Estadística Multivariada

Examen parcial

Instrucciones:

- Resuelve de forma individual los siguientes problemas
- Utiliza el software de su preferencia para realizar los cálculos cuando sea necesario
- Subir a la plataforma en un zip que contenga un archivo pdf con las respuestas documentadas y el código (en los ejercicios que lo requieran).
- Fecha límite de entrega: Viernes 23 de abril, 23:00

1. Los datos del archivo *datosavehembra.dat* corresponden a mediciones de x_1=longitud de cola y x_2=longitud de ala, para una muestra de $n = 45$ aves.
   (a) Construye y muestra una región de confianza (elipse) del 95% para μ_1 y μ_2. Supón que se sabe que $\mu_1 = 190$mm y $\mu_2 = 275$mm son valores estándar para las aves macho. ¿Son datos plausibles para las mediciones de las aves hembra?
   (b) Construye intervalos de confianza T^2 simultáneos de 95% para μ_1 y μ_2. ¿Hay alguna ventaja sobre los de bonferroni?

2. Muchos inversionistas están buscando dividendos que se pagarán de los beneficios futuros. Los datos del archivos ***cash hi tech.tx*** enumeran una serie de características sobre su situación financiera, hasta septiembre del 2010, de varias empresas de tecnologías e información. Las variables resultado a explicar son los dividendos actuales y futuros (current, Y_1 y 60% payout, Y_2).
   (a) Desarrolle un modelo de regresión multivariada usando la capitalización de mercado (market cap, z_1), efectivo neto (net cash, z_2) y flujo de efectivo (cash flow, z_3) como las variables independientes.
   (b) Analice el efecto que tiene el flujo de efectivo (z_3) respecto a los dividendos en conjunto. Comente los resultados
   (c) Dado $z_0 = [1, 21296, 7850, 15.2]$, obtenga una elipse de confianza al 95% para Bz_0 e interpretala
   (d) Obtenga una elipse de predicción al 95% para $Y_0 = [Y_{01}, Y_{02}]$ dado el valor del inciso anterior e interpretala.

3. 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 **datosvendedores**.

(a) Asumiendo un modelo ortogonal de factores para las variables estandarizadas. Obtén la solución por máxima verosimilitud de $\boldsymbol{L}$ y $\boldsymbol{\Psi}$ para $m = 2$ y $m = 3$ factores, considerando una rotación varimax, e interpreta las soluciones con $m = 2$ y $m = 3$ factores.

(b) A partir de las estimaciones de los parámetros obtén las comunalidades, las varianzas específicas y $\widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}^{'} + \widehat{\boldsymbol{\Psi}}$ para las soluciones en $m = 2$ y $m = 3$ factores. Compara los resultados. Qué elección de m prefieres en este punto? ¿Por qué?

(c) Realiza una prueba de $H_0 : \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^{'} + \boldsymbol{\Psi}$ Vs $H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{L}\boldsymbol{L}^{'} + \boldsymbol{\Psi}$ para $m = 2$ y $m = 3$. A partir de estos resultados y de la parte b), que elección de m parece ser la adecuada?

(d) De acuerdo al número de factores elegido en c), calcula las puntuaciones de los factores (factor scores) para los vendedores mediante: i) mínimos cuadrados ponderados y ii) mediante el enfoque de regresión. ¿Existe algun patron de agrupamiento de los vendedores de acuerdo a sus puntuaciones factoriales?, si es así ¿como se caracterizan los vendendedores de cada grupo, de acuerdo a la interpretación de los factores?

4. Considere que los vectores X_1,X_2,X_3 y X_4, son independientes, aleatorios e idénticamente distribuidos normalmente con parámetros

$$\mu = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}, \Sigma = \begin{pmatrix} 3 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

Considere las siguientes combinaciones lineales de los vectores aleatorios anteriores

$$\frac{1}{2}X_1 + \frac{1}{2}X_2 + \frac{1}{2}X_3 + \frac{1}{2}X_4$$

$$X_1 + X_2 + X_3 - 3X_4$$

(a) Obtenga el vector de medias y su matriz de covarianzas para cada uno de ellos

(b) Calcule la covarianza entre ellos

5. (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 Ψ con $\Sigma = LL^{'} + \Psi$, pero que $\psi_3 < 0$, por lo que la elección no es admisible.