Curso: Estadística Multivariada
Tarea 5
Fecha de entrega: 27 de abril de 2018

Instrucciones:

- Los primeros 4 ejercicios se entregan en pdf
- Los restantes ejercicios se entregarán con el script de R que generen y con un reporte en pdf que incluya los resultados obtenidos y las respuestas solicitadas.
- El script debe correrse de tal manera que los datos (si ocupa) estén situados en la carpeta donde se encuentra el script.
- Los ejercicios de R se revisarán en base a los resultados provenientes de correr el script, por lo que verifiquen y marquen la parte donde se encuentra el resultado.
- Todos los resultados se subirán a la plataforma

Ejercicios

1. Muestra que la matriz de covarianza

$$\boldsymbol{\rho} = \begin{pmatrix} 1.0 & .63 & .45 \\ .63 & 1.0 & .35 \\ .45 & .35 & 1.0 \end{pmatrix}$$

para las p=3 variables aleatorias estandarizadas $Z_1, Z_2\ y\ Z_3$, puede ser generada por el modelo de factores con m=1

$$Z_1 = .9F_1 + \varepsilon_1$$
$$Z_2 = .7F_1 + \varepsilon_2$$
$$Z_3 = .5F_1 + \varepsilon_3$$

donde $Var(F_1) = 1,\ Cov(\boldsymbol{\varepsilon}, F_1) = \mathbf{0},$ y

$$\boldsymbol{\Psi} = Cov(\boldsymbol{\varepsilon}) = \begin{pmatrix} .19 & 0 & 0 \\ 0 & .51 & 0 \\ 0 & 0 & .75 \end{pmatrix}.$$

Esto es, escribe $\boldsymbol{\rho}$ en la forma $\boldsymbol{\rho} = \boldsymbol{LL}' + \boldsymbol{\Psi}$

2. Usa la información del ejercicio anterior

   a) Calcula las comunalidades ${h_i}^2, i = 1,2,3$ e interpreta estas cantidades.

b) Calcula $Corr(Z_i, F_1)$ para i=1,2,3. Cual variable podría llevar el mayor peso en la interpretación del factor común ? Porqué?

3. Los valores y vectores propios de la matriz de correlaciones $\boldsymbol{\rho}$ en el ejercicio 1 son

$$\lambda_1 = 1.96, \quad e'_1 = [.625, .593, .507]$$
$$\lambda_2 = .68, \quad e'_2 = [-.219, -.491, .843]$$
$$\lambda_3 = .36, \quad e'_3 = [.749, -.638, -.177]$$

a) Asumiendo un modelo de factores con m=1, calcula la matriz de cargas **L** y la matriz de varianzas específicas $\boldsymbol{\Psi}$ usando el método por componentes principales. Compara los resultados con los del ejercicio 1.

b) Qué proporción de la varianza poblacional total es explicada por el primer factor común?

4. *(Solución única pero impropia: caso Heywood).* Considere un modelo factorial con m=1 para la población con matriz de covarianza

$$\Sigma = \begin{pmatrix} 1 & .4 & .9 \\ .4 & 1 & .7 \\ .9 & .7 & 1 \end{pmatrix}.$$

Muestra que existe una única elección de **L** y $\boldsymbol{\Psi}$ con $\boldsymbol{\Sigma} = \boldsymbol{LL'} + \boldsymbol{\Psi}$, pero que $\psi_3 < 0$, por lo que la elección no es admisible.

Ejercicios con R

5. El Proyecto de Evaluación de la Apertura Sintética de la Personalidad (SAPA) es una colección de datos psicológicos basada en la web.2 Un subconjunto de los datos está disponible en R como bfi en la biblioteca ”psych”.

Este subconjunto contiene datos en tres variables demográficas y 25 ítems de personalidad de2800 voluntarios. Como ejemplos de estos ítems, tenemos:
- Sé cómo consolar a los demás.
- Desperdicio mi tiempo.
- Hago amigos con facilidad.

Cada ítem es clasificado en una escala de 1-7, en si el encuestado siente que él o ella está de acuerdo con la declaración mucho, no está de acuerdo mucho o cae en algún lugar intermedio. Consulta el archivo de ayuda de bfi para obtener más detalles.

a) Utilice el comando complete.cases () para eliminar individuos en bfi con cualquier valor faltantes
b) Utilice el análisis de factores para agrupar elementos de naturaleza similar. Trata de interpretar la naturaleza de los ítems que se agrupan. Este es un ejercicio útil en psicología. El test de ji cuadrado para el número de factores puede no ser apropiado con una muestra tan grande.
c) Identifica las preguntas que tienen una preponderancia de acuerdo extremo y / o en desacuerdo las respuestas. Del mismo modo, identifica casos atípicos tales como personas que parecen responder de manera extrema. Es decir, las personas que tienden a estar totalmente de acuerdo o en desacuerdo con la mayoría de las preguntas.

6. El conjunto de datos Harmon23.cor en el paquete "datasets" es una matriz de correlación de ocho mediciones físicas realizadas en 305 niñas entre las edades de 7 y 17 años.

a) Realiza un análisis factorial de estos datos.
b) Varía el número de factores para encontrar un ajuste adecuado del modelo e interprete las cargas factoriales resultantes.

7. La matriz de correlación dada a continuación proviene de las puntuaciones de 220 chicos en seis asignaturas escolares: 1) Francés, 2) Inglés, 3) Historia, 4) Aritmética, 5) Álgebra y 6) Geometría.

$$\mathbf{R} = \begin{array}{l} \text{French} \\ \text{English} \\ \text{History} \\ \text{Arithmetic} \\ \text{Algebra} \\ \text{Geometry} \end{array} \begin{pmatrix} 1.00 & & & & & \\ 0.44 & 1.00 & & & & \\ 0.41 & 0.35 & 1.00 & & & \\ 0.29 & 0.35 & 0.16 & 1.00 & & \\ 0.33 & 0.32 & 0.19 & 0.59 & 1.00 & \\ 0.25 & 0.33 & 0.18 & 0.47 & 0.46 & 1.00 \end{pmatrix}.$$

a) Encuentre la solución de dos factores de un análisis de factor de máxima verosimilitud.
b) Mediante una inspección de las cargas, encuentre una rotación ortogonal que permite una interpretación más fácil de los resultados.