir acerca de la bondad del ajuste del modelo estimado.

Análisis Estadístico Multivariado | UTDT

Problem Set 6

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Ejercicio 1

Estime el modelo factorial para la base de datos *eurosec.dta*, de acuerdo con el número de factores resultantes del proceso de selección de modelos introducido por Schwarz (1978). Obtenga, luego, la matriz ortogonal $C_{m \times m}$ verificando que las cargas asociadas a cada factor presenten la máxima varianza posible. Finalmente, estime el vector de factores correspondiente al ajuste efectuado.

Ejercicio 2

Aplice el contraste empleado en el primer ejercicio a los datos correspondientes al Problem Set anterior. Explique brevemente el resultado obtenido. ¿Cuáles son sus conclusiones al respecto?

Ejercicio 3

Estime un modelo factorial empleando la base *sachs.dta*, extracto de Gallup, Sachs & Mellinger (1999).

Ejercicio 4

En base a los datos empleados por la Fundación Heritage para construir el *Freedom Index 2010*, contenidos en la base *heritage.dta*, ajustar un modelo factorial y comentar brevemente los resultados obtenidos.

Análisis Estadístico Multivariado | UTDT

Problem Set 7

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Ejercicio 1

La base de datos *firmas.dta* contiene información financiera referida a dos tipos de firmas: 12 firmas consideradas de buena performance, y 12 firmas no tan buenas en ese sentido. En la base se podrán observar dos características relevadas en la totalidad de las firmas: EBITASS (ganancias después de impuestos e intereses sobre activos) y ROTC (retorno sobre el capital).

- (a) Realizar un gráfico de las variables relevadas. Interpretarlo.
- (b) Realizar un test de medias multivariado.
- (c) Realizar un análisis discriminante.
- (d) Obtener predicciones para toda la muestra y comparar con la información observada.
- (e) Considere una firma con un EBITASS y un ROTC de 0,1. ¿En qué grupo la clasificaría?

Ejercicio 2

La base de datos *muestra_engh.dta* contiene información proveniente de la Encuesta Nacional de Gastos de los Hogares para tres regiones: Metropolitana, Noreste (NEA) y Sur. Considere responder a las siguientes consignas con las variables desagregadas por capítulo de gasto.

Variable	Descripción
Region	1 = Metropolitana; 4 = NEA; 6 = Sur
cap1	Capítulo 1 - Alimentos y bebidas
cap2	Capítulo 2 - Indumentaria y calzado
cap3	Capítulo 3 - Vivienda
cap4	Capítulo 4 - Equipamiento y funcionamiento del hogar
cap5	Capítulo 5 - Atención médica y gastos para la salud
cap6	Capítulo 6 - Transporte y comunicaciones
cap7	Capítulo 7 - Esparcimiento
cap8	Capítulo 8 - Educación
cap9	Capítulo 9 - Bienes y servicios varios
gctotal	Gasto total en consumo
intotal	Ingreso neto del hogar

- (a) Realice un test de igualdad de medias entre regiones.

- (b) Realice un análisis discriminante.
- (c) Obtenga las predicciones a partir de la/s estimación/es de la función/es discriminante/s.
- (d) Obtenga las predicciones a partir de las distancias de Mahalanobis.
- (e) Complete una tabla de predicción-realización.
- (f) ¿Cómo realizaría el análisis en caso de no poder suponer que las matrices de varianzas y covarianzas son iguales entre grupos? ¿Cuántos parámetros debería estimar?