

Curso: Estadística Multivariada
Tarea 8
Fecha de entrega: 22 de abril de 2020

Instrucciones:

Subir la tarea a la plataforma en un zip que contenga un archivo pdf con las respuestas documentadas y el código (en los ejercicios que lo requieran).

Ejercicios

1. Dada 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 :

a) Muestre que la matriz 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, escriba $\boldsymbol{\rho}$ en la forma $\boldsymbol{\rho} = \boldsymbol{LL}' + \boldsymbol{\Psi}$

b) Calcule las comunidades $h_i^2, i = 1,2,3$ e interpreta estas cantidades.

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

2. Con respecto a la matriz de correlaciones del ejercicio anterior:

i. Calcule los valores y vectores propios de la matriz de correlaciones $\boldsymbol{\rho}$ y asumiendo un modelo de factores con $m=1$, calcule la matriz de cargas \mathbf{L} y la matriz de varianzas específicas $\boldsymbol{\Psi}$ usando el método por componentes principales a partir de los vectores y valores propios. Compare los resultados del modelo con los del ejercicio anterior

- ii. ¿Que proporción de la varianza poblacional total es explicada por el primer factor común?
3. (Solución única pero impropia: caso Heywood). Considere un modelo factorial con $m=1$ para la población con matriz de covarianza

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

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

4. Una empresa está evaluando la calidad de su personal de ventas para lo cual seleccionó una muestra aleatoria de 50 vendedores y evaluó en cada uno de ellos 3 medidas de rendimiento: crecimiento de ventas, rentabilidad de ventas y ventas de nuevas cuentas. Estas medidas se han convertido a una escala, en la que 100 indica desempeño "promedio". Además, a los 50 individuos se les aplicaron 4 pruebas, que pretendían medir la creatividad, el razonamiento mecánico, el razonamiento abstracto y la capacidad matemática, respectivamente. Las $n = 50$ observaciones sobre las $p=7$ variables se muestran en el archivo "datos_vendedores".
- Asumiendo un modelo ortogonal de factores para las variables estandarizadas, obtenga la solución por máxima verosimilitud de \mathbf{L} y ψ para $m=2$ y $m=3$ factores, considerando una rotación varimax, e interprete las soluciones con $m=2$ y $m=3$ factores.
 - A partir de las estimaciones de los parámetros obtenga las comunialidades, las varianzas específicas y $\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\psi}$ para las soluciones en $m=2$ y $m=3$ factores. Compare los resultados. Qué elección de m prefiere en este punto? ¿Por qué?
 - Realice una prueba de $H_0: \Sigma = \mathbf{L}\mathbf{L}' + \Psi$ versus $H_1: \Sigma \neq \mathbf{L}\mathbf{L}' + \Psi$ para $m=2$ y $m=3$. A partir de estos resultados y de la parte b), que elección de m parece ser la adecuada?
 - De acuerdo al número de factores elegido en c), calcule las puntuaciones de los factores (*factor scores*) para los vendedores mediante: i) mínimos cuadrados ponderados y ii) mediante el enfoque de regresión. ¿Existe algún patrón de agrupamiento de los vendedores de acuerdo a sus puntuaciones factoriales?, si es

así, ¿cómo se caracterizan los vendedores de cada grupo, de acuerdo a la interpretación de los factores?

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 de 2800 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. Consulte 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 análisis de factores para agrupar elementos de naturaleza similar. Trate 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) Identifique las preguntas que tienen una preponderancia de acuerdo extremo y / o en desacuerdo las respuestas. Del mismo modo, identifique 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.